2000

Relationship of Topeka shiner distribution to geographic features of the Des Moines Lobe in Iowa

Steven John Clark
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/rtd

Part of the Aquaculture and Fisheries Commons, and the Terrestrial and Aquatic Ecology Commons

Recommended Citation
Clark, Steven John, "Relationship of Topeka shiner distribution to geographic features of the Des Moines Lobe in Iowa" (2000). Retrospective Theses and Dissertations. 16803.
http://lib.dr.iastate.edu/rtd/16803
Relationship of Topeka shiner distribution to geographic features of the Des Moines Lobe in Iowa

by

Steven John Clark

A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Fisheries Biology

Major Professor: Bruce W. Menzel

Iowa State University

Ames, Iowa

2000
Graduate College
Iowa State University

This is to certify that the Master's thesis of

Steven John Clark

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy
# TABLE OF CONTENTS

**CHAPTER 1: GENERAL INTRODUCTION**  
1

- Introduction  
1
- Thesis Organization  
3
- Literature Cited  
3

**CHAPTER 2: RELATIONSHIP OF TOPEKA SHINER DISTRIBUTION TO GEOGRAPHIC FEATURES OF THE DES MOINES LOBE IN IOWA**  
5

- Summary  
5
- Introduction  
6
- Materials and Methods  
8
- Results  
17
- Discussion  
21
- References  
28

**CHAPTER 3: GENERAL CONCLUSIONS**  
44

**APPENDIX 1: MAP OF GENERAL MODEL – TERMINAL NODE 3**  
46

**APPENDIX 2: MAP OF GENERAL MODEL – TERMINAL NODE 2**  
47

**APPENDIX 3: METHODS FOR GIS USERS**  
48

**ACKNOWLEDGMENTS**  
54
CHAPTER 1: GENERAL INTRODUCTION

Introduction

The Topeka shiner (*Notropis topeka* Gilbert) is a North American Plains stream minnow that has experienced a recent and rapid decline of populations throughout portions of its range (USFWS, 1996). It was identified as a candidate for endangered species listing in 1996 by the United States Fish and Wildlife Service (USFWS, 1996), and was effectively listed in January, 1999 (USFWS, 1998). Its historic range is the central prairie region, including portions of Iowa, Kansas, Minnesota, Missouri, Nebraska, and South Dakota. It once ranged widely throughout the north central region of Iowa known as the Des Moines Lobe (DML), and records of the fish exist at other locations including the northwest and southeast corners of the state. Historical records indicate that the shiner most likely entered the DML region from the northwest, through headwater connections with the Missouri River system (Bailey & Allum, 1962).

The Topeka shiner is a slab-sided cyprinid with moderately small eyes and a small oblique mouth (Pflieger, 1997). Its back is straw-colored, with prominently dark-edged scales, and its belly is silvery-white. It has silvery sides with a well-defined dusky stripe, and the base of the tail fin has a dark wedge-shaped spot. Breeding males have red-orange fins and a similar tinge on the head and body. Its closest relative is the sand shiner (*Notropis ludibundus* Girard) (Schmidt & Gold, 1995), with which it is broadly sympatric.

The Topeka shiner is commonly believed to prefer pools of small, moderately clear, prairie streams (Minckley & Cross, 1959; Pflieger, 1997). Several years of intensive collecting effort in the mid-1990’s in such habitats in Iowa produced few individuals at less than ten widely distributed sites. Similar lack of success prevailed in southwest Minnesota until it was discovered that the fish exists, often abundantly, in off-channel habitats such as oxbows. This information led to more success during subsequent efforts in Iowa in 1998 and 1999; however, known locations for the fish were still few, and only four sites were reproductive habitats.
To facilitate conservation programs for the species, there is need for better understanding of its habitat requirements. In consideration of its broad, but disjunct pattern of distribution in Iowa, habitat knowledge on both macro- and micro-scales is desirable. Because distributional information necessary for micro-habitat studies was limited for Iowa, I chose a landscape-level approach using Geographic Information Systems (GIS) technologies to identify the relationship between known Topeka shiner sites and landscape features. Such technologies allow researchers to study important ecological relationships at the geographic level, something previously difficult or impossible. GIS have been used to discover relationships between landscapes and lotic systems at stream-reach, watershed, and state-wide levels (Richards & Host, 1994; Allan, Erickson & Fay, 1997; Johnson et al., 1997, Wang et al. 1997).

Moreover, the use of GIS, in conjunction with statistical analysis, can elucidate complex ecological relationships (Johnson & Gage, 1997). Generally referred to as predictive modeling, this technique has been used in wildlife studies to identify suitable habitat for black bears (*Ursus americanus*), elk (*Cervus elaphus*), kangaroos (*Macropus sp.*), and the endangered Mt. Graham red squirrel (*Tamiasciurus hudsonicus grahamensis*) (Donovan, Rabe & Olson, 1987; Walker, 1990; Pereira & Itami, 1991; Bian & West, 1997; van Manen & Pelton, 1997).

There are three basic steps for developing predictive models with GIS. First, existing conditions are identified by measuring the values of landscape variables for each of the study units within GIS. Second, this information is subjected to statistical analysis to distinguish the key relationships which become the descriptive model. Finally, the model is applied at the landscape level within GIS to identify places with similar characteristics as the original study units.

There are several statistical techniques suitable for predictive modeling such as linear and logistic regression. While these methods are effective in some situations, they have limitations, especially when applied to ecological data (Michaelsen et al., 1994; Bell, 1996). Classification and regression trees (CART) provide an alternative to these traditional methods and have many benefits (Michaelsen et al., 1994; Bell, 1996; Borgognone et al., 1999). CART can express non-additive and non-linear relationships in a simple form, and there is no need to select variables or interactions in
advance. CART is non-parametric, and can use a mix of categorical and continuous data. Also, the analysis results in a tree diagram, showing relationships in a more intuitive way than a regression model, especially if the data used in the regression were transformed to fit normality.

Because of the current limitations of distributional information for the Topeka shiner in Iowa, a predictive habitat model might be a useful tool for discovering new populations of the fish. Maps resulting from the model could be used in the field to identify locations that warrant more intensive collecting efforts. Also, the model could be used to identify areas suitable for habitat improvement and re-introductive efforts. Finally, it may provide insights to critical habitat conditions or requirements.

**Thesis Organization**

This thesis is comprised of a general introduction, one paper to be submitted to the aquatic journal *Freshwater Biology*, general conclusions, and appendices. The paper, entitled Relationship of Topeka shiner distribution to geographic features of the Des Moines Lobe in Iowa, contains a summary (as requested by the journal in place of an abstract), introduction, materials and methods, results, discussion, and references. Tables and figures are located after the references.

**Literature Cited**


Summary

1. The Topeka shiner (*Notropis topeka* Gilbert) is a plains stream minnow that was federally listed as an endangered species, effective January, 1999. For several years, efforts were made to determine its current range and habitat in Iowa with little success.

2. Geographic Information Systems (GIS) technology was used to measure landscape-level variables at locations of fish collections made from 1970 to 1999. Classification trees were used to identify relationships between these variables and Topeka shiner habitat.

3. Two models were empirically derived to identify Topeka shiner habitat in the Des Moines Lobe geographic region. The general model was based on features of all Topeka shiner collection sites, whereas the specific model was based on characteristics of “typical-quality” shiner sites.

4. Characteristics identified as consistent with shiner habitat by the models were: a general lack of forest/woody riparian cover, and stream reaches with relatively frequent flooding (at least once every 2 years). Field collecting efforts revealed that off-channel habitats often have deep silt bottoms, frequently coincide with cattle grazing, and are a source of young-of-the-year shiners.

5. The two models were applied within the GIS to develop maps identifying areas of the Des Moines Lobe with characteristics similar to known “suitable” habitat. These maps will be used in future search efforts, and possibly to identify places suitable for habitat improvement and reintroductions to ensure the future of *N. topeka* in Iowa.
Introduction

Continental landscapes of the United States have been substantially altered since the initial influx of European settlers; and in midwestern states, especially by agricultural activities. Because rivers and streams are ecologically linked to the landscapes they drain, they have also been substantially altered in regions of intensive agriculture (Menzel, Barnum & Antosch, 1984). Changes in physical, chemical and hydrologic features of streams induced by agricultural practices are well known (Bulkley, 1975; Bishop, 1981; Menzel, 1983; Karr, Toth & Dudley, 1985; Allan, Erickson & Fay, 1997; Johnson et al., 1997).

Because of these alterations, many species of fish have declined in Iowa and some have been extirpated (Menzel, 1981; Harlan, Speaker & Mayhew, 1987). The Topeka shiner (*Notropis topeka* Gilbert) is one such fish that has experienced a recent and rapid decline of populations throughout its range (USFWS, 1996). It was identified as a candidate for endangered species listing in 1996 by the United States Fish and Wildlife Service (USFWS, 1996), and was formally listed in January, 1999 (USFWS, 1998).

The Topeka shiner is a rather chubby, slab-sided cyprinid with moderately small eyes and a small oblique mouth. It has a straw-colored back with dark-edged scales and silvery sides with a dusky lateral stripe. There is a dark wedge-shaped spot at the base of the tail, and its belly is silvery-white. Breeding males have orange-red fins and an orange tinge on the head and body. The fish is not known to exceed 76 mm (3 inches) in length (Pflieger, 1997). Its closest relative is the sand shiner (*Notropis ludibundus* Girard) (Schmidt & Gold, 1995), with which it is broadly sympatric.

The historic range of the Topeka shiner is the central prairie region, including portions of Iowa, Kansas, Minnesota, Missouri, Nebraska, and South Dakota (Fig. 1). It once ranged widely throughout the north central region of Iowa known as the Des Moines Lobe (DML), and was reported from many locations in the upper Des Moines and North Raccoon river systems (Fig. 2) (Harlan et al., 1987). Historical records indicate that it most likely entered this region from the northwest, through headwater stream connections with the Missouri River system (Bailey & Allum, 1962). A few collections in southeast Iowa are probably explained by the downstream transport of the fish during
flooding events. Old reports from upper reaches of the Iowa and Cedar river systems, if correct, might be explained by former headwater connections with the upper Des Moines River system. The fish has never been reported from the Skunk River, which drains the southeastern portion of the DML, probably because of the river's topographic isolation from the Des Moines River system.

Little is known of the Topeka shiner's habitat requirements in Iowa. Past collections of the fish throughout its range indicate that it is rarely found in abundance, and most often its numbers are low (Minckley & Cross, 1959). For Iowa, Harlan and Speaker (1969) stated that one or two specimens often make up a single collection. In southern and western portions of its range the Topeka shiner prefers pools of small, moderately clear, prairie streams having substrates of sand, gravel, rubble, and bedrock (Minckley & Cross, 1959; Pflieger, 1997). However, recent collecting in such habitats in Iowa resulted in little success. Greater success occurred in 1998 and 1999 when search efforts focused on off-channel habitats such as oxbows. Still, the shiner's range seemed to be small in Iowa and few sites produced substantial numbers of fish.

To facilitate conservation programs for the species, there is need for better understanding of its habitat requirements. In consideration of its broad, but now disjunct pattern of distribution in Iowa (Fig. 1), habitat knowledge on both macro- and micro-scales is desirable. Because distributional information necessary for micro-habitat studies was limited, I chose a landscape-level approach using Geographic Information Systems (GIS) technologies to identify the relationship between Topeka shiner sites and landscape features. When linked with statistics this approach can elucidate complex relationships between habitat and landscape variables (Johnson & Gage, 1997), as has been demonstrated in many terrestrial wildlife studies (Donovan, Rabe & Olson, 1987; Walker, 1990; Pereira & Itami, 1991; Clark, Dunn & Smith, 1993; Bian & West, 1997; van Manen & Pelton, 1997).

GIS has also been useful in aquatic situations (Delong & Brusven, 1991; Hamlett et al., 1992; Miller et al., 1995; Narumalani et al., 1997), although its application is still underutilized (Isaak & Hubert, 1997). Specifically, GIS has been used to evaluate relationships between aquatic systems and the surrounding landscape (Roth, Allan and Erickson, 1996; Allan & Johnson, 1997). Two
studies found inverse and direct relationships, respectively, between the amount of agriculture and forested land adjacent to a stream and a fish IBI (Wang et al., 1997; Lammert & Allan, 1999).

The choice of landscape variables for the study was guided by field observations on known shiner habitat. Most sites had grassy riparian areas, and none were heavily wooded. Therefore, I included variables to reflect three types of landcover: forested/wooded, grassland, and row crop agriculture. Commonly, there were morainal features in the adjacent landscapes. These relatively steeply sloping features tend to discourage tillage and may also be shallow groundwater sources. To incorporate their potential role in Topeka shiner habitat, local topography was included as a variable. Also, many off-channel sites were shallow and could benefit from groundwater inputs, and one even had a visible seep; therefore, I included a variable for alluvial deposits, which can contribute large amounts of shallow groundwater (Prior, 1991). Because sites occurred in floodplains, a variable to reflect relative flood frequency was included. Finally, since presumptively permanent populations occurred in off-channel (wetland) habitats, I included a variable to quantify such habitat.

This study had two objectives: 1) to develop two GIS-based models of the association of Topeka shiner habitat to local landscape features on the Des Moines Lobe using existing GIS data layers, 2) to apply the models throughout the landform to identify likely Topeka shiner sites for future search efforts.

**Materials and Methods**

**Study Area:**

The aquatic and associated riparian areas included in this study are located in the region of northcentral Iowa known as the Des Moines Lobe (DML) (Fig. 3). This landform was created by a lobe of the Wisconsin glacier 12,000 to 14,000 years ago. It is marked by bands of small morainal ridges created by the stagnation of the retreating glacier, within a generally flat ground moraine landscape (Prior, 1991). These morainal areas have greater relief than the surrounding landscape, and consequently, are used less extensively for row crop agriculture. The stream network on the lobe is poorly developed, widely spaced, and contrasts greatly to the more developed networks of the
adjacent older landforms. The DML lies within the Western Corn Belt Plains ecoregion, and is currently under intensive row crop agriculture. Its elevation is 274-457 m, and annual precipitation is 71-79 cm (Griffith et al., 1994). Streams of the region are of relatively low gradient and have large annual fluctuations in discharge. Flash flooding is common throughout the warmer seasons.

Collection Data:

Topeka shiner distributional data were based on stream fish collections made from 1970 through 1999 by ISU personnel. Off-channel sites included natural oxbow lakes, stream reaches isolated from the main system by human activities such as bridge construction, and small gravel pits.

At each site, crews sampled about 200 m of channel or wetland with a 6 mm (1/4 in) or 3 mm (1/8 in) mesh seine, or a backpack electrofishing unit. Sampling was generally stopped when no new species were being collected, and when the general range of habitat conditions was covered. Site locations were marked on 1:126,700 scale Iowa Department of Transportation county maps, and were described in field notes.

The locality data base included 490 collection sites, 37 (7.6%) of which yielded Topeka shiners. Thirty-five shiner collections were in the North Raccoon River drainage, and two were in the Des Moines River drainage. When present, the fish was typically rare; only one specimen was collected at 20 of the sites and evidence of reproduction (at least two year-classes present) was found at only four sites. A substantially higher proportion of off-channel sites produced shiners than did in-channel sites (Table 1). Drainage areas for the 37 Topeka shiner sites ranged from 5 to 3206 km² (2 to 1238 mi²) (Table 2).

The fish was found in several types of off-channel habitats, including naturally formed oxbows. In one such place, there was evidence of reproduction and year-around presence. However, most of the natural oxbows were ephemeral and likely provided habitat only during spring and summer. Most commonly, Topeka shiners were found in human-created “oxbows” such as those formed when the existing channel was bypassed with a new one during bridge construction, and abandoned gravel pits. Most of the artificial oxbows were also ephemeral, but two were permanent and served as reproductive sites in 1999; one was fed by a visible groundwater seep. An abandoned
gravel pit, which was relatively deep (> 1.5 m) and permanent, served as another reproductive habitat.

**Digital Data Sources:**

I obtained existing landscape coverages as digital data maps from three sources (Table 3). I used ALLUV100 as an indicator of possible local shallow groundwater. NWIs were separate county-wide coverages of the National Wetlands Inventory which are records of wetland location and classification as defined by the United States Fish and Wildlife Service (USFWS). ELEVATION30, is a mosaic of individual USGS Digital Elevation Model quadrangles for Iowa that I used to evaluate topographic effects. SOILS is a coverage of soils data taken from the published county soil surveys of Iowa. Each soil mapping unit can be linked to a database containing many attributes for each soil type, including flood frequency. LCOV92 is a coverage of land use where each 30 m pixel (cell) was placed into one of seven different classes: row crops, grasslands, water cover, barren/flooded cover, artificial cover (urban), forest/woody cover, and cloud covered areas. LOBE9 is a coverage delineating boundaries, various advances, and end moraines of the DML that I used to define the study area. I used BASIN24 to delineate major watersheds and identify the streams within them during a model evaluation step.

I digitized field collection sites on-screen in ArcView 3.2 (Environmental Systems Research Institute, Inc.) as a new point coverage. Using RIVERS and PLSS as references, I marked site locations on-screen as points and assigned them unique identification numbers. In cases of multiple collections at a site, I included only one record, retaining the record with the shiner, if it was present. After the site locations were digitized, I joined an existing table containing field data for each site which included: number of Topeka shiners collected, drainage area, and type of habitat collected, either in-channel or off-channel. I converted the coverage to a grid with 30 m pixels and named it SITES (Fig. 4); therefore, sites were represented as individual 30 m pixels.
**Digital Data Processing:**

The next step was to process the data so that relationships between landscape variables and collection sites could be identified and extracted for statistical analysis. I used ArcView with the Spatial Analyst and XTools extensions for all GIS applications.

ALLUV100 was originally in a vector format and was converted to grid (raster) format with 30 m pixels for ArcView processing. NWIs were individual, county-wide coverages, which I merged into one contiguous coverage for the DML. I set pixel values to reflect wetland type.

I next clipped the variable coverages to the borders of the DML with LOBE9. This proved problematic because LOBE9 had some small disjunct pieces in the northeastern corner of the coverage. Therefore, I slightly modified it before the clip to include these pieces as one contiguous shape by extending its border as a straight line (see Fig. 4). As a result, a negligible area was included that extends outside the DML.

The next step was to process the clipped coverages to show the variables I wished to include in the model. I converted each cell value in ELEVATION30 into a slope value in degrees and named the new coverage DSLOPE. Within LCOV92, I converted each landcover type into a new coverage and named them: AG (row crops), GRASS (grasslands), and WOOD (forest).

I used a NWI variable as a measure of the amount of off-channel habitat at a site. The NWI "palustrine" classification (area < 8 hectares, depth < 2 m) best described off-channel Topeka shiner habitats. Therefore, I created a new coverage to reflect the presence of palustrine wetlands and named it PAL.

The SOILS coverage contains many attributes for each soil type. Flood frequency is an attribute that could play an important role in the permanence of off-channel habitats and recruitment of Topeka shiners to a nearby stream. It seemed reasonable to focus on soils that flooded at least once every two years; therefore, I created a new coverage to reflect soils that flood "commonly" and "frequently" and called it FLOOD.

All landscape data layers existed in a grid format of 30 m pixels. However, I was concerned with a larger area surrounding a collection site, which reflected broader landscape conditions. Similar
aquatic studies have described surrounding landscape variables within either catchments or buffers along stream reaches (Richards & Host, 1994; Richards, Johnson & Host, 1996; Roth et al., 1996; Lammert & Allan, 1999). In general, both approaches have performed adequately (Richards & Host, 1994; Richards et al., 1996; Lammert & Allan, 1999). In the present case, it is possible to delineate watersheds for the sample sites, but the predictive modeling step would require this to be done for every pixel of stream length, a computationally prohibitive process. Therefore, I used a site buffer approach and tested two buffer options: 120 m (four - 30 m pixels) and 210 m (seven - 30 m pixels) radii around sites to measure the landscape variables.

I processed each coverage at these two buffer sizes. For DSLOPE, I calculated the average slope of all pixels within each radius. This resulted in two different coverages, one for each radius: DSLOPE7, with a seven-pixel radius; and DSLOPE4, with a four-pixel radius.

I used the same basic approach for AG, GRASS, WOOD, ALLUV, PAL, and FLOOD, but did not calculate an average. Rather, within the circular area of the given radius, I calculated the proportion of pixels with the presence of a given variable type, to the total number of pixels present. This resulted in new coverages in which each pixel represented the percentage of the variable type within the circular area around that pixel. Then, for each new coverage, I converted the percentages into 11 decile classes where: 0 = 0, 1-10 = class 5, 11-20 = class 15, 21-30 = class 25, ..., and 91-100 = class 95. The new coverages were named in a similar manner as DSLOPE.

I next identified the value of all variables at each site and added these values to the original SITES table. This table then contained the site’s identification number, the number of shiners found at the site, the drainage area, the habitat type, and the single predominant value for each variable.

The dependent variable in the model, Topeka shiner occurrence, may be expressed in different ways, ranging from simple presence/absence to semi-quantitative estimates of abundance and other observed biological characteristics. I believed that exploring relationships based on different classification systems for this variable would provide insight into characteristics of “better” habitats and may even be necessary for meaningful analyses due to the limited amount of available data. Therefore, I added three new dependent variables to the data table; CLASS1: absent = 0,
present = 1; CLASS2: absent = 0, 1-2 shiners collected = 1, \geq 3 collected = 3; CLASS3: absent = 0, 1-2 collected = 1, \geq 3 collected and no evidence of reproduction = 3, and evidence of reproduction (numerous adults and juveniles collected) = 4.

In general, collections of *N. topeka* were associated with tributary streams rather than major rivers. Thus I examined drainage areas of sites where the fish was collected to determine a basin size range that described the majority of collections. About 70% of the sites were in the range of 52 to 777 km² (20 to 300 mi²). I considered that off-channel sites within this range typified most sites with higher numbers of fish (Table 2); i.e. “better” habitat. The four sites with evidence of reproduction met these criteria. I used this information as a screening tool in the model to limit the collection data to sites with higher probabilities of Topeka shiner occurrence. I could not use it as a modeling variable, however, because it would be inapplicable for the predictive modeling step.

**Classification and Regression Trees:**

There are many statistical methods for analyzing relationships between variables in complex ecological data sets. Two commonly used methods are multiple regression and logistic regression. While they are effective in certain situations, they have some limitations, especially when analyzing ecological data (Michaelsen et al., 1994; Bell, 1996). Classification and regression trees (CART) provide an alternative to these traditional methods and have some benefits in such situations (Borgogonne et al., 1999). CART can express non-additive and non-linear relationships in a simple form, there is no need to select variables or interactions in advance, it is non-parametric, and it can use a mix of categorical and continuous data. Also, the analysis results in a “tree”, showing relationships in a more intuitive way than a regression model, especially if the original data were transformed to fit normality.

The main difference between classification trees and regression trees is the response variable. Classification trees predict class membership probabilities for categorical response variables, whereas regression trees predict average values for continuous response variables.

The basic reference on CART is Breiman et al. (1984) who developed many of the key ideas and provided explanations of the basic techniques. They also developed software for CART analysis.
with Salford Systems. Hereafter, CART refers to the Salford Systems software that was used in this study.

The CART methodology is known as binary recursive partitioning (Steinberg & Colla, 1997). It works by initially assigning all cases (data points) to one hypothetical group, called the parent node. CART then bifurcates the parent node iteratively at each possible value of every predictor variable. Cases with values of the predictor variable less than or equal to the splitting value go to the left child node and those with greater values go to the right. CART then measures the homogeneity of the response variable in the child nodes produced by each value-variable split and uses the one that most increases this value to make the first split or branch in the tree. Subsequently, child nodes are treated as new parent nodes and the process is repeated until further splitting is impossible or stopped by design. Splitting is impossible if there is only one case in a node or if all cases have identical values of the response variable. CART allows splitting to be stopped for a number of reasons including insufficient cases in a parent or child node. The default lower limit for the parent node is ten, and for a child node it is one.

Once a terminal node is found, it is given a "classification" (i.e. named). The simplest way that CART does this is by using the plurality rule: the group with the greatest representation determines class assignment. CART can also adjust for over- or under-sampling from certain classes. For the data used here, CART assigned node classification to the value of the dependent variable with a higher representation than that found in the original data set. This was appropriate because of the low presence to absence collection ratio and the likelihood that the fish was not always collected in habitats it could populate.

CART continues operations until it is not possible to grow the tree any further. A set of sub-trees is derived from this maximal tree, and the best sub-tree is determined by testing for costs (error rates), which are simply the rates at which cases are misclassified. With sufficient data, this may be determined by dividing the data into learning and test sub-samples. The learning sample is used to grow the maximal tree, and the test sample is used to calculate the cost of the derived sub-trees. In many instances, however, including this one, there is not sufficient data for this procedure, and an
alternative must be used. Lacking sufficient data, CART can apply a cross-validation technique. First, a maximal tree is established with the entire data set. Then, for 10-fold cross-validation (the default testing procedure), the data are divided into ten equal parts by random sampling, stratified on the dependent variable. Nine of the ten subsets are used as a learning sample to build a maximal tree from which sub-trees are derived. The excluded subset is used as a test sample to calculate the costs for all the sub-trees. This process is replicated ten times with the original ten subsets so that each subset is used as a test sample. The error counts of each of the ten test samples are summed to obtain the costs of each of the sub-trees of the full-sample maximal tree. These costs are assumed to be what can be expected when applying the tree to new data. Breiman et al. (1984) reported that 10-fold cross validation usually gives very good reliability. Because the costs are based on trees built with only 90% of the data, they will tend to over-estimate the true cost of the tree built with the full data set, consequently giving conservative cost estimates.

One of the benefits of CART is that it produces a simple tree. Because of this, relatively few variables are used explicitly as splitting criteria. This does not mean that variables not used as primary splitters are unimportant, in fact, they can be very important. To account for this, CART keeps track of each variable's contribution as a surrogate as well as a primary splitter. A surrogate is a splitting decision that closely mimics that of the primary split. A good surrogate split will not only create nodes of similar composition as the nodes from the primary split, but it will closely match the primary split on a case-by-case basis. If no other variable can mimic the primary split, no surrogates exist. Surrogates can be used to make a decision for a case lacking information on the primary splitting variable, and this is how CART processes cases with missing values.

The ability of one variable to hide the significance of another in a CART analysis is called masking. This is addressed by CART's variable importance measure. To calculate a variable importance score, CART calculates the reduction of the cost (improvement) attributable to each variable in its role as a surrogate to the primary split. The values of these improvements are added over each node and are scaled relative to the best performing variable. The variable with the highest sum of improvements is scored 100, and all other variables are scored lower, toward zero. The
importance score measures a variable's ability to mimic the chosen tree and to act as a replacement for variables appearing in primary splits. It does not give any indication of the variable's importance in the construction of other trees.

**CART Analysis and Predictive Modeling:**

For this study, I followed several steps when building a CART tree. First, within CART, I opened the data table containing the dependent (response) and independent (predictor) variables. Then, I chose the pertinent dependent and independent variables and selected between classification or regression analysis. I left other settings at the defaults because, through preliminary testing, they seemed to produce the best results. I chose subsets of the data for different model assumptions in an attempt to create two well-performing models. One model was general and applied to all sites. The other model was more specific and was intended to identify relationships for “typical-quality” sites. The latter was based only on off-channel sites, where the highest numbers of shiners were found, and within the range of drainage areas typifying most collections.

I used the costs of each class to assess the quality of the trees, placing more emphasis on the cost for classifying Topeka shiner sites. I would expect a higher cost for absence sites because some of these may have been capable of supporting populations of the fish, and therefore, would be misclassified. Cost was only regarded as a way of determining the relative performance of each tree and was not used to generate a numeric probability for estimating Topeka shiner habitat quality in the field. This was an observational study, prohibiting such estimates.

After I developed two acceptable trees based on different assumptions, I performed the predictive modeling step. I did this by identifying the set of relationships that led to each node classified as a shiner presence. This resulted in two models for the general tree (one for each node predicted as a presence) and one model for the specific tree. I applied these models in ArcView and created maps highlighting places on the DML “likely” to be Topeka shiner habitat.

To further evaluate the ability of the general model to identify likely shiner habitat, I measured the proportion of the stream length predicted to have such habitat relative to total stream length, in each of three large river basins on the DML (see Fig. 4). This was done to determine if the model
predicted relatively more habitat in a basin known to have substantial populations of shiners (North Raccoon), versus two basins that did not (Middle Raccoon and Boone).

Results

CART Analysis:

The total of 37 Topeka shiner collections was insufficient for performing a satisfactory regression tree analysis, therefore, only classification analyses were applied. For the same reason, it was only possible to build classification trees for the two simplest classes of shiner occurrence, presence (class-1) and absence (class-0).

There were some differences in the costs of trees built with variables measured at the two radius distances around the sites. For the general model the differences were minor. The 120 m radius data produced a total cost of 0.66, and the 210 m radius data produced a total cost of 0.64. However, for the specific model based only on off-channel site data, there was a large difference. The 210 m radius data produced a tree with a total cost of 0.79, whereas the 120 m radius data failed to produce a tree due to an excessively high cost. Because I wanted to build both models with variables measured at the same scale, I subsequently pursued modeling based only on the 210 m radius. Field observations support this decision. At most sites, the floodplain is fully captured within this distance, and landuse beyond the floodplain is predominantly agriculture.

General Model:

I built the classification tree for the general model with all 490 sites. The dependent variable was CLASS1 (presence/absence), and the predictor variables were AG7, GRASS7, WOOD7, ALLUV7, DSLOPE7, FLOOD7, and PAL7 (Table 4a). The resulting tree had five terminal nodes, two of which were identified as class-1 (likely Topeka shiner habitat) (Fig. 5). Thirty-one of the 37 "presence" sites were included in the class-1 terminal nodes.

The original non-terminal node was split on WOOD7 at 20% (Table 5a). Eighty-two cases went to the right node; all were class-0 resulting in terminal node 5. The remaining 408 sites went to
the left node (non-terminal node 2). The first surrogate to the primary split was DSLOPE7 at 15.5° (Table 5a).

Non-terminal node 2 was then split on FLOOD7 at 2.5% with 159 cases going to the left (non-terminal node 3) and 249 going to the right (non-terminal node 4). Non-terminal node 4 received 31 of the class-1 sites (likely Topeka shiner habitat). The first surrogate of FLOOD7 was a reverse split on AG7 at 70% (Table 5a). Thus, cases with greater than 70% of their soils flooding at least every two years went to the left node and those less than or equal to it went to the right.

Non-terminal node 3 was split with variable DSLOPE7 at a value of 10.5°. The left node (terminal node 1) received a total of 149 cases, four being class-1, and was classified as class-0 (unlikely habitat). Three of these four class-1 sites were represented by a single fish and the other by two fish. The right node (terminal node 2) received ten cases, eight being class-0 (absence) and two being class-1 (presence). This terminal node was classified as 1 (likely habitat). One of the shiner sites produced four fish, whereas the other was represented by one fish.

Non-terminal node 4 was split on variable WOOD7 at a value of 10% with 201 cases going to the left node (terminal node 3) and 48 going to the right (terminal node 4). Terminal node 4 only received two class-1 cases, each represented by one fish, and was classified as 0 (unlikely habitat). Terminal node 3 received the remaining 29 class-1 cases and was classified as 1. The first surrogate to this split was DSLOPE7 at a value of 16.5° (Table 5a).

The total cross-validated cost for the general tree was 0.662. The cross-validated cost for class-0 was 0.364; 165 of the 453 class-0 cases were misclassified as likely habitat. The cross-validated cost for class-1 was 0.297; 11 of the 37 class-1 cases were misclassified.

The variable importance scores for the general model are given in Table 6a. Sample standard deviations and means for variables of cases predicted as shiner habitat are shown in Table 7a.

The two models derived from the class-1 terminal nodes of the general tree were: 1) If FLOOD7 > 2.5% and WOOD7 <= 10% then Topeka shiner habitat is likely to be present (terminal node 3). This model implies relatively high flood frequencies but low amounts of woody vegetation in
the riparian areas of these shiner sites. It accounted for 29 of the 37 Topeka shiner sites. 2) If WOOD7 <= 20% and FLOOD <= 2.5% and DSLOPE7 > 10.5° then Topeka shiner habitat is likely to be present (terminal node 2). This model implies relatively low amounts of woody vegetation and low flood frequencies, but relatively high average slopes in the riparian areas of these shiner sites. Two of the 37 Topeka shiner sites were described by this model.

I entered each model separately in ArcView (Appendix 1, 2) and overlaid them to produce a map identifying stream reaches on the DML more likely to be habitat for the fish (Fig. 6).

Had the terminal node 3 model not had FLOOD7 as a variable, I might have clipped the model coverage with a buffer of the RIVERS coverage. In that way, only areas close to a river or stream would have been identified. I did not do this because soils listed with a flood class must be near a river or stream. However, the terminal node 2 model does not use FLOOD7 as a mapping variable because it is essentially zero. I considered clipping this coverage with a buffer but decided not to because inspection of the map did not indicate that the terminal node 2 model predicts areas further from the stream than did the terminal node 3 model. This is likely a function of the DSLOPE variable, because on the DML there are very few places with relatively high slopes that are not near a stream or river.

Overall, the general model performed sufficiently to predict differences in habitat suitability among three drainage systems. The North Raccoon River drainage that produced the most Topeka shiner collections had a substantial proportion of stream length (35%) predicted as suitable habitat. In contrast, the neighboring Middle Raccoon River drainage has produced no shiner occurrences and had a lesser proportion of its overall stream length classified as shiner habitat (27%). To the northeast, the Boone River drainage has produced just two collections of one fish each, and only 6% of its overall stream length is classified as likely Topeka shiner habitat.

Specific Model:

I built the classification tree for the specific model with data from 40 off-channel sites, 12 having produced Topeka shiners, with drainage areas between 52 and 777 km², and using the same dependent and predictor variables as for the general model (Table 4b).
The resulting tree had five terminal nodes, one of which was classified as a likely habitat condition. The final split in this tree was on GRASS7 at 90% and only resulted in splitting out two class-0 (absence) cases in the right node. I believe this was the result of CART capitalizing on the peculiarities in this small data set. The analyst needs to be aware of this common problem with classification trees (Bell, 1996). To deal with the problem I simply selected the next smallest tree, having four terminal nodes, and used it for the specific model (Fig. 7).

The first split for the specific tree was on FLOOD7 at a value of 30%. The left node received 13 cases, all class-0, resulting in terminal node 1. The remaining cases went to the right (non-terminal node 2). The first surrogate to the primary split was ALLUV7 at 40% (Table 5b).

Non-terminal node 2 was split with variable FLOOD7 at 80%. The right node received four cases, all class-0, resulting in terminal node 4. The remaining cases went to the left node (non-terminal node 3). The first surrogate to this split was a reverse split on DSLOPE7 at 1.5° (Table 5b).

Non-terminal node 3 was split with GRASS7 at 50%. Its left node (terminal node 2) received six cases, one of which was class-1 represented by a single collection of 11 shiners. The remaining eleven shiner cases went to the right node, which was classified 1 (likely habitat). The first surrogate was a reverse split on AG7 at 50% (Table 5b).

The total cross-validated cost for the specific tree was 0.797. The cross-validated cost for class-0 was 0.464; 13 of the 28 class-0 cases were misclassified. The cross-validated cost for class-1 was 0.333; four of the 12 class-1 cases were misclassified. These larger costs relative to the general model were expected because the trees built during cross-validation testing with a smaller data set would have deviated further from the original tree than if they were based on a larger data set. Also, it is simply more difficult to develop good models with limited data.

The variable importance scores for the specific model are given in Table 6b. The sample statistics for variables of cases predicted as shiner habitat are shown in Table 7b.

I derived one model from terminal node 3 of the specific tree: If 30% < FLOOD7 ≤ 80% and GRASS7 > 50% then Topeka shiner habitat is more likely to be present. This model implies relatively high flood frequencies and a high proportion of grass in the riparian area.
The ArcView map resulting from this model identifies places on the DML more likely to have Topeka shiner habitat if the assumptions of sampling in off-channel habitat with drainage areas between 52 and 777 km² are met (Fig. 8). My intention for building this model was to identify areas most likely to have permanent populations of the fish. Therefore, it was expected that the model would identify fewer areas than the general model, which is the case.

**Discussion**

**Predictive Models:**

The two models developed here reveal some consistent relationships between landscape features and Topeka shiner occurrence on the DML. The general model indicates that stream areas with few riparian trees are more likely to have Topeka shiner habitat (Fig. 5). The surrogate to both splits on WOOD7 is DSLOPE, thus revealing that sites with greater tree cover have higher than average slopes in the riparian area. The Topeka shiner is a plains species, and therefore, probably is adapted to the relatively low gradient and treeless riparian conditions indicated by this split.

Another primary split in the general model indicates that shiner sites are surrounded by soils that flood frequently. The surrogate split was an inverse relationship on AG7, implying that sites which flood frequently are not commonly adjacent to row crop agriculture. Many DML headwater streams are channelized drainage ditches surrounded by agricultural fields and bordered by a narrow grassy riparian fringe. These tend to be deeply incised, and flood rarely.

The final primary split in the general model is DSLOPE. It leads to a terminal node that indicates that Topeka shiner habitat may be present in areas with few riparian trees and infrequent flooding, as long as the average slope of the surrounding area is relatively steep (terminal node 2). This split describes stream reaches with narrow valleys and floodplains. A reverse surrogate split on AG7 indicates these areas also have little cropland present near the stream. Because there are few trees present in the model landscape, it could be assumed that the riparian land cover at these sites is primarily grassland.
Several landscape feature associations were evident in the specific model, derived for off-channel situations (Fig. 7). The first split on FLOOD7 showed that Topeka shiners were found chiefly in wetlands surrounded by flooding soils. The first surrogate to this split, ALLUV7, indicates that such areas are likely to be associated with alluvial deposits. This is consistent with the fact that larger floodplains of DML streams were formed by glacial outwash materials. A likely effect of the underlying alluvium is provision of groundwater to off-channel habitats. This input could prevent summer desiccation and winter freeze-out and, therefore, help support semi-isolated Topeka shiner populations. Elsewhere in its range the shiner survives drought conditions in stream refuges provided by groundwater inputs (Pflieger, 1997).

The next split was also on FLOOD7, and its surrogate was a reverse split on DSLOPE7. This association is reasonable because within this generally low-relief landscape, there tends to be more "flooded" soils in areas of lower average slope. However, the primary split is counter-intuitive if flooding is actually supportive of Topeka shiner habitat. Because there were only 40 cases in the data set, it is not unreasonable to believe that the split could be a result of CART exploiting a peculiarity in the data. I considered removing this split but did not for two reasons. First, the split that follows does make sense and is useful, and second, there could be some underlying relationship that is important but unclear.

The last split in this tree was on GRASS, and its first surrogate was a reverse split on AG7. Within this agroecosystem, this association is logical because there is less cropping in places with more grass cover, which are used chiefly for livestock pasturing. Off-channel habitats which flood frequently and are surrounded by grass may provide better Topeka shiner habitat for two reasons. First, these areas are more likely to be permanent water bodies. Second, riparian grass may buffer the system from the effects of agriculture by filtering silt and agricultural chemicals from surface runoff (Williams & Nicks, 1988).

Regarding variable importance scores, the first and second highest scored variables for the general and specific models were either primary splits, or the first surrogate to a primary split, respectively. It is noteworthy that the most important variable in the general model, WOOD7, is the
least important in the specific model (Table 6). This is probably because the difference in the amount of riparian wooded cover between shiner presence and absence sites is greater in the general model and is more useful as a splitter (Table 4). PAL7 is a relatively unimportant variable in both models, possibly for two reasons. First, most off-channel habitats in floodplains may not have been properly identified in the NWI either because they were too small or because they were dry at the time the NWI aerial photographs were taken. Second, the ubiquitous nature of these wetlands on the DML could have prevented them from being useful for differentiating sites.

For the four most important variables in the model (WOOD7, DSLOPE7, FLOOD7, GRASS7), standard deviations were lower in the terminal node subsamples than in the original, complete data sets (Table 7). This is an expected consequence of CART analysis since the process selects the most informative and reliable range of the variables to develop the model.

Several assumptions potentially affect the accuracy of these models. First, the models predict habitat for the Topeka shiner but not its actual presence. Within the DML, the Topeka shiner does not occur, and perhaps never did, in some apparently suitable habitat. For example, it has never been reported from the Skunk River on the DML, which is similar to streams from which it is known (Fig. 2). This is likely attributable to geographic isolation between the Skunk and Des Moines river systems during the Topeka shiner’s residency in Iowa. Also, stochastic events both natural and anthropogenic, have probably destroyed isolated populations throughout that history.

Second, the coverages used to build the models were assumed to accurately reflect conditions at the collection sites, but sources of error do exist. For example, the positional accuracy for most of the coverages is considered to be within +/- 50 m. However, it is believed to be only +/- 300 m for ALLUV100, although it may be better for the DML data which was digitized at a smaller scale. Additionally, attribute classification could be a source of error for LCOV92 because land cover types were interpreted from spectral classes with limited ground-truthing and because vegetative cover can change over time. Finally, classification error is probably present in the flooding attribute of the soils data. This is the result of the use of different flooding classification methods among county soil surveys, an example of a common problem whenever independent classification efforts are joined.
for one purpose. These likely sources of error can not be corrected without extensive effort, and must be accepted as a methodological weakness.

Third, the sites used in the CART analysis were assumed to be spatially independent. Because occurrence of the shiner at one site on a stream increases the likelihood of its presence at another site on the same stream, this assumption is false. The violation of this assumption is one of the chief reasons why the analysis cannot provide probability estimates for habitat suitability.

Finally, an assumption is that the models will be applied to field conditions under which they were developed. For the general model this only means that habitat sites must be in places accessible to and generally adequate for survival of small fishes during the warmer months. Assumptions for the specific model include those conditions but also restrict sites to off-channel habitats where the nearby channel has a drainage area of 52 to 777 km².

Judging by what is already known of Topeka shiner distribution and ecology in Iowa, several lines of evidence suggest that the models are useful predictors of shiner habitat. First, the general model predicts more habitat than the specific model (Figs. 6, 8). I expected this because assumptions in the specific model limited it to what I a priori identified as some characteristics of "preferred" habitat. Second, the general model predicts relatively more habitat in drainages that are known to have more extant populations (Fig. 4). Third, neither model predicts habitat in or near artificial drainage ditches. There are few records of shiners from drainage ditches, and all such collections produced few fish, probably waifs.

However, there are two problems evident in the models as well. First, discrepancies in the soil flooding classifications caused differences in the amount of predicted habitat between some adjacent counties. This created discernible county borders in a map created from terminal node 3 of the general model (Appendix 1). This is not a problem in a map created from terminal node 2 of the general model (Appendix 2) because it does not utilize the soils data. Second, there is a void in all maps due to the lack of digital soils data for Humboldt county.
Protection of the Topeka Shiner:

In general, these results describe several landscape elements of Topeka shiner habitat on the DML. Suitable areas include low gradient reaches of natural, meandering low-order streams, as well as off-channel areas of natural and artificial origin. Riparian areas are predominantly grassy and flood frequently. The valleys are underlain by glacial alluvium, and morainal deposits are a common feature of the local landscape. At least temporary habitat is afforded in floodplain depressions along major rivers. Constructed drainage ditches are not suitable long-term habitat for the Topeka shiner although waifs occur in them occasionally.

Field observations suggest that many, perhaps most, sites at which the shiner occurs are seasonally ephemeral and do not provide year-around habitat. In this regard, a distinction may be made between source and sink habitats for the species. A source habitat is one in which annual reproductive gains exceed mortality losses (Pulliam 1988) and a sink habitat is the opposite. Available evidence on reproduction and over-winter survival suggests that only a small proportion of the known Topeka shiner sites are source habitats. Permanently watered off-channel areas such as oxbows seem to be the most important of these habitats. Within this geographical region, therefore, *N. topeka* has characteristics of a metapopulation, a set of genetically connected local populations living in semi-isolated habitat areas within a larger region (Hanski & Simberloff, 1997).

Flooding seems to be the primary mechanism for maintaining DML Topeka shiner metapopulations. On the one hand, it probably disperses many fish into ephemeral sink habitats where they ultimately perish without reproducing. However, in a region where flash flooding is common, even temporary habitats occasionally may be sufficient to permit reproduction and subsequent recruitment to other areas. Under this scenario, over the long-term, a given area may act both as sink and source habitat according to local weather and hydrologic events.

This conceptual viewpoint has implications for protection of the species in this region. Conservation efforts might best be directed toward maintaining or even enhancing source habitats, especially oxbows, abandoned floodplain gravel pits, and similar areas known to support permanent populations. Because many such areas are now perched 1 m to 2 m above the water table due to
agricultural land drainage, deepening them by excavation may be feasible for maintaining an adequate permanent water supply.

Additionally, appropriate management of local riparian areas may contribute to protection of the fish’s habitat as well as achieving more generalized environmental benefits. Federal and state subsidy programs in place today encourage landowners and farmers to establish vegetated riparian buffer strips to reduce inputs of sediments and agricultural chemicals to surface waters (Williams & Nicks, 1988; Arora et al., 1996; Herron & Hairsine, 1998). Such buffers also reduce overland flow volume and increase infiltration near the channel (Munoz-Carpena, Parsons & Gilliam, 1993); which, if applied at a sufficiently large scale, could offer potential for stabilizing stream hydrologic conditions and raising local water tables. Plantings of native grasses and forbes in such areas would seem to most closely achieve the natural riparian vegetation conditions in which the Topeka shiner evolved.

**Applications of the Models:**

The models developed here have at least two field applications in Topeka shiner conservation programs. First, the general model can be used to determine if a site is “likely” habitat for the Topeka shiner. This would be useful in situations where it is important to judge that the fish is not present, without conducting time-consuming and expensive collecting efforts. The general model would be applicable in this situation because it was developed for all sites, and therefore covers a broad range of conditions. As an example, potential harmful impacts on the fish by proposed drainage ditch maintenance may be evaluated even when no direct fish collection evidence is available.

Second, the specific model could be used to determine if a site is “likely” Topeka shiner habitat worthy of protection. Because this model was developed for “typical-quality” off-channel habitats, it is better suited to identify areas that may be able to provide reproductive, or source-habitat. This may be useful for making a decision as to whether or not a site is suitable for habitat improvement work or reintroductive efforts.

Applying these models in a field situation is rather simple. A county soil survey can provide estimates of flooding frequency and average slopes of the soils in the area of interest. Aerial
photographs or on-site evaluation can be used to estimate the relative amounts of riparian woody or grass cover within about a 200 m radius of the site. The specific model, however, only applies for off-channel situations in drainages between 52 and 777 km² (20 to 300 mi²). Using the values estimated for these four variables, the user follows the splits in the applicable tree until a terminal node is reached. The class of this node corresponds to the predicted site class (habitat present/absent).

As an example, consider a case of proposed drainage ditch maintenance. The soil survey and an on-site evaluation resulted in: wooded cover = 0%, no soils have flood classes of “Common” or “Frequent”, and the average slope of the surrounding soils is 2-6%. To use the general model, start with node 1 which was split on WOOD7 (Fig. 5). Because the wooded cover of the site is less than 20%, follow the left branch of the tree to node 2. At node 2, note that the site does not contain soils that have a flood class of “Common” or “Frequent”; and, follow the left branch to node 3. This node is split on slope and its value is in degrees. Because the soil survey gives slope in percent, I converted the 10.5° criterion (slopepercent = tan (slopedegrees) * 100) to 18.5%. Since this site has slopes less than 18.5% (10.5°), follow the left branch to terminal node 1 which classifies the site as an absence (0), i.e., the model predicts that no suitable shiner habitat is present.

It is emphasized, however, that accuracy of probability estimates of the result are not possible. Such an application should be applied cautiously and only considered to give a reasonable idea of what to expect at a site.

Conclusions:

The technique applied here utilizing GIS and CART has led to the successful development of two predictive Topeka shiner habitat models. Furthermore, classification tree analysis has proven to be a viable alternative to more traditional statistical methods, especially for its ease of interpretation during model development and during subsequent field use by those unfamiliar with statistics.

The models developed here have several field applications for Topeka shiner protection and enhancement efforts. The general, conservative model is perhaps best applied to determine where the fish is not likely to occur. Regulatory agencies, for example, might use it to screen proposed projects involving highway bridge construction, drainage ditch maintenance, and similar operations in
waterways for potential impacts on local Topeka shiner populations. The specific model can be applied to identify likely areas of habitation. Such predictions can also be useful for evaluating possible impacts of waterway construction projects, aiding the search for additional habitat areas, and evaluating the suitability of sites for habitat improvement.

References


Table 1: Occurrence of the Topeka shiner in Des Moines Lobe localities, 1970 to 1999.

<table>
<thead>
<tr>
<th>Habitat Type</th>
<th>Number of Sites</th>
<th>Proportional Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TS Present</td>
<td>TS Absent</td>
</tr>
<tr>
<td>Meandering Stream Channel</td>
<td>14</td>
<td>338</td>
</tr>
<tr>
<td>Channelized Stream or Drainage Ditch</td>
<td>4</td>
<td>66</td>
</tr>
<tr>
<td>Naturally Formed Oxbow</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Artificially Formed Oxbow</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>Gravel Pit or Man-made Pond</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Other or Unknown</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>453</td>
</tr>
</tbody>
</table>

Table 2: Number of Topeka shiners collected at 37 sites on the Des Moines Lobe ordered by drainage area. Reproductive sites (at least two year-classes present) denoted as \(\geq 25\) shiners.

<table>
<thead>
<tr>
<th>Number of Shiners</th>
<th>Habitat (in- or off-channel)</th>
<th>Drainage Area (km²)</th>
<th>Number of Shiners</th>
<th>Habitat (in- or off-channel)</th>
<th>Drainage Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>IN</td>
<td>5</td>
<td>(\geq 25)</td>
<td>OFF</td>
<td>267</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>16</td>
<td>11</td>
<td>OFF</td>
<td>269</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>34</td>
<td>(\geq 25)</td>
<td>OFF</td>
<td>269</td>
</tr>
<tr>
<td>1</td>
<td>OFF</td>
<td>39</td>
<td>1</td>
<td>OFF</td>
<td>269</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>44</td>
<td>5</td>
<td>OFF</td>
<td>272</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>49</td>
<td>2</td>
<td>IN</td>
<td>287</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>52</td>
<td>5</td>
<td>IN</td>
<td>293</td>
</tr>
<tr>
<td>7</td>
<td>IN</td>
<td>60</td>
<td>1</td>
<td>OFF</td>
<td>293</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>62</td>
<td>1</td>
<td>OFF</td>
<td>303</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>62</td>
<td>(\geq 25)</td>
<td>OFF</td>
<td>324</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>72</td>
<td>1</td>
<td>OFF</td>
<td>334</td>
</tr>
<tr>
<td>(\geq 25)</td>
<td>OFF</td>
<td>129</td>
<td>3</td>
<td>IN</td>
<td>365</td>
</tr>
<tr>
<td>2</td>
<td>IN</td>
<td>129</td>
<td>1</td>
<td>IN</td>
<td>549</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>132</td>
<td>1</td>
<td>OFF</td>
<td>1836</td>
</tr>
<tr>
<td>2</td>
<td>IN</td>
<td>140</td>
<td>6</td>
<td>OFF</td>
<td>2893</td>
</tr>
<tr>
<td>2</td>
<td>IN</td>
<td>160</td>
<td>1</td>
<td>OFF</td>
<td>2893</td>
</tr>
<tr>
<td>6</td>
<td>OFF</td>
<td>189</td>
<td>4</td>
<td>OFF</td>
<td>3170</td>
</tr>
<tr>
<td>1</td>
<td>IN</td>
<td>232</td>
<td>1</td>
<td>OFF</td>
<td>3206</td>
</tr>
<tr>
<td>1</td>
<td>OFF</td>
<td>264</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Initial GIS databases used in study. Sources: Iowa Department of Natural Resources NRGIS Library website - www.igsb.uiowa.edu/nrgis/gishome.htm (IDNR); Geographic Information Systems facility at Iowa State University (ISU); Iowa Department of Natural Resources - Geological Survey Bureau (GSB). * Created using imagery from 1992. ** Created using data from 1970 - 1999.

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Source Scale</th>
<th>Form</th>
<th>Created</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALLUV100</td>
<td>alluvial deposits</td>
<td>1:24,000</td>
<td>vector</td>
<td>1991</td>
<td>IDNR</td>
</tr>
<tr>
<td>PLSS</td>
<td>section lines</td>
<td>1:24,000</td>
<td>vector</td>
<td>1990</td>
<td>IDNR</td>
</tr>
<tr>
<td>RIVERS</td>
<td>rivers &amp; streams</td>
<td>1:100,000</td>
<td>vector</td>
<td>1992</td>
<td>IDNR</td>
</tr>
<tr>
<td>BASIN24</td>
<td>drainage basins</td>
<td>1:24,000</td>
<td>vector</td>
<td>1996</td>
<td>IDNR</td>
</tr>
<tr>
<td>NWI</td>
<td>wetlands</td>
<td>1:24,000</td>
<td>vector</td>
<td>1996</td>
<td>IDNR</td>
</tr>
<tr>
<td>LOBE9</td>
<td>glacial moraines</td>
<td>1:250,000</td>
<td>vector</td>
<td>1995</td>
<td>GSB</td>
</tr>
<tr>
<td>LCOV92</td>
<td>vegetative cover</td>
<td>30x30 meters</td>
<td>raster</td>
<td>1998*</td>
<td>GSB</td>
</tr>
<tr>
<td>SOILS</td>
<td>county soils</td>
<td>0.81 hectares</td>
<td>raster</td>
<td>1999</td>
<td>ISU</td>
</tr>
<tr>
<td>ELEVATION30</td>
<td>digital elevation</td>
<td>30x30 meters</td>
<td>raster</td>
<td>1999</td>
<td>ISU</td>
</tr>
<tr>
<td>SITES</td>
<td>collection locations</td>
<td>varied</td>
<td>vector</td>
<td>1999**</td>
<td>Original</td>
</tr>
</tbody>
</table>

Table 4: Means and sample standard deviations (SD) for variables used in CART models.

(a) General Habitat Model (all sites)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shiner Absent</th>
<th>Shiner Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>WOOD7</td>
<td>10.9</td>
<td>14.1</td>
</tr>
<tr>
<td>GRASS7</td>
<td>43.1</td>
<td>24.0</td>
</tr>
<tr>
<td>AG7</td>
<td>40.8</td>
<td>27.6</td>
</tr>
<tr>
<td>ALLUV7</td>
<td>66.6</td>
<td>33.6</td>
</tr>
<tr>
<td>DSLOPE7</td>
<td>5.9</td>
<td>4.2</td>
</tr>
<tr>
<td>PAL7</td>
<td>6.3</td>
<td>11.1</td>
</tr>
<tr>
<td>FLOOD7</td>
<td>25.9</td>
<td>25.4</td>
</tr>
</tbody>
</table>

(b) Specific Habitat Model (off-channel sites with drainage areas from 5 - 3206 square kilometers)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shiner Absent</th>
<th>Shiner Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>WOOD7</td>
<td>2.7</td>
<td>5.5</td>
</tr>
<tr>
<td>GRASS7</td>
<td>56.1</td>
<td>22.8</td>
</tr>
<tr>
<td>AG7</td>
<td>33.6</td>
<td>21.4</td>
</tr>
<tr>
<td>ALLUV7</td>
<td>72.0</td>
<td>29.3</td>
</tr>
<tr>
<td>DSLOPE7</td>
<td>4.9</td>
<td>2.8</td>
</tr>
<tr>
<td>PAL7</td>
<td>5.5</td>
<td>5.2</td>
</tr>
<tr>
<td>FLOOD7</td>
<td>37.9</td>
<td>32.3</td>
</tr>
</tbody>
</table>
Table 5: Primary splitters and surrogates for the CART trees. Surrogate values denoted as (r) are reverse splits.

(a) General Habitat Model (all sites)

<table>
<thead>
<tr>
<th>Node (non-terminal)</th>
<th>Primary Splitter</th>
<th>Value</th>
<th>Surrogate</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WOOD7</td>
<td>≤ 20%</td>
<td>DSLOPE7</td>
<td>≤ 15.5°</td>
</tr>
<tr>
<td>2</td>
<td>FLOOD7</td>
<td>≤ 2.5%</td>
<td>AG7</td>
<td>≥ 70% (r)</td>
</tr>
<tr>
<td>3</td>
<td>DSLOPE7</td>
<td>≤ 10.5°</td>
<td>AG7</td>
<td>≥ 10% (r)</td>
</tr>
<tr>
<td>4</td>
<td>WOOD7</td>
<td>≤ 10%</td>
<td>DSLOPE7</td>
<td>≤ 16.5%</td>
</tr>
</tbody>
</table>

(b) Specific Habitat Model (off-channel sites with drainage areas from 5 - 3206 square kilometers).

<table>
<thead>
<tr>
<th>Node (non-terminal)</th>
<th>Primary Splitter</th>
<th>Value</th>
<th>Surrogate</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FLOOD7</td>
<td>≤ 30%</td>
<td>ALLUV7</td>
<td>≤ 40%</td>
</tr>
<tr>
<td>2</td>
<td>FLOOD7</td>
<td>≤ 80%</td>
<td>DSLOPE7</td>
<td>≥ 1.5° (r)</td>
</tr>
<tr>
<td>3</td>
<td>GRASS7</td>
<td>≤ 50%</td>
<td>AG7</td>
<td>≥ 50% (r)</td>
</tr>
</tbody>
</table>

Table 6: Relative variable importance scores for the CART trees.

(a) General Habitat Model (all sites)

<table>
<thead>
<tr>
<th>Variable</th>
<th>WOOD7</th>
<th>FLOOD7</th>
<th>DSLOPE7</th>
<th>AG7</th>
<th>ALLUV7</th>
<th>PAL7</th>
<th>GRASS7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>100</td>
<td>54</td>
<td>40</td>
<td>39</td>
<td>7</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

(b) Specific Habitat Model (off-channel sites with drainage areas from 5 - 3206 square kilometers)

<table>
<thead>
<tr>
<th>Variable</th>
<th>FLOOD7</th>
<th>DSLOPE7</th>
<th>GRASS7</th>
<th>ALLUV7</th>
<th>AG7</th>
<th>PAL7</th>
<th>WOOD7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>100</td>
<td>27</td>
<td>23</td>
<td>23</td>
<td>15</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 7: Comparison of sample means and standard deviations (SD) for model variables of all cases, to those groups of cases predicted as Topeka shiner habitat (terminal nodes).

(a) General Habitat Model (all sites)

<table>
<thead>
<tr>
<th>Variable</th>
<th>All Cases (n = 490)</th>
<th>Terminal Node 3 (n = 201)</th>
<th>Terminal Node 2 (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
</tr>
<tr>
<td>WOOD7</td>
<td>10.3 13.7</td>
<td>3.1 2.4</td>
<td>8.0 6.3</td>
</tr>
<tr>
<td>DSLOPE7</td>
<td>5.9    4.2</td>
<td>- 42.7</td>
<td>- 20.2</td>
</tr>
<tr>
<td>FLOOD7</td>
<td>26.7   25.4</td>
<td>20.2</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) Specific Habitat Model (off-channel sites with drainage areas from 5 - 3206 square kilometers)

<table>
<thead>
<tr>
<th>Variable</th>
<th>All Cases (n = 40)</th>
<th>Terminal Node 3 (n = 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
</tr>
<tr>
<td>GRASS7</td>
<td>58.5 21.8</td>
<td>71.5 14.1</td>
</tr>
<tr>
<td>FLOOD7</td>
<td>41.8 28.4</td>
<td>50.9 10.6</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Rangewide distribution of the Topeka shiner as of 1997. The Iowa distribution has been updated to reflect collections through 1999.

Figure 2: Topeka shiner distribution in Iowa.

Figure 3: Location of study region (Des Moines Lobe) as it relates to Iowa.

Figure 4: Fish survey stream collections on the Des Moines Lobe, 1970-1999.

Figure 5: Classification tree developed for general model using all sites without limits on drainage area. (class 0 = Topeka shiner absent, class 1 = Topeka shiner present).

Figure 6: Map identifying areas on the Des Moines Lobe predicted by terminal nodes 2 and 3 of the general model to be suitable habitat for the Topeka shiner.

Figure 7: Classification tree developed for specific model using only off-channel sites near stream reaches with drainage areas from 52 to 777 square kilometers (20 to 300 square miles). (class 0 = Topeka shiner absent, class 1 = Topeka shiner present).

Figure 8: Map identifying areas on the Des Moines Lobe predicted by the specific model to be suitable habitat for the Topeka shiner.
Figure 1

- Extant in 1990's
- Presumed Extirpated
Middle Raccoon River Drainage

Boone River Drainage

North Raccoon River Drainage

- Topeka Shiner Present (37)
- Topeka Shiner Absent (453)

Figure 4
Figure 6

- Predicted Habitat
- Humboldt County - No Data
Figure 7
Figure 8

- Predicted Habitat
- Humboldt County - No Data
CHAPTER 3: GENERAL CONCLUSIONS

The Topeka shiner was recently placed on the federal endangered species list; therefore, steps must be taken to ensure its survival. Previous efforts to locate extant populations of the shiner in Iowa, based on historical collection sites and descriptions of preferred habitat, resulted in little success. It became clear that finding the fish would require a new approach. Because lotic systems are intimately tied to the landscapes they drain, application of Geographic Information Systems (GIS) technologies, coupled with spatial data sets, could lend a better understanding of the shiner’s habitat. This approach has been used in many wildlife studies, and to a lesser extent in aquatic ones. However, studies such as the one here will become more commonplace as researchers recognize the benefits of this technique and as GIS software becomes more powerful and easier to use.

The objectives of this study were: 1) to develop GIS-based models of the association of Topeka shiner habitat to local landscape features on the Des Moines Lobe using existing GIS data layers, and 2) to apply the models throughout the landform to identify likely Topeka shiner sites for future search efforts.

The two models I developed, along with information from new collecting efforts, revealed some characteristics common to existing Topeka shiner habitat on the Des Moines Lobe. First, I found that the shiner exists in off-channel habitats in Iowa, sometimes abundantly. Second, Topeka shiner habitat is typically found in or near streams that approximate natural conditions. Channelized reaches are generally not suitable habitat. Third, the immediate riparian environments are typically grassland, and often grazed; row crop agriculture and forest/woody cover is limited. Finally, flooding of the off-channel habitats on a regular basis and the presence of groundwater-bearing alluvial deposits could be critical for maintaining permanent populations.

Because modeling is an imperfect science, the maps created with the models have some inaccuracies. First, they probably over-predict the actual presence of shiner habitat. Also, soils data is lacking for Humboldt county, preventing modeling in this area. Finally, there are inconsistencies in the soils data between counties.
However, the models do seem to perform reasonably well and at least serve to eliminate many areas as likely habitat. This alone can be useful for focusing the search for extant populations of the fish in Iowa. Also, the models could be used to identify areas that may be suitable for habitat improvement and re-introductive efforts for the fish.

This study has shown that a landscape-level approach can be useful for identifying characteristics of the aquatic habitats for a single species. Its products are offered as a tool in the effort to protect the future of the Topeka shiner in Iowa.
APPENDIX 1. MAP OF GENERAL MODEL - TERMINAL NODE 3

Predicted Habitat
Humboldt County - No Data
APPENDIX 2. MAP OF GENERAL MODEL - TERMINAL NODE 2

Predicted Habitat
Humboldt County - No Data
APPENDIX 3. METHODS FOR GIS USERS

Digital Data Sources:

I obtained existing digital data maps from three sources. I downloaded coverages of alluvial deposits (ALLUV100), section lines from the public land survey system (PLSS), rivers and streams (RIVERS), drainage basins (BASIN24), and the National Wetlands Inventory (NWI) from the Iowa Department of Natural Resources (IDNR) website - www.igsb.uiowa.edu/nrgis/gishome.htm. ALLUV100 is a state-wide vector coverage of alluvial deposits digitized on 1:24,000 scale maps based on county soil surveys. RIVERS are separate county-wide vector coverages of rivers and streams created from 1:100,000 scale digital line graph files from the United States Geological Survey (USGS). PLSSs are separate county-wide vector coverages of section lines digitized from USGS 7.5' topographic maps. BASIN24 is a coverage of drainage basins in Iowa delineated from 1:24,000 scale base maps. NWIs are separate county-wide vector coverages of the National Wetlands Inventory which are records of wetland location and classification as defined by the United States Fish and Wildlife Service (USFWS). Wetland features were interpreted from color-infrared photos taken in 1983 and 1984 and digitized at a scale of 1:24,000. Further information regarding these coverages can be found on the website.

I acquired coverages for vegetative cover (LCOV92), county soil survey data (SOILS), and a digital elevation model (elevation30) from the Geographic Information Systems facility at Iowa State University (ISU). Elevation30, developed by the IDNR, is a mosaic of individual USGS Digital Elevation Model quadrangles for Iowa. Additional information on the original data sources can be found on the USGS website - http://edcwww.cr.usgs.gov/glis/hyper/guide/usgs_dem. SOILS is a raster coverage of soils data developed by the Iowa Cooperative Soil Survey and the IDNR, Geological Survey Bureau (GSB). It is a mosaic of rasterized township polygons of soil mapping units taken from the published county soil survey report of Iowa. The minimum size delineation of soil polygons is 0.8 hectares (2 acres). Each soil mapping unit can be linked to the Iowa Soil Properties and Interpretations Database (ISPAID) containing many attributes for each soil type. Additional
information on the original soil coverages can be found on the Iowa Cooperative Soil Survey website - http://icss.agron.iastate.edu/data.htm.

LCOV92 is a raster coverage of land use produced in 1992 by the GSB with satellite imagery from the Landsat Thematic Mapper sensor. Each 30 meter pixel in this coverage was placed into one of seven different classes: row crops, grasslands, water cover, barren/flooded cover, artificial cover (urban), forest/woody cover, and cloud covered areas. The GSB provided a coverage of the DML (LOBE9) delineating boundaries, various advances, and end moraines. LOBE9 was digitized in 1995 by the GSB at a scale of 1:250,000. Further information for LCOV92 and LOBE9 can be obtained from the GSB at 109 Trowbridge Hall, Iowa City, IA.

I digitized field collection sites on-screen in ArcView 3.2 as a new point coverage. To do this, I first identified site locations from field maps and notes. Then, using RIVERS and PLSS as references, I marked site locations on-screen and assigned them unique identification numbers. In cases of multiple collections at a site, I included only one record, retaining the record with the shiner, if it was present. After the site locations were digitized, I joined an existing table containing field data for each site to the attribute table of the point coverage. The existing table contained information about the number of Topeka shiners collected, the drainage area for the site, and the type of habitat collected. The habitat type was identified as either in-channel or off-channel. I named the coverage SITES and converted it to a grid coverage at the same cell size (30 meters) and extent as elevation30. I set cell size and extent equal to elevation30 to ensure that the grids would align correctly and also did so for all vector-to-grid conversions.

Digital Data Processing:

The next step was to process the data so that relationships between the landscape variables and the collection sites could be identified and extracted for statistical analysis. I used ArcView with the Spatial Analyst and XTools extensions for all GIS applications.

I first clipped existing coverages to the borders of the DML. The elevation30, LCOV92, and SOILS coverages were in a raster format, and because of certain software constraints, I converted the vector coverage LOBE9 to a graphic before clipping the grids. Subsequent clipping with this
graphic proved problematic for the software because the coverage had some small disjunct pieces in
the northeastern part of the DML. Therefore, I slightly modified the graphic to include these pieces as
one contiguous shape by extending its border. As a result, a negligible area is included in the study
area that was not glaciated by the DML. I used the new graphic to clip the elevation30, LCOV92, and
SOILS coverages. I also converted the graphic to a shapefile (EXTENT) to clip vector coverages. I
clipped ALLUV100 with EXTENT and then converted it to a grid with cell values equal to alluvium
(present or absent) and named it ALLUV. Because the NWIs were county-wide coverages, I first
merged them into one coverage (NWI). I clipped NWI with EXTENT and converted it to a grid with
cell values equal to system (wetland type).

The next step was to process the clipped coverages to show the variables I wished to include
in the model. I used Spatial Analyst's Derive Slope function to return a slope value (in degrees) for
each cell from elevation30 and named it DSLOPE. Within LCOV92, I was interested in the effects of
row crops, grasslands, and forest/woody cover. Because I wanted each cover type to represent its
own variable in the model, I used Spatial Analyst's Reclassify function to change all cell values to
zero except the variable of interest. Repeating this process three times resulted in the coverages AG
(row crops), GRASS (grasslands), and WOOD (forest).

I wanted to use the NWI as a measure of the amount of off-channel habitat within an area.
The NWI "palustrine" classification (area < 8 hectares, depth < 2 m) best described off-channel
Topeka shiner habitats. Therefore, I used the Reclassify function to set all wetland types but
"palustrine" to zero and named the new grid PAL.

The SOILS coverage contains many attributes for each soil type. Flood frequency is an
attribute that could play an important role in the permanence of off-channel habitats and recruitment
of Topeka shiners to a nearby stream. Thus, it was reasonable to be interested in soils that flooded
at least once every two years; therefore, I reclassified all soils attributes to zero except for soils that
flood "commonly" and "frequently" and called the new coverage FLOOD. I did not include the
"ponded" soil classification because it refers to soils in closed depressions which contain water
primarily after precipitation and runoff, such as those found in former glacial potholes that are now part of agricultural fields.

I also had to change the scale at which the landscape variables were represented. All landscape data layers existed in a raster format with 30 meter grid cells. However, I was interested in a larger area surrounding a collection site, which reflected broader landscape conditions. Similar aquatic studies have described landscape variables (Richards & Host, 1994; Richards et al., 1996; Roth et al., 1996; Lammert & Allan, 1999) within catchments or different sized buffer widths along stream reaches. In the present case, it would be possible to delineate watersheds for the sample sites, but it is not possible to do the same for every grid cell of stream length in the predictive modeling step. The same is true for buffering stream reaches; therefore, I tested two buffer options: 120 m (four - 30 m grid cells) and 210 m (seven - 30 m grid cells) radii around sites for which to measure the landscape variables.

I used Spatial Analyst’s Neighborhood Statistics function to measure the mean slope around each cell of DSLOPE at the 210 m and 120 m radii. I used Spatial Analyst’s Map Calculator function to convert the average slope values to integers, and produce a table of the values. I called the resulting grids DSLOPE7 and DSLOPE4.

I used the same neighborhood analysis approach for AG, GRASS, WOOD, ALLUV, PAL, and FLOOD, but with sum rather than mean. This resulted in a count of the number of cells in the neighborhood around a site, designated as a particular variable type. I divided this sum by the total number of cells in the neighborhood and multiplied it by 100. This resulted in a percentage of cells with the variable type within 210 meters of a site. I then re-classified these proportions into 11 classes where: 0 = 0, 1-10 = class 5, 11-20 = class 15, 21-30 = class 25,..., and 91-100 = class 95. I named the resulting grids in a similar manner to DSLOPE.

The next step was to identify the value of each variable at each site. To do this, I used Spatial Analyst’s Summarize Zones command to summarize the selected grid (variable) within the zone of the SITES grid. This returned a table that contained the site’s identification number and the single predominant value for each variable. I joined this table to the SITES table. Because such
joins are not permanent, I created a new field in the SITES table and named it as the variable being measured. I set this new field equal to the predominant value in the joined table. I did this repeatedly until the values of all seven variables were included in the SITES table.

The dependent variable in the model, Topeka shiner occurrence, may be expressed in different ways, ranging from simple presence/absence to semi-quantitative estimates of abundance and other observed biological characteristics. I believed that exploring relationships based on different classification systems for the dependent variable would provide insight into the characteristics of “better” habitats and may even be necessary for meaningful analyses due to the limited amount of available data. Therefore, I added three new dependent variables to the data table;

CLASS1: absent = 0, present = 1; CLASS2: absent = 0, 1-2 shiners collected = 1, ≥3 collected = 3; CLASS3: absent = 0, 1-2 collected = 1, ≥3 collected and no evidence of reproduction = 3, and evidence of reproduction (numerous adults and juveniles collected) = 4.

In general, collections of *N. topeka* were associated with tributary streams rather than major rivers. Thus I examined the drainage areas of the sites where the fish was collected to determine a basin size range that described the majority of collections. I used this information as a possible assumption in the model to limit the collection data to sites with higher probabilities of Topeka shiner occurrence. I could not use it as a variable in the model, however, because this information would not be available for the predictive modeling step.

**Predictive Modeling:**

After I developed two acceptable trees based on different modeling assumptions, I performed the predictive modeling step. I did this by identifying the set of relationships that led to each node classified as a shiner presence. This resulted in two models for the general tree and one for the specific. In ArcView, I added the theme for each variable in the model to the active view and separately entered each model into Spatial Analyst’s Map Query function. This resulted in coverages for each model, identifying every grid cell on the DML “likely” to be Topeka shiner habitat. If a model utilized FLOOD7 it was not necessary to process the model coverage any further to eliminate
predicted areas unreasonably far from stream channel. If this was not the case, the map was examined to determine if further processing was necessary.

To further evaluate the ability of the general model to identify likely shiner habitat, I measured the proportion of the stream length predicted to have such habitat to the total stream length, in each of three large river basins on the DML. The reason for doing this was to see if the model was predicting relatively more habitat in a basin that I knew had substantial populations of shiners (North Raccoon), relative to two basins that did not (Middle Raccoon and Boone). To do this I first merged all the RIVER coverages into one large coverage and clipped it with EXTENT. I selected the smaller basins of the North Raccoon river in the BASIN24 coverage which included the majority of the shiner collections. I converted these basins to a shapefile and used it to clip RIVERS. I used the Calculate function in the clipped RIVERS table to recalculate the length of the individual stream segments since ArcView does not do this automatically. The Statistics function returned the sum of the segment lengths, thus giving the total stream length in the basin. Following this, I used a shapefile of the general predictive model to clip out all the stream segments predicted as habitat for the fish in that basin. Using the same method to measure stream length as above, I calculated the total stream length predicted to have the fish. I followed the same procedures for the Boone and Middle Raccoon rivers, except that I selected all basins for each river on the DML, and the basin for the Middle Raccoon river had to be clipped to the extent of the lobe.
ACKNOWLEDGMENTS

To my major professor Dr. Bruce W. Menzel, I will be forever grateful. He allowed me the opportunity to achieve this goal when I thought no one would. Through his guidance I have learned many things that will serve me well in life.

For serving as committee members, I would like to thank Drs. Diane M. Debinski, Joseph E. Morris, and Steven E. Jungst. Dr. Debinski, provided much needed support in my choice of statistical analysis and in my development as an ecologist. Dr. Morris was a source of assistance whenever I needed it.

A special thanks to Kevin Kane, Kathy Anderson, and Robin McNeely for assistance with GIS programming. Also to Deb Quade, from the University of Iowa, for providing me with the coverage of the Des Moines Lobe and for some useful discussion. Dr. Philip Dixon, an excellent statistics professor, deserves thanks for introducing me to, and assisting me with CART.

I would like to thank Ryan Lane, Schuyler Sampson, Ben LaFrentz, and Travis Paul for many hours searching the streams of Iowa for the elusive Topeka shiner. I would also like to thank Jeff Kopaska for general advice and assistance with field equipment.

To my parents, Carson and Kathleen, I not only want to thank you for your financial and emotional support throughout my collegiate career, but mostly for good values and the belief that I can accomplish anything I desire. Dad, thanks for introducing me to the outdoors and for buying that first boat that started me on this path.

To my wife and best friend Jennifer, thank you for all of your sacrifices and support. Without you, the rewards gained from this experience would be meaningless.

Finally I would like to thank Iowa State University, the Iowa Department of Natural Resources, and the United States Fish and Wildlife Service for providing the financial support which made this project possible.