# FINAL REPORT

# EVALUATION OF AQUATIC GAP ANALYSIS FISH DISTRIBUTION MODELS, WITH EMPHASIS ON RARE FISH SPECIES IN SOUTH DAKOTA

By

Cari-Ann Hayer and Steven S. Wall
Department of Wildlife and Fisheries Sciences

and

Charles R. Berry Jr.
U.S. Geological Survey (BRD)
South Dakota Cooperative Research Unit

South Dakota State University
Box 2140B, NPBL 138
Brookings, SD 57007

Final report for State Wildlife Grant program project number T-9, study number 2409

To

South Dakota Department of Game Fish and Parks
Pierre, South Dakota

2006

## EXECUTIVE SUMMARY

Aquatic gap analysis is a geographic planning approach for conserving biodiversity. Predicting fish species distribution and determining accuracy of predictive models are important in identifying gaps between location of a species and the location of conservation activities to protect the species. The purpose of this study was to use the new and improved aquatic gap models to predict the presence of the species of concern (e.g., central mudminnow, finescale dace, banded killifish, northern redbelly dace, pearl dace, longnose sucker, trout-perch, hornyhead chub, carmine shiner and sturgeon chub) in the streams of South Dakota, and assess the accuracy of the models. Specific objectives were to: assess accuracy of distribution models for species of concern, and increase data on present distributions of the species of concern. Other fish were collected also so we expanded the research to include model testing for all species.

Fishes were collected by seining and electrofishing at 143 sites. Detection probabilities were used as a weighting factor to reduce false negative errors. Presence and absence data were compared with predictions to evaluate model accuracy using Cohen's kappa statistic (range = 0.0 – 1.0) and correct classification rates. A high kappa (k) indicates better model accuracy than a low kappa. Another way to interpret a kappa value is as follows: when k = 0.43, the model can correctly predict the occurrence of a species 43% more often than by chance.

Model accuracy is synthesized by ecoregional drainage unit or EDU (k = 0.32 – 0.61). An EDU is a land area with uniform features that fashion a unique fish assemblage. Six EDUs were: East River, Minnesota, Black Hills, Cheyenne, White, and Grand/Moreau. Prediction models varied somewhat for the same species among EDUs.

The Northern redbelly dace had the highest kappa statistic of all the target species in the East River EDU (k = 0.605) indicating substantial model agreement (k between 0.60-0.80). Model predictions for other target species were moderately accurate (k 0.40-0.60), except for longnose sucker in the Black Hills EDU and central mudminnow in the Minnesota River EDU that had fair model agreement (k = 0.318 and 0.364, respectively). Appendices supply model accuracy statistics for all species.

The accuracy results were comparable to those of other studies, even though we tested the models against independent data, which is the most rigorous test possible. However, the models are affected by factors that affect the accuracy of all animal distribution models. We minimized some of the problems and improved model accuracy by a new computational approach that uses species detection probabilities.

The results were immediately useful in improving knowledge about the distribution of fish species of concern in South Dakota. The models can be helpful in planning other surveys. Assessing model accuracy is an important first step in understanding the relationships between fishes and habitat. Fish distribution models also reveal inadequacies in data about fish locations and habitat affinities.

The next step in using gap analysis results is to begin applying conservation practices to land where there is a high probability of the target species presence. The model accuracy assessments made in this study provide managers with a measure of confidence about the value of the models in decisions about where to place conservation practices. Administrators can have more confidence in the potential value of conservation practices in areas where the models are more accurate than where the models are less accurate.

# ACKNOWLEDGEMENTS

The South Dakota Department of Game, Fish and Parks (SDGF&P) funded this project as State Wildlife Grant Project T-9, Study Number 2409. The South Dakota Gap Analysis Project, particularly the Upper Missouri River Basin Gap Analysis Project (UMRBGAP) provided many of the GIS layers, species locations and modeling techniques. We thank those who allowed us to add their fish collection data to ours for testing our models, specifically Sheila Thomson – South Dakota State University (SDSU), Chris Hoagstrom - SDSU, Steve Freeling -Vermillion River water Development District, Jeff Shearer - SDGF&P, Nathan Morey – South Dakota Department of Transportation, Jason Kral – U.S. Geological Survey EMAP project, and Ryan Sylvester and Steve Freeling – UMRBGAP. Field technicians who helped collect data for this project are Jason Kral, Brian Abel, Wes Bouska and Dan Doris. We thank landowners who allowed us to study streams on their land. The South Dakota Cooperative Fish and Wildlife Research Unit is jointly sponsored by the SDGF&P, Wildlife Management Institute, US Fish and Wildlife Service, SDSU, and the US Geological Survey.

## TABLE OF CONTENTS

**Executive Summary** ..... i

**Acknowledgements** ..... iii

**Table of Contents** ..... iv

**List of Tables** ..... vi

**List of Figures** ..... viii

**Introduction** ..... 1

**Methods** ..... 7
- Classification of Riverine Ecosystems ..... 7
- Modeling ..... 8
- Fish collection data ..... 8
- Fish habitat models ..... 13
- Predicted GIS distribution ..... 17
- Independent sampling ..... 17
- Modeling accuracy assessment ..... 19

**Results** ..... 23
- East River EDU ..... 25
- Minnesota River EDU ..... 37
- Black Hills EDU ..... 46
- Cheyenne EDU ..... 53
- Grand and Moreau EDU ..... 57
- White EDU ..... 60

**Discussion** ..... 65

**REFERENCES** ..... 74

**APPENDICES** ..... 86
- Appendix 1. A brief description of the ten species of concern in South Dakota targeted within this study. ..... 86
- Appendix 2. Habitat variables associated with analysis of fish-habitat models ..... 89
  - Appendix 2.1 Relative gradient variables used for analysis in fish-habitat models ..... 89

Appendix 2.2 Relative elevation variables used for analysis in fish-habitat models ..... 90
Appendix 2.3. Gradient and elevation by EDU used for analysis in fish-habitat models ..... 91
Appendix 2.4. Gradient and elevation by region used for analysis in fish-habitat models. ..... 92
Appendix 2.5 Expanded Stream Sizes From Shreve Link used for analysis in fish-habitat models. ..... 93

Appendix 3. Independent sampling study site locations collected by the "rare fish study" ..... 94

Appendix 4. List of sampling locations from additional sources used in independent data set to access accuracy of fish distribution models. ..... 95

Appendix 5. List of species collected in five ecological drainage units with corresponding prevalence (frequency of occurrence). A dash indicates that the species does not occur or was not predicted within that EDU. ..... 97

Appendix 6. East River EDU accuracy metrics for all species collected and/or predicted. ..... 99

Appendix 7. Habitat variables associated with presence of fish species of concern used in predictive models ..... 101

Appendix 8. Minnesota EDU accuracy metrics for all species collected and/or predicted ..... 102

Appendix 9. Black Hills EDU accuracy metrics for all species collected and/or predicted ..... 104

Appendix 10. Cheyenne River EDU accuracy metrics for all species collected and/or predicted ..... 105

Appendix 11. Grand Moreau EDU accuracy metrics for all species collected and/or predicted ..... 107

Appendix 12. White River accuracy metrics for all species collected and/or predicted ..... 109

## LIST OF TABLES


TABLE 1 List of South Dakota fish species of concern. Species with asterisk (*) were not examined in this study. .....1

TABLE 2 Valley Segment habitat variables, GIS data source and methods used to delineate the variable .....10

TABLE 3 Data sources for fish collection sites used in modeling and in the independent dataset (italics), the number of collection sites contributed from each source and the Ecoregional Drainage Units (EDUs) represented from each source. .....12

TABLE 4 List of project/data sources used in independent data set validating accuracy of fish species model predictions from 2004-2005. .....19

TABLE 5 Two by two confusion matrix used to calculate model accuracy metrics .....21

TABLE 6 Possible measures for assessing the accuracy of prediction models .....21

TABLE 7 South Dakota fish species at risk kappa values by ecological drainage unit. .....25

TABLE 8 East River ecoregional drainage unit (EDU) rare fish species accuracy metrics from independent data set .....28

TABLE 9 Minnesota River Ecoregional Drainage Unit (EDU) rare fish species accuracy metrics from independent data set .....41

TABLE 10 Black Hills ecoregional drainage unit (EDU) rare fish species accuracy metrics from independent data set .....50

TABLE 11 Cheyenne River ecological drainage unit (EDU) rare fish species accuracy metrics from independent data set .....55

TABLE 12 Grand and Moreau River ecological drainage unit (EDU) rare fish species accuracy metrics from independent data set .....58

TABLE 13 White River ecological drainage unit (EDU) rare fish species accuracy metrics from independent data set .....62

TABLE 14 Range of correct classification rates (CCR) and kappa (k) statistics for rare species predicted to occur in each ecological drainage unit (EDU) ..... 64

## LIST OF FIGURES

FIGURE 1 Study area showing Ecoregional Drainage Units (EDUs), fish collection sites used in modeling, and independent test sites ..... 9

FIGURE 2 The number of new sites visited (independent data) and the number of species collected per ecological drainage unit (EDU). ..... 24

FIGURE 3 Correct classification rate, specificity, and sensitivity of presence-absence models for 46 species in relation to their occurrence in 62 streams in the East River ecoregional drainage unit ..... 27

FIGURE 4 Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in East River ecoregional drainage unit (EDU) ..... 28

FIGURE 5 Central mudminnow (*Umbra limi*) potential 11-digit hydrological unit distribution from model results ..... 30

FIGURE 6 Central mudminnow (*Umbra limi*) predicted valley segments and model test results ..... 31

FIGURE 7 Northern redbelly dace (*Phoxinus eos*) 11-digit hydrological unit distribution from model results ..... 32

FIGURE 8 Northern redbelly dace (*Phoxinus eos*) predicted valley segments and model test results for the Big Sioux and Minnesota River Basins ..... 33

FIGURE 9 Trout perch (*Percopsis Omiscomaycus*) potential 11-digit hydrological unit distribution from model results ..... 34

FIGURE 10 Banded killifish (*Fundulus diaphanus*) potential 11-digit hydrological unit distribution from model results ..... 34

FIGURE 11 Trout perch (*Percopsis Omiscomaycus*) predicted valley segments (red colored streams) for the Big Sioux and Minnesota River Basins. Does not include independent dataset ..... 35

FIGURE 12 Banded killifish (*Fundulus diaphanus*) predicted valley segments and model test results ..... 36

FIGURE 13 Correct classification rate, sensitivity and specificity of presence-absence models for 30 species in relation to their occurrence in 22 streams in the Minnesota ecological drainage unit ..... 39

FIGURE 14 Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Minnesota River Ecological Drainage Unit (EDU) ..... 40

FIGURE 15 Hornyhead chub (*Nocomis biguttatus*) 11-digit hydrological unit distribution from model results ..... 42

FIGURE 16 Carmine shiner (*Notropis percobromus*) potential 11-digit hydrological unit distribution from model results ..... 43

FIGURE 17 Hornyhead chub (*Nocomis biguttatus*) predicted valley segments and model test results. ..... 44

FIGURE 18 Carmine shiner (*Notropis percobromus*) predicted valley segments and model test results ..... 45

FIGURE 19 Correct classification rate, sensitivity, and specificity of presence-absence models for 18 species in relation to their occurrence in 15 streams in the Black hills ecological drainage unit. ..... 47

FIGURE 20 Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Black hills River ecological drainage unit (EDU) ..... 47

FIGURE 21 Finescale dace (*Phoxinus neogaeus*) 11-digit hydrological unit distribution from model results ..... 49

FIGURE 22 Finescale dace (*Phoxinus neogaeus*) predicted valley segments and model test results for the Black Hills ..... 50

FIGURE 23 Longnose sucker (*Catostomus catostomus*) 11-digit hydrological unit distribution from model results ..... 50

FIGURE 24 Longnose sucker (*Catostomus catostomus*) predicted valley segments and model test results for the Black Hills ..... 51

FIGURE 25 Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Cheyenne River ecological drainage unit (EDU)..... 53

FIGURE 26 Sturgeon chub (*Macrhybopsis gelida*) 11-digit hydrological unit distribution from model results ..... 54

FIGURE 27 Sturgeon chub (*Macrhybopsis gelida*) predicted valley segments and model test results for the South Dakota..... 54

FIGURE 28 Correct classification rate, sensitivity and specificity of presence-absence models for 28 species in relation to their occurrence in nine streams in the Grand and Moreau ecological drainage unit ..... 56

FIGURE 29 Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Grand and Moreau ecological drainage unit (EDU) ..... 57

FIGURE 30 Northern redbelly dace (Phoxinus eos) predicted valley segments and model test results for the Grand and Moreau River Basins..... 58

FIGURE 31 Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in White River ecological drainage unit (EDU)..... 60

FIGURE 32 Northern redbelly dace (*Phoxinus eos*) predicted valley segments and model test results for the White River Basin and Crow Creek watershed..... 61

FIGURE 33 Pearl dace (*Margariscus margarita*) 11-digit hydrological unit distribution from model results ..... 62

FIGURE 34 Pearl dace (*Margariscus margarita*) predicted valley segments and model test results for the White River Basin and Missouri River tributaries..... 62

## Introduction

Models that predict the potential distribution of South Dakota's fishes are an important component of the aquatic gap analysis process. An aquatic gap analysis has been conducted twice for South Dakota's fishes. Preliminary models describing the potential distribution of 116 fish species were first used by Smith et al. (2002) who found that 19 species were protected in >10% of their range. Researchers had the opportunity to improve on these models when South Dakota was included in the aquatic gap analysis study of the Upper Missouri River Basin (Wall et al. 2005). Among the new distribution models were models for South Dakota's fish species of concern (Table 1).

**Table 1:** List of South Dakota fish species of concern. Species with asterisk (*) were not examined in this study. Details of each species discussed in this report can be found in Appendix 1.

| Species Code | Family | Common Name | Scientific Name |
|---|---|---|---|
| P_PAS | Acipenseridae | Pallid Sturgeon * | *Scaphirhynchus albus* |
| P_CEM | Umbridae | Central Mudminnow | *Umbra limi* |
| P_BES | Cyprinidae | Blacknose shiner * | *Notropis heterolepis* |
| P_FID | | Finescale dace | *Phoxinus neogaeus* |
| P_HOC | | Hornyhead chub | *Nocomis biguttatus* |
| P_NRD | | Northern redbelly dace | *Phoxinus eos* |
| P_PED | | Pearl dace | *Margariscus margarita* |
| P_ROS | | Carmine shiner | *Notropis percobromus* |
| P_SIC | | Sicklefin chub * | *Hybopsis meeki* |
| P_SNC | | Sturgeon chub * | *Macrhybopsis gelida* |
| P_TOS | | Topeka shiner | *Notropis topeka* |
| P_LOS | Catostomidae | Longnose sucker | *Colostomies catostomus* |
| P_TRP | Percopsidae | Trout-perch | *Percopsis omiscomaycus* |
| P_BAK | Fundulidae | Banded killifish | *Fundulus diaphanus* |

## Purpose of the Study

The purpose of this study was to use the new and improved aquatic gap models to predict the presence of the species of concern in the streams of South Dakota, and conduct an accuracy assessment of the models. Specific objectives were to:

1) assess accuracy of models to validate their use as decision support tools, and
2) increase data on distributions of fish species focusing on species of concern.

Additionally, we assessed the accuracy of models for all species predicted to occur in South Dakota to increase the robustness of the predictive models. We also used a novel approach not included in the proposal which was to incorporate detection probabilities in the analysis of model accuracy assessment.

## Gap Analysis

The National Gap Analysis Program (GAP) was initiated in 1998 to provide a coarse-filter assessment strategy for identifying and prioritizing biodiversity and conservation needs. The Program uses predictive modeling to identify elements of biodiversity of land and water areas that are not sufficiently represented in conservation areas (Jennings 2000). A fundamental assumption of gap analysis is that in order to decrease the probability of extinction of a species as a result of anthropogenic impacts, identifying species and conservation areas before the population is diminished to the point of endangerment is necessary (Jennings 2000). Traditionally, the focus of GAP analysis has been on terrestrial systems and species. Recently, methods and concepts of terrestrial

GAP have been applied to identify areas of high diversity in aquatic systems (Sowa 1998; Meixler and Bain 1999).

Aquatic gap analysis uses biological data, remote sensing and geographic information systems to predict areas where fish species might find suitable habitat and where conservation programs can best be used to protect current species of concern and to keep common species common. An increasing focus on conserving rare and endangered species has related a need to obtain information on their abundance, distribution, and habitat use (Kral et al. 2005).

**Predictive models**

Predictive models have numerous applications for the conservation and management of fish populations (Olden and Jackson 2002) and play important roles in prioritizing surveys and monitoring programs for fish populations. Limitations of resources often preclude exhaustive and continual sampling of sites. Existing data on fish-habitat relationships can be used in the construction of predictive models. Predictive models provide scientists with the ability to assess the distribution and suitability of habitat for threatened species and prioritize areas for conservation (Oakes et al. 2005).

Predictive models of species occurrence are important conservation tools (Ruston et al. 2004) that identify: 1) areas where species have been poorly sampled, 2) the status of species in comparison to historical distributions, 3) areas for reintroduction, and 4) conservation areas that may be needed in the future as a result of anthropogenic impacts

(Angermeier et al. 2002). The biological and ecological integrity of river ecosystems is at risk from extreme land and hydrological alterations as a result of a rapidly increasing human population (Poff et al. 2003), resulting in the need for biogeographic information that is both spatially comprehensive and of appropriate resolution necessary for effective management of our biological resources (Jennings 2000). The use of statistical models to predict the likely occurrence or distribution of species is becoming increasingly important in conservation planning and resource management (Pearce and Ferrier 2000).

Fish distribution models may be used by conservation agencies to aid future management and conservation planning efforts. Uses of these predictive models include: 1) estimate habitat suitability, 2) forecast habitat change as a result of altered land-use patterns, 3) identify potential locations for species introductions, 4) predict the likelihood of exotic species invasion, and 5) predict "hotspots" for species and biodiversity conservation (Olden and Jackson 2002). The utility of predictive models is extensive; however their usefulness and validity have been recently debated (Short and Hestbeck 1995; Davis 1996; Edwards 1996; Scott et al. 1996; Guisan and Zimmerman 2000; Karl et al. 2000). Distribution models can be valuable tools but proper validation and accuracy assessment of predictions is necessary to quantify model performance, especially if the models are used for future conservation planning.

Model validation refers to the application of a model, the generation of predictions, and the quantification of predictive performance (e.g. the correct classification rate for predicting species presence or absence). Validation is an important

step in the modeling process because it quantifies our confidence in the predictions and potential applications of the model. To obtain unbiased estimates of the predictive accuracy of a model, independent data collected from sites other than those used to develop the models should be used (Pearce and Ferrier 2000; Olden et al. 2002).

**Sampling rare species**

Sampling rare species can be challenging and problematic to biologists and managers (McArdle 1990; Scott et al. 2002; Wall et al. 2004; Mackenzie et al. 2005) and relatively few predictive models have been applied to rare and endangered species (Miller 1986; Myatt 1987; Carey and Brown 1995; Godown and Peterson 2000; Elith and Burgman 2002). A species may be considered rare because of low abundance across a broad spatial scale, cryptic or nocturnal behavior and/or locally abundant (Gaston 1994; McDonald 2004; Mackenzie et al. 2005) all of which are relative to detectability.

Estimates of species presence are generally derived from catch data and are assumed to be representative of actual species abundance in the sampled area (Schmidt 2003); however, it is typically impossible to capture or detect all species present in a particular sampling area (Pollock et al. 2002; Gu and Swihart 2004). This is especially difficult for mobile aquatic animals such as fishes (Boulinier et al. 1998; Nichols et al. 1998a) and becomes even more difficult if the fishes are rare. Rare species have low model sensitivity implying that it is more difficult to predict the occurrence of organisms whose conservation and management may be the most critical (Olden and Jackson 2002).

Detection of a species is defined as the probability of detecting at least one individual of a species, assuming the species is present at the sampling site at that particular time (Boulinier et al. 1998; Nichols et al. 1998; Bayley and Peterson 2001) and is a function of abundance and effectiveness of sampling methods and effort (Nichols et al. 1998b; Gu and Swihart 2004). Failure to account for detection probabilities less than one may lead to biases in model applicability and it is important to incorporate these estimates into monitoring plans to predict species presence and help generate baseline data for long-term conservation and management of species (Mackenzie et al. 2002; Moilanen 1999; Hayer 2005; Mackenzie et al. 2005).

Resource agencies require decision support tools for managing aquatic resources. The models developed in this project will provide a basis for developing tools for making decisions regarding future land management and sampling/monitoring decisions. For example, these models can be used to assess risk to populations/communities as defined by the regulatory agency (Turner et al. 2004). In addition, Wall et al. (2004) provided recommendations for conservation of the topeka shiner, suggesting areas for new reserves or conservation easements to fill gaps between high-priority topeka shiner habitat and established conservation lands and identified sites where special permits might be needed to make land-use changes (Wall et al. 2004).

## Methods

### Classification of Riverine Ecosystems

We used GIS technology to classify riverine ecosystems at multiple spatial scales according to landscape variables important in determining the distribution of aquatic communities (Lammert et al. 1996, Higgins et al. 1998) based on a hierarchy system (Frissell et al. 1986). The classification system included three ecosystem scales: ecoregions (Bailey 1995), major drainages, and valley segment types. GIS analysis was performed using Arc View 3.2 (Environmental Systems Research Institute 1999), ARC/INFO version 8.2 (Environmental Systems Research Institute 2000), and ARC/INFO AML's (Arc Macro Language) supplied by The Nature Conservancy (The Nature Conservancy Freshwater Initiative 2000).

The state was divided into 19 ecoregional drainage units (Figure 1) according to the South Dakota Gap analysis project (Smith et al. 2002). An ecoregional drainage unit (EDU) is part of a major drainage within an ecoregional province, which Bailey (1995) defines as a large landscape area with uniform biotic and abiotic features. An EDU is considered to have a unique fish community because of geographical isolation by linear stream distance and natural barriers (e.g. waterfalls).

The National Hydrography Database (U.S. Geological Survey and Environmental Protection Agency 2001) at a scale of 1:100,000 formed the base hydrography layer from which valley segment types were delineated. Ten stream condition variables (Table 2)

were used to classify valley segments using GIS data (Smith et al. 2002). Numerical values were assigned to stream-habitat features (Table 2), and the numerical codes were concatenated to form a ten-digit code for each valley segment.

## Modeling

The modeling process included three sections: fish collection data, fish-habitat models, and predicted GIS distributions. Our models associated fish with ten "enduring" spatial habitat features (e.g., stream size, slope, and intermittency) and assumed fish-habitat associations were temporally consistent at this scale. We predicted potential fish distributions using pristine (i.e., no anthropogenic effects) habitat conditions as attributed to valley segments using the ten enduring physical features (McKenna et al. 2006).

### Fish collection data

We gathered fish collection data from several agencies (Table 3) representing watersheds within South Dakota. A large portion of our fish collection data came from the SD Gap Analysis Project (Smith et al. 2002) where data were already configured for application to our project. For other states we collected data from literature, reports, and State agencies (Table 3).

Sample points and associated data were converted to a spatially referenced layer for use in GIS. Locations of points were verified for accuracy in the GIS by comparing site descriptions to geo-referenced layers of: latitudinal and longitudinal position within

the GIS, roads, public land survey system, county, and stream name. Samples where point locations could not be verified or corrected to the assessment scale were not used.

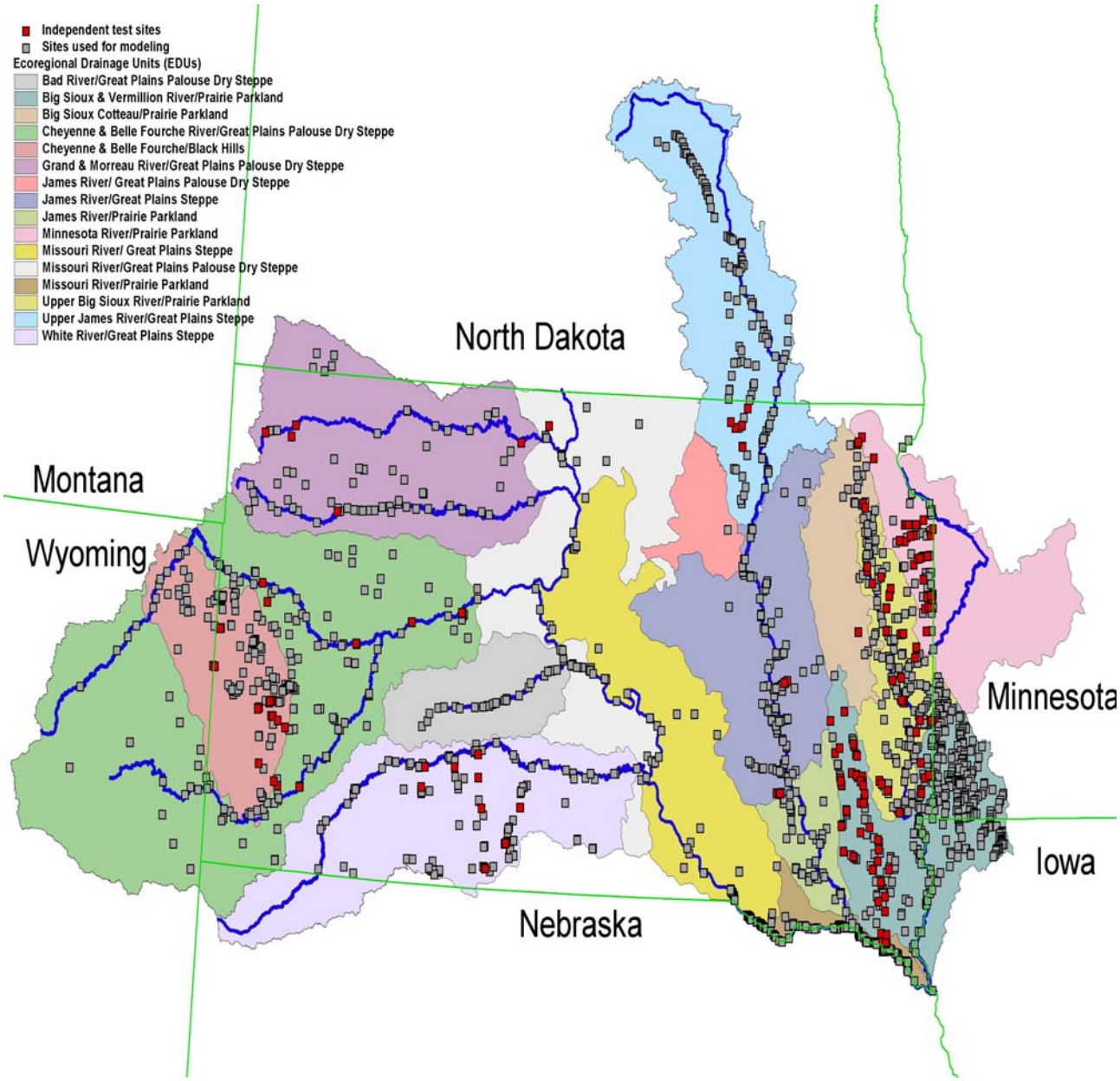


**Figure 1:** Study area showing Ecoregional Drainage Units (EDUs), fish collection sites used in modeling (gray squares), and independent test sites (red squares).

**Table 2:** Valley Segment habitat variables, GIS data source and methods used to delineate the variable

| Habitat variable and numeric code[a] | GIS data source | Method (refer to Smith et al. 2002) |
|---|---|---|
| Stream size class<br><br>1 = headwater (link = 0 to 9)<br>2 = creek (link = 10 to 75)<br>3 = small river (link = 76 to 1500)<br>4 = large river (link > 1500) | National Hydrography Database (NHD) | Shreve link |
| Channel slope class[b]<br><br>1 = low<br>2 = medium<br>3 = high | Digital Elevation Model (DEM)<br>(U.S. Geological Survey, EROS Data Center, Sioux Falls, SD) | elevation change between valley segment t-node and f-node relative to stream size and EDU |
| Network position class<br><br>0 = none (no discrepancy – e.g., headwater to headwater)<br>1 = headwater to creek (link = 0 to 9 and d-link = 10 to 75)<br>2 = headwater to small river (link = 0 to 9 and d-link = 76 to 1500)<br>3 = headwater to large river (link = 0 to 9 and d-link = 10 to 75)<br>4 = creek to small river (link = 10 to 75 and d-link = 76 to 1500)<br>5 = creek to large river (link = 10 to 75and d-link > 1500)<br>6 = small river to large river (link = 76 to 1500 and d-link > 1500)<br>7 = headwater to extra large river (Link = 0 to 9 and d-link > 10000)<br>8 = Creek to extra large river (10 to 75 and d-link > 10000)<br>9 = Small river to extra large river (Link = 76 to 1500 and d-link ><br>10 = large river to extra-large river (Link = 1501 to 10000 and d-link > | National Hydrography Database (NHD) | downstream Shreve link |

**Table 2 continued:** Valley Segment habitat variables, GIS data source and methods used to delineate the variable

| Habitat variable and numeric code[a] | GIS data source | Method (refer to Smith et al. 2002) |
|---|---|---|
| Parent material class<br>1 = alluvium<br>4 = glacial till<br>9 = outwash<br>8 = loess<br>6 = Sioux quartzite<br>14 = shale<br>other (i.e., eolian sands (3), silty glacial till (5), , lacustrine (7), ponded soils (13)) | State Soil Geographic | dominant parent materials by |
| Stream flow<br>1 = perennial<br>2 = intermittent | National Hydrography | classified according to National Hydorgraphy Database (NHD) codes for perennial and intermittent |
| Groundwater potential<br>1 = low (groundwater velocity 0 – 0.35 m/day)<br>2 = medium (groundwater velocity 0.36 – 0.1.5 m/day)<br>3 = high (>0.15 m/day) | DEM and parent material classification | MRI-DARCY model (Baker et al. 2003) – groundwater velocity to segments derived from hydraulic conductivity of parent materials and slope |
| Connectivity to Lake<br>0 = no<br>1 = yes | National Hydrography | valley segments connected to water body polygons > 4 ha |
| Floodplain influence<br>1 = yes<br>2 = no | DEM | slope from DEM to determine valley wall, valley segments > 249 m within valley wall of small-large rivers classed as 1 |
| Elevation[b]<br>1 = low<br>2 = medium<br>3 = high | DEM | maximum elevation of valley |

[a] A hypothetical concatenated code for first three variables is 12030123013, meaning headwater,
[b]Channel slope and elevation classifications varied among regions (i.e. east and west of Missouri River

**Table 3**: Data sources for fish collection sites used in modeling and in the independent dataset (italics), the number of collection sites contributed from each source, and the Ecoregional Drainage Units (EDUs) represented from each source.

| Source | # collection sites | Ecoregional Drainage Unit |
|---|---|---|
| SD Gap Analysis project | 460 | All EDUs |
| U.S. Geological Survey Missouri River Benthic Fish Project | 1057 | Missouri River/Great Plains Steppe, Missouri River/Prairie Parkland |
| ND Game and Fish Department | 97 | Upper James River/Great Plains Steppe, Grand & Moreau Rivers/Great Plains Dry Palouse Steppe |
| WY Game and Fish Department | 47 | Cheyenne & Belle Fourche Rivers/ Black Hills, Cheyenne & Belle Fourche Rivers/ Greta Plains Palouse Dry Steppe |
| MN Department of Natural Resources | 186 | Big Sioux & Vermillion Rivers/Prairie Parkland |
| IA Department of Natural Resources | 76 | Big Sioux & Vermillion Rivers/Prairie Parkland |
| South Dakota State University | 103 | All EDUs except Upper James River/Great Plains Steppe |
| Vermillion River Watershed Development District | 28 | Big Sioux & Vermillion Rivers/Prairie Parkland |
| SD Game, Fish & Parks | 41 | Minnesota River/ Prairie Parkland, Big Sioux & Vermillion Rivers/Prairie Parkland, James River/ Prairie Parkland, James River/Great Plains Steppe |
| Upper Missouri River Aquatic Gap Analysis Project | 5 | Upper James River/Great Plains Steppe |
| SD Department of Transportation | 14 | Big Sioux & Vermillion Rivers/Prairie Parkland |
| US Geological Survey EMAP | 24 | Minnesota River/ Prairie Parkland, Big Sioux & Vermillion Rivers/Prairie Parkland, White River/Great Plains Dry Palouse Steppe, Cheyenne & Belle Fourche Rivers/ Greta Plains Palouse Dry Steppe, Grand & Moreau Rivers/Great Plains Dry Palouse Steppe |

[a]Refer to Smith et al. 2002 for a complete list of all South Dakota contributors.

A total of 2,026 fish sample sites were used for modeling (Figure 1). We spatially linked the fish collection data from each site to the valley segment layer, and to an 11-digit hydrologic unit (HUC) layer (Smith et al. 2002). We consolidated all fish

collections where a site had multiple samples at different times. Fish from all sites within a valley segment were considered as being present in that valley segment. Fish were considered present in a valley segment if it occurred there at any time (i.e., historic samples were used [< 1970]) to maximize sample size for analyses.

**Fish-habitat models**

Once fish collection sites were spatially linked to valley segments, a table that attributed fish presence and absence to valley segment physical features was created for use in modeling. Our modeling approach was to use decision trees similar to the approach used by the Missouri Aquatic Gap Analysis project (Sowa et al. 2005). Decision trees are nonlinear/nonparametric models that use recursive-partitioning algorithms to produce a set of decision rules from mutually exclusive subsets of input variables that best predict the target variable (SPSS 2001). Because decision trees are non-linear, they more closely represent the real or actual distribution of a species (ter Braak and Prentice 1988), and show more agreement with wildlife-habitat relationships (i.e., niche theory [O'Connor 2002]) than do linear modeling techniques. We used AnswerTree ® 3.0 statistical software (SPSS 2001) to execute the decision tree analysis, and applied the exhaustive CHAID (Chi-squared Automated Interaction Detector) algorithm (Biggs et al. 1991), which uses chi-square statistics to identify optimal splits for each predictor. Exhaustive CHAID was found to be the most effective algorithm to predict fish species occurrence offered in AnswerTree® by the Missouri Aquatic Gap

Analysis project, and was easily interpreted (Sowa et al. 2005). Exhaustive CHAID compares independent variables using Pearson chi-square statistics to determine the one that provides the best prediction of the dependant variable by splitting the data into subgroups (SPSS 2001). The process is then applied recursively to subgroups to define sub-subgroups, and so on, until the tree is finished by a defined stopping criteria. The result is a dendrogram showing a set of decision rules identifying predictor variable combinations (valley segment habitat features) that are significantly associated with the dependent variable (fish species).

We modeled fish species separately according to three faunal regions to account for regional variations in species and habitat associations: 1) Missouri River Basin east of the Missouri River, 2) Missouri River Basin west of, and including, the Missouri River, and 3) the Minnesota River Basin. Presence and absence samples were chosen from 8-digit HUC's within the species range. We used a combination of variables that differed among regions. For example, we used temperature and elevation in the Black Hills where these features vary significantly, but not in the plains, as these features showed little variation in relation to fish distribution. We used the regional gradient and elevation variables for analysis (Appendix 2.1-2.4), except when a species distribution was restricted to a major drainage, where upon we used gradient relative to that drainage.

The variable for stream size was separated into smaller categories to increase variability and potentially, model accuracy. Essentially, we divided the stream size classifications in half to create the variable strmsize10 (Appendix 2.5). For example a

valley segment classified as a creek (stream size = 2, link = 10 to 75) was divided into two stream size classes (stream size10 = 3, link = 10 to 37; and stream size10 = 4, link = 38 to 75). We minimized the use of collinear independent variables (i.e., only one attribute representing stream size: stream size, stream size10 and stream order was used).

We selected variables to produce models that correctly predicted the most known presence sites, as we thought that it was better to err on the side of commission rather than omission. This selection would reduce the chance of errors of omission when predicting rare species of concern, and increase errors of commission. We considered it better to over predict the occurrence of species of concern rather than miss potential conservation opportunities (Olden and Jackson 2001). However, model predictions were restricted to known species ranges to the 8-digit HUC level or to the 11-digit HUC level for relict species occurring in isolated watersheds.

We used modeling criteria for AnswerTree ® 3.0 as was used by Sowa et al. (2005). We set the stopping criteria for the tree size to 10 levels, and minimum number of collection records that occur in a child node = 1. We used habitat affinities in literature to assess the validity of child nodes with low numbers of collections. We set the alpha level for splitting and merging equal to 0.10 and used the Bonferoni alpha adjustment to account for the increased likelihood of a type one error associated with multiple comparisons. An alpha level of 0.10 was found by Sowa et al. (2005) to produce optimal models. For models that did not produce an output using alpha = 0.10 (e.g., small sample size), we increased alpha in increments of 0.01 up to 0.20 to produce

an output. If we could not obtain an output from AnswerTree®, we used valley segment affinities and habitat affinities found in literature that matched the species occurrence to predict valley segments for fish presence similar to the approach used in South Dakota Gap Analysis for aquatic systems (Smith et al. 2002). For species that were recorded only once in a faunal region, we considered as present where they occurred and did not predict potential valley segments.

We used the "relative 50%-approach" to prune decision tree outputs (Sowa et al. 2005). For each model, we first identified the node with the highest occurrence percentage that also contained at least 4% of all the collection records from the overall input dataset. For example, if the input dataset contained 1,000 total collection records, we would identify the highest occurrence percentage for those nodes having 40 or more total collection records, to account for the fact that terminal nodes with < 4% of samples generally provided grossly inflated or deflated occurrence percentages. We then divided the highest occurrence percentage by 2 and selected all nodes having occurrence percentages greater than or equal to this percentage. For example, if the highest occurrence percentage was 80% we would select all nodes with occurrence percentages $\geq$ 40% and if the highest occurrence percentage is 50% we would select all nodes with occurrence percentages $\geq$ 25%. This proved to be an efficient and standardized approach that accounted for differences in species prevalence or commonness, which were not accounted for by the other model evaluation tools included in AnswerTree® 3.0.

**Predicted GIS distributions**

Decision rules from the model output for each fish species were entered into a GIS query to select valley segment types containing habitat features associated with the occurrence of the species. Fish ranges by 11-digit HUC were coded as: 1 = fish was recorded in the HUC and 2 = fish was predicted in HUC using valley segment species prediction codes. HUC distribution codes were defaulted to present (i.e., fish recorded in HUC) if there was an occurrence recorded in the HUC regardless of predictions.

**Field methods: Independent sampling**

Gap distribution maps were used to locate potential sites where species at risk were predicted to occur at the reach level. Fieldwork occurred during summer 2004 (June-September) and summer 2005 (July-September) for a total of 40 sites (Table 4; Appendix 3), which focused on the Big Sioux River and Black Hills drainages (Figure 1). We focused on these two areas to increase sampling effort in areas where several rare species were predicted, rather that sample statewide with less effort because of logistics. Fish were collected at each site by seining (seine length was not constant due to different habitat, but varied between 15-50 meters) and/or backpack electrofishing with a Smith-Root LR-24 electrofisher (depending on available habitat and flow conditions) were identified and released. We included in the analysis an additional 103 sites that were visited by other researchers conducting projects unrelated to ours but conducted during

the same time (Table 4, Appendix 4). We did some of the projects (e.g., EMAP) and some were done by others (e.g., Vermillion River Basin TMDL).

We recorded the presence and absence of fishes sampled. Presence/absence data has been used as an efficient sampling technique to examine environmental factors related to species distribution and community structure (Burnham and Overton 1979; Mingoti and Meeden 1992; Peterson and Rabeni 1995; Royle and Nichols 2003; Gu and Swihart 2004) and has been used for a variety of monitoring programs (Mingoti and Meeden 1992; Conway and Simon 2003; Peterson and Dunham 2003; Royle and Nichols 2003; Selmi and Bouliner 2003; Stauffer et al. 2003) being especially useful for evaluating rare species (Chao et al. 1992; Rushton et al. 2004). Presence/absence surveys allow for more time to sample additional sites and for the computation of detection probabilities. We used presence/absence data to evaluate model accuracy and also to calculate detection probabilities.

Detection probabilities were calculated for all species collected in the independent dataset based on maximum likelihood methods described by Mackenzie et al. (2002) and recently used by others (Turner et al. 2004; Wintle et al. 2004; Hayer 2005). Low detection rates of fish species and especially rare species effects model accuracy (Tyre et al. 2003) and have been suggested to be included as an error estimate (Hoeting et al. 2000) to improve model performance and minimize bias in parameter estimates and associated uncertainty (Mackenzie et al. 2002; Martin et al. 2005).

**Table 4:** List of project/data sources used in independent data set validating accuracy of fish species model predictions from 2004-2005**.**

| Sampling Source | # sites |
|---|---|
| SD Rare Fish project | 40 |
| SD GF&P Minnesota River Basin Survey | 17 |
| EMAP | 22 |
| Upper Missouri River Basin Aquatic Gap Project | 5 |
| SDSU SD Stream Project | 5 |
| SD DOT Project | 14 |
| SD GF&P Topeka shiner monitoring project | 24 |
| Vermillion Basin Water Development District | 14 |
| USGS picked | 2 |
| **Total Sites Visited** | **143** |

## Model Accuracy Assessment

Model accuracy was evaluated for ecological drainage units (EDUs) where independent data test sites occurred (Figure 1). Models were tested separately for each EDU because species prevalence varies by EDU, which affects model accuracy. For example, fathead minnows are very abundant in streams east of the Missouri River, whereas the species is rare in the Black Hills. Also, species composition varies by EDU. For example, longnose dace occur west of the Missouri River but not east of the Missouri River, where the western blacknose dace becomes the dominant *Rhinichthys* species. EDUs east of the Missouri River were combined into one test region (denoted as East river EDU) as these EDUs have similar species and habitat characteristics (Wall et al. 2004). For species with restricted distributions, such as the northern redbelly dace (*Phoxinus eos*) and blackside darter (*Percina maculata*), only test sites within the species respective distributions were used.

Fish collection sites from the independent dataset were spatially linked to the valley segment layer. Models were assessed at the valley segment scale, thus fish from all sites within a valley segment were considered as being found in that valley segment. Valley segments were coded into four classes: A = true presence (i.e. fish was both predicted to occur in and captured in the valley segment); B = error of commission (i.e. fish was predicted to occur in the valley segment, but was not detected); C = error of omission (i.e. fish was captured in the valley segment, but was not predicted to occur there); and D = true absence (i.e. fish was not captured, nor predicted to occur in the valley segment). These classes were entered into a 2x2 confusion matrix (Table 5).The matrix was used to calculate the following accuracy metrics (Table 6) according to Fielding and Bell (1997) for each fish species sampled in the independent dataset. Some metrics cannot be calculated if A, B, C, or D = 0 (e.g. odds-ratio).

The number of true presences, true absences, omission errors, commission errors, correct classification rate, sensitivity, specificity, and Cohen's Kappa are the only accuracy metrics discussed in the results and discussion of this report, although others were calculated and may provide additional insight into model performance (Table 6). Prevalence (proportion of sites occupied) and detection probabilities (proportion of sites where the species was detected after taking into account presence or absence) were also reported for the independent dataset.

**Table 5:** Two by two confusion matrix used to calculate model accuracy metrics

| | | Actual Survey | | |
|---|---|---|---|---|
| | | Present | Absent | |
| Model Prediction | Present | A | B | |
| | Absent | C | D | |
| | | | | N |

**Table 6:** Possible measures for assessing the accuracy of prediction models. The formulae are applied to assessments of correctly predicted positive occurrences (a), falsely predicted positive occurrences (b), falsely predicted negative occurrences (c), and correctly predicted negative cases (d). N is the overall number of cases.

| **Performance Measure** | **Definition** | **Calculation** |
|---|---|---|
| Prevalence | Percentage of sites occupied by species | (a + c) / N |
| Overall diagnostic power | Percentage of true and false negatives | (b + d) / N |
| Correct classification rate | Percentage of all cases correctly predicted | (a + d )/ N |
| Sensitivity | Percentage of true positives correctly predicted | a / (a + c) |
| Specificity | Percentage of true negatives correctly predicted | d / (b + d) |
| False positive rate | Percentage of true negatives incorrectly predicted | b / (b + d) |
| False negative rate | Percentage of true positives incorrectly predicted | c / (a + c) |
| Positive predictive power | Percentage of predicted presences that were real | a / (a + b) |
| Negative predictive power | Percentage of predicted absences that were real | d / (c + d) |
| Misclassification rate | Percentage of all cases incorrectly predicted | (b + c) / N |
| Kappa | Proportion of specific agreement | [(a + d ) - (((a + c) (a + b) + (b + d) (c + d)) / N] / [N - (((a + c) (a + b) + (b + d) (c + d)) / N)] |

We used Cohen's kappa (k) as the main indicator of model accuracy (Fielding and Bell 1997, Manel et al.2001). Kappa is the ratio of the proportion of agreement (corrected for chance) divided by the maximum number of times predictions could agree with observations (corrected for chance). Kappa measures model agreement with what is observed, rather than model association with what is observed (Fielding and Bell 1997). Kappa values of 0.0-0.4 are considered to indicate slight to fair model performance, values of 0.4-0.6 moderate, 0.6-0.8 substantial and 0.8-1.0 almost perfect (Manel et al. 2001). Negative values are possible and indicate poor model agreement (Manel et al. 2001). Models that classify everything wrong can have a high value of association with the independent dataset, but no agreement. In addition, overall accuracy is affected by species prevalence and detection, and it is possible to achieve high accuracy by chance when prevalence is high or low. For example, a species with 5% prevalence and a model that classifies everything as negative as prediction rule, regardless of associated habitats, would achieve 95% accuracy. Thus a measure of accuracy such as k, which measures the proportion of specific agreement and accounts for chance, is appropriate for testing model accuracy and is less affected by prevalence (Manel et al. 2001; Fielding and Bell 1997). The value calculated by k is the probability above that of chance that the model predicts the presence or absence of a species in a valley segment. For example, k = 0.43 means the model can correctly predict the occurrence of a species in a valley segment 43% more often than by randomly choosing a valley segment.

We used detection probabilities as a means to minimize the false negative error rate (i.e., commission error = the model predicted presence, but actual field surveys did not collect the species). Cell B in the confusion matrix (Table 5) was recalculated by multiplying the actual number of false negatives from the field surveys by the detection probability of the species. Accuracy measures were recalculated using the weighted value for commission error (Cell B). Only the weighted values for accuracy measures were reported unless otherwise specified. To increase predictive inference, in addition to species of concern, all other species predicted to occur within the sampling area was analyzed for model accuracy.

## Results

Sixty-two species were collected from 143 independent sampling sites across six ecoregional drainage units in South Dakota (Figure 2). East River sampling effort far exceeded the other five EDUS (62 sites). Model accuracy for all rare species was moderate to substantial indicating the models performed well (Table 7). The central mudminnow and northern redbelly dace were collected in the East River EDU and the Minnesota River EDU and exhibited low prevalence (Appendix 5). The trout-perch was collected in East river only. The hornyhead chub was collected in Minnesota EDU and exhibited high prevalence and detection probability. The sturgeon chub was captured in the Cheyenne River EDU. The most prevalent (P) fishes were the fathead minnow in East River and Grand Moreau (P= 0.943 and 1.00 respectively), largemouth bass in

Minnesota River EDU (P=1.000), longnose dace in Black Hills (P=0.800), white sucker in Cheyenne River (P=0.857), and sand shiner and fathead chub in white river (P= 0.889; Appendix 5).

Model accuracy results are presented for each EDU sampled. Each section contains model accuracy results for all species followed by results for specific rare species collected and/or predicted within the EDU.

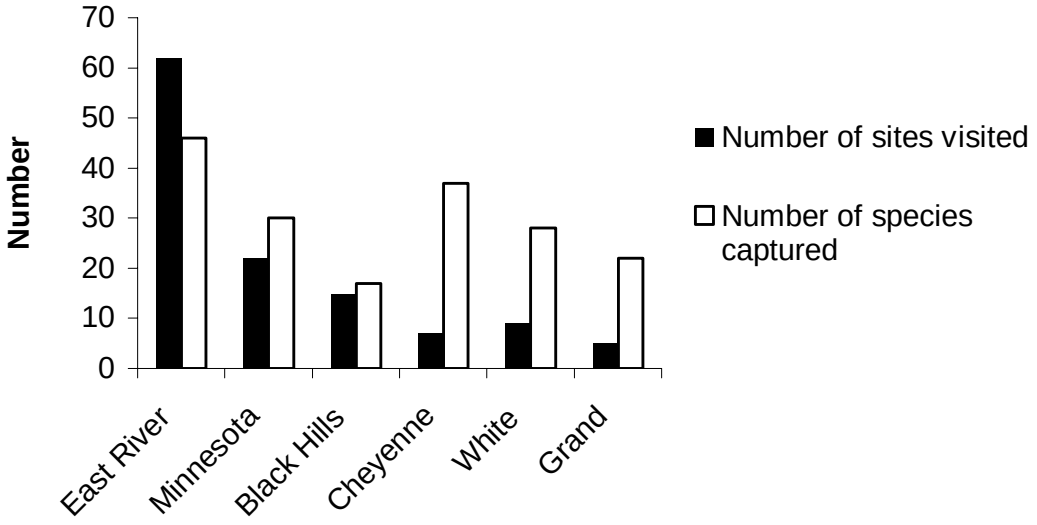


**Figure 2:** The number of new sites visited (independent data) and the number of species collected per ecological drainage unit (EDU).

**Table 7.** South Dakota fish species at risk kappa statistics by ecological drainage unit. The x's indicate that the species was not predicted within that EDU.

| Species Name | East River | Minnesota | Black Hills | Cheyenne | Grand Moreau | White |
|---|---|---|---|---|---|---|
| Banded killifish | 0.488 | 0.488 | x | x | x | x |
| Central Mudminnow | 0.545 | 0.364 | x | x | x | x |
| Northern redbelly dace | 0.605 | 0.476 | x | x | 0.444 | 0.400 |
| Finescale dace | x | x | 0.444 | x | x | x |
| Hornyhead chub | x | 0.441 | x | x | x | x |
| Pearl dace | x | x | x | x | x | 0.400 |
| Carmine shiner | x | 0.448 | x | x | x | x |
| Sturgeon chub | x | x | x | 0.561 | x | x |
| Longnose sucker | x | x | 0.318 | x | x | x |
| Trout-perch | 0.424 | x | x | x | x | x |

<u>East River EDU</u>

*All species*

We collected 44 of 46 species predicted to occur at 62 sites within this EDU (Appendix 5). Prevalence ranged from 0.000 (banded killifish and plains topminnow) to 0.943 (fathead minnow). The black bullhead, common shiner, common carp, creek chub, fathead minnow, johnny darter, sand shiner, and white sucker were found in over 60% of samples (Appendix 6). Detection probabilities were also variable ranging from 0.000 (banded killifish plains topminnow, and golden shiner) to 0.316 (fathead minnow; Appendix 6). Correct classification rates varied from 0.185 (walleye) to 0.986 (golden shiner; Appendix 6) and 15 species had rates exceeding 75% indicating acceptable model performance (Hurley 1986). Correct classification rates varied curvi-linearly and

systematically with prevalence (Figure 3) indicating that it is easier to correctly predict presences or absences of rarer species and abundant species. Correctly predicted presences (sensitivity) had a positive relationship with prevalence whereas correctly predicted absences (specificity) had a negative relationship (Figure 3). Correctly predicted presences (0.831 ± 0.040) were generally higher than correctly predicted absences (0.772 ± 0.045).

Including detection probabilities substantially improved model performance for almost all species (Figure 4). The mean kappa value (± SE) before adjusting with detection probability was 0.229 ± 0.038 and ranged between -0.127 and 1.00 (common carp and smallmouth buffalo, respectively). Kappa values adjusted with detection probability ranged from -0.023 (smallmouth buffalo) to 0.82 (gizzard shad) and the mean was 0.42 ± 0.026. Ninety-six percent of species models were better than chance at predicting presence or absence of a species (Figure 4) within the East River EDU and 65% of models had moderate, substantial, or almost perfect model agreement (Figure 4) according to the kappa statistic.

*Rare Species*

*Central Mudminnow*

The central mudminnow was predicted to occur at seven sites but was only collected at one site resulting (Figures 5 and 6) in six commission errors. Predictive modeling associated presence of central mudminnow with creeks and headwaters

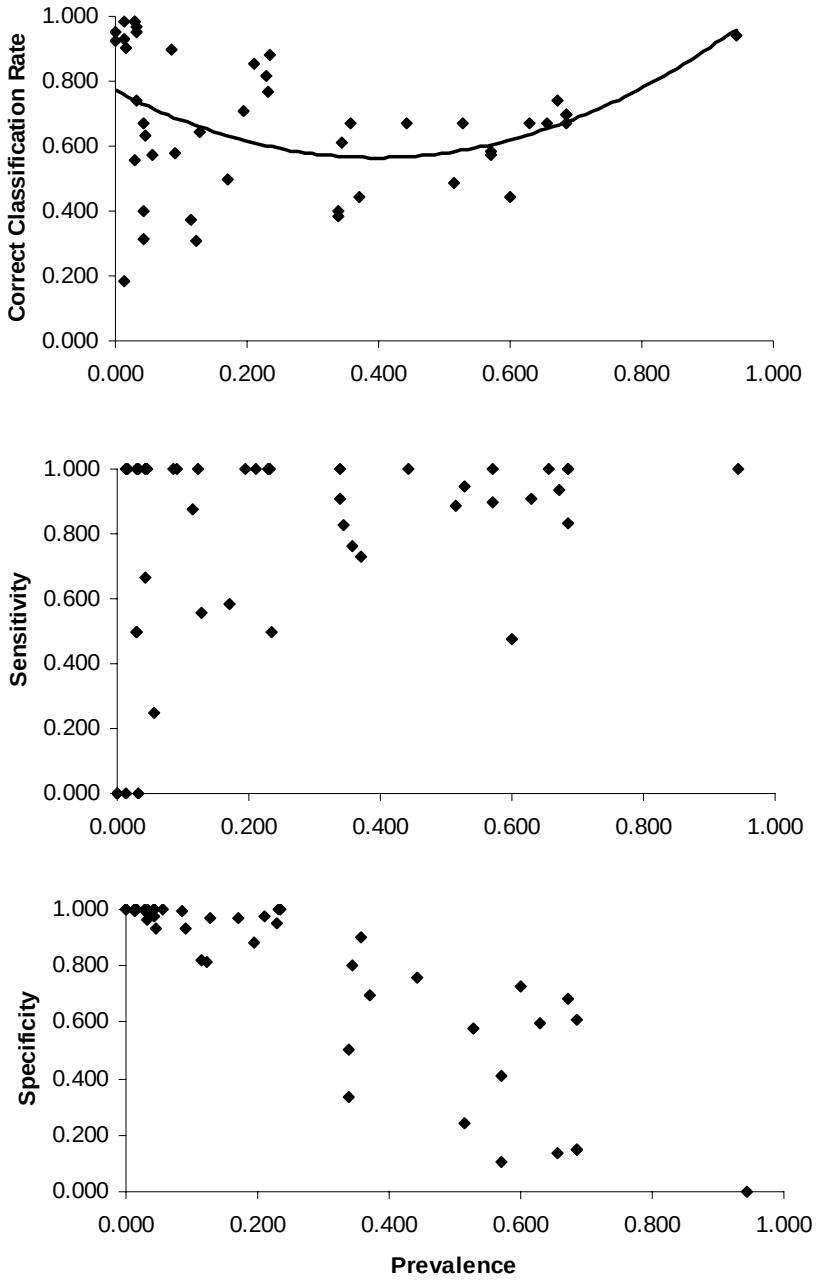


**Figure 3:** Correct classification rate (top), specificity (middle), and sensitivity (bottom) of presence-absence models for 46 species in relation to their occurrence in 62 streams in the East River ecoregional drainage unit.

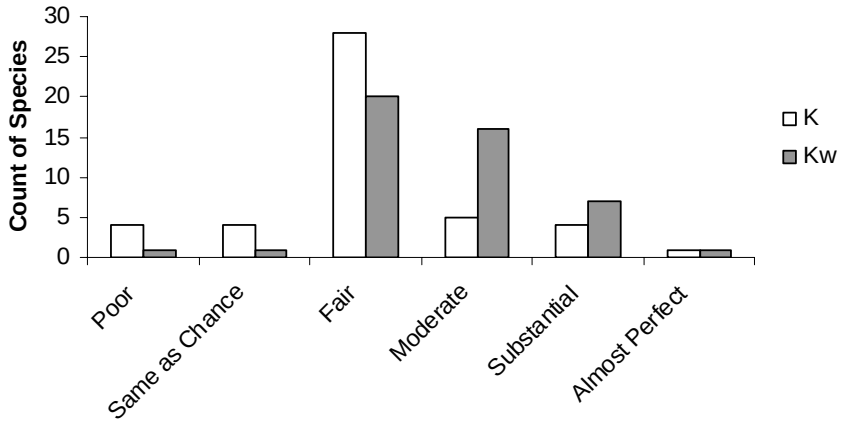


**Figure 4:** Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in East River ecoregional drainage unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

**Table 8:** East River ecoregional drainage unit (EDU) rare fish species accuracy metrics from independent data set. Metrics are weighted with detection probability. Species include P_BAK (banded killifish), P_CMB (Central mudminnow), P_NRD (Northern redbelly dace), and P_TRP (trout-perch). A (.) indicates that the accuracy metric could not be calculated

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_BAK | 0 | 3 | 0 | 59 | 0.000 | 0.000 | 0.952 | 0.952 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.488 |
| P_CEM | 1 | 6 | 0 | 55 | 0.001 | 0.016 | 0.887 | 0.903 | 1.000 | 1.000 | 0.000 | 0.000 | 0.992 | 1.000 | 0.000 | 0.545 |
| P_NRD | 2 | 0 | 2 | 13 | 0.213 | 0.235 | 0.765 | 0.882 | 0.500 | 1.000 | 0.000 | 0.500 | 1.000 | 0.867 | 0.118 | 0.605 |
| P_TRP | 3 | 14 | 0 | 16 | 0.084 | 0.091 | 0.521 | 0.576 | 1.000 | 0.931 | 0.069 | 0.000 | 0.718 | 1.000 | 0.036 | 0.424 |

connected to a lake or slough (Appendix 7). Detection probability for the central mudminnow was 0.001 and prevalence was also low (0.016). Modeling correctly predicted 90.3% of all cases corresponding to correctly predicting 100% of true negatives (specificity; Table 8) and positives (sensitivity). Model performance was moderate with a kappa value of 0.545 (Table 8) indicating that the model was able to predict the occurrence of the central mudminnow in a valley segment 54.5% more often than by randomly choosing a valley segment.

*Northern redbelly dace*

The northern redbelly dace was collected at two sites predicted by models (Peg Murky Run; Figures 7 and 8) and at two sites that were not predicted by models (Peg Murky Run and Stray Horse Creek). Habitat modeling predicted this species to occur in segments adjacent to valley segments where northern redbelly dace was recorded (Appendix 7). Detection probability for the northern redbelly dace was 0.213 and prevalence was moderate (0.235). Modeling correctly predicted 88.2% of all cases corresponding to correctly predicting 100% of true negatives (specificity; Table 8) and only 50% of true positives (sensitivity). Model performance was substantial with the highest kappa value of all the rare fishes across all EDUs (k = 0.605; Table 8).

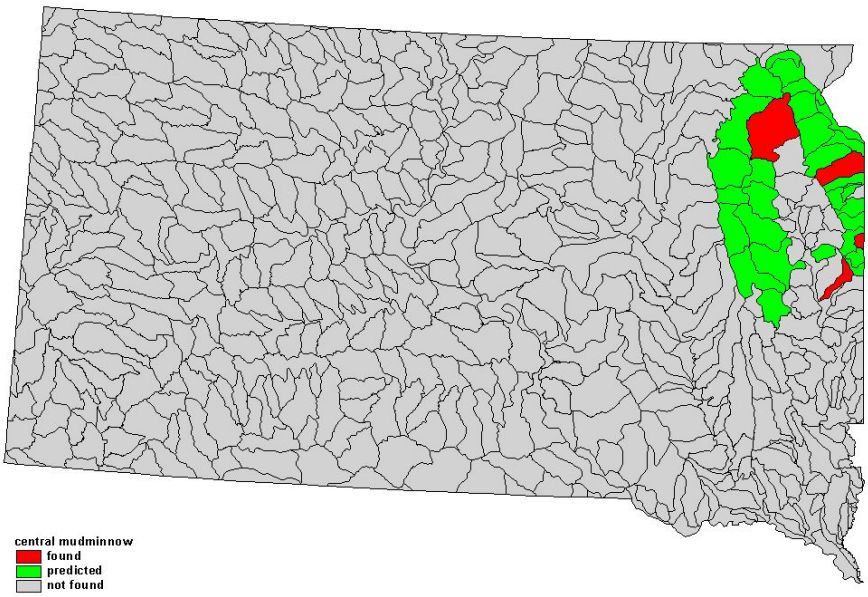


**Figure 5:** Central mudminnow (*Umbra limi*) potential 11-digit hydrological unit distribution from model results. Does not include independent dataset.

*Trout-perch*

The trout-perch was collected at three of 17 predicted sites (Big Sioux River, Medary Creek, and West Pipestone Creek; Figures 9 and 11) resulting in 14 commission errors. Predictive modeling associated presence of trout-perch with 4$^{th}$ and 5$^{th}$ order streams within the Big Sioux River Basin (Appendix 7). Detection probability for trout-perch was 0.084 and prevalence was also low (0.091). The model correctly predicted the presence or absence of trout-perch for 58% of sites, correctly predicting species presence 100% of the time and correctly predicting species absence 93% of the time (Table 8) after

taking detection into account. Model performance was moderate with a kappa value of 0.424 (Table 8).

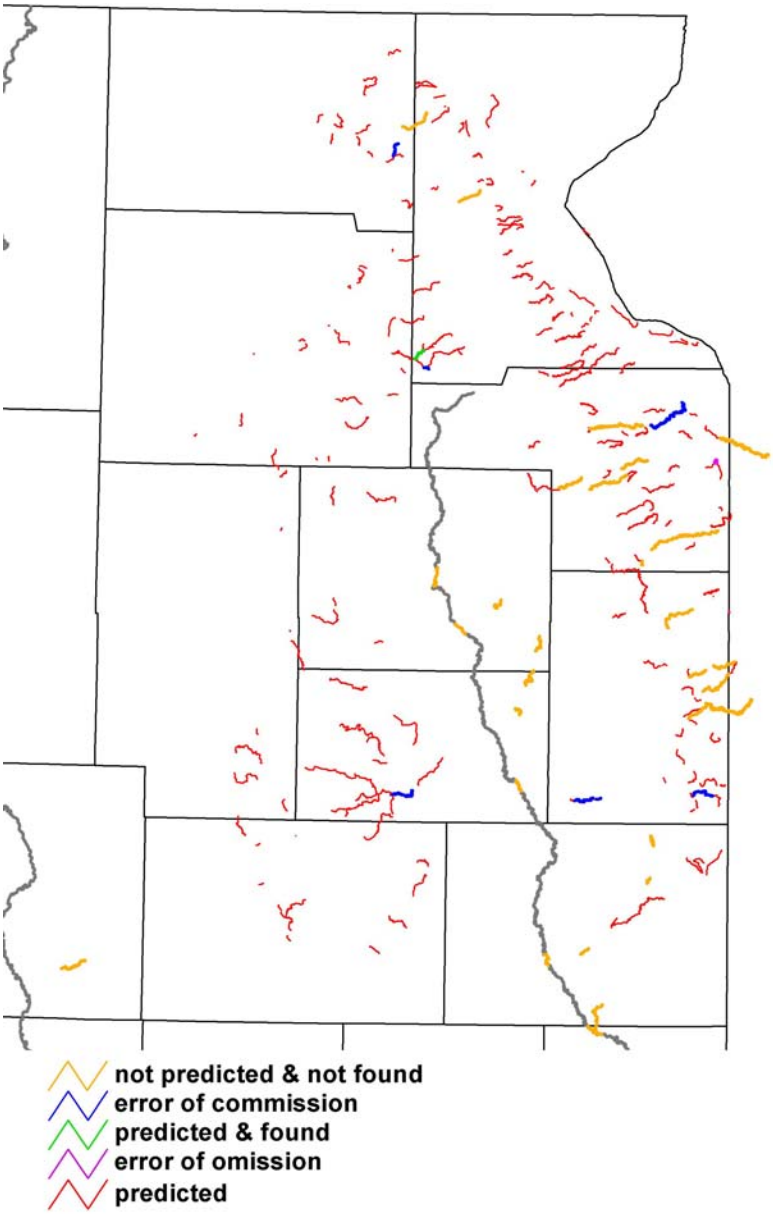


**Figure 6:** Central mudminnow (*Umbra limi*) predicted valley segments (red colored streams) and model test results.

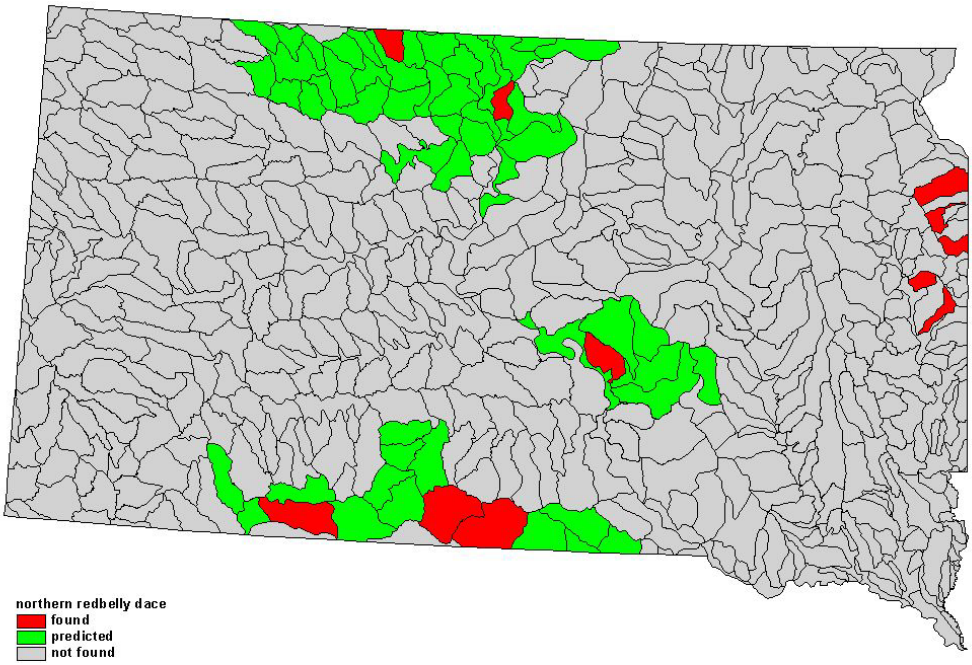


**Figure 7:** Northern redbelly dace (*Phoxinus eos*) 11-digit hydrological unit distribution from model results. Does not include independent dataset.

*Banded killifish*

The banded killifish was predicted to occur at three sites but was not captured in the survey (Figures 10 and 12) resulting in three commission errors. Predictive modeling indicated that presence of banded killifish was associated with stream segments connected to glaciated lakes or sloughs on the Prairie Cotteau (Appendix 7). The model correctly predicted the presence or absence of the species for 95% of sites surveyed. Even though the species was not collected, model performance was moderate (kappa = 0.488; Table 8).

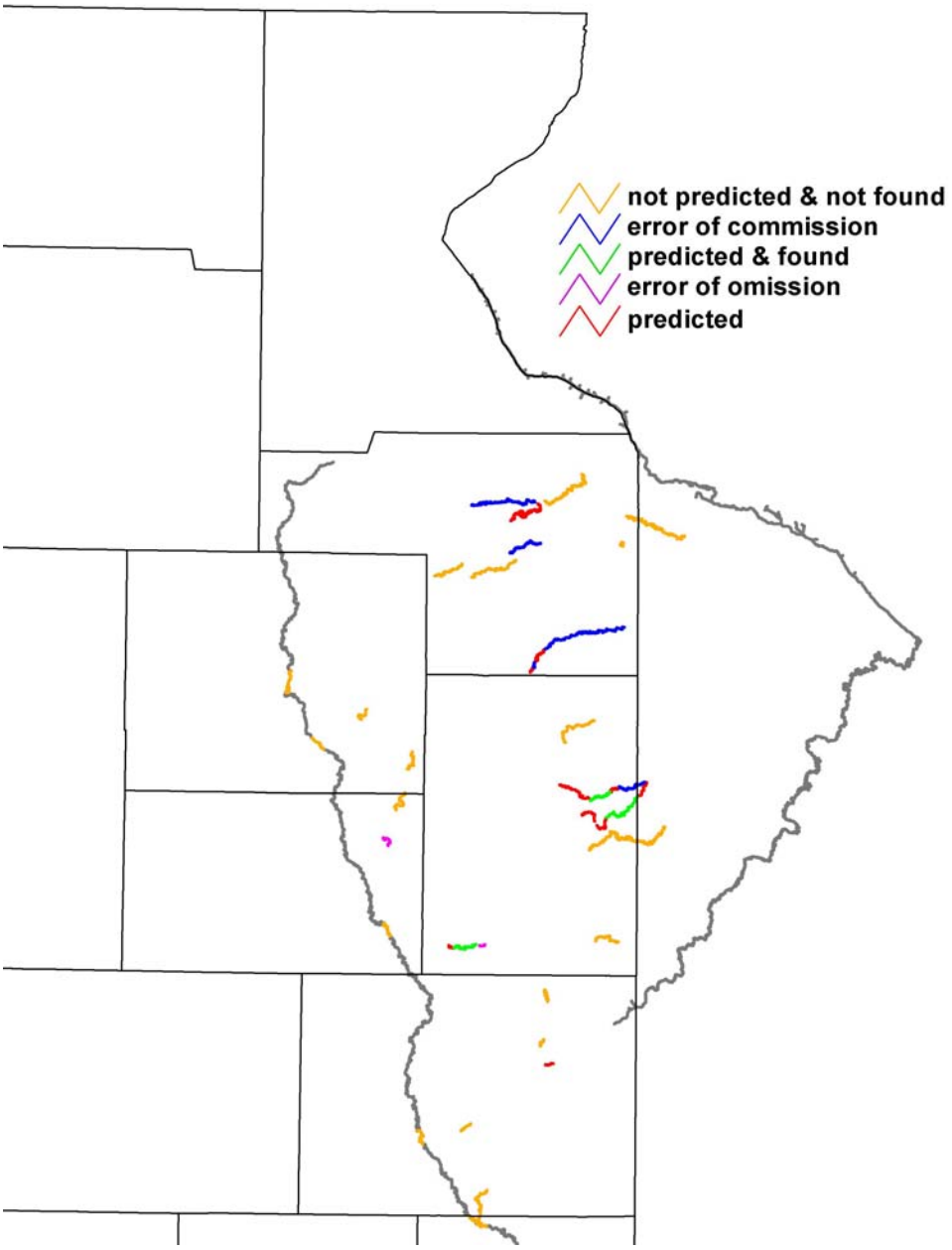


**Figure 8:** Northern redbelly dace (*Phoxinus eos*) predicted valley segments (red colored streams) and model test results for the Big Sioux and Minnesota River Basins.

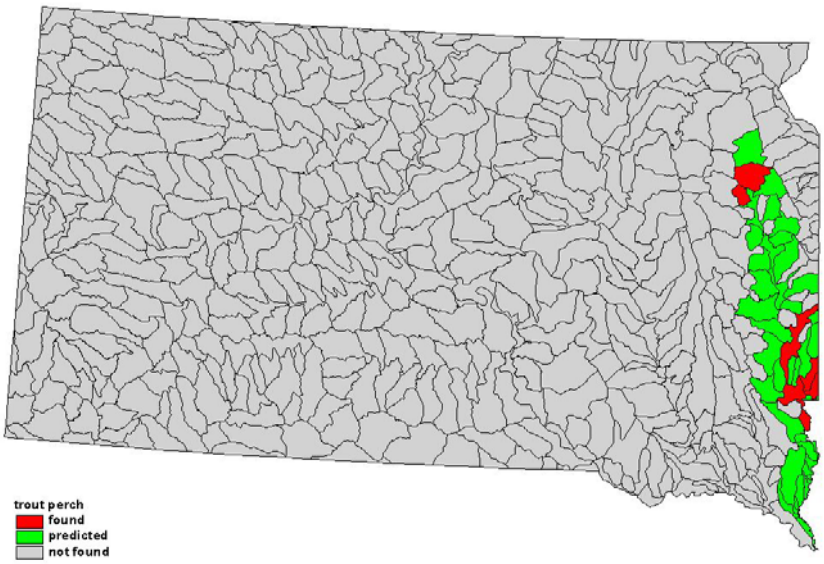


**Figure 9:** Trout perch (*Percopsis Omiscomaycus*) potential 11-digit hydrological unit distribution from model results. Does not include independent dataset.

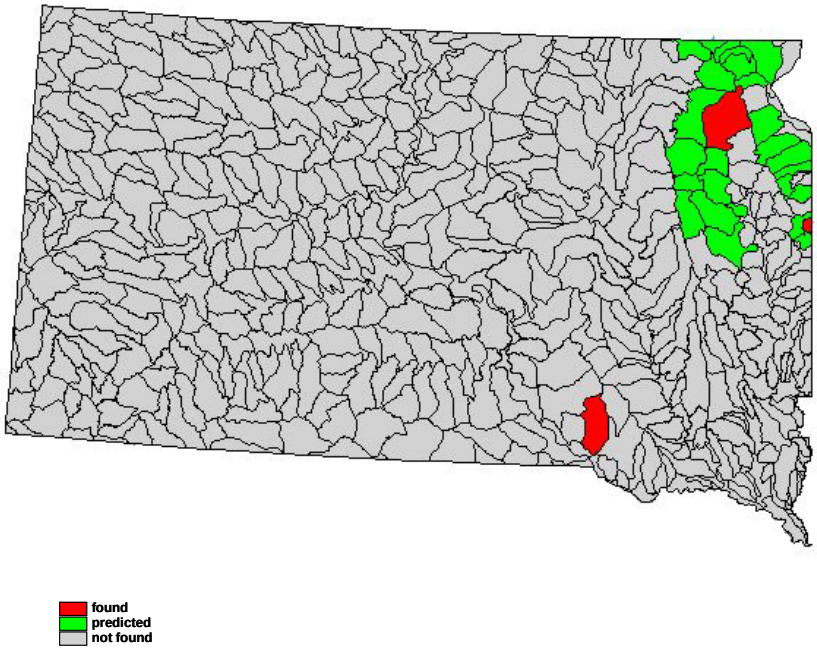


**Figure 10:** Banded killifish (*Fundulus diaphanus*) potential 11-digit hydrological unit distribution from model results. Does not include independent dataset.

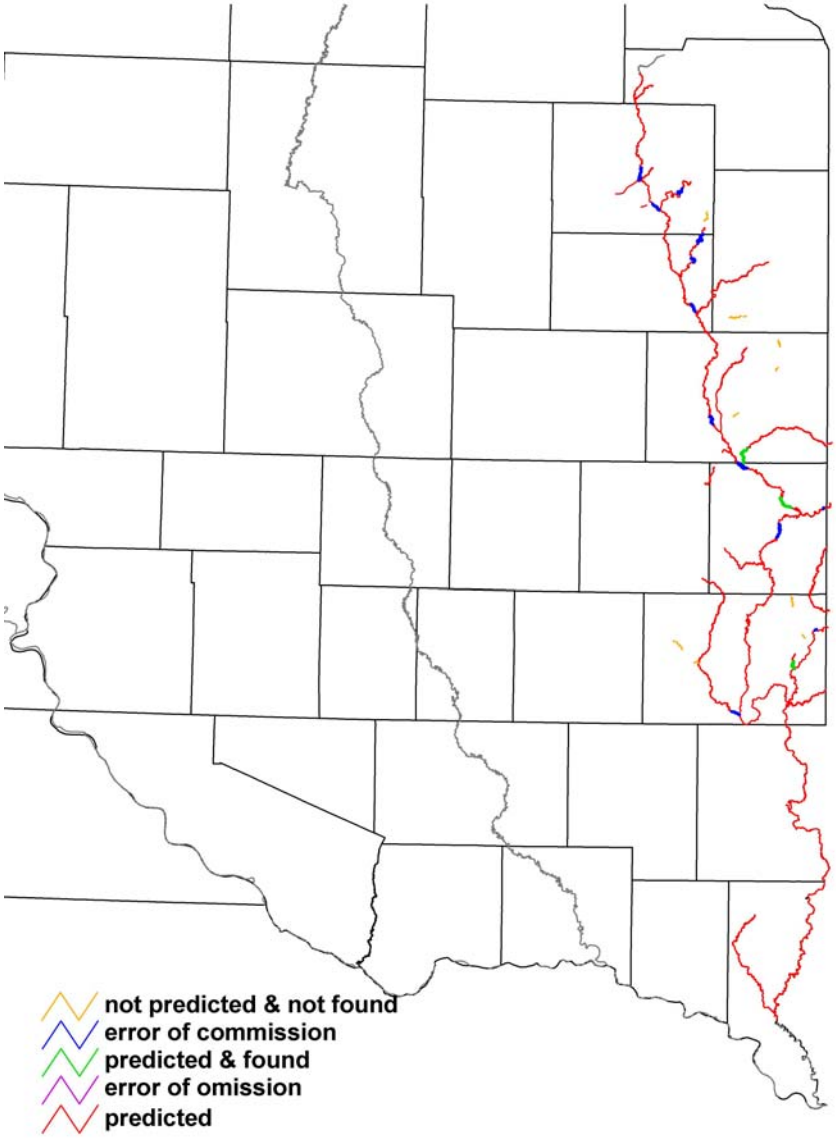


**Figure 11:** Trout perch (Percopsis Omiscomaycus) potential 11-digit hydrological unit distribution from model results. Does not include independent dataset.

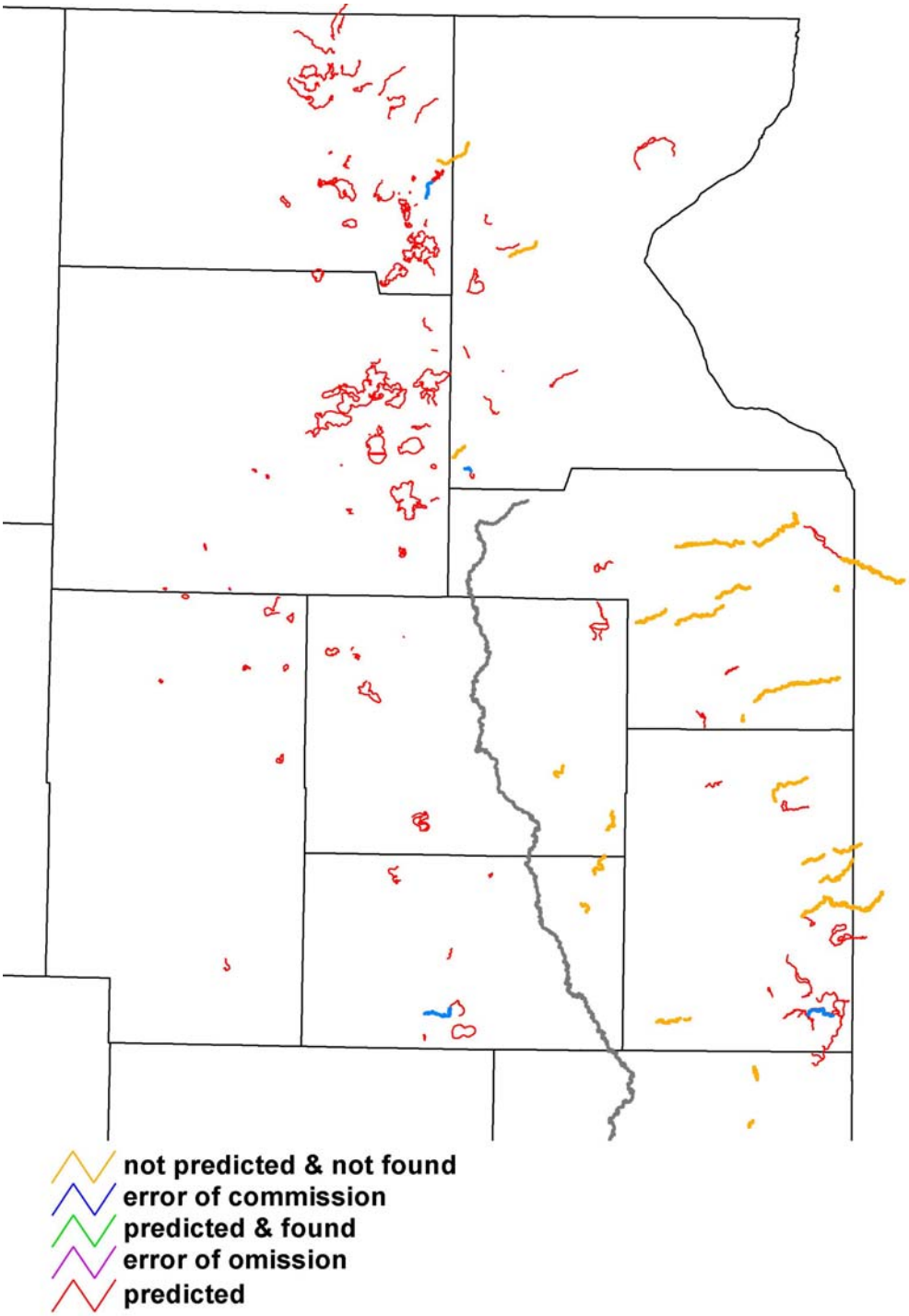


**Figure 12:** Banded killifish (*Fundulus diaphanus*) predicted valley segments (red colored streams) and model test results.

## Minnesota River Results

*All species*

We collected 27 of 30 species predicted to occur at 22 sites within this EDU (Appendix 5). Species occurrence ranged from 0.000 (banded killifish, largemouth bass, and walleye) to 0.818 (blacknose dace). The blacknose dace, common shiner, creek chub, fathead minnow, johnny darter, largemouth bass, central stoneroller, and white sucker were found in over 60% of sampling locations. Detection probabilities were also variable (Appendix 8) ranging from 0.000 (banded killifish, largemouth bass, and walleye) to 0.747 (johnny darter). Correct classification rates ranged from 0.136 (sand shiner) to 1.000 (brown trout; Appendix 8) and eight species had rates exceeding 75% indicating acceptable model performance. Correct classification rates varied curvi-linearly and systematically with prevalence (Figure 13) indicating that it is easier to correctly predict occurrence of rare species and abundant species. Correctly predicted presences (sensitivity) had a positive relationship with prevalence whereas correctly predicted absences (specificity) had a negative relationship (Figure 13). Correctly predicted presence (0.851 ± 0.051) was generally higher than correctly predicted absence (0.679 ± 0.052).

Including detection probabilities substantially improved model performance for almost all species (Figure 14). The mean kappa value (± SE) before adjusting with detection probability was 0.199 ± 0.045 and ranged between -0.091 and 1.00 (northern pike and brown trout, respectively). Kappa values that were adjusted with detection

probability ranged from 0.127 (sand shiner) to 1.000 (brown trout) and the mean was 0.406 ± 0.031. One hundred percent of species models were better than chance at predicting presence or absence of a species (Figure 14) within the Minnesota River EDU and 43 % of models had moderate, substantial, or almost perfect model agreement (Figure 14) according to the kappa statistic.

*Rare species*

*Central mudminnow*

The central mudminnow was not collected at two predicted sites, (Figures 5 and 6) resulting in two commission errors, and collected in the Yellow Bank River where it was not predicted (omission error; Table 9). Predictive modeling related presence of central mudminnow to creeks and headwaters with perennial flow, creeks and headwaters with intermittent flow and connected to a lake or slough, or creeks and headwaters with intermittent flow and groundwater potential (Appendix 7). Detection probability for the central mudminnow was 0.004 and prevalence was also low (0.045). Modeling correctly predicted 0.864% of all cases indicating acceptable model performance, corresponding to correctly predicting 100% of true absences (specificity; Table 9) and never predicting true presences. Model performance however was fair (kappa = 0.364; Table 9) indicating that the model was able to predict the occurrence of the central mudminnow 36.4% more often than by randomly choosing a valley segment.

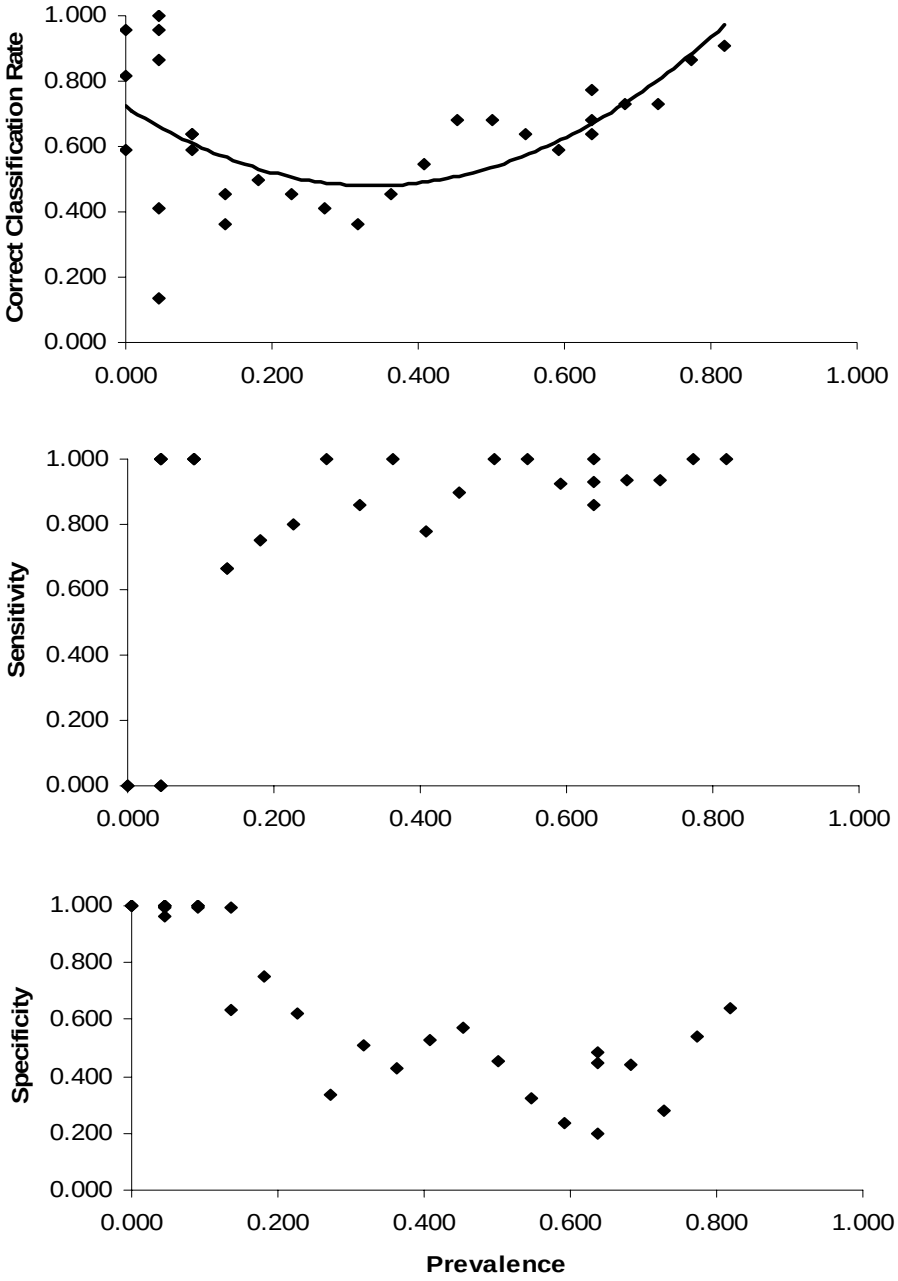


**Figure 13:** Correct classification rate (top), sensitivity (middle) and specificity (bottom) of presence-absence models for 30 species in relation to their occurrence in 22 streams in the Minnesota ecological drainage unit.

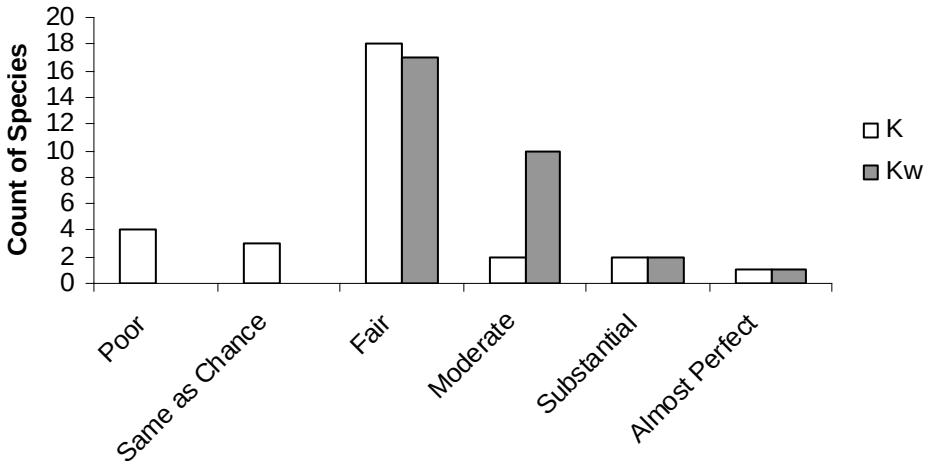


**Figure 14:** Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Minnesota River Ecological Drainage Unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

<u>*Rare Species*</u>

*Northern redbelly dace*

The northern redbelly dace was collected at two of 10 predicted sites (La Qui Parle River and Monighan Creek; Figures 7 and 8) resulting in eight commission errors (Table 9). Headwaters and creeks with perennial flow or headwaters and creeks with intermittent flow and medium to high groundwater potential were habitats where modeling predicted presence of northern redbelly dace (Appendix 7). Detection probability for the northern red belly dace was 0.004 and prevalence was also low (0.091). Modeling correctly predicted 63.6% of all cases indicating moderate model

performance, corresponding to correctly predicting almost 100% of true absences (specificity; Table 9) and predicting 97% of true positives. The kappa statistic also indicated that model performance was moderate (kappa = 0.476; Table 9).

**Table 9:** Minnesota River Ecoregional Drainage Unit (EDU) rare fish species accuracy metrics from independent data set. Metrics are weighted with detection probability. Species include P_BAK (banded killifish), P_CEB (central mudminnow), P_NRD (northern redbelly dace), P_ROS (carmine shiner), and P_HOC (hornyhead chub). A (.) indicates that the accuracy metric was could not be calculated.

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_BAK | 0 | 1 | 0 | 21 | 0.000 | 0.000 | 0.955 | 0.955 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.488 |
| P_CEM | 0 | 2 | 1 | 19 | 0.004 | 0.045 | 0.864 | 0.864 | 0.000 | 1.000 | 0.000 | 1.000 | 0.000 | 0.950 | 0.046 | 0.364 |
| P_NRD | 2 | 8 | 0 | 12 | 0.004 | 0.091 | 0.547 | 0.636 | 1.000 | 0.997 | 0.003 | 0.000 | 0.984 | 1.000 | 0.001 | 0.476 |
| P_ROS | 2 | 9 | 0 | 11 | 0.004 | 0.091 | 0.502 | 0.591 | 1.000 | 0.997 | 0.003 | 0.000 | 0.983 | 1.000 | 0.002 | 0.448 |
| P_HOC | 11 | 7 | 0 | 4 | 0.680 | 0.500 | 0.398 | 0.682 | 1.000 | 0.457 | 0.543 | 0.000 | 0.698 | 1.000 | 0.216 | 0.441 |

*Hornyhead chub*

The Hornyhead chub was collected at 11 of 18 predicted sites, within Florida Creek, Monighan Creek, and Yellow Bank River (Figures 15 and 17). The hornyhead chub was predicted to occur in headwaters, creeks and small rivers (Appendix 7). Detection probability and prevalence for the hornyhead chub were high (0.680 and 0.500 respectively). Modeling correctly predicted 68.2% of all cases corresponding to correctly predicting 45.7 % of true absences (specificity; Table 9) and 100% of true presences. The

kappa statistic also indicated that model performance was moderate (kappa = 0.476; Table 9).

*Banded killifish*

The banded killifish was predicted to occur at one site (Figures 11 and 12) but was not captured in the survey resulting in one commission error (Table 9). Habitat variables associated with presence of banded killifish were stream segments connected to lakes and sloughs (Appendix 7). The model correctly predicted the presence or absence of the species for 95% of sites surveyed. Even though the species was not collected, model performance was moderate (kappa = 0.488; Table 9).

**Figure 15:** Hornyhead chub (*Nocomis biguttatus*) 11-digit hydrological unit distribution from model results. Does not include independent dataset.

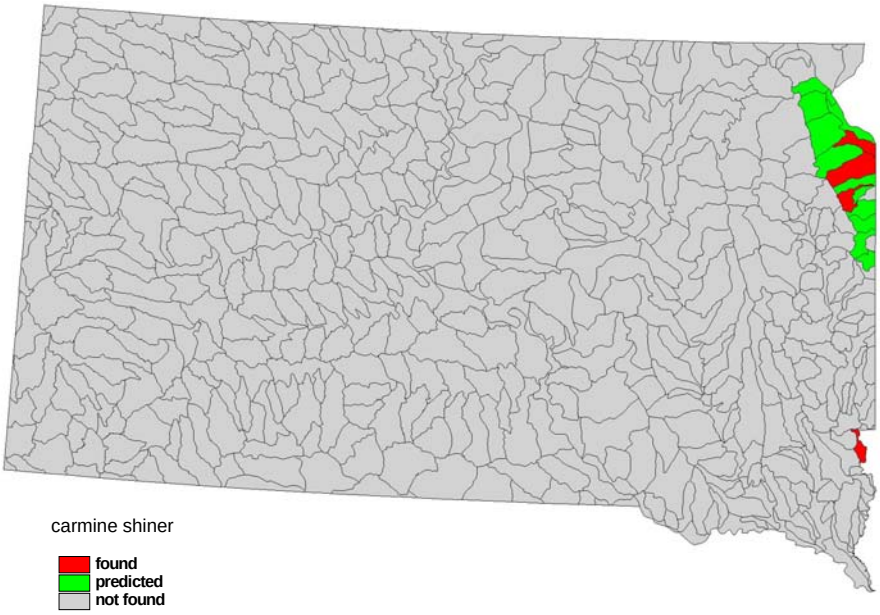


**Figure 16:** Carmine shiner (*Notropis percobromus*) potential 11-digit hydrological unit distribution from model results. Does not include independent dataset.

*Carmine shiner*

The carmine shiner was found at 2 of 11 predicted sites (Figures 16 and 18) resulting in nine commission errors (Table 9). Carmine shiners were predicted to occur in creeks and small rivers with perennial flow and no size discrepancy, or in creek segments with perennial flow and confluence with a small river (Appendix 7). Correct classification rate was moderate (0.591) correctly predicting presence 100% of the time and correctly predicting species absence 97% of the time (Table 9). The models predicted 48% better than would be expected by chance according to the kappa statistic indicating moderate model performance (Table 9).

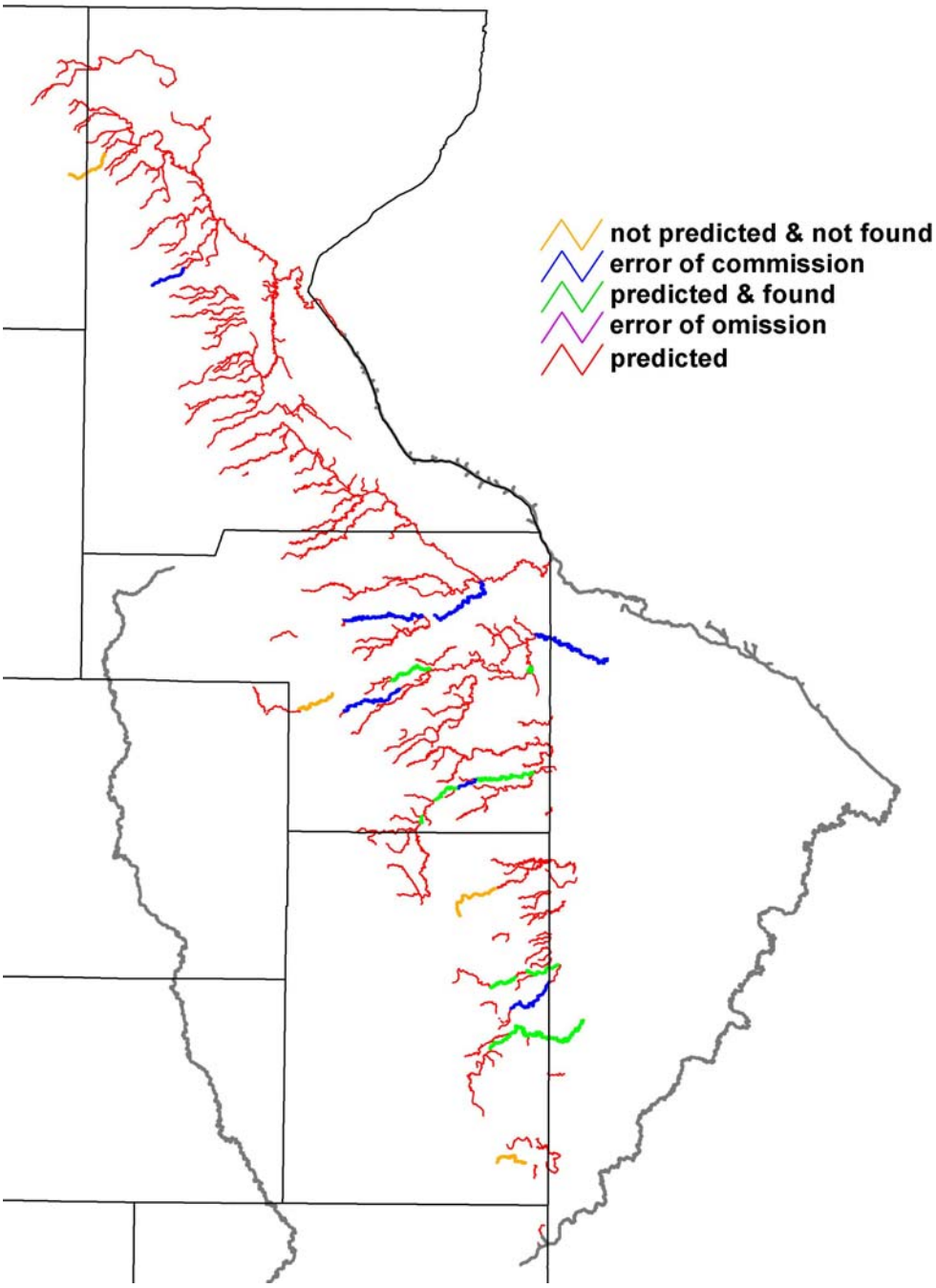


**Figure 17:** Hornyhead chub (*Nocomis biguttatus*) predicted valley segments (red colored streams) and model test results.

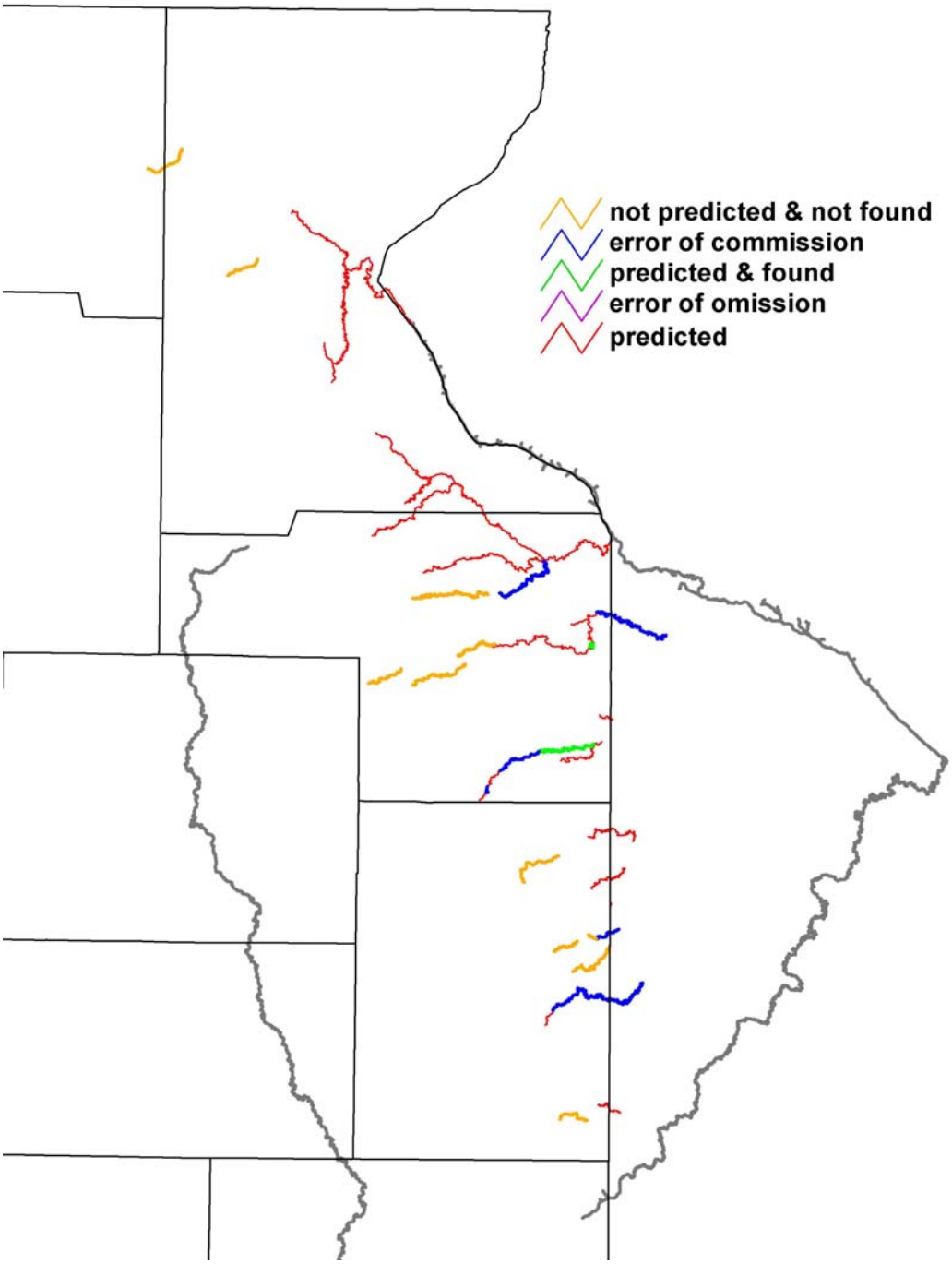


**Figure 18:** Carmine shiner (*Notropis percobromus*) predicted valley segments (red colored streams) and model test results.

## Black Hills Results

*All species*

We collected eight of 18 species predicted to occur at 15 sites within this EDU (Appendix 5). Species occurrence ranged from 0.000 (black crappie, common carp, finescale dace, golden shiner, green sunfish, longnose sucker, northern pike, and plains topminnow) to 0.800 (longnose dace; Appendix 9). The creek chub and longnose dace were found in over 60% of sampling locations. Detection probabilities were also variable ranging from 0.000 (black crappie, common carp, finescale dace, golden shiner, green sunfish, longnose sucker, northern pike, and plains topminnow) to 0.593 (rock bass). Correct classification rates ranged from 0.266 (brook trout) to 0.954 (rock bass; Appendix 9) and varied prevalence (Figure 19) showing a general trend of higher correct classification rates for rare species and abundant species. Incorporating detection probabilities into the analysis substantially improved model performance for almost all species (Figure 20). The mean kappa value (± SE) before adjusting with detection probability was 0.062 ± 0.041 and ranged between -0.111 and 0.645 (rainbow trout and rock bass, respectively). Kappa values that were adjusted with detection probability ranged from 0.069 (creek chub) to 0.688 (rock bass) and the mean was 0.377 ± 0.035 (Appendix 9). One hundred percent of species models were better than chance at predicting presence or absence of a species (Figure 20) within the Black Hills EDU and 47 % of models had moderate, substantial, or almost perfect model agreement (Figure 20).

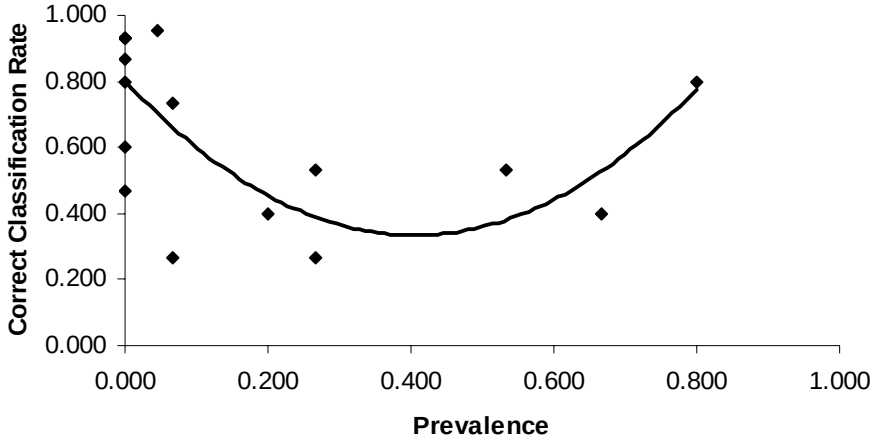


**Figure 19:** Correct classification rate of presence-absence models for 18 species in relation to their occurrence in 15 streams in the Black Hills ecological drainage unit.

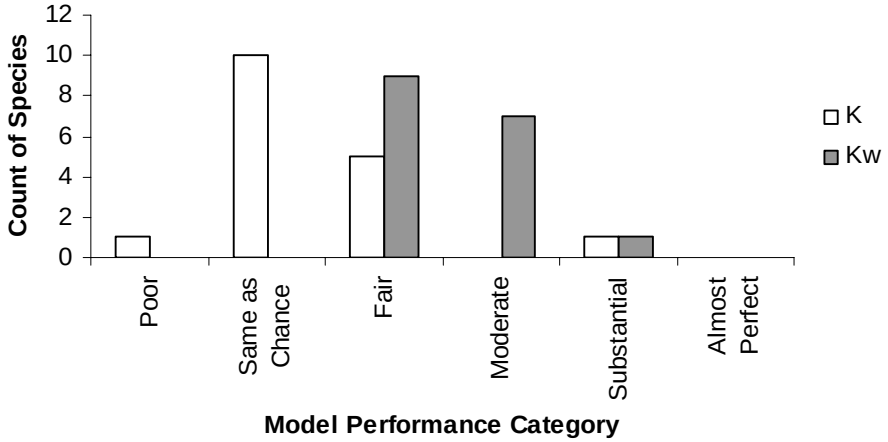


**Figure 20:** Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Black Hills River ecological drainage unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

**Table 10:** Black Hills ecoregional drainage unit (EDU) accuracy metrics for P_FID (finescale dace) and P_LOS (longnose sucker) from independent data set. Metrics are weighted with detection probability. A (.) indicates that the accuracy metric could not be calculated.

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_FID | 0 | 3 | 0 | 12 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |
| P_LOS | 0 | 8 | 0 | 7 | 0.000 | 0.000 | 0.467 | 0.467 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.318 |

*Rare species*

*Finescale dace*

The finescale dace was not found at three predicted sites (Figure 21 and 22) resulting in three commission errors. Fine scale dace were predicted to occur in stream segments with intermittent flow, medium to high gradients and stream orders greater than one, or stream segments with perennial flow through bedrock geology (Appendix 7). The model correctly predicted the presence or absence of the species for 80% of sites surveyed. Even though the species was not collected, model performance was moderate (kappa = 0.444; Table 10).

*Longnose sucker*

The longnose sucker was not captured at eight predicted sites (Figures 23 and 24) resulting in eight commission errors. Longnose suckers were predicted to occur in

headwaters, creeks, and small rivers at medium elevations (Appendix 7). The model correctly predicted the presence or absence of the species for only 46.7% of sites surveyed. Even though the species was not collected, model performance was fair (kappa = 0.31.8; Table 10).

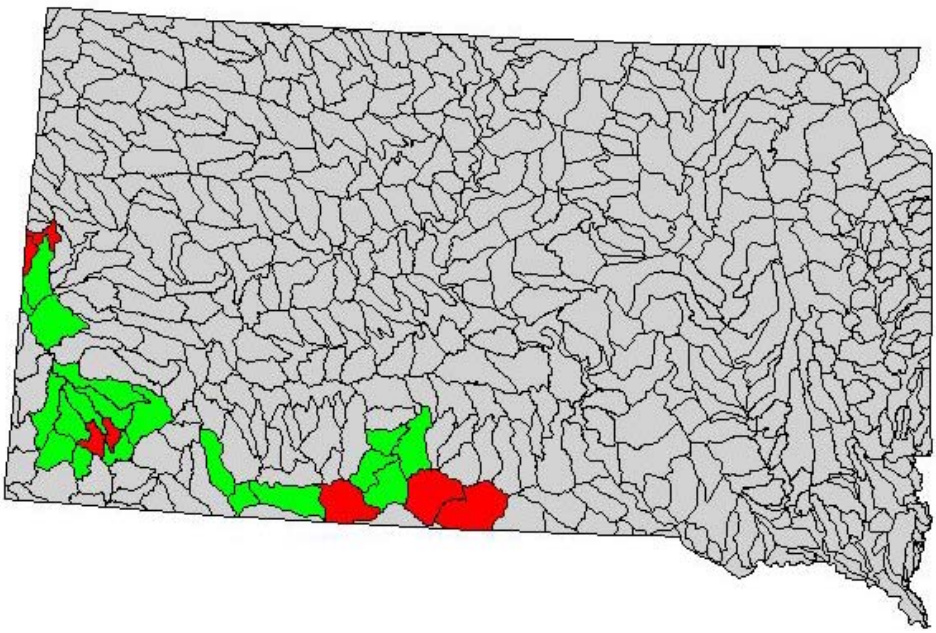

**Figure 21:** Finescale dace (*Phoxinus neogaeus*) 11-digit hydrological unit distribution from model results. Does not include independent dataset.

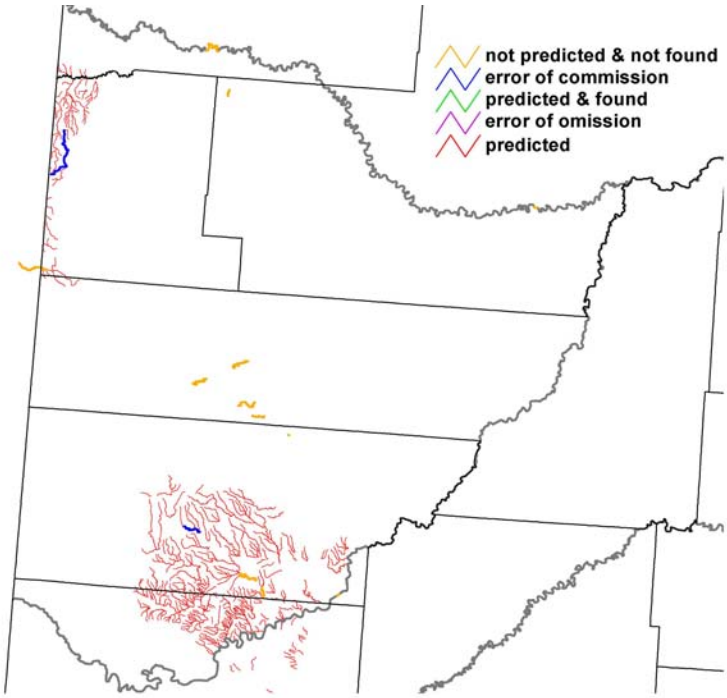


**Figure 22:** Finescale dace (*Phoxinus neogaeus*) predicted valley segments (red colored streams) and model test results for the Black Hills.

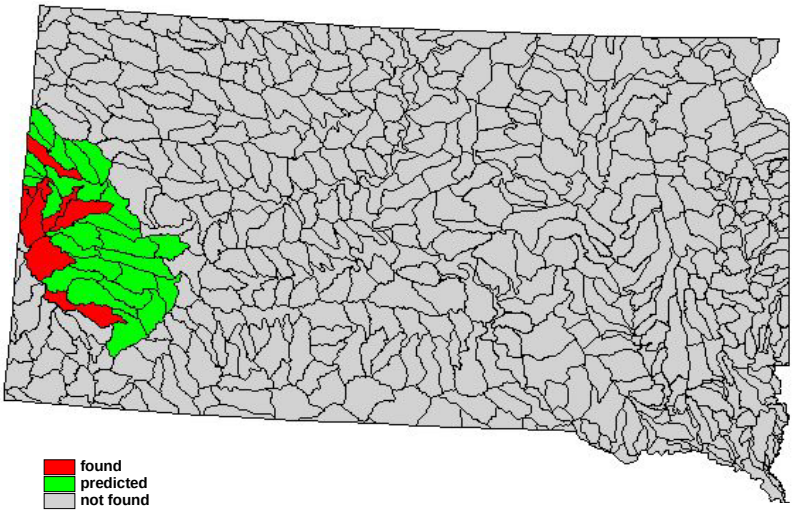


**Figure 23:** Longnose sucker (*Catostomus catostomus*) 11-digit hydrological unit distribution from model results. Does <u>not</u> include independent dataset.

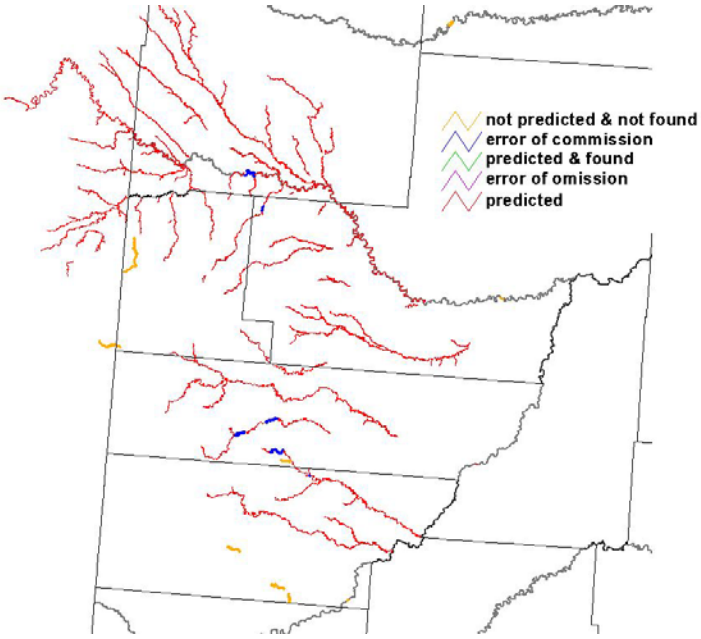


**Figure 24:** Longnose sucker (*Catostomus catostomus*) predicted valley segments (red colored streams) and model test results for the Black Hills.

## Cheyenne River Results

*All species*

We collected 27 of 37 species predicted to occur at seven sites in this EDU (Appendix 5). Species occurrence ranged from 0.000 (black crappie, brook trout, emerald shiner, freshwater drum, golden shiner, Iowa darter, largemouth bass, longnose sucker, walleye and white bass) to 0.850 (white sucker; Appendix 10). The creek chub, fathead minnow, longnose dace, stonecat, and white sucker were found in over 60% of sampling locations. Detection probabilities were also variable ranging from 0.000 (black crappie, brook trout, emerald shiner, freshwater drum, golden shiner, Iowa darter,

largemouth bass, longnose sucker, walleye and white bass) to 0.667 (channel catfish, common carp, goldeye, gizzard shad, and shorthead redhorse). Correct classification rates ranged from 0.142 (common carp) to 1.000 (bluegill, brown trout and mountain sucker; Appendix 10).

Including detection probabilities substantially improved model performance for almost all species (Figure 25). The mean kappa value (± SE) before adjusting with detection probability was 0.212 ± 0.062 and ranged between -0.615 (common carp) and 1.000 (bluegill, mountain sucker, and brown trout). Kappa values (Appendix 10) that were adjusted with detection probability ranged from -0.286 (common carp) to 1.000 (brown trout, bluegill and mountain sucker) and the mean was 0.412 ± 0.048. Eighty-six percent of species models were better than chance at predicting presence or absence of a species (Figure 25) within the Cheyenne River EDU and 54 % of models had moderate, substantial, or almost perfect model agreement (Figure 25).

*Rare species*

*Sturgeon Chub*

The sturgeon chub was predicted to occur at three sites but was only captured at one site in the Cheyenne River (Figures 26 and 27) corresponding to two commission errors. Predictive modeling associated 5$^{th}$ order streams flowing through alluvial outwash or 6$^{th}$ order streams with occurrence of sturgeon chub (Appendix 7). The model correctly predicted the presence or absence of the species for 71.4 % of sites surveyed.

Detection probability was 0.012 and the species occurred in over 14 % of sites surveyed in this EDU. Model performance was moderate (kappa = 0.561; Table 11).

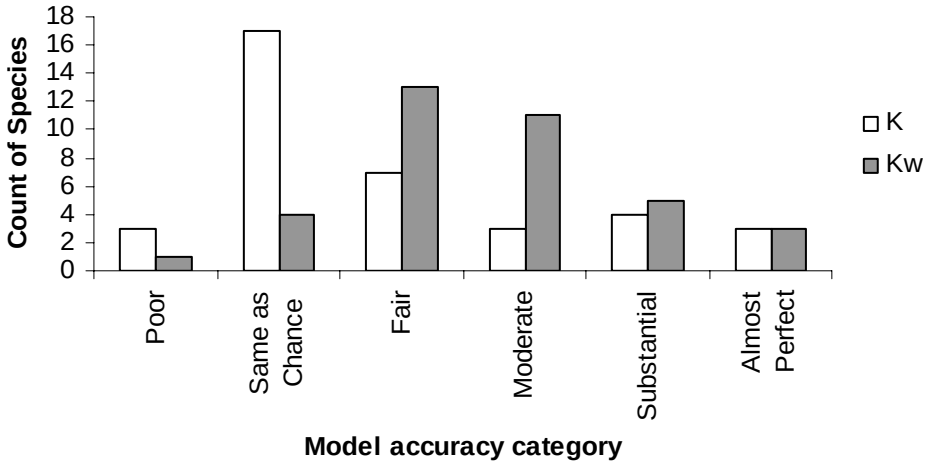


**Figure 25:** Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Cheyenne River ecological drainage unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

**Table 11**: Cheyenne River ecological drainage unit (EDU accuracy metrics for P_SNC (sturgeon chub) from independent data set. Metrics are weighted with detection probability. A (.) indicates that the accuracy metric could not be calculated.

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_SNC | 1 | 2 | 0 | 4 | 0.012 | 0.143 | 0.575 | 0.714 | 1.000 | 0.994 | 0.006 | 0.000 | 0.976 | 1.000 | 0.003 | 0.561 |

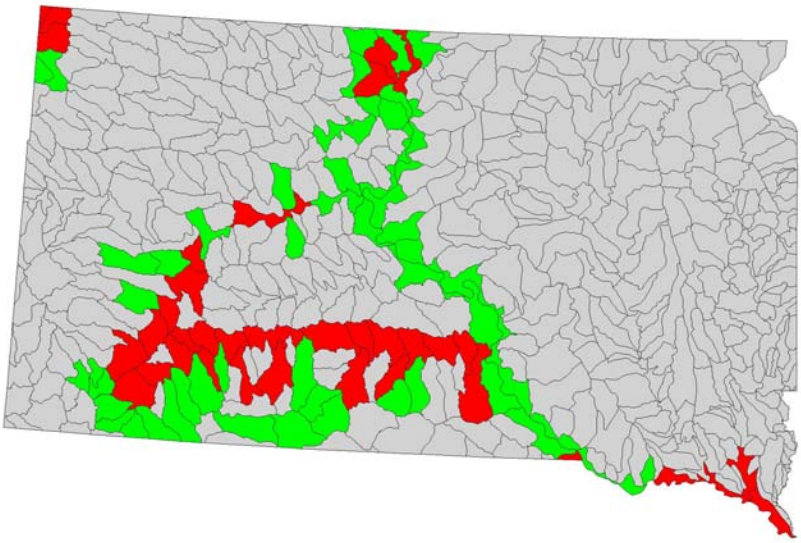

**Figure 26:** Sturgeon chub (*Macrhybopsis gelida*) 11-digit hydrological unit distribution from model results. Does not include independent dataset

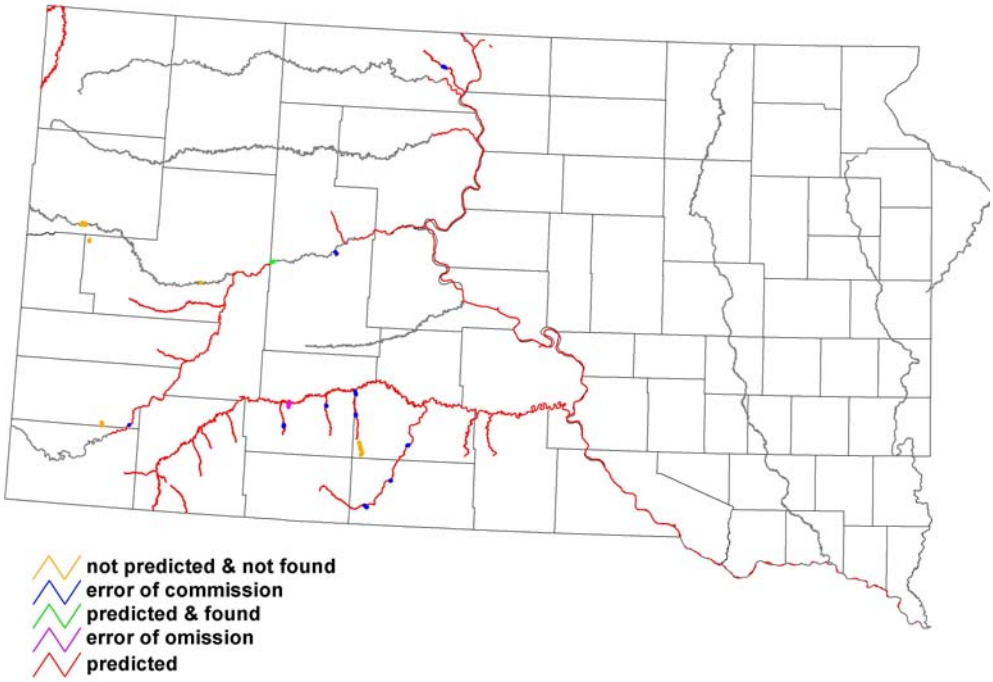


**Figure 27:** Sturgeon chub (*Macrhybopsis gelida*) predicted valley segments (red colored streams) and model test results for the South Dakota.

## Grand and Moreau Results

*All species*

We collected 22 of 26 species predicted to occur at five sites within this EDU (Appendix 5). Species occurrence ranged from 0.000 (brassy minnow, Iowa darter, johnny darter, northern redbelly dace, plains minnow, and yellow perch) to 1.000 (fathead minnow). The black bullhead, brook stickleback, fathead minnow, flathead chub, green sunfish, longnose dace, sand shiner, shorthead redhorse, stonecat, and white sucker occurred in over 60 % of samples (Appendix 11). Detection probabilities were also variable ranging from 0.000 (brassy minnow, Iowa darter, johnny darter, northern redbelly dace, plains minnow, and yellow perch) to 0.667 (flathead chub and longnose dace; Appendix 11). Correct classification rates ranged from 0.200 (yellow perch) to 1.000 (fathead minnow, longnose dace and white sucker; Appendix 11) and varied curvilinearly with prevalence (Figure 28).

Incorporating detection probabilities into the confusion matrix substantially improved model performance for almost all species (Figure 29). The mean kappa value (± SE) before adjusting with detection probability was 0.153 ± 0.076 and ranged between -0.154 (creek chub and flathead chub) and 1.000 (longnose dace and white sucker). Kappa values adjusted with detection probability ranged from 0.000 (emerald shiner, flathead chub, goldeye, and orange sunfish) to 1.000 (longnose dace and white sucker) and the mean was 0.0.365 ± 0.062. Eighty-one percent of species models were better

than chance at predicting presence or absence of a species (Figure 29) and 47 % of models had moderate, substantial, or almost perfect model agreement (Figure 29).

*Rare species*

**Table 12**: Grand/Moreau River ecoregional drainage unit (EDU) accuracy metrics for P_NRD (Northern redbelly dace) from independent data set. Metrics are weighted with detection probability. A (.) indicates that the accuracy metric could not be calculated.

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_NRD | 0 | 1 | 0 | 4 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |

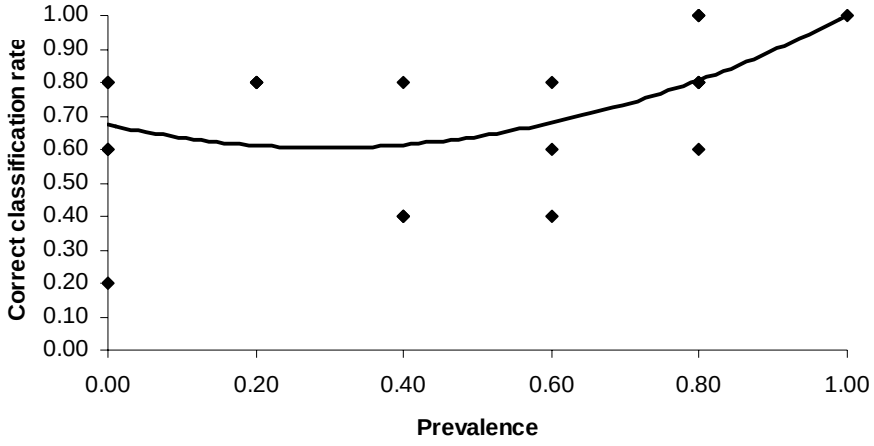


**Figure 28:** Correct classification rate of presence-absence models for 28 species in relation to their occurrence in 9 streams in the Grand and Moreau ecological drainage unit.

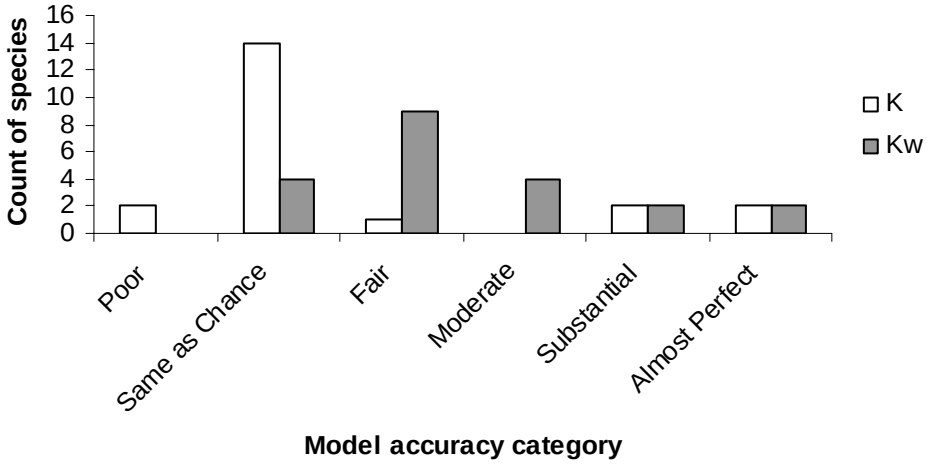


**Figure 29:** Count of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in Grand and Moreau ecological drainage unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

*Northern Redbelly dace*

The northern redbelly dace was not collected at one site of five where it was predicted to occur (Figures 7 and 30). Habitats related to northern redbelly dace presence were stream segments with intermittent flow and medium to low groundwater potential, or 2$^{nd}$ to 4$^{th}$ order streams with perennial flow, or 5$^{th}$ order streams with perennial flow and medium to high gradients (Appendix 7). The model correctly predicted the presence or absence of the species for only 80 % of sites surveyed. Even though the species was not collected, model performance was moderate (Kappa = 0.444; Table 12).

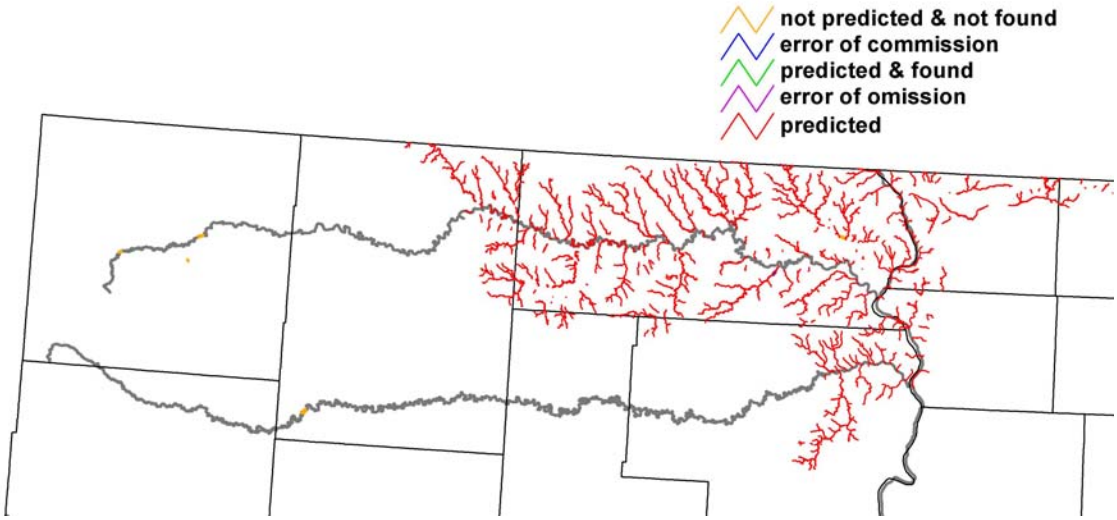


**Figure 30:** Northern redbelly dace (*Phoxinus eos*) predicted valley segments (red colored streams) and model test results for the Grand and Moreau River Basins.

## White River Results

*All species*

We collected 14 of 29 species predicted to occur at nine sites in this EDU (Appendix 5). Species occurrence ranged from 0.000 (black crappie, bluegill, common carp, goldeye, Iowa darter, northern pike, northern redbelly dace, plains minnow, plains topminnow, river carpsucker, sauger, shorthead redhorse, sturgeon chub, central stoneroller, and white sucker) to 0.889 (sand shiner). The flathead chub and sand shiner occurred in over 60% of samples (Appendix 12). Detection probabilities were also variable ranging from 0.000 (black crappie, bluegill, common carp, goldeye, Iowa darter, northern pike, northern redbelly dace, plains minnow, plains topminnow, river

carpsucker, sauger, shorthead redhorse, sturgeon chub, central stoneroller, and white sucker) to 1.000 (brassy minnow and red shiner; Appendix 12). Correct classification rates ranged from 0.000 (common carp) to 1.000 (bigmouth shiner; Appendix 12) and varied with prevalence.

Incorporating detection probabilities into the analysis substantially improved model performance for almost all species (Figure 31). The mean kappa value (± SE) before adjusting with detection probability was 0.097 ± 0.045 and ranged between -0.235 (fathead minnow) and 1.000 (bigmouth shiner). Kappa values that were adjusted with detection probability ranged from 0.000 (common carp) to 1.000 (bigmouth shiner) and the mean was 0.342 ± 0.038. Ninety-three percent of species models were better than chance at predicting presence or absence of a species (Figure 31) and 21.4 % of models had moderate, substantial, or almost perfect model agreement (Figure 31).

*Rare species*

*Northern Redbelly dace*

The northern redbelly dace was not collected at three predicted sites (Figures 7 and 32). The northern redbelly dace was predicted to occur in either: 1) stream segments with intermittent flow and medium to low groundwater potential, or 2) second to fourth order streams with perennial flow, or 3) fifth order streams with perennial flow and medium to high gradients (Appendix 7). The model correctly predicted the presence or

absence of the species for only 66.7 % of sites surveyed. Even though the species was not collected, model performance was moderate (kappa = 0.400; Table 13).

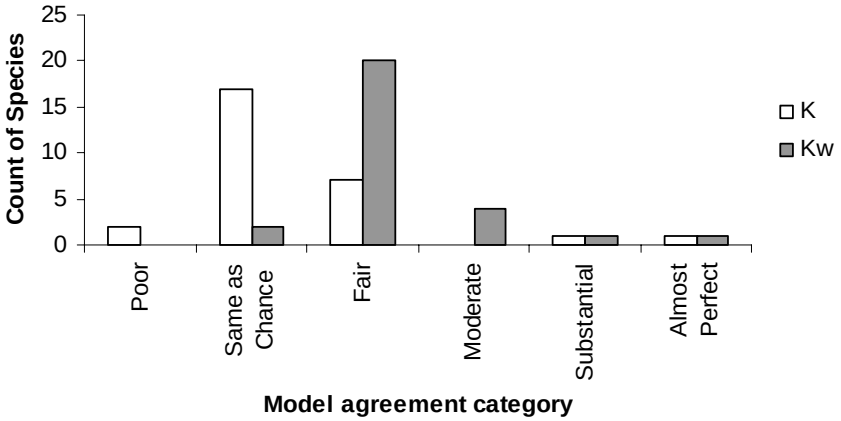


**Figure 31:** Number of species corresponding to model accuracy metric kappa (K) and kappa weighted with detection probability (Kw) in White River ecological drainage unit (EDU). Model agreement groups are Poor (k<0), same as chance (k=0), fair (0>k<0.40), moderate (0.40>k<0.60), substantial (0.60<k>0.80) and almost perfect (k>0.80).

**Table 13:** White river ecological drainage unit accuracy metrics for P_NRD (northern redbelly dace) and P_PED (pearl dace) from independent data set. Metrics are weighted with detection probability. A (.) indicates that the accuracy metric could not be calculated.

| Species Code | True presences | Commission errors | Omission errors | True absences | DP | Prevelence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P_NRD | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| P_PED | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |

*Pearl Dace*

The pearl dace was not collected at three predicted sites (Figures 33 and 34). Predictive modeling indicated that headwaters and creeks with stream orders >1 were associated with presence of pearl dace (Appendix 7). The model correctly predicted the presence or absence of the species for only 66.7 % of sites surveyed. Even though the species was not collected, model performance was moderate (Kappa = 0.400; Table 13).

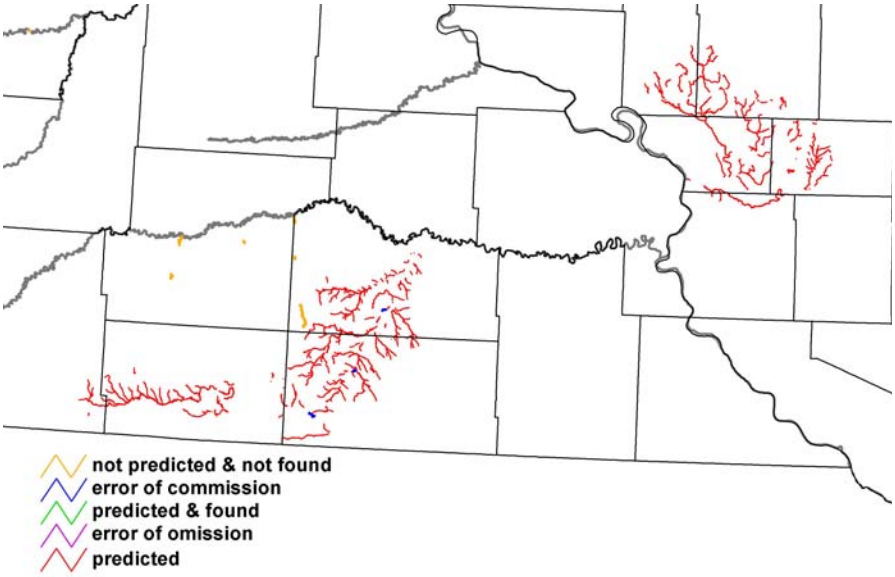


**Figure 32:** Northern redbelly dace (*Phoxinus eos*) predicted valley segments (red colored streams) and model test results for the White River Basin and Crow Creek watershed.

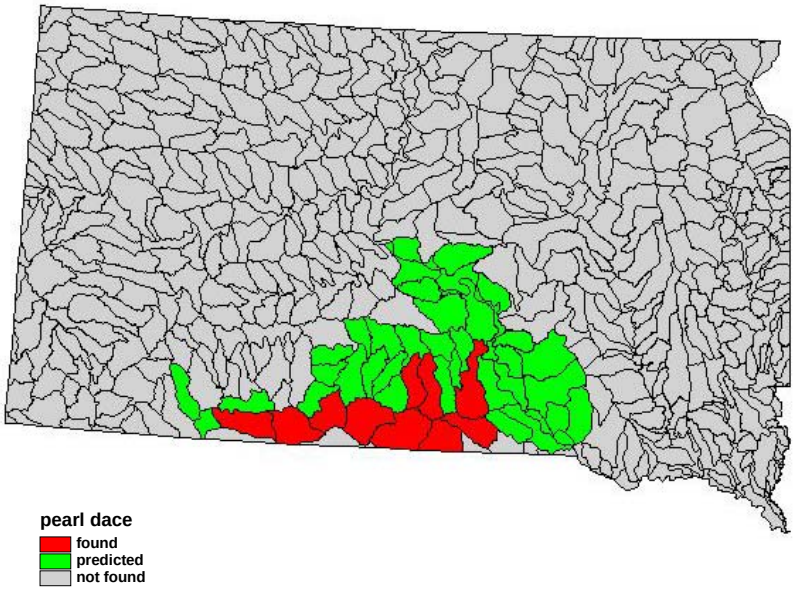


**Figure 33:** Pearl dace (*Margariscus margarita*) 11-digit hydrological unit distribution from model results. Does <u>not</u> include independent dataset.

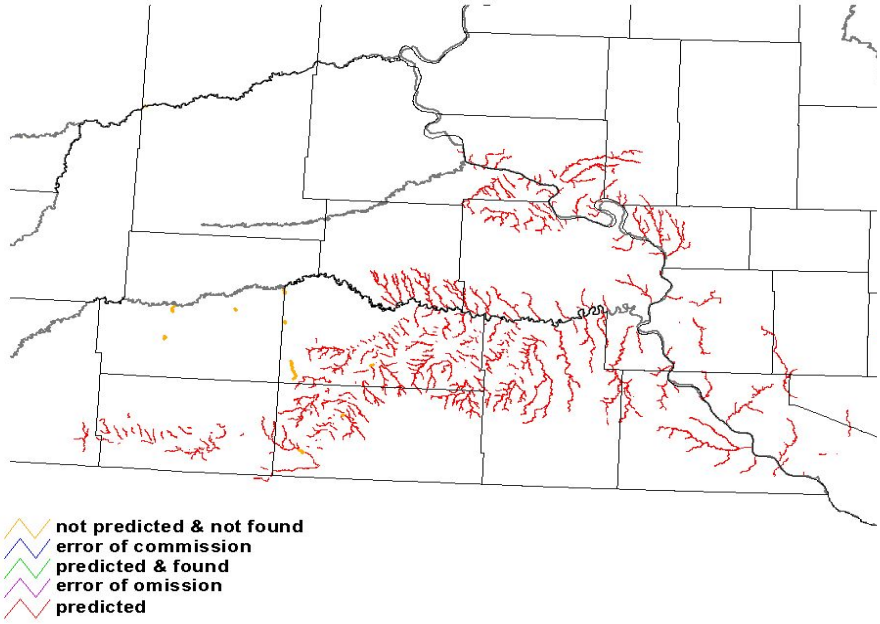


**Figure 34:** Pearl dace (*Margariscus margarita*) predicted valley segments (red colored streams) and model test results for the White River Basin and Missouri River tributaries.

## Discussion

The objective of this study was to assess the predictive accuracy of Aquatic GAP distribution models with independent field surveys for fish species and specifically rare fish across five ecological drainage units in South Dakota. The models presented had relatively high predictive power for most species, and were variable among drainages and among species (Tables 7 and 14).

The correct classification rates (Table 14) were comparable to other studies (>70% on average) assessing accuracy of fish prediction models (Porter et al. 2000; Manel et al. 2001; Olden and Jackson 2001; Filipe et al. 2002; Rashleigh et al. 2005; Oakes et al. 2005, Sylvester 2004). The kappa statistic, however, indicated substantial model agreement (kappa > 40%) for most species of concern, a finding that is slightly better than other studies (Manel et al. 2001; Oaks et al. 2005). Even though predictive models were fairly accurate, and several factors may have affected the accuracy and subsequent agreement between models and the independent surveys.

**Table 14: Range of correct classification rates (CCR) and kappa (k) statistics for rare species predicted to occur in each ecological drainage unit (EDU).**

| EDU | CCR | k |
|---|---|---|
| East River | 58-95 | 42-61 |
| Minnesota | 59-96 | 36-49 |
| Black Hills | 47-80 | 32-44 |
| Cheyenne | 71 | 42 |
| Grand/Moreau | 80 | 44 |
| White River | 67 | 40 |

There are two main sources of prediction errors associated with modeling: omission (false negatives) and commission (false positives). Omission errors refer to sites where a fish was captured but was not predicted to occur whereas commission errors refer to sites where a fish was predicted to occur but was not captured. Commission error consists of two components: true error in that the species does not occur at that location resulting from inaccurate models and apparent error in that the species does occur at the site but was missed in field surveys (Dedon et al. 1986, Scott et al. 1993; Edwards et al. 1996). A species may not have been detected in surveys because of inadequate sampling methods (Nichols et al. 1998a), local extinction and colonization rates, habitat preferences, spatial and temporal differences (Hayer 2005), as well as specific differences in species behavior (Bayley and Dowling 1993; Boulinier et al. 1998; Nichols et al. 1998b; Bayley and Peterson 2001; Gu and Swihart 2004). Heterogeneity in species detectability may exist among samples and among species within a sampling session (Boulinier et al. 1998; Nichols et al. 1998a; Hayer 2005).

The inability to detect a species can inflate commission error (Scott et al. 1993; Cassidy et al. 1994; Krohn 1996). In order to minimize commission error, other studies have applied a 0.5 threshold as a weighting factor of commission (e.g. Karl et al. 2000), which increased the predictive ability of the models, however, this arbitrary number did not account for heterogeneous detection rates among species. A method we used to increase the accuracy of our models and thus reduce commission error was to compute detection probabilities for each species and assign these as weights on commission error.

As a result of weighting commission errors, our models demonstrated high predictive accuracy and improved most species models and all targeted rare species models.

In addition to heterogeneous detection probabilities, several other factors could account for the incorrect classifications including sampling effort and size, biases and use of different sampling gears, temporal and spatial variation of fishes, and problems with the GAP species distribution models. Low sample sizes of species occurrence from independent field collections can greatly effect model accuracy and increase errors of omission and commission (Stockwell and Peterson 2002; Wall et al. 2004). Most species usually comprise a small portion of the total fish community (Braaten 1993; Cunningham 1999; Wall et al 2001), so it was difficult to identify true present sites. Conversely it was impossible to identify true absence sites (Wall et al. 2004). Sampling effort was variable between EDUs in this study with East River EDU and Minnesota EDU having the most effort. As a result, model accuracy metrics were variable among EDU: East River EDU had the highest correct classification rates (mean = 0.828) and kappa statistics (mean = 0.514) for species of concern. Our data for rare species demonstrated a general trend of increased predictive accuracy as sample sites increased. An increase in sampling effort in other EDUs may have increased predictive accuracy of models.

Increased sampling can lead to increased prevalence, which varied curvi-linearly with correct classification rates in our study reflecting effects of prevalence on sensitivity and specificity: positive occurrences were more effectively predicted as prevalence increased and negative occurrences as prevalence declined. This relationship has been

attributed to insufficient data points available to accurately classify habitats in which species are either present or absent (Manel et al. 2001; Olden et al. 2002; Scott et al. 2002; Wall et al. 2004) leading to the increased probability of correctly predicting the occurrence of rare and very abundant species (Olden et al. 2002).

It is important that researchers and managers choose the optimum sampling method to obtain the most accurate and precise estimate of fish populations (Tate et al. 2003). All fish sampling gears are subject to bias depending on the environmental and chemical conditions of the water, and behavioral, ecological and morphological characteristics of species present (Paller 1995; Meador et al. 2003; Sylvester 2004). Efficiency issues of sampling gears (i.e., seines and electrofishing) may be affected by habitat conditions and as a result species are potentially missed. This study utilized two gears, electrofishing and seining, depending on the habitat and we feel that the gears were effective in the small, wadeable streams that we sampled; however observation or capture of all species is usually impossible, especially for such mobile organisms as fish and their sometimes patchy distribution (Fausch and Bestigen 1996, Boulinier et al. 1998; Nichols et al. 1998b). The inability of sampling methods to capture all species present within a sampling unit may have affected model accuracy; however, we feel that incorporating detection probabilities into accuracy metrics accounted for gear efficiency issues.

Variation in predictive accuracy of models may be the result of behavioral differences among species, which may relate to habitat. Habitat has been recognized as a primary gradient for arrangement and structure of animal communities (Schoener 1974)

especially fishes (Schlosser 1987, Arterburn and Berry 2002, Wall et al. 2004, Berry et. al. 2004). Individual species preferences and behaviors may be the source of some variation in model accuracy, which may be compounded by other factors such as seasonal shifts in distribution (Filipe et al. 2002, Fausch and Bestigen 1996). For example, the banded killifish prefers habitats associated with lakes, ponds, and sluggish streams, whereas our sampling focused on wadeable streams sites. Sampling of lakes and ponds may have led to the collection of this fish and thus increased the predictive accuracy of models. The finescale dace prefers small lakes in cool, boggy environments associated with springs or beaver dams (Bailey and Allum 1962; Baxter and Stone 1995). Although finescale dace were collected in other habitats, targeting preferred habitat as we did for the stream fishes may have increased the predictive accuracy of the finescale dace model.

Temporal variation in abiotic conditions at several scales (e.g., seasonal, Harland and Berry 2004; historical, Dieterman and Berry 1998, Shearer and Berry 2003, Galat et al. 2005) can be extreme in small streams and can be important in structuring fish communities (Schlosser 1987; Magoulick 2000). Seasonal hydrologic variation, water temperature fluctuations, increased or decreased food, and life history functions (e.g., spawning) of fishes, can also influence fish distribution across seasons (Magoulick 2000; Angermeier et al. 2002; Dieterman and Berry 2000, Herbert and Gelwick 2003,). For example, many predicted sites were dry as a result of drought, being an intermittent stream, or other factors. These sites may have contained suitable habitat for a particular fish species at one time, but not during the study period. It is virtually impossible to

account for such factors in modeling. It is important that monitoring programs account for seasonal variability in order to accurately represent the target species and account for seasonal changes in distribution (Filipe et al. 2002).

We used classifications of temperature, stream size, stream flow (perennial or intermittent), ground water potential, channel slope, network position, parent materials, floodplain influence, connectivity to lakes and elevation to predict valley segments where fishes were likely to occur. We found stream size, stream flow, and groundwater potential were key habitat variables for predicting most fish species at this scale in prairie streams, as have others (Smith et al. 2002; Sowa et al. 2005; Wall et al. 2004). Connectivity to lakes and sloughs was used to predict suitable habitat for the central mudminnow and banded killifish. This is likely as a result that these species were found predominantly in permanent water bodies. Northern redbelly dace were rare in the East River EDU and were predicted to be in valley segments adjacent to known locations.

Model accuracy is affected by the inherent inaccuracies of GIS methods (Goodchild and Gopal 1989). We modeled fish distributions using broad-scale variables that were created using GIS layers, assuming that the variables we associated fish presence were surrogates for conditions needed by the species. GIS layers are not 100% accurate and errors accumulate with each layer that is added to create variables and added to the model. For example, if segment length accuracy is 90% and digital elevation accuracy is 90%, overall accuracy for creating the channel slope variable would be 81% (90% x 90%). Additionally, it is possible that variables not used in our GIS models (e.g.,

bedrock geology, soils), or factors at a finer scale, may be driving fish distributions and thus species distribution predictions are less accurate.

Spatial autocorrelation may also affect model accuracy (Smith 1994, Nash et al. 1999, Lichstein et al. 2002). Spatial autocorrelation (i.e., similarity between sites correlated with distance between sites) is a natural phenomenon of ecological systems at all spatial scales (Legendre 1993). Accounting for spatial autocorrelation in the modeling process may have improved accuracy (Hinch et al. 1994, Pearson 2002, Legendre et al. 2002, Segurado and Araujo 2004). Spatial corrections can also be applied to model results (i.e. post processing) to account for errors of omission by weighting false positives close to actual positive cases as less serious than those located at a distance from a real positive (Fielding and Bell 1997). We applied an approach similar to the ubiquitous Minesweeper computer game to weight false positives, as suggested in Fielding and Bell (1997), but results did not improve model performance.

Despite these factors that reduce model accuracy, most of our models had high predictive abilities. The accuracy assessment in this project quantifies performance of the Aquatic GAP fish distribution models against an independent dataset, which is the most rigorous method of testing any models. Our models for rare species contained very minimal omissions. From a conservation standpoint, models that produce low omission error are of the most interest, because they can be used to identify areas of suitable habitat for rare species across the landscape (Oakes et al. 2005). Moreover, species

presence data is considerably less biased compared with species absence data that can arise from failure to detect a species at a particular site (Scott et al. 2002).

Our results are the opposite of many studies reporting on models that predicted absence better than presence, which was attributed to the difficulty in predicting occurrence of rare species (Olden and Jackson 2002). Correctly predicted presences (83%) in our models were generally higher than correctly predicted absences (77%) thus emphasizing the accuracy of our models to predict occurrence of fish species.

## Conclusion and Application

Improved knowledge about the distribution of threatened and endangered fishes of South Dakota was an immediate benefit of this study. The data in part contributed to the decision to remove the trout-perch and central mudminnow from the State's list of fish species of concern in 2006.

Assessing the accuracy of the fish distribution models has less tangible benefits but is an important first step in understanding the relationships between fishes and the riverine habitat, which is the focus of modern river management (Williams et al. 1997, Wissmar and Bisson 2003). Fish distribution models also reveal inadequacies in data about fish locations and habitat affinities. Some model inaccuracy may be due to local habitat influences (e.g., riparian condition, competition), not the enduring landscape variables that are the focus of the gap analysis modeling methods.

We suggest that the next step in using gap analysis results for fishes of South Dakota (Wall et al. 2004, 2005) would be to begin applying conservation practices to land where there is a high probability of species presence. The model accuracy assessments made in this study provide managers with a measure of confidence about the value of the models to direct conservation activities. Choices must be made about where conservation practices would be most effective because opportunities and funding for conservation are limited. One factor in the decision is the accuracy of the fish distribution model. Administrators can have more confidence in funding conservation in areas where the models were more accurate than where they were less accurate. Determining the ultimate benefit to the species or fish assemblage must await future studies after conservation practices have been applied and are in place for several years.

## References


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, M. L. Morrison et al., editors. Predicting Species Occurrences. Issues of Accuracy and Scale. Washington D.C., Island Press.

Arterburn, J. and C. Berry. 2002. Effect of hook style, bait type, and river locations on trotline catches of flathead and channel catfish. North American Journal of Fisheries Management. 22:573-578.

Bailey, R.G. 1995. Description of the ecoregions of the United States. 2nd ed. Rev. and expanded (1st ed. 1980). Miscellaneous Publication No. 1391 (rev.), Washington DC: USDA Forest Service. 108pp with separate map at 1:7,500,00.

Bailey, R.M. and M.O. Allum. 1962. Fishes of South Dakota. University of Michigan, Museum of Zoology. Miscellaneous Publications Number 119, 131 pp.

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

Bayley, P.B. and D.C. Dowling. 1993. The effects of habitat in biasing fish abundance and species richness estimates when using various sampling methods in streams. Polski Archiwum Hydrobiologii 40:5-14.

Bayley, P. and J. T. Peterson. 2001. An approach to estimate probability of presence and richness of fish species. Transactions of the American Fisheries Society 130: 620-633.

Baxter, G.T. and M.D. Stone. 1995. Fishes of Wyoming. Wyoming Game and Fish Department, Cheyenne.

Berry, C., M. Wildhaber, and D. Galat. 2004. Fish distribution and abundance. Volume 3 of Population structure and habitat use of benthic fishes along the Missouri and Lower Yellowstone Rivers. Cooperative Fish and Wildlife Research Unit, South Dakota State University, Brookings, South Dakota.

Biggs, D., B. de Ville, and E. Suen. 1991. A method of choosing multiway partions for classification and decision trees. Journal of Applied Statistics 18: 49-62.

Boulinier, T., J. Nichols, J. Sauer, J. Hines and K. Pollock. 1998. Estimating species richness: the importance of heterogeneity in species detectability. Ecology 79 (3): 1018-1028.

Braaten, P.J. 1993. The influence of habitat structure and environmental variability on habitat use by fish in the Vermillion River, South Dakota. Masters thesis, South Dakota State University, Brookings, SD.

Burnham, K. and W. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60 (5): 927-936.

Carey, P.D. & Brown, N.J. 1995. The use of GIS to identify sites that will become suitable for a rare orchid, *Himantoglossum hircinum* L., in future changed climate. *Biodiversity Letters*, 2, 117–123.

Cassidy, K., E.O. Garton, W.B. Grohn, L.S. Mills, J.M. Scott, K. Williams, and B. Csuti. 1994. Assessing the predictive ability of gap analysis vertebrate distributions. Gap analysis handbook. Idaho Cooperative Fish and Wildlife Research Unit, Moscow, Idaho, USA.

Chao, A., S. M. Lee and S. L. Jeng. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. Biometrics 48 (1): 201-216.

Conway, C. J. and J. D. Simon. 2003. Comparison of detection probability associated with burrowing owl survey methods. Journal of Wildlife Management 67 (3): 501-511.

Cunningham, G.R. 1999. A survey for the Topeka shiner (Notropis topeka) within the Big Sioux, Vermillion, and James river basins in South Dakota. Eco-Centrics, 2041 N. 66th St., Omaha, NE 68104.

Davis, F.W. 1996. The nature of gap analysis. BioScience 46: 74-75.

Dedon, M.F., S.A. Laymon, and R.H. Barrett. 1986. Evaluating models of wildlife-habitat relationships of birds in black oak and mixed conifer habitats. Pages 115-116 *in* J. Verner, M.L. Morrison, and C.J. Ralph, editors. Wildlife 2000. Modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison, Wisconsin, USA.

Dieterman, D. and C. Berry. 2000. Summer habitat associations of rare fishes in South Dakota tributaries to the Minnesota River. Proceedings of the South Dakota Academy of Science 79:103-110.

Dieterman, D. and C. Berry. 1998. Fish community and water quality changes in the Big Sioux River. The Prairie Naturalist 30:199-244.

Edwards, T.C. 1996. Data defensibility and gap analysis. BioScience 46: 75-76.

Edwards, T.C., E.T. Deshler, D. Foster, and G.G. Moisen. 1996. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conservation Biology 10:263-270.

Elith, J. and M. Burgman. 2002. Predictions and their validation: rare plants in the central highlands, Victoria, Australia. Pages 303-313 *in* J.M. Scott, P.J. Heglund, J.B. Haufler, M. Morrison, M.G. Raphael, W.B. Wall & F. Samson, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, CA.

Environmental Systems Research Institute. 1999. ArcView GIS, Version 3.2, Environmental Systems Research Institute, Redlands, California.

Environmental Systems Research Institute. 2000. ARC/INFO GIS, Version 8.0.2, Environmental Systems Research Institute, Redlands, California.

Fielding, A. H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24 (1): 38-49.

Filipe, A., I. Cowx and M. Collares-Pereira. 2002. Spatial modeling of freshwater fish in semi-arid river systems: a tool for conservation. River research and applications 18: 123-136.

Fausch, Kurt D., and K. R. Bestgen. 1996. Ecology of fishes indigenous to the central and southwestern Great Plains. Pages 131-166 in F. Knopf and F. Samson, editors. Ecology and conservation of Great Plains vertebrates. Springer-Verlag, New York.

Frissell, C. A., J. L. William, C.E. Warren, and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10:199-214.

Galat, D., C. Berry, et al. 2005. Spatiotemporal patterns and changes in Missouri River fishes. Pages 249-292 in J. Rinne, R. Hughes and B. Calamusso, editors. Historical changes in large river fish assemblages of the Americas. American Fisheries Society, Symposium 45, Bethesda, Maryland.

Gaston, K.J. 1994. Rarity. Chapman and Hall, London, UK.

Godown, M.E. and A.T. Peterson. 2000. Preliminary distributional analysis of US endangered bird species. Biodiversity and Conservation 9: 1313–1322.

Goodchild, M.F., and Gopal, S. 1989. Accuracy of spatial databases. Tayler and Francis, London.

Gu, W. and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation 116: 195-203.

Guisan, A. and N.E. Zimmerman. 2000. Predictive habitat distribution models in ecology. Ecological Modeling 135: 147-186.

Harland, B. and C. Berry. 2004. Fishes and habitat characteristics of the Keya Paha River, South Dakota-Nebraska. Journal of Freshwater Ecology 19:169-177.

Hayer, C.A. 2005. Influence of spatial, temporal, and abiotic factors on detection probabilities for fishes in the Mobile River Drainage, Alabama. Thesis submitted to Auburn University, Auburn, Alabama.

Herbert, M. E. and F. P. Gelwick. 2003. Spatial variation of headwater fish assemblages explained by hydrologic variability and upstream effects of impoundment. Copeia 2: 273-284.

Higgins, J., M. Lammert, M. Bryer, M. DePhilip, and D. Grossman. 1998. Freshwater conservation in the Great Lakes Basin: Development and application of an aquatic community classification framework. The Nature Conservancy, Great Lakes Program, 8 South Michigan Ave., Suite 2301, Chicago, IL 60603.

Hinch S. G., K. M. Somers, and N. C. Collins. 1994. Spatial autocorrelation and assessment of habitat-abundance relationships in littoral zone fish. Canadian Journal of Fisheries and Aquatic Sciences 51:701-712.

Hoeting, J.A., M. Leecaster, and D. Bowden. 2000. An improved model for spatially correlated binary responses. Journal of Agricultural, Biological, and Environmental Statistics 5: 102-114.

Hurley, J.F. 1986. Summary, development, testing, and application of wildlife-habitat models: the manager's viewpoint. Pages 151-153 *in* J.A. Verner, M.L. Morrison, and C.J. Ralph, editors. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison.

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

Karl, J.W., P.J. Heglund, E.O. Garton, J.M Scott, N.M. Wright, R.L. Hutto. 2000. Sensitivity of species habitat-relationship model performance to factors of scale. Ecological Applications 10(6): 1690-1675.

Kral, J.G., C.R. Berry, Jr., and S.S. Wall. 2005. Evaluation of a decision support tool (i.e. aquatic gap analysis) to help conserve fish species at risk in South Dakota streams. Annual Report to South Dakota Department of Game, Fish and Parks, Project # T-9, study # 2409.

Krohn, W.B. 1996. Predicted vertebrate distributions from gap analysis: considerations in the designs of statewide accuracy assessments. Pages 147-162 *in* J.M. Scott, T.H. Tear, and F.W. Davis, editors. Gap analysis: a landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA.

Lammert, M., J. Higgins, D. Grossman, and M. Bryer. 1996. A classification framework for freshwater communities. Proceedings of The Nature Conservancy's Aquatic Community Classification Workshop; New Haven, Missouri; April 9-11, 1996. The Nature Conservancy, Arlington, Virginia.

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm. Ecology 74: 1659-1673.

Legendre P., M. R. T. Dale, M.-J. Fortin, J Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25:601-615.

Lichstein J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 73:445-463.

MacKenzie, D., J. D. Nichols, G. Lachman, S. Droege, A. Royle and C. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83 (8): 2248-2255.

MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi and L. L. Bailey. 2005. Improving inferences in population studies of rare species that are detected imperfectly. Ecology 86 (5): 1101-1113.

Magoulick, D. D. 2000. Spatial and temporal variation in fish assemblages of drying stream pools: The role of abiotic and biotic factors. Aquatic Ecology 34: 29-41.

Manel, S., H. C. Williams and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38: 921-931.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. Ecology Letters 8: 1235-1246.

McArdle, B. H. 1990. When are rare species not there? Oikos 57 (2): 276-277.

McDonald, L.L. 2004. Sampling rare populations. *In* Sampling rare or elusive species. Thompson, W.L editor. Washington D.C., Island Press: 11-42.

McKenna J. E., Jr., R. P. McDonald, C. Castiglione, S. Morrison, K. Kowalski, and D. Passino-Reader. 2005. A process to model habitats and fish-environment relationships in support of aquatic gap analysis in the Great Lakes. American Fisheries Society Symposium 48. Bethesda, MD *in press*.

Meador, M., J. McIntyre and K. Pollock. 2003. Assessing the efficacy of Single-Pass Backpack Electrofishing to Characterize Fish Community Structure. Transactions of the American Fisheries Society 132: 39-46.

Meixler, M.S. and M. B. Bain. 1999. Update on New York aquatic GAP project. GAP Analysis Program Bulletin 8: 39.

Miller, R.I. 1986. Predicting rare plant distribution patterns in the southern Appalachians of the south-eastern U*SA*. Journal of Biogeography. 13: 293–311.

Mingoti, S. A. and G. Meeden. 1992. Estimating the total number of distinct species using presence and absence data. Biometrics 48: 863-875.

Moilanen, A. 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. Ecology 80 (3): 1031-1043.

Morey, NM and CR Berry Jr. 2004. New distributional record of the northern redbelly dace in the northern Great Plains. The Prairie Naturalist 36(4):257-260.

Myatt, M.M. 1987. Predicting the habitat geography of sensitive plants and community types. Conservation and Management of Rare and Endangered Plants (ed. T.S. Elias), pp. 173 –179. The California Native Plants Society, Sacramento, CA.

Nash C. H., J. S. Richardson, and S. G. Hinch. 1999. Spatial autocorrelation and fish production in freshwaters: a comment on Randal et al. (1995). Canadian Journal of Fisheries and Aquatic Sciences 56:1696-1699.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock and J. R. Sauer. 1998a. Estimating rates of local species extinction, colonization, and turnover in animal communities. Ecological applications 8 (4): 1213-1225.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock and J. R. Sauer. 1998b. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology 12 (6): 1390-1398.

Oakes R. M., K. B. Gido, J. A. Falke, J. D. Olden, and B. L. Brock. 2005. Modeling of stream fishes in the Great Plains, USA. Ecology of Freshwater Fish 14: 361-374.

O'Connor, R. J. 2002. The conceptual basis of species distribution modeling: time for a paradigm shift. Pages 25-33 in Scott, J. M., P. J. Heglund & M.. L. Morrison (eds.), Predicting Species Occurrences: Issues of Accuracy and Scale, Island Press, Washington DC.

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

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

Olden, J. D., D. A. Jackson and P. R. Peres-Neto. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. Transactions of the American Fisheries Society 131: 329-336.

Page, L.M. and B.M. Burr, 1991. A field guide to freshwater fishes of North America north of Mexico. Houghton Mifflin Company, Boston. 432 pp.

Paller, M. 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. North American Journal of Fisheries Management 15: 110-120.

Pearce, J. and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modeling 133: 225-245.

Pearson D. M. 2002. The application of local measures of spatial autocorrelation for describing pattern in north Australian landscapes. Journal of Environmental Management 64:85-95.

Peterson, J. T. and J. Dunham. 2003. Combining inferences from models of capture efficiency, detectability, and suitable habitat to classify landscapes for conservation of threatened bull trout. Conservation Biology 17: 1070-1077.

Peterson, J. T. and C. Rabeni. 1995. Optimizing sampling effort for sampling warmwater stream fish communities. North American Journal of Fisheries Management 15: 528-541.

Poff, N. L., J. D. Allen, M. A. Palmer, D. D. Hart, B. D. Richter, A. H. Arthington, K. H. Rogers, J. L. Meyer and J. A. Stanford. 2003. River flows and water wars: emerging science for environmental decision making. Frontiers environmental ecology 1 (6): 298-306.

Pollock, K.H., J.D. Nichols, T.R. Simons, G.L. Farnsworth, L. L. Bailey, and J.R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13: 1-15.

Porter, M. S., J. Rosenfeld and E. A. Parkinson. 2000. Predictive models of fish species distribution in the Blackwater Drainage, British Columbia. North American Journal of fisheries management 20: 349-359.

Rashleigh, B., R. Parmar, J. M. Johnston and M. C. Barber. 2005. Predictive habitat models for the occurrence of stream fishes in the mid-Atlantic highlands. North American Journal of fisheries management 25: 1353-1366.

Royle, J. A. and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84 (3): 777-790.

Rushton, S. P., S. J. Ormerod and G. Kerby. 2004. New paradigms for modeling species distributions. Journal of Applied Ecology 41: 193-200.

Schmidt, B.R. 2003. Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians. Comptes Rendus Biologies 326: 119-124.

Schlosser, I. J. 1987. The role of predation in age-related and size-related habitat use by stream fishes. Ecology 68 (3): 651-659.

Schoener, T. W. 1974. Resource partitioning in ecological communities. Science 185: 27-39.

Scott, J.M., F. Davis, B. Csuti, R. Noss, B.Butterfield, C. Groves, H. Anderson, S. Caicco, F.D. D'ercjoa, T.C. Edwards, J. Ulliman, and G. Wright. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs Number 123.

Scott, J.M., M. Jennings, R.G. Wright, and B. Csuti. 1996. Landscape approaches to mapping biodiversity. BioScience 46: 77-78.

Scott, J.M., P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Rafael, W.A. Wall, and F.B. Samson, editors. 2002. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C., USA.

Segurado P., and M. B. Araujo. 2004. An evaluation of methods for modeling species distributions. Journal of Biogeography 31:1555-1568.

Selmi, S. and T. Boulinier. 2003. Does time of season influence bird species number determined from point-count data? A capture-recapture approach. Journal of Field Ornithology 74 (4): 349-356.

Shearer, J. and C. Berry. 2003. Fish community persistence in eastern North and South Dakota rivers. Great Plains Research 13:139-159.

Short, H.L. and J.B. Hestbeck. 1995. National biotic resource inventories and GAP analysis. Bioscience 45: 535-539.

Smith P. A. 1994. Autocorrelation in logistic regression modeling of species distributions. Global Ecology and Biogeography Letters 4:47-61

Smith, V.J., J.A. Jenks, C.R. Berry, Jr., C.J. Kopplin, and D.D. Fecske. 2002. The South Dakota Gap Analysis Project, final report. US Department of the Interior, US Geological Survey, Reston, Virginia.

Sowa, S. P. 1998. Gap Analysis in riverine environments. Gap Analysis Bulletin 7: 18-20.

Sowa S. P., D. D. Diamond, R. Abbitt, G. M. Annis, T. Gordon, M. E. Morey, G. R. Sorenson, and D. True. 2005. A GAP analysis for riverine ecosystems of Missouri. Final report to USGS National Gap Analysis Project. 1657 pp. http://www.cerc.usgs.gov/morap/projects.asp.

SPSS. 2001. AnswerTree® 3.0 User's Guide. SPSS Incorporated. Chicago, IL

Stauffer, H.B. C.J. Ralph, and S.L. Miller. 2002. Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. *In* Predicting species occurrences. Issues of accuracy and scale. J.M. Scott, P.J. Heglund, M.L. Morrison et al. editors. Washington D.C. Island Press: 357-365.

Stockwell, D. R. and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modeling 148: 1-13.

Sylvester, R.M. 2004. Upper Missouri River Basin aquatic GAP fish distribution model accuracy assessment and white sucker, Catostomus commersonii, population characteristics in the upper Missouri river basin. Masters thesis. South Dakota State University. Brookings, SD.

Tate, W., R. Myers and J. Estes. 2003. Comparison of Electrofishing and Rotenone for Sampling Largemouth Bass in Vegetated Areas of Two Florida Lakes. North American Journal of Fisheries Management 23 (1): 181-188.

ter Braak, C. J. F. and Prentice, I. C. 1988. A theory of gradient analysis. Advances in Ecological Research 18: 271-317

The Nature Conservancy Freshwater Initiative. 2000. GIS tools for aquatic macrohabitat classification. Internet: http://www.freshwaters.org/ccwp/home.html, (Accessed February 14, 2000).

Turner, G., E. Irwin, J. T. Peterson, B. Freeman, M. Freeman, E. Kramer, C. Straight, A. Wimberley, D. Moosman, C. Hayer and J. C. Ruiz. 2004. Aquatic GAP: A GIS and modeling approach for analysis of biodiversity in the Tallapoosa River Basin, Alabama and Georgia. GIS Symposium.

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

U.S. Department of Agriculture. 1995. State soil geographic (STATSGO) database – data use information Miscellaneous Publication Number 1492. National Cartography and GIS Center, USDA, Natural Resources Conservation Service. Fort Worth, TX.

U.S. Fish and Wildlife Service. 2001. Endangered and threatened wildlife and plants; 12-month finding for a petition to list the sicklefin chub (*Macrhybopsis meeki*) and the sturgeon chub (*Macrhybopsis gelida*) as endangered. Federal Register 66:19910-19914.

U.S. Geological Survey /Environmental Protection Agency. 2002. National Hydrography Dataset. 1400 Independence Road, Rolla, MO 65401 http://nhd.usgs.gov/index.html (Accessed October 2002).

Wall, S.S., C.M. Blausey, J.A. Jenks, and C.R.Berry, Jr. 2001. Topeka shiner (*Notropis topeka*) population status and habitat conditions in South Dakota streams. South Dakota State University, Brookings SD.

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.

Williams, J. E., C. Wood, and M. Dombeck. 1997. Watershed restoration: Principles and practices. American Fisheries Society, Bethesda, Maryland.

Wintle, B. A., M. A. McCarthy, K. M. Parris and M. A. Burgman. 2004. Precision and bias of methods for estimating point survey detection probabilities. Ecological Applications 14 (3): 703-712.

Wissmar, R.C., and P. A. Bisson, editors. 2003. Strategies for restoring river ecosystems: sources of variability and uncertainty in natural and management systems. American Fisheries Society, Bethesda, Maryland.

**Appendix 1: A brief description of the ten species of concern in South Dakota targeted within this study.**

**Central Mudminnow *Umbra limi***

The central mudminnow is common throughout the St. Lawrence-Great Lakes region, north onto the Canadian Shield, west into Manitoba and the Red River Basin, and south into Missouri and southern Illinois. In South Dakota isolated populations have recently been recorded in Blue Dog Lake, Owen's Creek (a tributary to Blue Dog Lake), Lake Cochrane, and the North Fork of the Yellowbank River (Figures 5,6). The species was also recorded in a tributary to Sixmile Creek in 1952, but recent surveys have not produced the fish.

**Northern Redbelly Dace *Phoxinus eos***

The northern redbelly dace is fairly common throughout the upper Great Lakes region east to the St. Lawrence and Atlantic drainages in New England. This species is less common in the upper Missouri River basin and Mackenzie drainage of British Columbia/Northwest Territories. Several scattered populations remain in South Dakota (Figures 7,8, 32). A glacial relict population occurs in several streams of the Sand Hills region of Nebraska and extreme southern South Dakota, similar to the blacknose shiner. This population of northern redbelly dace is disjunct from any other population of this species. A few peripheral populations to the Great Lakes region have been documented in tributaries to the Big Sioux River and Minnesota River. This species also has been recorded in Crow Creek in 1933 and Blue Blanket Creek in 1973, two tributaries that flow into the Missouri River off the eastern edge of the Missouri Cotteau (Bailey and Allum 1962). More recently a population of northern redbelly dace was documented in Stink Creek, a tributary to the Grand River (Morey and Berry 2004). This stream was characterized as having an abundance of aquatic macrophytes and a heavy groundwater influence, typical of other northern redbelly dace streams in South Dakota. Since other Grand River tributaries share these characteristics, it is likely more northern redbelly dace populations exist in this region as evidenced by their presence in the Cannonball, Knife and Heart Rivers in North Dakota.

**Finescale Dace *Phoxinus neogaeus***

The finescale dace exhibits a wide distribution from the St. Lawrence drainage of New England west through the upper Great Lakes region north to the Mackenzie drainage and Arctic Circle. Disjunct populations have been documented in the upper Missouri drainage, Sand Hills of Nebraska and South Dakota, and Black Hills of South

Dakota (Figures 21, 22). Currently the only viable population of finescale dace in South Dakota occurs in Cox Lake north of the Black Hills. Current management efforts are underway to reintroduce the finescale dace to other historic locations throughout the state. Hybridization does occur with northern redbelly dace and may complicate species identification in some areas.

**Sturgeon Chub *Macrhybopsis gelida***

The sturgeon chub is endemic to the Missouri River and its tributaries. Sturgeon chub has declined from 45% of its historical range in the Missouri and Yellowstone rivers (U.S. Fish and Wildlife Service 2001). In South Dakota viable populations still exist in the White River, Cheyenne River, and Little Missouri River (Figures 26,27). The species was also recorded in the Grand River near its confluence with the Missouri River in 1952 (Bailey and Allum 1962). This species prefers turbid waters with swift currents and has declined through much of its range due to impoundment of the Missouri River.

**Pearl Dace *Margariscus margarita***

The pearl dace share a similar distribution to the finescale and northern redbelly dace, ranging from the St. Lawrence and Atlantic drainages throughout the upper Great Lakes region and scattered across Canada. The pearl dace also exists in the upper Missouri River basin of Montana and North Dakota. Two disjunct populations occur, one in the tributaries to the Chesapeake Bay area of Maryland and Virginia, and the other in the Sand Hills region of Nebraska and southern South Dakota. A handful of streams (including LaCreek National Wildlife Refuge) in Todd, Tripp, and Shannon counties contain the only known populations of pearl dace in South Dakota (Figures 33, 34).

**Hornyhead chub *Nocomis biguttatus***

The hornyhead chub is located in the Minnesota drainage and has been reported in the Red and Big Sioux River systems in South Dakota (Figure 15, 17; Hoagstrom and Berry 2006, unpublished data). It is widely distributed in previously glaciated areas including the Great Lakes basin, upper Mississippi basin and Southern part of Ohio basin.

**Carmine shiner *Notropis percobromus***

The carmine shiner is currently found in South Dakota in the Minnesota River Basin (Figures 16, 18), but has been reported from the Big Sioux and Red River drainages (Hoagstrom and Berry 2006, unpublished data). The native range includes the Atlantic Slope from St. Lawrence River drainage, Quebec, to James River drainage,

Virginia. Also includes the Red River and Mississippi River basins from Quebec to Manitoba and south to North Carolina, Alabama, Arkansas and Oklahoma (Page and Burr 1991). The carmine shiner prefers clear and cool fast flowing streams with clean gravel bottoms and is intolerant of turbid or silty water.

**Longnose Sucker *Colostomies catostomus***

The longnose sucker is perhaps the most widespread sucker in the northern United States and Canada. Ranging from the St. Lawrence and Atlantic drainages in New England west throughout the upper Great Lakes region, throughout Canada and into Alaska. The longnose sucker inhabits a variety of clear, coldwater streams in the Rocky Mountain region of the upper Missouri and Platte River basins. In South Dakota the longnose sucker exists in several streams in the northern Black Hills region (Figures 23, 24), where it is thought to be an isolated and disjunct population. Historic records indicate this species occurred in French Creek and Castle Creek. Land use changes and water quality problems have likely reduced this species' range in the Black Hills.

**Trout-Perch *Percopsis omiscomaycus***

The range of the trout-perch extends throughout most of Canada, the Great Lakes region, upper Mississippi River basin, and eastern portion of the Missouri River basin. In South Dakota the trout-perch occurs in the Big Sioux River and its tributaries, including Lake Kampeska and Lake Pelican (Figures 9, 11). These populations are peripheral to populations in Minnesota and Iowa where the species is more common.

**Banded Killifish *Fundulus diaphanus***

The banded killifish occurs in drainages of the middle and upper Atlantic Coast states, Great Lakes region west into eastern North and South Dakota. The species has only been reported in several eastern glacial lakes within South Dakota (Figures 10, 12). It was once thought to be an isolated population, but more recent information indicates the distribution is peripheral as the banded killifish is much more common in Minnesota.

**Appendix 2.1:** Relative gradient variables used for analysis in fish-habitat models

| | Slope | | | | | |
|---|---|---|---|---|---|---|
| Major Drainage | Headwater | Creek | Small River | Large River | Extra Large River | Gradient |
| East River | <0.006 | < 0.002 | | < 0.001 | | 1 |
| | ≥ 0.006 < 0.0100 | ≥ 0.002 < 0.004 | | ≥ 0.001 < 0.002 | | 2 |
| | ≥ 0.010 | ≥ 0.004 | | ≥ 0.002 | | 3 |
| Missouri Plains | < 0.010 | < 0.004 | < 0.003 | < 0.002 | < 0.0003 | 1 |
| Bad River | ≥ 0.010 < 0.018 | > 0.004 < 0.009 | ≥ 0.003 < 0.006 | ≥ 0.002 < 0.005 | ≥ 0.0003 < 0.0006 | 2 |
| | ≥ 0.018 | ≥ 0.009 | ≥ 0.006 | ≥ 0.005 | ≥ 0.0006 | 3 |
| Grand/Moreau | < 0.01 | < 0.005 | < 0.004 | < 0.005 | | 1 |
| | ≥ 0.01 < 0.016 | ≥ 0.005 < 0.08 | ≥ 0.004 < 0.008 | ≥ 0.005 < 0.009 | | 2 |
| | ≥ 0.016 | ≥ 0.008 | ≥ 0.008 | ≥ 0.009 | | 3 |
| Cheyenne/Belle Fourche | < 0.017 | < 0.008 | < 0.006 | < 0.005 | | 1 |
| | ≥ 0.017 < 0.028 | ≥ 0.008 < 0.013 | ≥ 0.006 < 0.010 | > 0.005 < 0.009 | | 2 |
| | ≥ 0.028 | ≥ 0.013 | ≥ 0.010 | > 0.009 | | 3 |
| White | < 0.014 | < 0.006 | | < 0.004 | | 1 |
| | ≥0.014 < 0.023 | ≥ 0.006 < 0.011 | | ≥ 0.004 < 0.007 | | 2 |
| | ≥ 0.023 | ≥ 0.011 | | ≥ 0.007 | | 3 |
| Minnesota River | < 0.039 | < 0.003 < 0.002 | | | | 1 |
| | ≥0.039 < 0.049 | ≥ 0.003 < 0.006 ≥ 0.002 < 0.003 | | | | 2 |
| | ≥ 0.049 | ≥ 0.006 ≥ 0.003 | | | | 3 |

**Appendix 2.2:** Relative elevation variables used for analysis in fish-habitat models

| | DEM | | | | | |
|---|---|---|---|---|---|---|
| Major Drainage | Headwater | Creek | Small River | Large River | Extra Large River | Elevation |
| East River/ | < 439 | < 417 | < 408 | < 348 | | 1 |
| | ≥ 439 < 542 | ≥ 417 < 499 | ≥ 408 < 473 | ≥ 348 < 363 | | 2 |
| | ≥ 542 | ≥ 499 | ≥ 473 | ≥ 363 | | 3 |
| Missouri River/ Bad River | < 521 | < 493 | < 487 | < 420 | | 1 |
| | ≥ 521 < 709 | ≥ 493 < 655 | ≥ 487 < 627 | ≥ 420 <490 | | 2 |
| | ≥ 709 | ≥ 655 | ≥ 627 | ≥ 490 | | 3 |
| White River | < 770 | < 734 | < 659 | < 542 | | 1 |
| | ≥ 770 < 1104 | ≥ 734 < 998 | ≥ 659 < 835 | ≥ 542 < 667 | | 2 |
| | ≥ 1104 | ≥998 | ≥ 835 | ≥ 667 | | 3 |
| Cheyenne/Belle Fourche | < 0.017 | < 0.008 | < 0.006 | < 0.005 | 1 | Cheyenne |
| | ≥ 0.017 < 0.028 | ≥ 0.008 < 0.013 | ≥ 0.006 < 0.010 | > 0.005 < 0.009 | 2 | |
| | ≥ 0.028 | ≥ 0.013 | ≥ 0.010 | > 0.009 | 3 | |
| Grand/Moreau | < 704 | < 663 | < 638 | < 553 | | 1 |
| | ≥ 704 < 915 | ≥ 663 < 829 | ≥ 638 < 770 | ≥ 553 < 615 | | 2 |
| | ≥ 915 | ≥ 829 | ≥ 770 | ≥ 615 | | 3 |
| Minnesota River | < 545 | < 356 | < 338 | | | |
| | ≥ 545 < 740 | ≥ 356 < 380 | ≥ 338 < 381 | | | |
| | ≥ 740 | ≥ 380 | ≥ 381 | | | |

**Appendix 2.3:** Gradient and elevation by EDU used for analysis in fish-habitat models

| EDU | *Slope* | Gradient | DEM | Elevation |
|---|---|---|---|---|
| Upper James River – Great Plains Steppe | < 0.005 | 1 | < 445 | 1 |
| | ≥ 0.005 < 0.010 | 2 | ≥ 445 < 486 | 2 |
| | ≥ 0.010 | 3 | ≥ 486 | 3 |
| James River – Great Plains Steppe/Palouse Dry Steppe | < 0.004 | 1 | < 427 | 1 |
| | ≥ 0.004 < 0.008 | 2 | ≥ 427 < 491 | 2 |
| | ≥ 0.008 | 3 | ≥ 491 | 3 |
| Big Sioux/Vermillion River – Prairie Parkland | < 0.004 | 1 | < 450 | 1 |
| | ≥ 0.004 < 0.008 | 2 | ≥ 450 < 503 | 2 |
| | ≥ 0.008 | 3 | ≥ 503 | 3 |
| | | | | |
| Big Sioux Cotteau/Prairie Parkland | < 0.004 | 1 | < 450 | 1 |
| | ≥ 0.004 < 0.008 | 2 | ≥ 450 < 503 | 2 |
| | ≥ 0.008 | 3 | ≥ 503 | 3 |
| | | | | |
| Missouri River – Great Plains Steppe/ Palouse Dry Steppe/Prairie Parkland | < 0.009 | 1 | < 522 | 1 |
| | ≥ 0.009 < 0.016 | 2 | ≥ 522 < 625 | 2 |
| | ≥ 0.016 | 3 | ≥ 625 | 3 |
| White River – Great Plains Steppe | < 0.012 | 1 | < 741 | 1 |
| | ≥ 0.012 < 0.021 | 2 | ≥ 741 < 888 | 2 |
| | ≥ 0.021 | 3 | ≥ 888 | 3 |
| Cheyenne River/Belle Fourche – Great Plains Dry Palouse/Black Hills | < 0.014 | 1 | < 977 | 1 |
| | ≥ 0.014 < 0.024 | 2 | ≥ 977 < 1228 | 2 |
| | ≥ 0.024 | 3 | ≥ 1228 | 3 |
| Grand/Moreau – Great Plains Palouse Dry Steppe | < 0.009 | 1 | < 720 | 1 |
| | ≥ 0.009 < 0.016 | 2 | ≥ 720 < 818 | 2 |
| | ≥ 0.016 | 3 | ≥ 818 | 3 |
| | | | | |
| Minnesota River – Prairie Parkland | < 0.010 | 1 | < 410 | 1 |
| | ≥ 0.010 < 0.019 | 2 | ≥ 410 < 476 | 2 |
| | ≥ 0.019 | 3 | ≥ 476 | 3 |

**Appendix 2.4:** Gradient and elevation by region used for analysis in fish-habitat models

| Region | Slope | Gradient | DEM | Elevation |
|---|---|---|---|---|
| East Missouri River Region | < 0.006 | 1 | < 439 | 1 |
| | ≥ 0.006 < 0.010 | 2 | ≥ 439 < 542 | 2 |
| | ≥ 0.010 | 3 | ≥ 542 | 3 |
| West Missouri River Region | < 0.013 | 1 | < 922 | 1 |
| | ≥ 0.013 < 0.021 | 2 | ≥ 922 < 1510 | 2 |
| | ≥ 0.021 | 3 | ≥ 1510 | 3 |
| Minnesota River Region | < 0.010 | 1 | < 410 | 1 |
| | ≥ 0.010 < 0.019 | 2 | ≥ 410 < 476 | 2 |
| | ≥ 0.019 | 3 | ≥ 476 | 3 |

**Appendix 2.5:** Expanded Stream Sizes from Shreve Link used for analysis in fish-habitat models

| **Stream Size** | **Link #** | **Strmsize10** |
|---|---|---|
| Headwater1 (plains) | 1-4 | 1 |
| Headwater2 (plains) | 5-9 | 2 |
| Headwater1 (Black Hills) | 1-2 | 1 |
| Headwater2 (Black Hills) | 3-4 | 2 |
| Creek1 (plains) | 10-42 | 3 |
| Creek2 (plains) | 43-75 | 4 |
| Creek1 (Black Hills) | 5 -40 | 3 |
| Creek2 (Black Hills) | 41-75 | 4 |
| Small River1 | 76-788 | 5 |
| Small River2 | 789-1500 | 6 |
| Large River1 | 1501-5751 | 7 |
| Large River2 | 5752-10,000 | 8 |
| Extra Large River | > 10,000 | 9 |

**Appendix 3:** Independent sampling study site locations collected by the "rare fish study"

| STREAM | X-COORD | Y-COORD |
|---|---|---|
| Owens Creek | -97.21133 | 45.34953 |
| Unknown Creek S of Ortley | -97.18856 | 45.32482 |
| Gary Creek | -96.46202 | 44.79047 |
| Mob Creek | -96.59913 | 45.03175 |
| Big Sioux River, Egan | -96.62935 | 44.00806 |
| Big Sioux River, Flandreau | -96.58855 | 44.07911 |
| Split Rock Creek, Leo | -96.57028 | 43.65053 |
| Split Rock Creek, Olsen | -96.49123 | 43.75416 |
| Big Sioux River | -96.76033 | 44.18115 |
| Medary Creek | -96.76503 | 44.19469 |
| Skunk Creek | -96.79872 | 43.53462 |
| Big Sioux River, N Watertown | -97.16163 | 44.95097 |
| Monighan Creek | -96.47904 | 44.81294 |
| Unknown Creek N WPA N o | -96.52593 | 44.60055 |
| Unnamed Creek | -97.27451 | 45.69951 |
| Unnamed Creek | -97.26928 | 45.70911 |
| Willow Creek | -97.01348 | 44.91728 |
| Big Sioux River | -97.10576 | 44.87873 |
| Big Sioux River | -97.16116 | 44.97674 |
| Unnamed Creek | -97.10255 | 45.61845 |
| Big Sioux River | -96.78472 | 44.19549 |
| Big Sioux River | -96.87995 | 44.31091 |
| Dolph Creek | -97.24898 | 44.59127 |
| Big Sioux River | -96.95132 | 44.60163 |
| Bell Fourche River | -103.59458 | 44.67345 |
| Beaver Creek | -104.00373 | 44.38252 |
| Beaver Creek | -104.01169 | 44.38064 |
| Cold Creek | -104.06304 | 44.15732 |
| Cold Creek | -104.03957 | 44.15363 |
| Spring Creek | -103.40960 | 43.98879 |
| Spring Creek | -103.42497 | 43.98537 |
| Spring Creek | -103.44166 | 43.98131 |
| Spring Creek | -103.54649 | 43.94017 |
| Spring Creek | -103.56222 | 43.93206 |
| Battle Creek | -103.41895 | 43.89797 |
| Battle Creek | -103.40079 | 43.89749 |
| Iron Creek | -103.34453 | 43.87378 |
| Battle Creek | -103.26672 | 43.83755 |
| Beaver Creek | -103.51728 | 43.61352 |
| Beaver Creek | -103.35028 | 43.51725 |

**Appendix 4:** List of sampling locations from additional sources used in independent data set to access accuracy of fish distribution models.

| Source | STREAM | X_COORD | Y_COORD |
|---|---|---|---|
| SD GF&P Minnesota River Basin Survey | Yellowbank River, N. Fk | -96.66104 | 45.16706 |
| SD GF&P Minnesota River Basin Survey | Yellowbank River, N. Fk | -96.76194 | 45.12946 |
| SD GF&P Minnesota River Basin Survey | Yellowbank River, N. Fk | -96.48487 | 45.16760 |
| SD GF&P Minnesota River Basin Survey | Whetstone River, S. Fk | -96.76224 | 45.22766 |
| SD GF&P Minnesota River Basin Survey | Whetstone River, S. Fk | -96.74264 | 45.22736 |
| SD GF&P Minnesota River Basin Survey | Yellowbank River, S. Fk | -96.66285 | 44.99296 |
| SD GF&P Minnesota River Basin Survey | Yellowbank River, S. Fk | -96.62414 | 45.02656 |
| SD GF&P Minnesota River Basin Survey | Yellowbank River, S. Fk | -96.51654 | 45.04294 |
| SD GF&P Minnesota River Basin Survey | Monigan Creek | -96.51473 | 44.80656 |
| SD GF&P Minnesota River Basin Survey | Monigan Creek | -96.52072 | 44.80366 |
| SD GF&P Minnesota River Basin Survey | Cobb Creek | -96.54074 | 44.73295 |
| SD GF&P Minnesota River Basin Survey | Cobb Creek | -96.50623 | 44.74826 |
| SD GF&P Minnesota River Basin Survey | Cobb Creek | -96.48043 | 44.74265 |
| SD GF&P Minnesota River Basin Survey | Crow Timber Creek | -96.55883 | 44.90606 |
| SD GF&P Minnesota River Basin Survey | Gary Creek | -96.46542 | 44.78887 |
| SD GF&P Minnesota River Basin Survey | Sica Hollow (Standfast) | -97.23167 | 45.74015 |
| SD GF&P Minnesota River Basin Survey | Monigan Creek | -96.47244 | 44.81535 |
| EMAP | Moreau River S. Fk | -102.84173 | 45.13343 |
| EMAP | Grand River S. Fk | -103.34156 | 45.61845 |
| EMAP | Clarks Fork Creek | -103.38092 | 45.54903 |
| EMAP | Grand River S. Fk | -103.66366 | 45.55743 |
| EMAP | Pass Creek | -101.46860 | 43.69060 |
| USGS | Blackpipe Creek, | -101.16746 | 43.45970 |
| USGS | Blackpipe Creek, Low | -101.22741 | 43.77624 |
| EMAP | Blackpipe Creek | -101.21529 | 43.63970 |
| EMAP | Eagle Nest Creek | -101.77920 | 43.68012 |
| EMAP | Bear-in-the-lodge Cr | -101.80576 | 43.56054 |
| EMAP | Plum Creek | -101.46237 | 44.59970 |
| EMAP | High Bank Creek | -100.90623 | 45.62580 |
| EMAP | Oak Creek | -100.61474 | 45.73478 |
| EMAP | Whetstone River, S. | -96.64024 | 45.22736 |
| EMAP | Whetstone River, S. | -96.56055 | 45.25566 |
| EMAP | Whetstone River, S. | -96.72213 | 45.22887 |
| EMAP | Yellowbank River, N. | -96.70234 | 45.16156 |
| EMAP | Yellowbank River, N. | -96.46093 | 45.20366 |
| EMAP | Yellowbank River, N. | -96.81705 | 45.13396 |
| EMAP | Flandreau Creek | -96.46486 | 44.07447 |
| EMAP | Little White River | -100.90123 | 43.26534 |
| EMAP | Coffee Creek | -101.10364 | 43.10853 |
| EMAP | Little White River | -100.76452 | 43.48089 |
| EMAP | Vermillion River, W. | -97.35517 | 43.54569 |
| Upper Missouri River Basin Aquatic Gap Project | Maple River, | -98.46904 | 45.89398 |
| Upper Missouri River Basin Aquatic Gap Project | Elm River, S | -98.60039 | 45.77671 |
| Upper Missouri River Basin Aquatic Gap Project | Elm River, S | -98.65140 | 45.80821 |
| Upper Missouri River Basin Aquatic Gap Project | Maple River, | -98.53031 | 45.79118 |
| Upper Missouri River Basin Aquatic Gap Project | Elm River, S | -98.51818 | 45.66578 |
| SDSU SD Stream Project | Whitewood Creek | -103.53120 | 44.56465 |
| SDSU SD Stream Project | Beaver Creek | -103.30640 | 43.49110 |
| SDSU SD Stream Project | Belle Fourche R | -102.56527 | 44.37030 |
| SDSU SD Stream Project | Upper Cheyenne | -103.08048 | 43.49833 |
| SDSU SD Stream Project | Lower Cheyenne | -101.98898 | 44.52320 |
| SD DOT Project | East Fork Vermillion River | -97.26054 | 43.87349 |

**Appendix 4 continued:** List of sampling locations from additional sources used in independent data set to access accuracy of fish distribution models.

| Source | STREAM | X_COORD | Y_COORD |
|---|---|---|---|
| SD DOT Project | Little Vermillion River | -97.23920 | 43.73806 |
| SD DOT Project | Little Vermillion River | -97.30188 | 43.76282 |
| SD DOT Project | Little Vermillion River | -97.25104 | 43.75021 |
| SD DOT Project | Little Vermillion River | -97.43379 | 43.90764 |
| SD DOT Project | East Fork Vermillion River | -97.28155 | 43.92085 |
| SD DOT Project | East Fork Vermillion River | -97.28884 | 43.93578 |
| SD DOT Project | East Fork Vermillion River | -97.19206 | 43.74299 |
| SD DOT Project | West Branch Skunk Creek | -97.01460 | 43.71999 |
| SD DOT Project | West Branch Skunk Creek | -96.92504 | 43.66374 |
| SD DOT Project | Strayhorse Creek | -96.90802 | 44.85062 |
| SD DOT Project | Strayhorse Creek | -96.95296 | 44.73520 |
| SD DOT Project | West Branch Skunk Creek | -96.98644 | 43.70214 |
| SD DOT Project | East Fork Vermillion River | -97.16321 | 43.64203 |
| SD GF&P Topeka shiner monitoring project | Peg Murky Run | -96.80466 | 44.58220 |
| SD GF&P Topeka shiner monitoring project | Peg Murky Run | -96.78801 | 44.58377 |
| SD GF&P Topeka shiner monitoring project | Peg Murky Run | -96.76738 | 44.58578 |
| SD GF&P Topeka shiner monitoring project | Six-mile Creek | -96.78828 | 44.32713 |
| SD GF&P Topeka shiner monitoring project | Six-mile Creek | -96.64494 | 44.44448 |
| SD GF&P Topeka shiner monitoring project | West Fork Vermillion River | -97.42210 | 43.73246 |
| SD GF&P Topeka shiner monitoring project | West Fork Vermillion River | -97.41801 | 43.71416 |
| SD GF&P Topeka shiner monitoring project | West Fork Vermillion River | -97.35309 | 43.54051 |
| SD GF&P Topeka shiner monitoring project | Middle Pearl Creek | -98.01176 | 44.28750 |
| SD GF&P Topeka shiner monitoring project | Middle Pearl Creek | -97.99182 | 44.29632 |
| SD GF&P Topeka shiner monitoring project | Middle Pearl Creek | -98.02625 | 44.28224 |
| SD GF&P Topeka shiner monitoring project | Six-mile Creek | -96.63145 | 44.50729 |
| SD GF&P Topeka shiner monitoring project | Strayhorse Creek | -96.95431 | 44.73292 |
| SD GF&P Topeka shiner monitoring project | Strayhorse Creek | -96.90880 | 44.84930 |
| SD GF&P Topeka shiner monitoring project | Strayhorse Creek | -96.92306 | 44.80460 |
| SD GF&P Topeka shiner monitoring project | West Pipestone Creek | -96.54017 | 43.73971 |
| SD GF&P Topeka shiner monitoring project | West Pipestone Creek | -96.57742 | 43.82782 |
| SD GF&P Topeka shiner monitoring project | West Pipestone Creek | -96.57319 | 43.66569 |
| SD GF&P Topeka shiner monitoring project | Turkeyridge Creek | -97.37940 | 43.30216 |
| SD GF&P Topeka shiner monitoring project | Turkeyridge Creek | -97.27745 | 43.27202 |
| SD GF&P Topeka shiner monitoring project | Turkeyridge Creek | -97.07340 | 43.19752 |
| SD GF&P Topeka shiner monitoring project | Enemy Creek | -98.08804 | 43.63549 |
| SD GF&P Topeka shiner monitoring project | Enemy Creek | -98.06305 | 43.63185 |
| SD GF&P Topeka shiner monitoring project | Enemy Creek | -98.05160 | 43.63180 |
| Vermillion Basin Water Development District | East Fork Vermillion River | -97.24887 | 43.86499 |
| Vermillion Basin Water Development District | Little Vermillion River | -97.20900 | 43.71515 |
| Vermillion Basin Water Development District | East Fork Vermillion River | -97.18060 | 43.69196 |
| Vermillion Basin Water Development District | East Fork Vermillion River | -97.10976 | 43.44536 |
| Vermillion Basin Water Development District | West Fork Vermillion River | -97.20517 | 43.41564 |
| Vermillion Basin Water Development District | Camp Creek | -97.02010 | 43.37339 |
| Vermillion Basin Water Development District | Long Creek | -96.90427 | 43.19979 |
| Vermillion Basin Water Development District | Vermillion River | -96.98109 | 43.11184 |
| Vermillion Basin Water Development District | Turkey Ridge Creek | -96.98834 | 43.12658 |
| Vermillion Basin Water Development District | Frog Creek | -97.01377 | 43.01110 |
| Vermillion Basin Water Development District | Vermillion River | -96.94417 | 43.04034 |
| Vermillion Basin Water Development District | Vermillion River | -96.94238 | 42.95339 |
| Vermillion Basin Water Development District | Clay Creek Ditch | -96.98123 | 42.82267 |
| Vermillion Basin Water Development District | Vermillion River | -96.92442 | 42.81737 |

**Appendix 5:** List of species collected in five ecological drainage units (EDUs) with corresponding prevalence (frequency of occurrence) from independent data set. A dash indicates that the species does not occur or was not predicted within that EDU.

| | | | | Ecological Drainage Unit (EDU) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Family** | **Species Code** | **Common Name** | **Scientific Name** | **East River** | **Minnesota River** | **Black Hills** | **Cheyenne River** | **Grand Moreau River** | **White River** |
| Lepisosteidae | P_SHG | Shortnose gar | *Lepisosteus osseus* | 0.043 | - | - | - | - | - |
| Clupeidae | P_GZD | Gizzard shad | *Dorosoma cepedianum* | 0.029 | - | - | 0.429 | - | - |
| Hiodontidae | P_GOE | Goldeye | *Hiodon alosoides* | 0.029 | - | - | 0.429 | 0.200 | 0.000 |
| Salmonidae | P_BKT | Brook trout | *Salvelinus fontinalis* | - | - | 0.267 | 0.000 | - | - |
| | P_BNT | Brown trout | *Salmo trutta* | - | 0.045 | 0.267 | 0.143 | - | - |
| | P_RBT | Rainbow trout | *Oncorhynchus mykiss* | - | - | 0.067 | - | - | - |
| Umbridae | P_CEM | Central mudminnow | *Umbra limi* | 0.016 | 0.045 | - | - | - | |
| Esocidae | P_NOP | Northern pike | *Esox lucius* | 0.171 | 0.045 | 0.000 | - | - | 0.000 |
| Cyprinidae | P_BIS | Bigmouth shiner | *Notropis dorsalis* | 0.443 | 0.545 | - | - | - | 0.222 |
| | P_BLD | Blacknose dace | *Rhinichthys atratulus* | 0.194 | 0.818 | - | - | - | - |
| | P_BLM | Bluntnose minnow | *Pimephales notatus* | 0.210 | 0.455 | - | - | - | - |
| | P_BRM | Brassy minnow | *Hybognathus hankinsoni* | 0.571 | 0.591 | - | - | 0.000 | 0.333 |
| | P_STR | Central stoneroller | *Campostoma anomalum* | 0.338 | 0.682 | - | - | - | 0.000 |
| | P_COC | Common carp | *Cyprinus carpio* | 0.600 | 0.091 | 0.000 | 0.429 | 0.400 | 0.000 |
| | P_CNS | Common shiner | *Luxilus cornutus* | 0.629 | 0.636 | - | - | - | |
| | P_CRC | Creek chub | *Semotilus atromaculatus* | 0.657 | 0.773 | 0.667 | 0.714 | 0.400 | 0.222 |
| | P_EMS | Emerald shiner | *Notropis atherinoides* | 0.046 | - | - | 0.000 | 0.200 | - |
| | P_FHM | Fathead minnow | *Pimephales promelas* | 0.943 | 0.727 | 0.200 | 0.714 | 1.000 | 0.111 |
| | P_FID | Finescale dace | *Phoxinus neogaeus* | - | - | 0.000 | - | - | - |
| | P_FLC | Flathead chub | *Platygobio gracilis* | - | - | - | 0.429 | 0.600 | 0.889 |
| | P_GOS | Golden shiner | *Notemigonus crysoleucas* | 0.014 | - | 0.000 | 0.000 | - | - |
| | P_HOC | Hornyhead chub | *Nocomis biguttatus* | - | 0.500 | - | - | - | - |
| | P_LOD | Longnose dace | *Rhinichthys cataractae* | - | - | 0.800 | 0.714 | 0.800 | 0.444 |
| | P_NRD | Northern redbelly dace | *Phoxinus eos* | 0.235 | 0.091 | - | - | 0.000 | 0.000 |
| | P_PLM | Plains minnow | *Hybognathus placitus* | - | - | - | 0.143 | 0.000 | 0.000 |
| | P_RES | Red shiner | *Cyprinella lutrensis* | 0.529 | - | - | 0.571 | - | 0.111 |
| | P_RIC | River carpsucker | *Carpiodes carpio* | 0.114 | - | - | 0.571 | 0.400 | 0.111 |
| | P_ROS | Carmine shiner | *Notropis rubellus* | - | 0.091 | - | - | - | - |
| | P_SAS | Sand shiner | *Notropis stramineus* | 0.671 | 0.045 | - | 0.571 | 0.800 | 0.889 |
| | P_SNC | Sturgeon chub | *Macrhybopsis gelida* | - | - | - | 0.143 | - | 0.000 |
| | P_TOS | Topeka shiner | *Notropis topeka* | 0.338 | - | - | - | - | - |
| Catostomidae | P_BIB | Bigmouth buffalo | *Ictiobus cyprinellus* | 0.043 | - | - | - | - | - |

**Appendix 5 continued:** List of species collected in five ecological drainage units (EDUs) with corresponding prevalence (frequency of occurrence) from independent data set. A dash indicates that the species does not occur or was not predicted within that EDU.

| | | | | Ecological Drainage Unit (EDU) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Family** | **Species Code** | **Common Name** | **Scientific Name** | **East River** | **Minnesota River** | **Black Hills** | **Cheyenne River** | **Grand Moreau River** | **White River** |
| Catostomidae | P_LOS | Longnose sucker | *Colostomies catostomus* | - | - | 0.000 | 0.000 | - | - |
| | P_MOS | Mountain sucker | *Catostomus platyrhynchus* | - | - | 0.067 | 0.286 | - | - |
| | P_SHR | Shorthead redhorse | *Moxostoma macrolepidotum* | 0.129 | - | - | 0.429 | 0.800 | 0.000 |
| | P_SAB | Smallmouth buffalo | *Ictiobus bubalus* | 0.031 | - | - | - | - | - |
| | P_WHS | White sucker | *Catostomus commersonii* | 0.686 | 0.636 | 0.533 | 0.857 | 0.800 | 0.000 |
| Ictaluridae | P_BLB | Black bullhead | *Ameiurus melas* | 0.686 | 0.364 | - | 0.286 | 0.600 | 0.111 |
| | P_CCF | Channel catfish | *Ictalurus punctatus* | 0.229 | - | - | 0.429 | 0.600 | 0.222 |
| | P_STC | Stonecat | *Noturus flavus* | 0.123 | - | - | 0.714 | 0.800 | 0.111 |
| | P_TAM | Tadpole madtom | *Noturus gyrinus* | 0.371 | 0.182 | - | - | - | 0.111 |
| Percopsidae | P_TRP | Trout-perch | *Percopsis omiscomaycus* | 0.091 | - | - | - | - | - |
| Fundulidae | P_BAK | Banded killifish | *Fundulus diaphanus* | 0.000 | 0.000 | - | - | - | - |
| | P_PLK | Plains killifish | *Oncorhynchus mykiss* | - | - | - | 0.429 | - | - |
| | P_PLT | Plains topminnow | *Fundulus sciadicus* | 0.000 | - | 0.000 | 0.286 | - | 0.000 |
| Gasterosteidae | P_BRS | Brook stickleback | *Culaea inconstans* | 0.343 | 0.409 | - | - | - | - |
| Moronidae | P_WHB | White bass | *Morone chrysops* | 0.031 | - | - | 0.000 | - | - |
| Centrarchidae | P_BLC | Black crappie | *Pomoxis nigromaculatus* | 0.086 | - | 0.000 | 0.000 | - | 0.000 |
| | P_BLG | Bluegill | *Lepomis macrochirus* | 0.014 | - | - | 0.143 | - | 0.000 |
| | P_GSF | Green sunfish | *Lepomis cyanellus* | 0.514 | 0.273 | 0.000 | 0.571 | 0.800 | 0.111 |
| | P_LMB | Largemouth bass | *Micropterus salmoides* | 0.057 | 1.000 | - | 0.000 | - | - |
| | P_OSF | Orangespotted sunfish | *Lepomis humilis* | 0.571 | 0.136 | - | 0.286 | 0.200 | - |
| | P_ROB | Rock bass | *Ambloplites Rupestris* | - | 0.045 | 0.045 | 0.143 | - | - |
| | P_SMB | Smallmouth bass | *Micropterus dolomieu* | 0.029 | - | - | 0.286 | - | - |
| Percidae | P_BED | Blackside darter | *Percina maculata* | 0.231 | 0.227 | - | - | - | - |
| | P_IOD | Iowa darter | *Etheostoma exile* | 0.357 | 0.318 | - | 0.000 | 0.000 | 0.000 |
| | P_JOD | Johnny darter | *Etheostoma nigrum* | 0.686 | 0.636 | - | - | 0.000 | - |
| | P_SAR | Sauger | *Stizostedion canadense* | - | - | - | 0.286 | - | 0.000 |
| | P_WAE | Walleye | *Stizostedion vitreum* | 0.014 | 0.000 | - | 0.000 | - | - |
| | P_YEP | Yellow perch | *Perca flavescens* | 0.043 | 0.136 | - | - | 0.000 | - |
| Sciaenidae | P_FRD | Freshwater drum | *Aplodinotus grunniens* | 0.031 | - | - | 0.000 | - | - |

**Appendix 6:** East River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lepisosteidae | P_SHG | 3 | 4 | 0 | 19 | 0.120 | 0.043 | 0.278 | 0.314 | 1.000 | 0.975 | 0.025 | 0.000 | 0.862 | 1.000 | 0.007 | 0.257 |
| Clupeidae | P_GZD | 2 | 1 | 0 | 67 | 0.213 | 0.029 | 0.960 | 0.986 | 1.000 | 0.997 | 0.003 | 0.000 | 0.904 | 1.000 | 0.003 | 0.822 |
| Hiodontidae | P_GOE | 1 | 3 | 1 | 38 | 0.001 | 0.029 | 0.543 | 0.557 | 0.500 | 1.000 | 0.000 | 0.500 | 0.996 | 0.974 | 0.014 | 0.365 |
| Umbridae | P_CEM | 1 | 6 | 0 | 55 | 0.001 | 0.016 | 0.887 | 0.903 | 1.000 | 1.000 | 0.000 | 0.000 | 0.992 | 1.000 | 0.000 | 0.545 |
| Esocidae | P_NOP | 7 | 34 | 5 | 28 | 0.027 | 0.171 | 0.413 | 0.500 | 0.583 | 0.968 | 0.032 | 0.417 | 0.882 | 0.848 | 0.085 | 0.364 |
| Cyprinidae | P_BIS | 31 | 23 | 0 | 59 | 0.225 | 0.443 | 0.302 | 0.671 | 1.000 | 0.756 | 0.244 | 0.000 | 0.857 | 1.000 | 0.074 | 0.532 |
| | P_BLD | 12 | 18 | 0 | 32 | 0.241 | 0.194 | 0.586 | 0.710 | 1.000 | 0.881 | 0.119 | 0.000 | 0.734 | 1.000 | 0.070 | 0.551 |
| | P_BLM | 13 | 9 | 0 | 40 | 0.099 | 0.210 | 0.660 | 0.855 | 1.000 | 0.978 | 0.022 | 0.000 | 0.936 | 1.000 | 0.014 | 0.725 |
| | P_BRM | 40 | 9 | 0 | 1 | 0.298 | 0.571 | 0.138 | 0.586 | 1.000 | 0.104 | 0.896 | 0.000 | 0.822 | 1.000 | 0.123 | 0.311 |
| | P_STR | 20 | 37 | 2 | 6 | 0.158 | 0.338 | 0.182 | 0.400 | 0.909 | 0.506 | 0.494 | 0.091 | 0.774 | 0.750 | 0.121 | 0.288 |
| | P_COC | 20 | 17 | 22 | 11 | 0.240 | 0.600 | 0.215 | 0.443 | 0.476 | 0.730 | 0.270 | 0.524 | 0.831 | 0.333 | 0.372 | 0.195 |
| | P_CNS | 40 | 19 | 4 | 7 | 0.248 | 0.629 | 0.167 | 0.671 | 0.909 | 0.598 | 0.402 | 0.091 | 0.895 | 0.636 | 0.124 | 0.426 |
| | P_CRC | 46 | 23 | 0 | 1 | 0.273 | 0.657 | 0.104 | 0.671 | 1.000 | 0.137 | 0.863 | 0.000 | 0.880 | 1.000 | 0.090 | 0.353 |
| | P_EMS | 3 | 24 | 0 | 38 | 0.120 | 0.046 | 0.629 | 0.631 | 1.000 | 0.929 | 0.071 | 0.000 | 0.510 | 1.000 | 0.044 | 0.412 |
| | P_FHM | 66 | 4 | 0 | 0 | 0.316 | 0.943 | 0.018 | 0.943 | 1.000 | 0.000 | 1.000 | 0.000 | 0.981 | . | 0.018 | 0.392 |
| | P_GOS | 0 | 0 | 1 | 69 | 0.000 | 0.014 | 0.986 | 0.986 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.986 | 0.014 | 0.000 |
| | P_NRD | 2 | 0 | 2 | 13 | 0.213 | 0.235 | 0.765 | 0.882 | 0.500 | 1.000 | 0.000 | 0.500 | 1.000 | 0.867 | 0.118 | 0.605 |
| | P_RES | 35 | 21 | 2 | 12 | 0.223 | 0.529 | 0.297 | 0.671 | 0.946 | 0.577 | 0.423 | 0.054 | 0.799 | 0.857 | 0.154 | 0.461 |
| | P_RIC | 7 | 43 | 1 | 19 | 0.099 | 0.114 | 0.332 | 0.371 | 0.875 | 0.817 | 0.183 | 0.125 | 0.622 | 0.950 | 0.075 | 0.291 |
| | P_SAS | 44 | 15 | 3 | 8 | 0.246 | 0.671 | 0.167 | 0.743 | 0.936 | 0.684 | 0.316 | 0.064 | 0.923 | 0.727 | 0.096 | 0.502 |
| | P_TOS | 22 | 32 | 0 | 3 | 0.184 | 0.338 | 0.137 | 0.385 | 1.000 | 0.337 | 0.663 | 0.000 | 0.789 | 1.000 | 0.091 | 0.275 |
| Catostomidae | P_BIB | 3 | 23 | 0 | 44 | 0.001 | 0.043 | 0.629 | 0.671 | 1.000 | 0.999 | 0.001 | 0.000 | 0.989 | 1.000 | 0.000 | 0.455 |
| | P_SHR | 5 | 21 | 4 | 40 | 0.065 | 0.129 | 0.591 | 0.643 | 0.556 | 0.967 | 0.033 | 0.444 | 0.786 | 0.909 | 0.077 | 0.421 |

**Appendix 6 continued:** East River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Catostomidae | P_SAB | 0 | 43 | 2 | 63 | 0.001 | 0.031 | 0.970 | 0.969 | 0.000 | 0.999 | 0.001 | 1.000 | 0.000 | 0.969 | 0.031 | -0.023 |
| Ictaluridae | P_WHS | 48 | 21 | 0 | 1 | 0.270 | 0.686 | 0.095 | 0.700 | 1.000 | 0.150 | 0.850 | 0.000 | 0.894 | 1.000 | 0.081 | 0.366 |
| | P_BLB | 48 | 21 | 0 | 1 | 0.273 | 0.686 | 0.096 | 0.700 | 1.000 | 0.149 | 0.851 | 0.000 | 0.893 | 1.000 | 0.082 | 0.365 |
| | P_CCF | 16 | 13 | 0 | 40 | 0.169 | 0.229 | 0.617 | 0.814 | 1.000 | 0.949 | 0.051 | 0.000 | 0.879 | 1.000 | 0.031 | 0.679 |
| | P_STC | 8 | 45 | 0 | 12 | 0.039 | 0.123 | 0.227 | 0.308 | 1.000 | 0.813 | 0.187 | 0.000 | 0.743 | 1.000 | 0.043 | 0.262 |
| | P_TAM | 19 | 32 | 7 | 12 | 0.166 | 0.371 | 0.247 | 0.443 | 0.731 | 0.693 | 0.307 | 0.269 | 0.782 | 0.632 | 0.176 | 0.307 |
| Percopsidae | P_TRP | 3 | 14 | 0 | 16 | 0.084 | 0.091 | 0.521 | 0.576 | 1.000 | 0.931 | 0.069 | 0.000 | 0.718 | 1.000 | 0.036 | 0.424 |
| Fundulidae | P_BAK | 0 | 3 | 0 | 59 | 0.000 | 0.000 | 0.952 | 0.952 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.488 |
| | P_PLT | 0 | 5 | 0 | 60 | 0.000 | 0.000 | 0.923 | 0.923 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.480 |
| Gasterosteidae | P_BRS | 19 | 22 | 4 | 22 | 0.246 | 0.343 | 0.409 | 0.612 | 0.826 | 0.802 | 0.198 | 0.174 | 0.778 | 0.846 | 0.141 | 0.458 |
| Moronidae | P_WHB | 2 | 16 | 0 | 46 | 0.120 | 0.031 | 0.737 | 0.738 | 1.000 | 0.960 | 0.040 | 0.000 | 0.510 | 1.000 | 0.030 | 0.451 |
| Centrarchidae | P_BLC | 6 | 7 | 0 | 57 | 0.045 | 0.086 | 0.819 | 0.900 | 1.000 | 0.995 | 0.005 | 0.000 | 0.951 | 1.000 | 0.004 | 0.693 |
| | P_BLG | 1 | 5 | 0 | 64 | 0.001 | 0.014 | 0.914 | 0.929 | 1.000 | 1.000 | 0.000 | 0.000 | 0.993 | 1.000 | 0.000 | 0.564 |
| | P_GSF | 32 | 32 | 4 | 2 | 0.195 | 0.514 | 0.118 | 0.486 | 0.889 | 0.243 | 0.757 | 0.111 | 0.837 | 0.333 | 0.146 | 0.275 |
| | P_LMB | 1 | 27 | 3 | 39 | 0.001 | 0.057 | 0.558 | 0.571 | 0.250 | 0.999 | 0.001 | 0.750 | 0.964 | 0.929 | 0.043 | 0.355 |
| | P_OSF | 36 | 26 | 4 | 4 | 0.222 | 0.571 | 0.140 | 0.571 | 0.900 | 0.409 | 0.591 | 0.100 | 0.862 | 0.500 | 0.140 | 0.333 |
| Percidae | P_SMB | 1 | 0 | 1 | 68 | 0.001 | 0.029 | 0.971 | 0.986 | 0.500 | 1.000 | 0.000 | 0.500 | 1.000 | 0.986 | 0.014 | 0.660 |
| | P_BED | 3 | 3 | 0 | 7 | 0.001 | 0.231 | 0.539 | 0.769 | 1.000 | 0.999 | 0.001 | 0.000 | 0.999 | 1.000 | 0.000 | 0.649 |
| | P_IOD | 19 | 17 | 6 | 28 | 0.187 | 0.357 | 0.445 | 0.671 | 0.760 | 0.898 | 0.102 | 0.240 | 0.857 | 0.824 | 0.131 | 0.510 |
| | P_JOD | 40 | 17 | 8 | 7 | 0.265 | 0.686 | 0.164 | 0.671 | 0.833 | 0.609 | 0.391 | 0.167 | 0.899 | 0.467 | 0.179 | 0.379 |
| | P_WAE | 1 | 27 | 0 | 12 | 0.001 | 0.014 | 0.172 | 0.186 | 1.000 | 0.997 | 0.003 | 0.000 | 0.964 | 1.000 | 0.001 | 0.161 |
| Sciaenidae | P_YEP | 2 | 41 | 1 | 26 | 0.001 | 0.043 | 0.372 | 0.400 | 0.667 | 0.998 | 0.002 | 0.333 | 0.972 | 0.963 | 0.015 | 0.298 |
| | P_FRD | 2 | 3 | 0 | 60 | 0.213 | 0.031 | 0.933 | 0.954 | 1.000 | 0.989 | 0.011 | 0.000 | 0.758 | 1.000 | 0.010 | 0.665 |

**Appendix 7:** Habitat variables associated with presence of fish species of concern used in predictive models

| Fish | Habitat variables | Distribution restrictions | EDU |
|---|---|---|---|
| Central mudminnow | Creeks and headwaters with perennial flow; OR creeks and headwaters with intermittent flow and connected to a lake or slough; OR creeks and headwaters with intermittent flow and groundwater potential | none | Minnesota |
| Central mudminnow | Creeks and headwaters connected1 to a lake or slough | Big Sioux Cotteau | |
| Trout perch | 4th and 5th order streams | Big Sioux River Basin | Big Sioux/Vermillion |
| Northern redbelly dace | Headwaters and creeks with perennial flow; **OR** headwaters and creeks with intermittent flow and medium to high groundwater potential | none | Minnesota River |
| Northern redbelly dace | Segments adjacent to valley segments where fish was recorded | 11-digit HUCs where fish was recorded | Big Sioux/Vermillion |
| Northern redbelly dace | Stream segments with intermittent flow and medium to low groundwater potential; **OR** 2nd to 4th order streams with perennial flow; **OR** 5th order streams with perennial flow and medium to high gradients | 11-digit HUCs where fish was recorded and surrounding HUCs | White River, Missouri River Great Plains, and Grand and Moreau Rivers |
| Pearl dace | Headwaters and creeks with stream orders > 1 | 11-digit HUCs where fish was recorded and surrounding HUCs | White River, Missouri River Parkland and Missouri River Great Plains |
| Finescale dace | Stream segments with intermittent flow, medium to high gradients and stream orders > 1; **OR** stream segments with perennial flow through bedrock geology | 11-digit HUCs where fish was recorded and adjacent 11-digit HUCs | White River and Black Hills |
| Banded killifish | Stream segments connected to lakes and slough; **OR** glaciated lakes in Big Sioux River Basin | Big Sioux Cotteau in Big Sioux River Basin | Minnesota and Big Sioux/Vermilion Rivers |
| Carmine shiner | Creeks and small rivers with perennial flow and no size discrepancy or at confluence of a creek with a small river | none | Minnesota River |
| Hornyhead chub | Headwaters, creeks and small rivers | none | Minnesota River |
| Longnose sucker | Headwaters, creeks and small rivers at medium elevations | none | Black Hills |
| Sturgeon chub | 5th order streams flowing through alluvial outwash; **OR** 6th order streams | none | White River, Cheyenne River, Little Missouri River, Grand and Moreau River, Missouri River Parkland and Missouri River Great Plains |

**Appendix 8:** Minnesota EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Salmonidae | P_BNT | 1 | 0 | 0 | 21 | 0.004 | 0.045 | 0.955 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| Umbridae | P_CEM | 0 | 2 | 1 | 19 | 0.004 | 0.045 | 0.864 | 0.864 | 0.000 | 1.000 | 0.000 | 1.000 | 0.000 | 0.950 | 0.046 | 0.364 |
| Esocidae | P_NOP | 0 | 12 | 1 | 9 | 0.004 | 0.045 | 0.411 | 0.409 | 0.000 | 0.995 | 0.005 | 1.000 | 0.000 | 0.900 | 0.048 | 0.273 |
| Cyprinidae | P_BIS | 12 | 8 | 0 | 2 | 0.521 | 0.545 | 0.280 | 0.636 | 1.000 | 0.324 | 0.676 | 0.000 | 0.742 | 1.000 | 0.189 | 0.366 |
| | P_BLD | 18 | 2 | 0 | 2 | 0.562 | 0.818 | 0.142 | 0.909 | 1.000 | 0.640 | 0.360 | 0.000 | 0.941 | 1.000 | 0.051 | 0.670 |
| | P_BLM | 9 | 6 | 1 | 6 | 0.747 | 0.455 | 0.476 | 0.682 | 0.900 | 0.572 | 0.428 | 0.100 | 0.668 | 0.857 | 0.249 | 0.442 |
| | P_BRM | 12 | 8 | 1 | 1 | 0.398 | 0.591 | 0.190 | 0.591 | 0.923 | 0.239 | 0.761 | 0.077 | 0.790 | 0.500 | 0.190 | 0.288 |
| | P_CNS | 14 | 5 | 0 | 3 | 0.639 | 0.636 | 0.282 | 0.773 | 1.000 | 0.484 | 0.516 | 0.000 | 0.814 | 1.000 | 0.145 | 0.510 |
| | P_COC | 2 | 8 | 0 | 12 | 0.004 | 0.091 | 0.547 | 0.636 | 1.000 | 0.997 | 0.003 | 0.000 | 0.984 | 1.000 | 0.001 | 0.476 |
| | P_CRC | 17 | 3 | 0 | 2 | 0.562 | 0.773 | 0.168 | 0.864 | 1.000 | 0.543 | 0.457 | 0.000 | 0.910 | 1.000 | 0.077 | 0.585 |
| | P_FHM | 15 | 5 | 1 | 1 | 0.511 | 0.727 | 0.161 | 0.727 | 0.938 | 0.281 | 0.719 | 0.063 | 0.855 | 0.500 | 0.161 | 0.327 |
| | P_HOC | 11 | 7 | 0 | 4 | 0.680 | 0.500 | 0.398 | 0.682 | 1.000 | 0.457 | 0.543 | 0.000 | 0.698 | 1.000 | 0.216 | 0.441 |
| | P_NRD | 2 | 8 | 0 | 12 | 0.004 | 0.091 | 0.547 | 0.636 | 1.000 | 0.997 | 0.003 | 0.000 | 0.984 | 1.000 | 0.001 | 0.476 |
| | P_ROS | 2 | 9 | 0 | 11 | 0.004 | 0.091 | 0.502 | 0.591 | 1.000 | 0.997 | 0.003 | 0.000 | 0.983 | 1.000 | 0.002 | 0.448 |
| | P_SAS | 1 | 19 | 0 | 2 | 0.004 | 0.045 | 0.094 | 0.136 | 1.000 | 0.963 | 0.037 | 0.000 | 0.929 | 1.000 | 0.003 | 0.127 |
| | P_STR | 14 | 5 | 1 | 2 | 0.511 | 0.682 | 0.207 | 0.727 | 0.933 | 0.439 | 0.561 | 0.067 | 0.846 | 0.667 | 0.161 | 0.406 |
| Catostomidae | P_WHS | 13 | 7 | 1 | 1 | 0.582 | 0.636 | 0.231 | 0.636 | 0.929 | 0.197 | 0.803 | 0.071 | 0.762 | 0.500 | 0.231 | 0.251 |
| Ictaluridae | P_BLB | 8 | 12 | 0 | 2 | 0.221 | 0.364 | 0.212 | 0.455 | 1.000 | 0.430 | 0.570 | 0.000 | 0.751 | 1.000 | 0.121 | 0.322 |
| | P_TAM | 3 | 10 | 1 | 8 | 0.266 | 0.182 | 0.485 | 0.500 | 0.750 | 0.750 | 0.250 | 0.250 | 0.530 | 0.889 | 0.167 | 0.338 |
| Fundulidae | P_BAK | 0 | 1 | 0 | 21 | 0.000 | 0.000 | 0.955 | 0.955 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.488 |
| Gasterosteidae | P_BRS | 7 | 8 | 2 | 5 | 0.562 | 0.409 | 0.432 | 0.545 | 0.778 | 0.527 | 0.473 | 0.222 | 0.609 | 0.714 | 0.295 | 0.300 |
| Centrarchidae | P_GSF | 6 | 13 | 0 | 3 | 0.456 | 0.273 | 0.406 | 0.409 | 1.000 | 0.336 | 0.664 | 0.000 | 0.503 | 1.000 | 0.270 | 0.258 |
| | P_LMB | 0 | 4 | 0 | 18 | 0.000 | 0.000 | 0.818 | 0.818 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.450 |
| | P_OSF | 2 | 11 | 1 | 8 | 0.004 | 0.136 | 0.366 | 0.455 | 0.667 | 0.995 | 0.005 | 0.333 | 0.979 | 0.889 | 0.047 | 0.349 |

**Appendix 8 continued:** Minnesota EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Centrarchidae | P_ROB | 1 | 1 | 0 | 20 | 0.004 | 0.045 | 0.909 | 0.955 | 1.000 | 1.000 | 0.000 | 0.000 | 0.996 | 1.000 | 0.000 | 0.735 |
| Percidae | P_BED | 4 | 11 | 1 | 6 | 0.335 | 0.227 | 0.440 | 0.455 | 0.800 | 0.619 | 0.381 | 0.200 | 0.520 | 0.857 | 0.213 | 0.301 |
| | P_IOD | 6 | 13 | 1 | 2 | 0.147 | 0.318 | 0.178 | 0.364 | 0.857 | 0.511 | 0.489 | 0.143 | 0.758 | 0.667 | 0.132 | 0.261 |
| | P_JOD | 12 | 5 | 2 | 3 | 0.747 | 0.636 | 0.306 | 0.682 | 0.857 | 0.445 | 0.555 | 0.143 | 0.763 | 0.600 | 0.261 | 0.331 |
| | P_WAE | 0 | 9 | 0 | 13 | 0.000 | 0.000 | 0.591 | 0.591 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.371 |
| | P_YEP | 2 | 13 | 1 | 6 | 0.266 | 0.136 | 0.430 | 0.364 | 0.667 | 0.634 | 0.366 | 0.333 | 0.366 | 0.857 | 0.203 | 0.233 |

**Appendix 9:** Black Hills EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Salmonidae | P_BKT | 4 | 11 | 0 | 0 | 0.112 | 0.267 | 0.082 | 0.267 | 1.000 | 0.000 | 1.000 | 0.000 | 0.765 | . | 0.082 | 0.191 |
| | P_BNT | 2 | 5 | 2 | 6 | 0.144 | 0.267 | 0.448 | 0.533 | 0.500 | 0.893 | 0.107 | 0.500 | 0.735 | 0.750 | 0.181 | 0.345 |
| | P_RBT | 0 | 3 | 1 | 11 | 0.003 | 0.067 | 0.734 | 0.733 | 0.000 | 0.999 | 0.001 | 1.000 | 0.000 | 0.917 | 0.067 | 0.354 |
| Esocidae | P_NOP | 0 | 1 | 0 | 14 | 0.000 | 0.000 | 0.933 | 0.933 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.483 |
| Cyprinidae | P_COC | 0 | 6 | 0 | 9 | 0.000 | 0.000 | 0.600 | 0.600 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.375 |
| | P_CRC | 1 | 0 | 9 | 5 | 0.356 | 0.667 | 0.333 | 0.400 | 0.100 | 1.000 | 0.000 | 0.900 | 1.000 | 0.357 | 0.600 | 0.069 |
| | P_FHM | 2 | 8 | 1 | 4 | 0.003 | 0.200 | 0.268 | 0.400 | 0.667 | 0.994 | 0.006 | 0.333 | 0.989 | 0.800 | 0.068 | 0.321 |
| | P_FID | 0 | 3 | 0 | 12 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |
| | P_GOS | 0 | 1 | 0 | 14 | 0.000 | 0.000 | 0.933 | 0.933 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.483 |
| | P_LOD | 11 | 2 | 1 | 1 | 0.399 | 0.800 | 0.120 | 0.800 | 0.917 | 0.556 | 0.444 | 0.083 | 0.932 | 0.500 | 0.120 | 0.436 |
| Catostomidae | P_LOS | 0 | 8 | 0 | 7 | 0.000 | 0.000 | 0.467 | 0.467 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.318 |
| | P_MOS | 1 | 11 | 0 | 3 | 0.003 | 0.067 | 0.202 | 0.267 | 1.000 | 0.989 | 0.011 | 0.000 | 0.968 | 1.000 | 0.002 | 0.232 |
| | P_WHS | 3 | 2 | 5 | 5 | 0.421 | 0.533 | 0.389 | 0.533 | 0.375 | 0.856 | 0.144 | 0.625 | 0.781 | 0.500 | 0.389 | 0.227 |
| Fundulidae | P_PLT | 0 | 1 | 0 | 14 | 0.000 | 0.000 | 0.933 | 0.933 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.483 |
| Centrarchidae | P_BLC | 0 | 1 | 0 | 14 | 0.000 | 0.000 | 0.933 | 0.933 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.483 |
| | P_GSF | 0 | 2 | 0 | 13 | 0.000 | 0.000 | 0.867 | 0.867 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.464 |
| | P_ROB | 1 | 1 | 0 | 20 | 0.593 | 0.045 | 0.936 | 0.955 | 1.000 | 0.971 | 0.029 | 0.000 | 0.628 | 1.000 | 0.027 | 0.688 |

**Appendix 10:** Cheyenne EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clupeidae | P_GZD | 0 | 0 | 3 | 4 | 0.667 | 0.429 | 0.571 | 0.571 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.571 | 0.429 | 0.000 |
| Hiodontidae | P_GOE | 3 | 2 | 0 | 2 | 0.667 | 0.429 | 0.476 | 0.714 | 1.000 | 0.600 | 0.400 | 0.000 | 0.692 | 1.000 | 0.190 | 0.523 |
| Salmonidae | P_BKT | 0 | 2 | 0 | 5 | 0.000 | 0.000 | 0.714 | 0.714 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.417 |
| | P_BNT | 1 | 0 | 0 | 6 | 0.012 | 0.143 | 0.857 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| Cyprinidae | P_COC | 1 | 4 | 2 | 0 | 0.667 | 0.429 | 0.381 | 0.143 | 0.333 | 0.000 | 1.000 | 0.667 | 0.273 | 0.000 | 0.667 | -0.286 |
| | P_CRC | 2 | 0 | 3 | 2 | 0.375 | 0.714 | 0.286 | 0.571 | 0.400 | 1.000 | 0.000 | 0.600 | 1.000 | 0.400 | 0.429 | 0.276 |
| | P_EMS | 0 | 2 | 0 | 5 | 0.000 | 0.000 | 0.714 | 0.714 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.417 |
| | P_FHM | 4 | 2 | 1 | 0 | 0.400 | 0.714 | 0.114 | 0.571 | 0.800 | 0.000 | 1.000 | 0.200 | 0.833 | 0.000 | 0.257 | 0.132 |
| | P_FLC | 3 | 2 | 0 | 2 | 0.011 | 0.429 | 0.289 | 0.714 | 1.000 | 0.989 | 0.011 | 0.000 | 0.992 | 1.000 | 0.003 | 0.610 |
| | P_GOS | 0 | 1 | 0 | 6 | 0.000 | 0.000 | 0.857 | 0.857 | . | 1 | 0 | . | . | 1 | 0 | 0.4615 |
| | P_LOD | 5 | 2 | 0 | 0 | 0.375 | 0.714 | 0.107 | 0.714 | 1.000 | 0.000 | 1.000 | 0.000 | 0.870 | . | 0.107 | 0.309 |
| | P_PLM | 1 | 4 | 0 | 2 | 0.012 | 0.143 | 0.293 | 0.429 | 1.000 | 0.976 | 0.024 | 0.000 | 0.954 | 1.000 | 0.007 | 0.362 |
| | P_RES | 4 | 1 | 0 | 2 | 0.500 | 0.571 | 0.357 | 0.857 | 1.000 | 0.800 | 0.200 | 0.000 | 0.889 | 1.000 | 0.071 | 0.731 |
| | P_RIC | 4 | 1 | 0 | 2 | 0.500 | 0.571 | 0.357 | 0.857 | 1.000 | 0.800 | 0.200 | 0.000 | 0.889 | 1.000 | 0.071 | 0.731 |
| | P_SAS | 4 | 3 | 0 | 0 | 0.500 | 0.571 | 0.214 | 0.571 | 1.000 | 0.000 | 1.000 | 0.000 | 0.727 | . | 0.214 | 0.222 |
| | P_SNC | 1 | 2 | 0 | 4 | 0.012 | 0.143 | 0.575 | 0.714 | 1.000 | 0.994 | 0.006 | 0.000 | 0.976 | 1.000 | 0.003 | 0.561 |
| Catostomidae | P_LOS | 0 | 2 | 0 | 5 | 0.000 | 0.000 | 0.714 | 0.714 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.417 |
| | P_MOS | 2 | 0 | 0 | 5 | 0.012 | 0.286 | 0.714 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| | P_SHR | 3 | 2 | 0 | 2 | 0.667 | 0.429 | 0.476 | 0.714 | 1.000 | 0.600 | 0.400 | 0.000 | 0.692 | 1.000 | 0.190 | 0.523 |
| | P_WHS | 6 | 1 | 0 | 0 | 0.333 | 0.857 | 0.048 | 0.857 | 1.000 | 0.000 | 1.000 | 0.000 | 0.947 | . | 0.048 | 0.364 |
| Ictaluridae | P_BLB | 2 | 4 | 0 | 1 | 0.012 | 0.286 | 0.150 | 0.429 | 1.000 | 0.955 | 0.045 | 0.000 | 0.977 | 1.000 | 0.007 | 0.362 |
| | P_CCF | 3 | 2 | 0 | 2 | 0.667 | 0.429 | 0.476 | 0.714 | 1.000 | 0.600 | 0.400 | 0.000 | 0.692 | 1.000 | 0.190 | 0.523 |
| | P_STC | 3 | 2 | 2 | 0 | 0.400 | 0.714 | 0.114 | 0.429 | 0.600 | 0.000 | 1.000 | 0.400 | 0.789 | 0.000 | 0.400 | 0.014 |
| Fundulidae | P_PLK | 1 | 0 | 2 | 4 | 0.011 | 0.429 | 0.571 | 0.714 | 0.333 | 1.000 | 0.000 | 0.667 | 1.000 | 0.667 | 0.286 | 0.364 |
| | P_PLT | 0 | 0 | 2 | 5 | 0.012 | 0.286 | 0.714 | 0.714 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.714 | 0.286 | 0.000 |

**Appendix 10 continued:** Cheyenne River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moronidae | P_WHB | 0 | 2 | 0 | 5 | 0.000 | 0.000 | 0.714 | 0.714 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.417 |
| Centrarchidae | P_BLC | 0 | 3 | 0 | 4 | 0.000 | 0.000 | 0.571 | 0.571 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.364 |
| | P_BLG | 1 | 0 | 0 | 6 | 0.012 | 0.143 | 0.857 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| | P_GSF | 3 | 0 | 1 | 3 | 0.010 | 0.571 | 0.4286 | 0.857 | 0.750 | 1.000 | 0.000 | 0.250 | 1.000 | 0.750 | 0.143 | 0.720 |
| | P_LMB | 0 | 1 | 0 | 6 | 0.000 | 0.000 | 0.857 | 0.857 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.462 |
| | P_OSF | 2 | 1 | 0 | 4 | 0.012 | 0.286 | 0.573 | 0.857 | 1.000 | 0.997 | 0.003 | 0.000 | 0.994 | 1.000 | 0.002 | 0.758 |
| | P_ROB | 0 | 0 | 1 | 6 | 0.012 | 0.143 | 0.857 | 0.857 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.857 | 0.143 | 0.000 |
| | P_SMB | 0 | 0 | 2 | 5 | 0.012 | 0.286 | 0.714 | 0.714 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.714 | 0.286 | 0.000 |
| Percidae | P_IOD | 0 | 1 | 0 | 6 | 0.000 | 0.000 | 0.857 | 0.857 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.462 |
| | P_SAR | 1 | 1 | 1 | 4 | 0.667 | 0.286 | 0.667 | 0.714 | 0.500 | 0.857 | 0.143 | 0.500 | 0.600 | 0.800 | 0.238 | 0.373 |
| | P_WAE | 0 | 2 | 0 | 5 | 0.000 | 0.000 | 0.714 | 0.714 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.417 |
| Sciaenidae | P_FRD | 0 | 3 | 0 | 4 | 0.000 | 0.000 | 0.571 | 0.571 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.364 |

**Appendix 11:** Grand Moreau River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | Detection Probability | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hiodontidae | P_GOE | 0 | 0 | 1 | 4 | 0.012 | 0.200 | 0.800 | 0.800 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.800 | 0.200 | 0.000 |
| Cyprinidae | P_BRM | 0 | 1 | 0 | 4 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |
| | P_COC | 2 | 3 | 0 | 0 | 0.012 | 0.400 | 0.007 | 0.400 | 1.000 | 0.000 | 1.000 | 0.000 | 0.982 | . | 0.007 | 0.283 |
| | P_CRC | 1 | 2 | 1 | 1 | 0.012 | 0.400 | 0.205 | 0.400 | 0.500 | 0.977 | 0.023 | 0.500 | 0.977 | 0.500 | 0.205 | 0.282 |
| | P_EMS | 0 | 0 | 1 | 4 | 0.012 | 0.200 | 0.800 | 0.800 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.800 | 0.200 | 0.000 |
| | P_FHM | 5 | 0 | 0 | 0 | 0.400 | 1.000 | 0.000 | 1.000 | 1.000 | . | . | 0.000 | 1.000 | . | 0.000 | . |
| | P_FLC | 1 | 1 | 2 | 1 | 0.667 | 0.600 | 0.333 | 0.400 | 0.333 | 0.600 | 0.400 | 0.667 | 0.600 | 0.333 | 0.533 | 0.000 |
| | P_LOD | 4 | 0 | 0 | 1 | 0.667 | 0.800 | 0.200 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| | P_NRD | 0 | 1 | 0 | 4 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |
| | P_PLM | 0 | 2 | 0 | 3 | 0.000 | 0.000 | 0.600 | 0.600 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.375 |
| | P_RIC | 2 | 1 | 0 | 2 | 0.012 | 0.400 | 0.402 | 0.800 | 1.000 | 0.994 | 0.006 | 0.000 | 0.994 | 1.000 | 0.002 | 0.705 |
| | P_SAS | 4 | 1 | 0 | 0 | 0.500 | 0.800 | 0.100 | 0.800 | 1.000 | 0.000 | 1.000 | 0.000 | 0.889 | . | 0.100 | 0.286 |
| Catostomidae | P_SHR | 4 | 1 | 0 | 0 | 0.012 | 0.800 | 0.002 | 0.800 | 1.000 | 0.000 | 1.000 | 0.000 | 0.997 | . | 0.002 | 0.441 |
| | P_WHS | 4 | 0 | 0 | 1 | 0.500 | 0.800 | 0.200 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| Ictaluridae | P_BLB | 3 | 2 | 0 | 1 | 0.500 | 0.600 | 0.200 | 0.600 | 1.000 | 0.000 | 1.000 | 0.000 | 0.750 | . | 0.200 | 0.231 |
| | P_CCF | 2 | 0 | 1 | 2 | 0.667 | 0.600 | 0.400 | 0.800 | 0.667 | 1.000 | 0.000 | 0.333 | 1.000 | 0.667 | 0.200 | 0.615 |
| | P_STC | 2 | 0 | 2 | 1 | 0.500 | 0.800 | 0.200 | 0.600 | 0.500 | 1.000 | 0.000 | 0.500 | 1.000 | 0.333 | 0.400 | 0.286 |
| Centrarchidae | P_GSF | 4 | 1 | 0 | 0 | 0.500 | 0.800 | 0.100 | 0.800 | 1.000 | 0.000 | 1.000 | 0.000 | 0.889 | . | 0.100 | 0.286 |
| | P_OSF | 0 | 0 | 1 | 4 | 0.012 | 0.200 | 0.800 | 0.800 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.800 | 0.200 | 0.000 |
| Percidae | P_IOD | 0 | 1 | 0 | 4 | 0.000 | 0.000 | 0.800 | 0.800 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.444 |
| | P_JOD | 0 | 2 | 0 | 3 | 0.000 | 0.000 | 0.600 | 0.600 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.375 |
| | P_YEP | 0 | 4 | 0 | 1 | 0.000 | 0.000 | 0.200 | 0.200 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.167 |

**Appendix 12:** White River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hiodontidae | P_GOE | 0 | 2 | 0 | 7 | 0.000 | 0.000 | 0.778 | 0.778 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.438 |
| Esocidae | P_NOP | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| Cyprinidae | P_BIS | 2 | 0 | 0 | 7 | 0.634 | 0.222 | 0.778 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| | P_BRM | 3 | 4 | 0 | 2 | 1.000 | 0.333 | 0.667 | 0.556 | 1.000 | 0.333 | 0.667 | 0.000 | 0.429 | 1.000 | 0.444 | 0.250 |
| | P_COC | 0 | 9 | 0 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | . | . | . | . | . | . | 0.000 | 0.000 |
| | P_CRC | 2 | 7 | 0 | 0 | 0.006 | 0.222 | 0.005 | 0.222 | 1.000 | 0.000 | 1.000 | 0.000 | 0.980 | . | 0.005 | 0.181 |
| | P_FHM | 0 | 6 | 1 | 2 | 0.006 | 0.111 | 0.226 | 0.222 | 0.000 | 0.982 | 0.018 | 1.000 | 0.000 | 0.667 | 0.115 | 0.158 |
| | P_FLC | 4 | 0 | 4 | 1 | 0.382 | 0.889 | 0.111 | 0.556 | 0.500 | 1.000 | 0.000 | 0.500 | 1.000 | 0.200 | 0.444 | 0.182 |
| | P_LOD | 4 | 4 | 0 | 1 | 0.275 | 0.444 | 0.233 | 0.556 | 1.000 | 0.476 | 0.524 | 0.000 | 0.784 | 1.000 | 0.122 | 0.385 |
| | P_NRD | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| | P_PLM | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| | P_RES | 1 | 2 | 2 | 4 | 1.000 | 0.111 | 0.778 | 0.667 | 1.000 | 0.714 | 0.286 | 0.000 | 0.333 | 1.000 | 0.222 | 0.372 |
| | P_RIC | 1 | 3 | 0 | 5 | 0.000 | 0.111 | 0.556 | 0.667 | 1.000 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 0.509 |
| | P_PED | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1 | 0 | 0.4 |
| | P_SAS | 7 | 1 | 1 | 0 | 0.320 | 0.889 | 0.036 | 0.778 | 0.875 | 0.000 | 1.000 | 0.125 | 0.956 | 0.000 | 0.147 | 0.186 |
| | P_SNC | 0 | 7 | 0 | 2 | 0.000 | 0.000 | 0.222 | 0.222 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.182 |
| | P_STR | 0 | 1 | 0 | 8 | 0.000 | 0.000 | 0.889 | 0.889 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.471 |
| Catostomidae | P_SHR | 0 | 5 | 0 | 4 | 0.000 | 0.000 | 0.444 | 0.444 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.308 |
| | P_WHS | 0 | 6 | 0 | 3 | 0.000 | 0.000 | 0.333 | 0.333 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.250 |
| Ictaluridae | P_BLB | 1 | 1 | 0 | 7 | 0.006 | 0.111 | 0.778 | 0.889 | 1.000 | 0.999 | 0.001 | 0.000 | 0.994 | 1.000 | 0.001 | 0.709 |
| | P_CCF | 2 | 5 | 0 | 2 | 0.006 | 0.222 | 0.226 | 0.444 | 1.000 | 0.985 | 0.015 | 0.000 | 0.985 | 1.000 | 0.003 | 0.383 |
| | P_STC | 1 | 6 | 0 | 2 | 0.006 | 0.111 | 0.226 | 0.333 | 1.000 | 0.982 | 0.018 | 0.000 | 0.965 | 1.000 | 0.004 | 0.288 |
| | P_TAM | 0 | 0 | 1 | 8 | 0.006 | 0.111 | 0.889 | 0.889 | 0.000 | 1.000 | 0.000 | 1.000 | . | 0.889 | 0.111 | 0.000 |

**Appendix 12 continued:** White River EDU accuracy metrics for all species collected and/or predicted. A (.) indicates that the metric was incalculable

| Family | Species Code | # of true presences | Commission errors | Omission errors | # of true absences | DP | Prevalence | Overall Diagnostic power | Correct classification rate | Sensitivity | Specificity | False positive rate | False negative rate | Positive predictive power | Negative predictive power | Misclassification rate | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fundulidae | P_PLT | 0 | 2 | 0 | 7 | 0.000 | 0.000 | 0.778 | 0.778 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.438 |
| Centrarchidae | P_BLC | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| | P_BLG | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| | P_GSF | 1 | 8 | 0 | 0 | 0.006 | 0.111 | 0.005 | 0.111 | 1.000 | 0.000 | 1.000 | 0.000 | 0.954 | . | 0.005 | 0.099 |
| Percidae | P_IOD | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |
| | P_SAR | 0 | 3 | 0 | 6 | 0.000 | 0.000 | 0.667 | 0.667 | . | 1.000 | 0.000 | . | . | 1.000 | 0.000 | 0.400 |