UNIVERSITY OF KENTUCKY
College of Engineering

Kentucky Transportation Center

A GIS-Based Expert Systems Predictive Habitat Model for Threatened and Endangered Species: Case Study Using Kentucky Arrow Darter

<u>Our Mission</u>

We provide services to the transportation community through research, technology transfer and education. We create and participate in partnerships to promote safe and effective transportation systems.

Kentucky Transportation Center

176 Oliver H. Raymond Building

Lexington, KY 40506-0281

(859) 257-4513

fax (859) 257-1815

www.ktc.uky.edu

Research Report
KTC-15-10/ SPR14-483-1F

A GIS-Based Expert Systems Predictive Habitat Model for Threatened and Endangered Species: Case Study Using Kentucky Arrow Darter

by

Benjamin Blandford, PhD
Research Scientist

Michael Shouse, PhD
Research Associate

and

John Ripy

Kentucky Transportation Center
College of Engineering
University of Kentucky
Lexington, Kentucky

in Cooperation with

Kentucky Transportation Cabinet
Commonwealth of Kentucky

and

Federal Highway Administration
U. S. Department of Transportation

August 2015

1. Report No. KTC-15-10/ SPR14-483-1F	2. Government Accession No.	3. Recipient's Catalog No		
4. **Title and Subtitle** A GIS-Based Expert Systems Predictive Habitat Model for Threatened and Endangered Species: Case Study Using Kentucky Arrow Darter		5. **Report Date**		
		6. **Performing Organization Code**		
7. **Author(s):** Benjamin Blanford, Michael Shouse, and John Ripy		8. **Performing Organization Report No.** KTC-15-10/ SPR14-483-1F		
9. **Performing Organization Name and Address** Kentucky Transportation Center College of Engineering University of Kentucky Lexington, Kentucky 40506-0281		10. **Work Unit No. (TRAIS)**		
		11. **Contract or Grant No.** SPR14-483		
12. **Sponsoring Agency Name and Address** Kentucky Transportation Cabinet State Office Building Frankfort, Kentucky 40622		13. **Type of Report and Period Covered** Final		
		14. **Sponsoring Agency Code**		
15. **Supplementary Notes**				

16. **Abstract**

This study presents a GIS-based predictive habitat suitability model for the Kentucky arrow darter (*Etheostoma spilotum*), a fish species of the upper Kentucky River basin that is a candidate for federal listing by the U.S. Fish and Wildlife Service. The model is based on previous work: the development of a similar predictive model for identifying the habitat of the blackside dace, a threatened minnow species of the upper Cumberland River basin in Southeastern Kentucky. The research describes a weighted, rules-based system which incorporates expert knowledge about habitat preferences for the arrow darter. For this model, five habitat factors were identified by experts as essential to modeling the habitat: stream gradient, canopy coverage, land cover, riparian zone width, and stream order. Using a GIS, the five habitat factors were parameterized and combined across the entire Kentucky River basin stream network. Experts evaluated combinations of habitat factors to determine habitat suitability. Using locational modeling statistics, the resulting model was tested against known Kentucky arrow darter occurrences. The analysis demonstrated successful identification of streams where the arrow darter was likely — and unlikely — to exist. Model results could be useful to transportation planners, particularly when determining sensitive landscape that could be impacted by transportation planning processes. This model may help planners save money on habitat mitigation when transportation initiatives take place in known unsuitable arrow darter habitats. A GIS model similar to the one developed in this study may be applicable to other endangered species.

17. **Key Words** Kentucky arrow darter, GIS model, rules-based expert system, habitat factors, transportation planning, stream network, environmentally sensitive locations		18. **Distribution Statement**		
19. **Security Classification (report)** Unclassified	20. **Security Classification (this page)** Unclassified	21. **No. of Pages** 28		19. **Security Classification (report)**

Acknowledgments

Table of Contents

List of Figures

List of Tables

1 Executive Summary

This research provides a decision support tool that will allow Kentucky Transportation Cabinet (KYTC) biologists to identify species-specific environmentally sensitive locations. As such, this report describes the development of a predictive habitat model — based on expert knowledge — for the Kentucky arrow darter (*Etheostoma spilotum*), a fish species endemic to the upper Kentucky River system. The arrow darter is a candidate for federal listing under the Endangered Species Act (ESA) by the U.S. Fish and Wildlife Service. This research builds on a pilot project that successfully modeled blackside dace habitat. Blackside dace, which is indigenous to the upper Cumberland River system of southeastern Kentucky, has been designated as a threatened fish species under the ESA. The blackside dace model was expert-driven because it relied on detailed habitat suitability information from biologists and specialists with knowledge of the species. Those experts reviewed the information before it was used to generate Geographic Information Systems (GIS) layers to describe important arrow darter habitat characteristics. The layers were then combined and applied across the study area to describe the unique attributes of every stream of interest. The model incorporated five habitat suitability factors: stream gradient, stream order, land cover, riparian width, and canopy cover. For each factor, 100 meter stream segments were ranked on a scale from one to four, with one indicating the lowest habitat suitability and four the highest. The resulting combinations of habitat layers and their relative parameterizations were analyzed by biologists in order to yield an overall habitat suitability score. A weighted rules system was then used to apply these habitat suitability evaluations across all potential combinations of the GIS layers.

To test the performance of the expert systems predictive habitat model, the results were analyzed (using locational statistics) against recorded observations of Kentucky arrow darters. Analysis demonstrated the model's success at identifying both the likely and unlikely sites for Kentucky Arrow darter occurrences. 97.33 percent of recorded observations occurred along stream segments that the model flagged as having suitable habitat. The model indicated there were no observed occurrences of Kentucky arrow darter along stream segments that the model identified as having low habitat suitability.

Overall, the results demonstrated the model's success at accurately predicting habitat suitability for the Kentucky arrow darter. The results met the research objective of developing a GIS-driven expert-systems model for predicting stream habitat suitability for the species in the upper Kentucky River system. This research will serve two purposes for stakeholders in Kentucky: 1) It will minimize the disturbance of the species during transportation planning, maintenance, and construction by helping transportation planners avoid high quality habitats, and 2) It will provide wildlife managers with a spatial representation and visualization of suitable habitat, thus improving the relationship between transportation agencies and regulatory agencies. This model may help planners save money on habitat mitigation when transportation initiatives take place in habitats

that are known to be unsuitable for arrow darters. A GIS model similar to the one developed in this study may be applicable to other endangered species.

2 Introduction

As part of the National Environmental Policy Act (NEPA) process, transportation planners must be responsive to both state and federal Departments of Fish and Wildlife when determining the impacts transportation projects potentially pose to threatened and endangered species. Since assessments hinge upon quantifying the impacts to habitats that house particular species, better understanding of the habitat combinations and locations that support threatened and endangered species will help transportation planners identify how initiatives would negatively impact their habitat, thus improving the practicality of planning decisions. The Kentucky arrow darter is a candidate for federal listing under the Endangered Species Act (ESA) by the U.S. Fish and Wildlife Service because of its population decline and sensitivity to habitat degradation.

The primary objective of the research was to develop a predictive habitat suitability model for the Kentucky arrow darter in the upper Kentucky River system. Using insights gained from the predictive modeling of pre-historic archeological sites in Kentucky (Mink et al. 2009), the research team developed a rules-based Geographic Information Systems (GIS) model that used existing and available data coverages, combined with expert knowledge about suitable habitat, to develop a comprehensive and accurate predictive habitat model that will apply to certain threatened and endangered species in Kentucky. As a prototype for this research, an expert systems-based predictive model was developed (Blandford et al. 2013). This model assessed habitat suitability of streams in the upper Cumberland River system for the blackside dace, a fish species endemic to southeastern Kentucky that is listed as threatened by the United States Fish and Wildlife Service (Biggins, 1988).

This previous modeling on the blackside dace was important to the design of this research because information on the Kentucky arrow darter is limited. State wildlife biologists used field observations and a priori knowledge to note that both species have similar ecological niches. In other words, both will thrive in landscapes with similar characteristics. Based on the success of the modeling effort, the expert-systems methods were modified and applied to the Kentucky arrow darter.

Conservation efforts have produced general guidelines that identify areas where Kentucky arrow darter populations may potentially thrive. These areas are contingent on the species' habitat preferences. However, spatially explicit modeling of arrow darter distributions among sub-watersheds within the natural habitat range was not previously conducted. To conserve the arrow darter and the riparian zones it depends on, researchers must accurately delineate areas that have suitable, high-quality habitat. Having this information allows transportation planners to account for the species' presence when planning transportation projects. Minimizing the disturbance to critical habitats when projects are implemented saves time and money for the state transportation agency.

3 Conceptual Approach

3.1 Rules-Based Expert Systems

A common approach for habitat modeling is to perform a regression analysis, where possible environmental factors and existing GIS data are used. The existing data may be only a crude representation of the imagined critical factor, and when combined with many other similarly lightly-theorized data strings in a model, can lead to a misinterpretation of relationships.

Aware of the potential problems of this approach, the project team instead adopted a "rules-based" philosophy, relying on existing expert knowledge and research to carefully describe the habitat features that best support survival of the arrow darter. To implement this approach, researchers worked closely with state biologists, who possess a large amount of data and knowledge on aquatic species and their habitat preferences. Drawing from interviews and literature reviews, the research team developed a list of critical habitat factors necessary for arrow darters to thrive. This list drove GIS modeling, as the team worked to most closely match data with the habitat conditions identified verbally or in the literature.

A rules-based system uses expert knowledge and published research to specify likely relationships between the spatially modeled factors and the occurrence of the species of interest. This approach lets models better represent the complex manner in which factors interact. However, because it is based on general knowledge about the species of interest, the model may lack some local specification or factors that are particularly relevant. Yet, the model, once well developed, may be applied to a broader range of landscapes.

Additionally, a system like this does not rely on the existence of an observed dataset to develop a spatial model. In cases where data are sparse, missing, or unreliable, it may be the only option to develop a model. With the arrow darter, a certain amount of observation data were provided from U.S. Fish and Wildlife Service to test the model's performance.

4 Kentucky Arrow Darter Habitat

The Kentucky River Basin (Figure 1) is approximately 18,000 sq. km and contains nearly 26,000 km of streams, the most significant of which is the 418 km Kentucky River. This drainage area experiences a change in relief from 990 m in the Appalachian Mountains of eastern Kentucky to 128 m at the Kentucky River's confluence with the Ohio River. Historically, the Kentucky arrow darter's habitat has been limited to the upper reaches of the basin. For this research, the upper Kentucky River system is defined as all segments upstream from the confluence of the Kentucky River and Red River. This location approximates where the river system transitions from the mountainous Cumberland Plateau upstream to the more level, rolling hills of the Bluegrass Region downstream. The Cumberland Plateau is in the southern section of the Appalachian Plateau province and is characterized by deeply incised drainages, narrow ridges, and steep slopes. The Cumberland Plateau has a humid subtropical climate and a mean annual precipitation of about 1,200 mm/yr. (Phillips & Marion, 2005). Modeling the upper half of the Kentucky River system covered approximately 13,600 km of streams, encompassing all the known existing and historic range of the Kentucky arrow darter (USFWS, 2014).
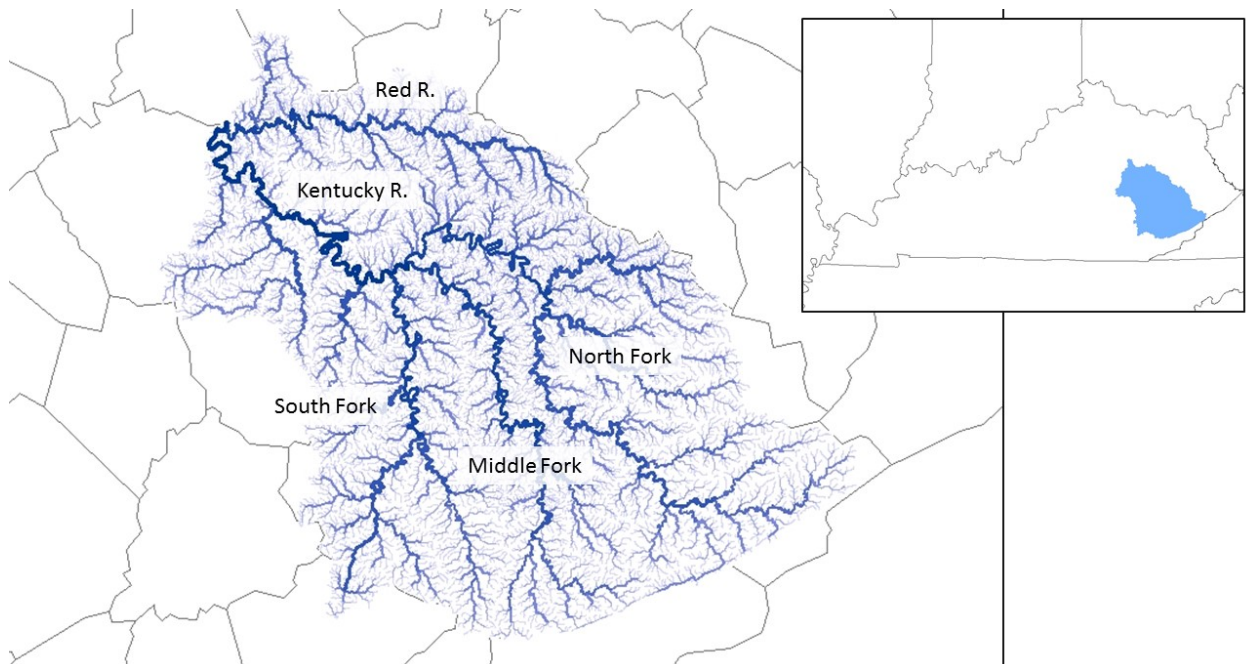
Figure 1: Study Area of the Upper Kentucky River System

The Kentucky arrow darter was first described by Gilbert (1887) and Woolman (1892) and is officially classified as a candidate for federal listing as threatened or endangered (USFWS, 2013). The fish is endemic to the upper Kentucky River system and is very similar in appearance

to the Cumberland arrow darter of the neighboring upper Cumberland River system. Each species has a sufficient number of unique traits to warrant their individual species classification (USFWS, 2013). Recent stream surveys have found the Kentucky arrow darter present in the upper Kentucky River system, though the species appears to have disappeared from significant portions of its historic range (USFWS, 2013). Recent surveys found arrow darter presence in just 34 of 68 streams it has historically been found in, representing a 50 percent decline in range. This disappearance correlates with an increased intensity of coal strip mining in southeastern Kentucky during the first decade of the 2000s (SFC, 2010). Coal mining activities have been identified as "the most imminent and substantial source of threats to the subspecies because [they] have the potential to significantly, and often permanently, alter instream water quality and cause physical habitat disturbance" (USFWS, 2013). Other factors contributing to the loss of suitable habitat include logging, increased in-stream sedimentation, oil and gas exploration, and other non-point pollutants from agriculture runoff and domestic sewage.

5 Habitat Suitability Factors Selected for Model

Several issues were taken under consideration when determining which habitat factors to include in the analysis. Of primary consideration was the availability of relevant data. A number of the suggested habitat factors obtained from published literature and through interviews with state biologists were either not contained in any available data sets or were inadequately described in available data sets. Table 1 lists the data sources used for this model. A second consideration was the relative likelihood that a habitat factor would impact the presence/absence of the arrow darter. Finally, the total number of habitat factors needed to be constrained for modeling purposes. The following habitat factors were included in the model: stream gradient, canopy coverage, land cover, water conductivity, riparian zone width, bridges/culvert density, and stream order. Iterative development of the model eventually led to the omission of water conductivity and bridges/culvert density. The following sections describe these habitat factors in more detail.

Table 1: Data sources for variables used in the expert systems model for Kentucky Arrow Darter habitat

Data Source	Description	Variable(s) Derived
Digital Elevation Model	10 m resolution digital elevation model from USGS	Stream Gradient
Geologic Structures 24k	This data set is derived from the 7.5-minute geologic quadrangle maps (scale 1:24,000)	Riparian (Alluvium) Width
National Hydrologic Dataset	Stream network for the Kentucky River Basin (1:24000)	All Variables
Gap Analysis	Land cover data from the Southeast Gap Analysis Project	Land Cover
LANDFIRE	Existing Vegetation Cover as percent cover of the live canopy layer for a 30-m grid cell.	Forest Canopy

5.1 Stream Gradient

Within the rugged terrain of the study area, Kentucky arrow darters are often found in pools between riffles (Thomas, 2008; USFWS, 2013). The model included stream gradient to account for the potential presence of such pools in the stream network. Stream gradient for each 100 m

segment was calculated as follows: the elevation difference between each segment endpoint was divided by the actual length of the segment (using the NHD and the 10 m resolution Digital Elevation Model (DEM)). To account for the potential offset between the NHD polyline dataset and the DEM, the endpoint elevation used was the minimum cell value -- found by using an 8-neighbor classification around each endpoint location.

5.2 Canopy

Canopy cover was selected as a habitat factor due to the importance of water temperature. A greater concentration of canopy cover obstructs incoming sunlight, protecting the water and regulating its temperature. Essentially, canopy cover can be used as a proxy for stream temperature (OWEB, 1999; Fiala et al., 2006). A dense canopy reduces the likelihood of water temperature fluctuations, which can affect fish species distributions (Huff et al., 2005). Canopy also lessens the likelihood of elevated water temperatures, which can result in decreased and unsuitable dissolved oxygen levels (Diana, 2004).

Forest canopy data were obtained from LANDFIRE, a shared federal program whose principal partners included the United States Forest Service, the United States Geological Survey, and The Nature Conservancy. The forest canopy cover data describe the portion of the forest floor covered by the vertical projection of tree crowns. Derived from Landsat imagery and spatially explicit biophysical gradients, these data estimate percent canopy cover at 30 m resolution.

5.3 Land Cover

Riparian corridors play an essential role in maintaining water quality and habitat within aquatic ecosystems (Naiman, Decamps, & Pollock, 1993). Effective riparian corridors filter out pollutants and other sediments, minimize flood events, and regulate water temperatures. Damage to the riparian corridor often stems from land use changes, such as increases in logging, mining, or agricultural practices. These changes can result in elevated temperatures and reductions in dissolved oxygen content, both of which place stress on aquatic species while decreasing their competitive ability (Black, 2007). Clearly, the literature stresses the importance of healthy riparian corridors and their effect on the habitability of streams.

The riparian vegetation habitat factor did not require significant data processing. The data acquired from USGS Gap Analysis Vegetation Cover contained assigned values for each cell pertaining to specific land cover types. For example, a value of 1204 indicated a high intensity of developed land; a value of 1402 indicated cultivated cropland; a value of 4126 indicated deciduous dominated forest and woodland; and so forth. The GAP dataset for the Kentucky River Basin contains land cover data at 30 m resolution. Riparian vegetation habitat was calculated by evaluating the land use type adjacent to the respective streams along the network.

5.4   Riparian Zone Width

A riparian zone is characterized by its vegetation type as well as its width, height, and bank slope (Delong & Brusven, 1991). Riparian zone width in particular is a critical factor. It can be used as a proxy for a riparian zone's effectiveness at filtering pollutants and sediments and as a means to minimizing the severity of flood events. Sufficiently wide riparian zones are likely to have a positive impact on habitat suitability. Riparian corridors are identified by both topographic and vegetation factors. Two aspects of the riparian corridors are covered by other habitat factors in this model: percent canopy and land cover. This variable concerns the topographic characteristics within the watershed area to identify the likelihood of riparian corridors adjacent to streams.

During model development, stream riparian zones were estimated to match the extent of mapped alluvium in the area, based on 24k scale geologic data. To measure the riparian width, alluvium polygons were converted to polylines along their perimeters. Polylines were then split into segments of approximately 10 meter using *ET GeoWizards* (exact + remainder). For each segment, a point file was generated. Then, perpendicular lines were created from the alluvium perimeter to the stream layer using *ET GeoWizards* perpendiculars. For each 100 meter stream segment, a spatial join was performed with each adjoining perpendicular line. After this, the average length of the lines per 100 meter stream segment was calculated. These distances were averaged for each 100 meter stream segment, yielding its average riparian width.

5.5   Stream Order (Strahler)

For modeling purposes, Strahler stream order provided a useful way to capture the relative size and volume of streams within the network (Strahler, 1957). Under this hierarchical system, each stream segment is treated as a node. Accordingly, when two first-order streams, or source streams, join together, they form a second-order stream. When two second-order streams join together, they form a third-order stream, and so on. This hierarchical system provides a useful way of understanding the water volume within a given stream network. Stream order (also known as link magnitude) is inversely related to Kentucky arrow darter presence/absence. Sampling by Thomas (2008) found arrow darter presence in streams of first to third order, with 60 percent of those in second order streams.

Stream order was calculated using a custom python script developed for this project. A simplified explanation of the procedure is as follows: 1) The stream network for the Kentucky River Basin was systematically dissected to remove first order streams from the network. These were identified by first converting each stream segment into endpoints. 2) Where stream segments overlapped (areas of confluence), multiple nodes joined at that geographic location (with one from each stream segment joining). The nodes were joined to the original stream

network and the number of joins were counted. 3) A line segment that was joined to 3 or fewer nodes was considered a first order stream. Line segments that joined with 4 nodes represented second order streams. First and second order streams were labeled as such and removed from the network. This procedure was duplicated until all streams were designated appropriately. All of this was calculated in relation to the Kentucky River, which is the highest order stream in the network.

5.6 Factors Not Included

The project team felt the list of the five representative factors sufficiently evaluated habitat suitability for the Kentucky arrow darter with a high degree of accuracy; however, a number of other variables identified either in the literature or by state biologists were not included in the model for a variety of reasons.

One factor not included in this model was a comprehensive measure of stream network connectivity. The arrow darter's range can be limited by network disconnectivity. This can be the product of disturbances, such as built or naturally occurring (beaver-made) dams, or of degraded or polluted streams, which hamper fish movement. Limited stream network connectivity is a major problem facing aquatic species (Detar, 2004; Jones, 2005). A density measure of bridges and culverts was considered as one way to address the problem of stream network connectivity. Unfortunately, there is no database that contains the location of all culverts in Kentucky. To sidestep this issue, the team used GIS to create a stream/road intersection database as a proxy for estimating the location of bridges and culverts in the study area. This proved unfeasible, due to the scalar limitations of the available data. At very small scales, the data were inaccurate, a multitude of false intersections were produced, and a number of other actual intersections were missed altogether. Due to these limitations, the stream/road intersection factor was left out of the final model.

A second factor not included in the final model was the specific conductivity of water in the upper Kentucky River systems streams. Elevated conductivity levels contribute significantly to the loss of suitable habitat for aquatic species in southeastern Kentucky. Black (2007) writes, "Abnormally high conductivity within the Upper Cumberland River drainage is most likely linked to resource extraction. This region is amidst vast coal, natural gas, and timber reserves; thus, land use disturbances are widespread and commonplace" (p. 43). In Southeastern Kentucky, resource extraction, particularly coal mining, is believed to be a major cause of habitat degradation and decreasing populations of aquatic species (Biggins, 1988; Black & Mattingly, 2007; Mattingly, 2005; Starnes & Starnes, 1978). Although strip mining has a number of detrimental effects on streams in the area, the increased specific conductivity of the water is most relevant for arrow darter distributions. Specific conductance is a measure of how effectively water conducts an electrical current. It is a function of the amount of dissolved solids contained in water, and it serves as a useful proxy for water quality (Wenner, Ruhlman, & Eggert, 2003).

Increased conductivity levels negatively and severely impact the habitat of aquatic species (Black & Mattingly, 2007).

For this project, sufficient water conductivity data were not available for the entire study area. Some data were available in the form of conductivity measurements along particular reaches in the system, but there was an insufficient number of points to extrapolate conductivity across the entire system. Without adequate conductivity data available, the team modeled mine density as a proxy measure of specific conductivity. Strip mines are known to result in elevated measures of specific conductivity in nearby streams, though this relationship is not easy to predict accurately.

After considerable deliberation between the project team and the state biologists, the mine density factor was calculated by considering the density (in terms of total area) of all mines per hydrological unit 14 (HUC14). This density measure included surface and underground mines, as both are viewed as potential sources of stream contamination. The HUC14 areas were selected because each delineated unique watersheds within the Upper Kentucky River drainage that were large enough to be relevant but small enough to adequately describe the features of individual streams within the network.

To calculate mine density, all the mine polygons acquired from the Kentucky Geospatial Data Clearinghouse were split by HUC14 using *XToolsPro* (split polygons by feature layer). The newly split polygons were then joined by HUC14. These joined polygons were then dissolved to create single mine polygons per HUC14. This process resulted in a shapefile (DensityAllMines) that contained the total area of mined out areas per HUC14.

The mine density was calculated by dividing the shape area of the newly formed mines per HUC14 by the total area of HUC14 (unit for both is square feet). Two measures of density were calculated: the first (TotalDensity) included the total area of all the mines in the HUC14. Because many of the mines overlay one another (some surface, some underground), this density measure in some cases was greater than 1, meaning that the total area of mines (surface and underground) exceeded the area of HUC14. The second density calculation (SurfaceDensity) only counted overlaying mines as a single polygon (i.e., it did not count overlapping polygons twice or more). As such, this density measure could not be greater than 1.

Despite these efforts, mine density was omitted from the final model. Comparing the mine density data to the conductivity data yielded inconsistent results, and ultimately it was concluded that model accuracy would be improved by leaving out this proxy measure.

6   Rules-Based Surface Generation

6.1   Parameterizing the Habitat Factors

To create an expert-based system, habitat factors were categorized, or parameterized, in terms of their impact on habitat suitability for the Kentucky arrow darter. The research team divided each habitat factor individually into classes with distinct break points. Each break point indicated a transition in habitat suitability. For this model, the habitat factors were divided into four classes, with a value of 1 indicating least suitable and a value of 4 indicating most suitable. This classification was created after consultations with state biologists and after studying previous habitat modeling research. For each habitat factor, a stream raster was created and classified based on the criteria in Table 2. Each habitat factor was parameterized into four categories of suitability, ranging from 1 (low) to 4 (high).

Table 2: Habitat factors used in the expert systems model for Kentucky Arrow Darter.

Habitat Factor	(low) ← Suitability → (high)			
	1	2	3	4
Gradient (100 m)	>6%	4-6%	2-4%	<2%
Canopy (percent coverage)	<50%	50-70%	70-90%	>90%
Land cover	Developed, Mined	Agriculture, Sparse	Grassland, Shrub	Forested
Riparian Width	0 meter	0-9 meter	9-18 meter	>18 meter
Stream Order (Strahler)	> 6	4-5	1	2-3

The five habitat factors were combined into a single layer containing the parameterized data for each factor (Figure 2). Over each 100 m stream segment, a field was created containing coded data for each factor. For example, a segment may be coded 43423, with each successive digit reflecting the parameterization of each habitat factor for that 100 m segment.

After each habitat factor was parameterized, they were combined into a single layer representing the stream network of the upper Kentucky River system (Figure 2). For each 100 m segment along the stream network, a unique value was generated that combined data from each of the respective habitat factors. This partitioned steam segments into 100 meter stretches, with corresponding arrow darter-related habitat information. For example, in a cell coded 43232, each successive digit reflects a habitat factor parameter for a 10 m cell. In this case, the gradient (4) is

<2 percent; the canopy (3) is 70 – 90 percent; the land cover (2) is agriculture, sparse; the riparian zone width (3) is 9–18 meters; and the stream order (2) is an order of 4-5. This information was computed for every 100 meter segment of the stream network.
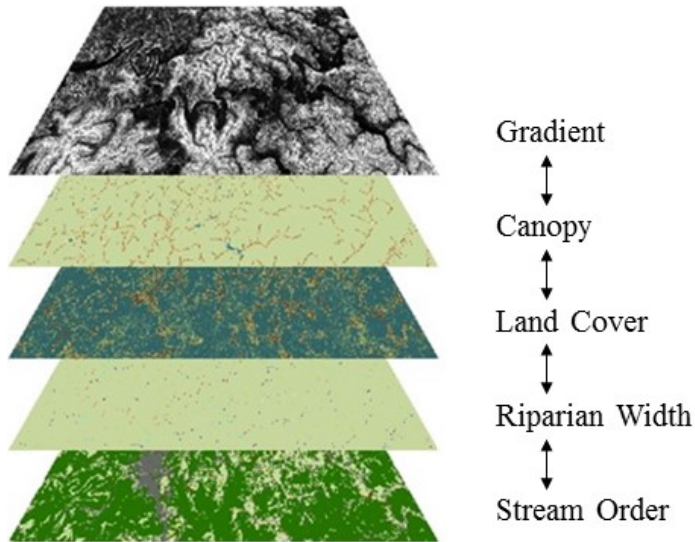

Figure 2: Five GIS layers are combined into one raster dataset representing combinations of habitat factors.

6.2 Expert evaluation of habitat scenarios

With each habitat factor identified, parameterized, and mapped in GIS, the team then evaluated the habitat factor combinations (scenarios) in terms of suitability to Kentucky arrow darter. Again, the project team solicited the expert knowledge of state biologists, where they the combined influence of each habitat factor for a given scenario. That scenario was then rated on a scale of one to four (one being not suitable, four being highly suitable).

Because the number of potential combinations of habitat factors (over a thousand) was too large to be assessed individually by experts, a weighted system was created. This system began with a set of rules identified by experts. For example, all cells having a Stream Order parameter of 1 (indicating a Strahler order of 6-7, per Table 2) received an overall score of 1 for habitat suitability. The experts deemed all large streams as highly unsuitable for arrow darter, regardless of the other habitat factors. Additionally, any cells within the Gradient parameter 1 (greater than 6 percent) and Stream Order parameter 3 (Strahler Order of 1) were also automatically assigned an overall score of 1 for habitat suitability because shallow streams along steep slopes are highly unsuitable for the arrow darter. Once these initial rules were accounted for, a weighted formula was applied across all remaining cells.

$$S = \sum Wi\, Xi$$

Where:
S = surface of total probability score
W = influence or weight factor of the ith factor
X = Criteria score for the ith parameter

So,

$S = (GR_w * GR_x + CA_w * CA_x + LC_w * LC_x + RW_w * RW_x + SO_w * SO_x)$

Where:
GR = Gradient
CA = Canopy
LC = Land Cover
RW = Riparian Zone Width
SO = Stream Order

Figure 3: Formula used to calculate habitat suitability.

In this formula, Gradient and Stream Order were given the most weight, at 30 percent each, Canopy was accorded a weight of 20 percent, and Riparian Vegetation and Riparian Width were each given 10 percent (Figure 3).

Initial calculations for each of the habitat factors were combined into a stream network defined by 100 meter segments. This resulted in a total of 132,783 stream segments containing coded habitat suitability data (Figure 4).
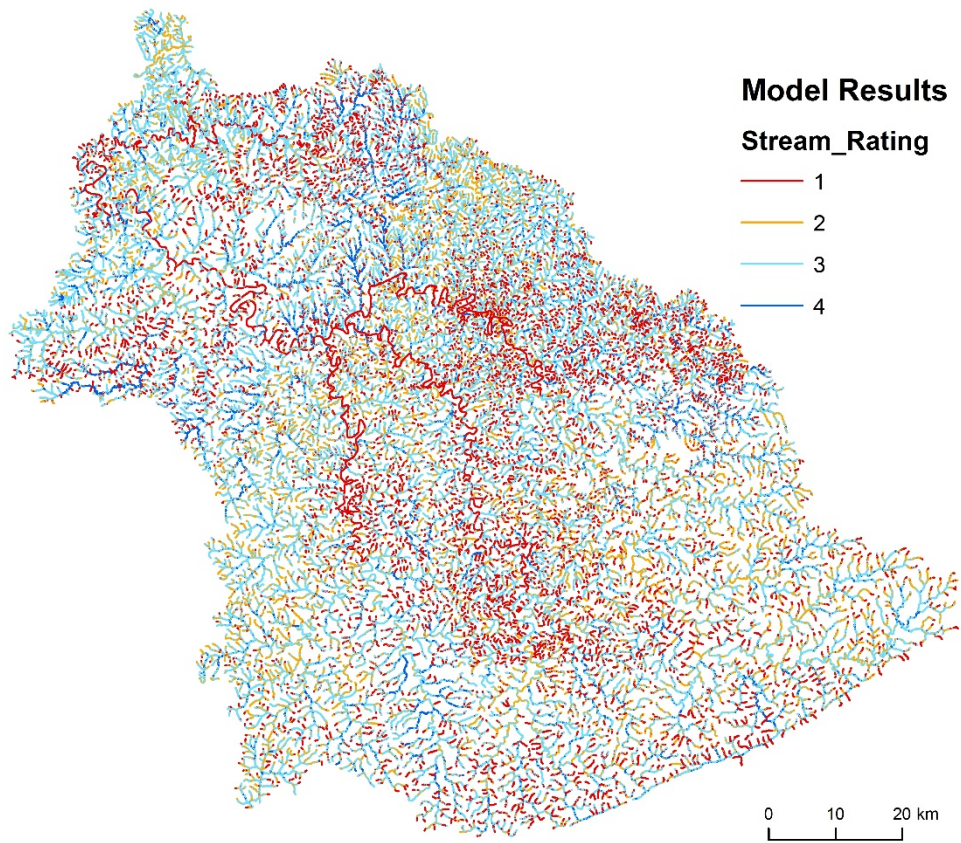
Figure 4: Map of the modeled results, with stream segments represented according to the predicted habitat suitability.

7 Results

7.1 Model Evaluation

Arrow darter presence data were obtained from the USFWS (via the Kentucky Transportation Cabinet) for testing of the model. The presence data, which had been produced through USFWS field testing, included 75 occurrences of arrow darter in the upper Kentucky River system. The data are contained in a spreadsheet, and include fields for the latitude, longitude, stream name, number of specimens found, date of testing, method of testing, and other relevant notes for each occurrence. Due to the dataset's sensitivity, it was provided to the study team under the condition that it could be used for testing the model, but the specific locations could not be shared at the time of publication. More specific information on Kentucky arrow darter presence can be found in USFWS (2013).

Using this presence data, the performance of the model was tested for accuracy using locational modeling statistics, as described by Kvamme (2006). This type of analysis is used in archaeology to assess the performance of GIS-based predictive archaeological site models (Ripy et al., 2014). This statistical technique yields several measures of model accuracy and model improvement that are useful to assess the quality of model results.

This expert systems model was developed entirely from expert knowledge — in this case, from state biologists and available literature on the subject. Presence data were not used to develop the model; they were used only when testing the model's performance. This follows Kvamme (2006), who argues, "it is uniformly agreed that a model must be tested before one can place reliance in it, and this stricture should apply to any model regardless of its means of derivation… The ultimate test, however, is against samples independent of those used to develop a model" (p. 26).

7.2 Model Performance

Table 3 is a snapshot of the calculations used to analyze model performance. The first three columns provide the base upon which further calculations were made. The first column, Habitat Rating, lists the four categories of stream habitat suitability, where 1 is the least suitable and 4 is the most suitable. The second column, M, contains the number of 100 meter stream segments assigned to each of the four Habitat Rating categories, as calculated using the weighted formula. The third column, S, contains the number of Kentucky arrow darter occurrences found on stream

segments that fell within each habitat suitability category. The results for each stream segment are based on 75 independent occurrence locations.

Table 3: Results from the expert systems model evaluation for Kentucky arrow darter.

Habitat Rating	M	S	1 p(M)	2 p(S)	3 p(M|S)	4 p(S|M)	5 p(M|S)-p(M)	6 p(S|M)/P(S|M')
1	27,761	0	20.91%	0.056%	0.00%	0.000%	-20.9%	0.00
2	24,105	2	18.15%	0.056%	2.67%	0.008%	-15.5%	0.12
3	70,817	43	53.33%	0.056%	57.33%	0.061%	4.0%	1.18
4	10,100	30	7.61%	0.056%	40.00%	0.297%	32.4%	8.10
Totals	132,783	75						

The following six columns in Table 3 provide various measures to assess the model's robustness.

1) *p(M)* is the base rate of the model indicating an occurrence. It is the percentage of stream segments in the model with each the four habitat ratings. A majority of stream segments, 53.33 percent, have a habitat rating of 3, which is slightly suitable, and 7.61 percent of segments have a rating of 4, which is very suitable. Combined, 60.94 percent of stream segments have a favorable suitability rating, whereas 39.06 have an unfavorable rating.

2) *p(S)* is the base rate that an arrow darter occurrence will be found at some location. It is the total number of occurrences (75) divided by the total number of modeled stream segments (132,783). This base rate indicates the random chance of an arrow darter occurring on any stream segment, regardless of the segment's classification.

3) *p(M/S)* is a common calculation used to assess model accuracy. It describes the probability of the model correctly predicting an occurrence. Table 3 shows that a combined 97.33 percent of arrow darter occurrences were found in stream segments modeled as having suitable habitat. Only 2.67 percent were found in stream segments modeled as unsuitable. Of those, 0 percent were found in stream segments classified as highly unsuitable.

4) *p(S/M)* indicates the probability of an occurrence within each of the four Habitat Rating categories. This column can be compared to column 2 (*p(S)*) to assess the model performance. For stream segments rated as unsuitable, *p(S/M)* should be less than *p(S)*, whereas for stream segments rated as suitable, it should be higher. If this is not the case, the model performs worse than random chance. Table 3 shows the *p(S/M)* for Categories 1 (0 percent) and 2 (0.008 percent) both as being well below the *p(S)* (0.056 percent). Category 3 (0.061 percent) is slightly higher than the base rate, and Category 4 (0.297 percent) is well above.

5) *p(M/S)-p(M)* is an indicator of the model's improvement over random chance at predicting arrow darter occurrences for each habitat suitability category. Table 3 shows negative percentages for the unsuitable habitat stream segments: -20.9 percent for Category 1 and -15.5 percent for Category 2. This indicates the model performs well at identifying stream segments that have a low probability of arrow darter occurrences. Category 3 yields a percentage of 4.0, which indicates a slight improvement over chance. Whereas Category 4 yields a percentage of 32.4, which indicates a significant improvement over chance.

6) *p(S/M)/p(S/M')* is a model improvement ratio. It indicates how many times more likely an arrow darter occurrence is where the model indicates one (*M*) — compared to where the model does not (*M'*) — for each habitat suitability rating. Table 3 shows how the model performed at correctly predicting where arrow darter occurrences are unlikely, as both Category 1 and Category 2 are less than one times as likely to have an occurrence. Category 3 is 1.18 times more likely, and Category 4 performs the best at 8.1 times more likely.

8 Conclusions and Recommendations

Information on the spatial distribution of suitable habitat for Threatened and Endangered Species is critical for conservation efforts. In this study, a GIS-derived expert system model was developed for the Kentucky arrow darter, a candidate species for federal listing. Results from this study showed that the model, although derived strictly from expert input and published research, performed very well against an independent dataset that recorded where the arrow darter occurred within the upper Kentucky River Basin. 40 percent of fish were found along stream segments that had the most suitable habitat; this contained just 7.61 percent of stream segments. When the model was divided into two categories, suitable vs. non-suitable, approximately 60 percent of stream segments in the Upper Kentucky River basin contained habitat suitable for the Kentucky arrow darter. When tested against independent presence locations, over 97 percent of occurrences occurred in habitat the model designated as suitable. It could be argued that overfitting contributed to the high model accuracy and to the high number of stream segments categorized as suitable. To counter that argument, overfitting is acceptable in this case for two primary reasons: 1) The arrow darter is a candidate for federal protection because of its population decline and sensitivity to habitat degradation. In this and similar cases, being less restrictive lets researchers examine larger areas of habitat, and therefore identify more opportunities for range expansion. 2) State transportation agencies are trying to identify areas of unsuitable habitat as an avoidance measure for sampling and mitigation in future projects. Being more restrictive about what qualifies as unsuitable habitat lends some credence to adjacent areas being selected as potential project locations. In the Kentucky River Basin for example, state transportation planners can now be confident that projects occurring at or near 20.91 percent of streams would minimize disturbances to the Kentucky arrow darter.

These results are important to environmental professionals and transportation planners. Because considerable work for this project went into generating GIS layers relevant to an aquatic habitat species, this model could be applied to similarly threatened and endangered fish or other aquatic species in Kentucky. Future research could be directed toward aquatic and terrestrial habitats alike. This type of model is most successful when sufficient expert knowledge about the species and its preferred habitat exists before modeling begins. In situations of high uncertainty regarding species habitat preference, this type of model may require many iterations to successfully parameterize it. But this type of model is particularly useful in situations where the exact range of a species has not yet been determined. Incorporating expert knowledge into predictive habitat modeling, rather than including only the known presence/absence data, can aid in the successful, precise identification of suitable habitats. Of course, each iteration of the model would be adjusted and applied uniquely for each species according to the specifications of expert biologists, but a considerable knowledge base has been put into place. This project demonstrates

a proof for predicting habitats of threatened and endangered species via expert systems modeling. Applying this methodology to similar species may prove to be a worthwhile endeavor.

9 Works Cited

Biggins, Richard G. (1988). *Recovery Plan for Blackside Dace (Phoxinus Cumberlandensis)*. Atlanta, GA: U.S. Fish and Wildlife Service.

Black, Tyler R. (2007). *Population Densities and Performance of Predictive Habitat Models for the Threatened Blackside Dace.* (Master's Thesis). Tennessee Technological University, Cookeville, TN.

Black, Tyler R., & Mattingly, Hayden T. (2007). *Conservation Status of the Threatened Blackside Dace in Selected Cumberland River Streams.* (Final Project Report). Tennessee Technological University, Cookeville, TN.

Blandford, Benjamin L., Ripy, J., & Grossardt, T. (2013). *GIS-Based Expert Systems Model for Predicting Habitat Suitability of Blackside Dace in Southeastern Kentucky*. (TRB 92nd Annual Meeting Compendium of Papers). Washington DC: Transportation Review Board.

Delong, Michael D., & Brusven, Merlyn A. (1991). Classification and Spatial Mapping of Riparian Habitat with Applications Toward Management of Streams Impacted by Nonpoint Source Pollution. *Environmental Management,* 15(4), 565–571.

Detar, Jason E. (2004). *Population Densities and Movement Patterns of the Threatened Blackside Dace*. (Master's Thesis). Tennessee Technological University, Cookeville, TN.

Diana, J. S. (2004). *Biology and Ecology of Fishes*. (2nd ed.). Traverse City, MI: Cooper.

Fiala, A. C., Garman, S. L., & Gray, A. N. (2006). Comparison of Five Canopy Cover Estimation Techniques in the Western Oregon Cascades. *Forest Ecology and Management,* 232 (1-3), 188–97.

Gilbert, C. H. (1887). Descriptions of New and Little Known Etheostomoids. *Proceedings of the U.S. National Museum,* 10, 47–64.

Huff, D. D., Hubler, S. L., & Borisenko, A. N. (2005). Using Field Data to Estimate the Realized Thermal Niche of Aquatic Vertebrates. *North American Journal of Fisheries Management,* 25(1), 346–60.

Jones, Brena K. (2005). *Predictive Habitat Models for Conservation of the Threatened Blackside Dace.* (Master's Thesis). Tennessee Technological University, Cookeville, TN.

Kvamme, Kenneth L. (2006). There and Back Again: Revisiting Archaelogical Locational Modeling. In Mark Mehrer & Konnie Wescott (Eds.). *GIS and Archaelogical Site Location Modelling*, (3–40). Boca Raton, FL; London: CRC.

Mattingly, Hayden T. (2005). *Factors Affecting the Distribution and Recovery of the Threatened Blackside Dace in Kentucky and Tennessee.* (Final Project Report). Cookeville, TN: U.S. Fish and Wildlife Service.

Mink, P., Grossardt, T., Ripy, J., & Bailey, K. (2009). Proceedings of the ESRI National User Conference: *Predictive Archaeological Modeling Using GIS-Based Fuzzy Set Estimation*. San Diego, CA.

Naiman, Robert J., Decamps, H., & Pollock, Michael. (1993). The Role of Riparian Corridors in Maintaining Regional Biodiversity. *Ecological Applications,* 3(2), 209–212.

OWEB. (1999). *Stream Shade and Canopy Cover Monitoring Methods, in Water Quality Monitoring: Technical Guide Book.* 775 Summer St. NE., Suite 360, Salem, OR, 97301-1290: Oregon's Watershed Enhancement Board.

Phillips, J. D. & Marion, D. A. (2005). Biomechanical effects, lithological variations, and local pedodiversity in some forest soils of Arkansas. *Geoderma*, 124(1), 73-89.

Ripy, J., Grossardt, T., Shouse, M., Mink, P., Bailey, K., & Shields, C. (2014). Expert Systems Archeological Predictive Model. *Transportation Research Record: Journal of the Transportation Research Board,* 2403(1), 37-44.

SFC. (2010). Southeastern Fishes Council Proceedings: *Southeastern Fishes Council State Reports: Kentucky.* Athens, GA.

Starnes, Wayne C., & Starnes, Lynn B. (1978). A New Cyprinid of the Genus Phoxinus Endemic to the Upper Cumberland River Drainage. *Copeia,* 1978 (3), 508–516.

Thomas, M. R. (2008). *Distribution and Status of the Kentucky Arrow Darter (Etheostoma Sagitta Spilotum).* (State and Tribal Wildlife Grant T-9-RSI-1). Kentucky Department of Fish and Wildlife Resources.

USFWS. (2013). *U.S. Fish and Wildlife Service Species Assessment and Listing Priority Assignment Form: Kentucky Arrow Darter*. United States Fish and Wildlife Service.

USFWS. (2014). Endangered and Threatened Wildlife and Plants; Review of Native Species That Are Candidates for Listing as Endangered or Threatened. (Annual Notice of Findings on Resubmitted Petitions; Annual Description of Progress on Listing Actions; Proposed Rule, 50 CFR Part 17). *Federal Register,* 79(234). United States Fish and Wildlife Service.

Wenner, David B., Ruhlman, M., & Eggert, S. (2003). *The Importance of Specific Conductivity for Assessing Environmentally Impacted Streams.* Athens, GA.

Woolman, A. J. (1890). Report of an Examination of the Rivers of Kentucky, with Lists of Fishes Obtained. *Bulletin of the U.S. Fish Commission* (10), 249–88.