Section 6 (Texas Traditional) Report Review

Form emailed to FWS S6 coordinator (mm/dd/yyyy): 11/15/2012

TPWD signature date on report: 10/31/2012

Project Title: Using ecological niche modeling to predict the probability of occurrence of rare fish and mussel species in East Texas

Final or Interim Report? Interim

Grant #: TX E-138-R-1

Reviewer Station: Clearlake ESFO

Lead station concurs with the following comments: Yes

Interim Report (check one):

- [x] Acceptable (no comments)
- [ ] Needs revision prior to final report (see comments below)
- [ ] Incomplete (see comments below)

Final Report (check one):

- [ ] Acceptable (no comments)
- [ ] Needs revision (see comments below)
- [ ] Incomplete (see comments below)

Comments: No comments or concerns with document.

INTERIM REPORT

As Required by

THE ENDANGERED SPECIES PROGRAM

TEXAS

Grant No. TX E-138-R

Endangered and Threatened Species Conservation

Using ecological niche modeling to predict the probability of occurrence of rare fish and mussel species in East Texas

Prepared by:

Dr. Lance Williams

TEXAS PARKS & WILDLIFE

Carter Smith
Executive Director

Clayton Wolf
Director, Wildlife

31 October 2012

INTERIM REPORT

STATE: ____Texas_______________ GRANT NUMBER: ___TX E-138-R-1__

GRANT TITLE: Using ecological niche modeling to predict the probability of occurrence of rare fish and mussel species in East Texas

REPORTING PERIOD: ____1 Sep 11 to 30 Sep 12_

OBJECTIVE(S). To use ecological niche modeling of landscape characteristics (e.g., geomorphic, geological, topographic) and fish and mussel distributions to predict the probability of occurrence for rare species in East Texas rivers.

Segment Objectives:

Task 1. Oct 2011 – Aug 2012 – *Compile GIS data layers and data necessary for modeling.*
Task 2. Sept 2012 – May 2013 – *Ecological niche modeling:* a tool that can be used to predict the distribution of our target species in other river systems in East Texas, or in similar types of streams in the southeastern United States where those species occur.

Significant Deviations:

None.

Summary Of Progress:

Please see Attachment A.

Location: Delta, Fannin, Lamar, Red River, Bowie, Cass, Morris, Titus, Camp, Upshur, Franklin, Hopkins, Delta, Rains, Wood, Van Zandt, Smith, Henderson, Cherokee, Anderson, Houston, Trinity, Polk, Tyler, Angelina, Nacogdoches, Panola, Harrison, and Gregg Counties, Texas.

Cost: ___Costs were not available at time of this report, they will be available upon completion of the Final Report and conclusion of the project.__

Prepared by: _Craig Farquhar_____________ Date: _31 October 2012

Approved by: ______________________________ Date:_____31 October 2012____
C. Craig Farquhar

ATTACHMENT A

Interim Report – Section 6

Title:

E-138-R - Using ecological niche modeling to predict the probability of occurrence of rare fish and mussel species in East Texas

Principal Investigators:

Lance R. Williams, Marsha G. Williams, and Neil B. Ford, Department of Biology, University of Texas at Tyler, Tyler, TX 75799. Phone: 903-565-5878; Fax: 903-566-7189; Email: lwilliams@uttyler.edu

Reporting Period:

1 October 2011 – 30 September 2012

Summary of Progress:

Task 1. Oct 2011 – Aug 2012 – *Compile GIS data layers and data necessary for modeling.* Layers required will include, but not be limited to, soils, geology, landuse/landcover, and DEM. We will create a GIS layer based on landscape-level geomorphic features (e.g., floodplain width, sinuosity). We will use the digital elevation model to calculate the topographic index (TOPMODEL) to predict areas of groundwater upwelling. We will use our georeferenced fish and mussel database (Ford et al. 2010) for predictive modeling using MAXENT. Additional, georeferenced historical data will also be incorporated into our database (e.g., Ford and Nicholson 2006).

Twelve GIS layers were incorporated in the model, specifically: infiltration excess and overland flow (TopModel), groundwater recharge, soil type, geology, vegetation type, sunlight, aquifers, spring source density, density of roads, density of dams, total nitrogen load, and landform. The sunlight layer provided information regarding solar radiation and mean annual cloud cover. It was obtained from CLIMond based on data from Kriticos et al. (2012). The roads layer was a series of one-dimensional lines. Therefore, we interpolated the density of roads across the landscape using the kernel density function in ArcMap version 9.3 (ESRI Inc. 2008). Hydrography features and the major aquifers were obtained from USGS, based on data from Hayes (2006). Landcover vegetation types were obtained from USGS, based on data from McMahan et al. (1984). Major soil types were obtained from USGS, based on data from various members of the Soil survey staff (2006). Landform data such as slope, local relief, profile type, percentage of area occupied by sand, ice and standing water, and patterns of major peaks were obtained from USGS, based on data from Hammond (1964). The reservoir data was obtained from the Texas Water Development Board, based on data from Ruddy and Hitt (1990) and springs data was obtained from USGS, based on data from Heitmuller and Williams (2006). The reservoir data and springs data were point coverages, and therefore we used the kernel density function to interpolate their densities across the landscape. The groundwater recharge layer, which provided the mean annual ground water recharge estimates (Wolock, 2003a), was obtained from USGS, based on data from Wolock (2003b). Information regarding infiltration excess, overland flow estimates , and total nitrogen loads were obtained from USGS, based on data from Smith et al. (1997).

Most environmental data were obtained as raster files; vector data were converted to raster format in ArcMap. Environmental layers were clipped in order to constrain them to lotic habitats. We did this by adding a 1000m buffer around water features (ponds, streams, rivers, canals, and dams), obtained from an environmental layer called "NHDFlowline" obtained from the US Geological Survey (USEPA and USGS, 2005), and clipping the environmental layers to match the lotic buffer. The environmental data were projected in the Universal Transverse Mercator coordinate system North American Datum 1983 (XY coordinate system "GCS_North_America_1983"), and the cell sizes were equalized to a resolution of 0.014 arc-seconds (approximately 4 m^2).

We are currently creating environmental layers for distance to dams and distance to dams of varying size. We are also revisiting our original resolutions and projections to improve the accuracy of all of our environmental layers for the final product. We are also continuing our search for applicable environmental layers to use in our modeling to improve the models further.

Task 2. Sept 2012 – May 2013 – *Ecological niche modeling.* We will use the GIS layers compiled in Task 1 and all validated historical and current biology data to model the probability of presence or absence of each species in each spatial cell in the rivers. Ecological niche modeling will be conducted using the MAXENT software package. MAXENT produces a predictive model, which can be displayed geospatially, that represents the relative probability of a species occurring in a particular cell, given a set of environmental conditions associated with that cell and known species distributions (Pineda and Lobo 2009, Urbina-Cardona and Flores-Villela 2010). Ecological niche modeling has been used to model spread of invasive species (Thuiller et al. 2005), impacts of climate change (Thomas et al. 2004), and spatial patterns of diversity (Graham et al. 2006). Recent evaluations have shown to MAXENT to be a robust method for modeling geographic distributions of species, especially with conservation implications (Phillips and Dudik 2008).

We restricted our analysis to locations falling within East Texas, with the Trinity River as the western boundary and including the Cypress, Sulphur, Sabine, Neches, and Angelina rivers and their associated watersheds. Habitat suitability models were built separately for each species. Species with less than five occurrence points were considered too poorly sampled to be modeled accurately (Pearson et al., 2007). In Maxent, we used AUC and "gain" to determine aspects of model fit. The area under the operator receiving curve, AUC (Fielding and Bell, 1997), measures the probability that a randomly chosen presence site will be ranked above a randomly chosen pseudoabsence site (Phillips and Dudik, 2008). Models with AUC > 0.75 are treated as good fits (Elith, 2002). Gain is the mean log probability of the occurrence samples, minus a constant that makes the uniform distribution have zero gain. Since gain is not bounded by zero or one, it is useful only for comparative purposes among nested models. For each species, we compared the gain of the full model (all variables included) to models based solely on one environmental variable. To avoid over-fitting, we dropped variables from the final model if the single-variable model had less than 5% of the test gain of the full model (i.e., the variable accounted for less than 5% of the fit of the full model). AUC and gain values were calculated first using "training data" and then using "test data." "Training data" are the known occurrence points that are used to generate the models. "Test data" are comprised of known occurrence points there are held back until after the models were developed, and they are not used to generate the models. The test data is plugged into the models only after they are created, and therefore can be viewed as quasi-independent verification of the models. We used a cross-validation approach (Pearson et al., 2007) to subdivide our datasets into the training data points and test data points.

For niche models that had a good fit to the data (AUC > 0.75), we further tested whether they were significantly different from one another. We did this using ENMTools, a software package that allows one to test whether the habitat suitability scores generated by niche modeling for two species exhibit statistically significant ecological differences (Warren et al., 2010). Specifically, for every possible pair of species' niche models, we used the "niche identity test" module. It asks whether niche models generated from two or more species are more different than expected if they were drawn from the same underlying distribution. It does this by pooling empirical occurrence points and randomizing (permuting) their identities to produce two new samples with the same numbers of observations as the empirical data (Warren et al., 2010). We repeated this procedure 100 times, generating niche similarity values based on the permuted data from each run. This gave us our distribution under the null hypothesis of no difference in the niches of the two species, which we then compared to the observed level of niche differentiation.

ENMtools output provides three different statistics to measure niche similarity: Schoener's D (Schoener, 1968), the *I* statistic (Warren et al., 2008), and relative rank, RR (Warren and Seifert, 2011). All three metrics range from zero to one; zero indicating that species have completely different niche models and one meaning that the pair of species have identical niche models. The *I* and D statistic are calculated by taking the difference between the species suitability score at each grid cell, after the suitabilities have been standardized so that they sum to one over the geographic space being measured. The relative rank is an estimate of the probability that the relative ranking of any two patches of habitat is the same for the two models. Although the statistics emphasize different aspects of the data, we chose to use the *I* statistic because it has been shown that RR, *I*, and D metrics are highly correlated (Warren et al. 2008). We considered two species to have significantly different niches if the observed *I* statistic was below the five percent quantile from the null distribution (corresponding to a 5% chance that two niche models would be that different if they were estimated from two species that actually had the same niche).

RESULTS

The training AUC values for mussels ranged from 0.9898-0.9976 and test AUC values ranged from 0.7788-0.9097, indicating that all of the models are good fits (Table 1). The relative contributions of the different environmental variables to the niche models (as measured by test gain when the model only included that particular environmental variable) varied depending on the particular species. For instance, the only variable that contributed substantially to the Southern hickorynut model was soil type

(Table 1). The variables aquifers, landform, total nitrogen load, soil type, TopModel, mean annual cloud cover, groundwater recharge, and vegetation type contributed more than five percent in most models (Lousiana pigtoe, Texas pigtoe, triangle pigtoe, and the sandbank pocketbook). Groundwater recharge and mean annual cloud cover contributed less than five percent to the models for the Texas pigtoe and the Texas heelsplitter and those variables were removed. Density of springs contributed to the model and was retained for both species. Reservoir density was incorporated into the models for triangle pigtoe and Texas heelsplitter and road density was incorporated into the model for Louisiana pigtoe. Soil type contributed the most information to niche models of all mussel species. Vegetation also had high test gain, when only this variable was incorporated, for Louisiana pigtoe, Texas pigtoe, triangle pigtoe, and Texas heelsplitter (Table 1).

Table 1. Summary information for the individual mussel species' niche models. The training AUC, test AUC, and test gains for the models are presented, as well as test gains for models fit with only the specified individual variables.

SPECIES	TRAINING AUC	TEST AUC	TEST GAIN (FULL MODEL)	TEST GAIN (AQUIFERS ONLY)	TEST GAIN (RESERVOIR DENSITY ONLY)
P. riddellii	0.9927	0.899	0.9787	0.2447	-
F. askewi	0.9898	0.8168	0.9766	0.2639	-
F. lananensis	0.9941	0.9097	1.5025	0.2013	0.1871
L. satura	0.9927	0.8703	1.2138	0.314	-
O. jacksoniana	0.9969	0.7788	1.229	-	-
P. amphichaenus	0.9976	0.8141	1.4583	0.1006	0.087

SPECIES	TEST GAIN (ROAD DENSITY ONLY)	TEST GAIN (SPRINGS DENSITY ONLY)	TEST GAIN (LAND FORM ONLY)	TEST GAIN (NITROGEN ONLY)	TEST GAIN (GROUND WATER RECHARGE ONLY)
P. riddellii	0.3854	-	0.347	0.3529	0.3254
F. askewi	-	0.0601	0.1454	0.2772	0.301
F. lananensis	-	-	0.4934	0.2155	0.1298
L. satura	-	-	0.1954	0.233	0.2553
O. jacksoniana	-	-	-	-	-
P. amphichaenus	-	0.0926	0.0988	0.1035	-

SPECIES	TEST GAIN (SOILS ONLY)	TEST GAIN (CLOUD COVER ONLY)	TEST GAIN (TOPMODEL ONLY)	TEST GAIN (VEGETATION ONLY)
P. riddellii	1.1133	0.2971	0.4258	0.894
F. askewi	0.955	0.2646	0.3026	0.7522
F. lananensis	0.9054	0.2902	0.145	1.3643
L. satura	1.1697	0.2743	0.2043	0.3838
O. jacksoniana	1.229	-	-	-
P. amphichaenus	0.8804	-	0.1481	1.0568

The final ***Outcome*** for this project will be georeferenced maps of probability of occurrence for each of the rare fish and mussel species collected in the four rivers of East Texas we are currently studying. The final ***Output*** for this project would be a modeling tool that can be used to predict the distribution of our target species in other river systems in East Texas, or in similar types of streams in the southeastern United States where those species occur. While the model will require regional recalibration, we believe it will be a very useful landuse planning tool to predict whether or not rare species would be impacted by a project (e.g., road, reservoir, etc.)

In summary, we have completed the initial modeling phase of the project and are now in the process of fine tuning the environmental layers (revisiting resolution and projection for each layer) to create ecologically accurate distribution maps for the final report.

REFERENCES

Elith, J. 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Quantitative methods for conservation biology (eds. Ferson S, Burgman M). Springer, New York.

Fielding, A.H. and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. –Environ. Conserv. 24: 38-49.

Hammond, E. 1964. Classes of land-surface form in the United States. U.S. Geological Survey. Accessed on November 17, 2011 at URL http://water.usgs.gov?GIS/metadata/usgswrd/XML/na70_landfrm.xml.

Hayes, M. 2006. Major Aquifers. Texas Water Development Board. Accessed on January 12, 2011 at URL www.twdb.state.tx.us/mapping/gisdata.asp.

Heitmuller, F.T. and Williams, I.P. 2006. Compilation of historical water-quality data for selected springs in Texas, by ecoregion: U.S. Geological Survey Data Series 230, 32.

Kriticos, D.J. et al. 2012. CliMond: global high resolution historical and future scenario climate surfaces for bioclimatic modeling. -Methods Ecol. Evol. 3: 53-64. DOI: 10.1111/j.2041-210X.00134.x

McMahan, C.A. et al. 1984. The vegetation types of Texas—including cropland. Texas Parks and Wildlife Department. Accessed on October 3, 2011 at URL http://www.tpwd.state.tx.us/landwater/land/maps/gis/tescp /index.phtml.

Pearson, R.G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. –J. Biogeogr. 34: 102-117.

Phillips, S.J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. -Ecography 31: 161-175.

Ruddy, B.C. and Hitt, K.J. 1990. Summary of selected characteristics of large reservoirs. U.S. Geological Survey. Accessed November 15, 2011 at URL http://water.usgs.gove/lookup/getspatial?reservoir

Schoener, T.W. 1968. Anolis lizards of Bimini: resource portioning in a complex fauna. -Ecology 49: 704-726.

Smith, R.A. et al. 1997. Regional interpretation of water-quality monitoring data. -Water Resour. Res.33: 2781-2798.

Soil Survey Staff. 2006. Digital general soil map of U.S. Natural Resources Conservation Service, United States Department of Agriculture. Accessed on November 20, 2011 at URL http://SoilDataMart.nrcs.usda.gov/>

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS). 2005. National Hydrography Dataset Plus. Horizon Systems Corporation. Accessed on July 18, 2011 at URL www.horizon-systems.com/NHDPlus.

Warren, D.L. et al. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. -Evolution 62: 2868-2883.

Warren, D.L. et al. 2010. ENMtools: a toolbox for comparative studies of environmental niche models. -Ecography 33: 607-611.

Wolock, D.M. 2003a. Estimated mean annual natural ground-water recharge in the conterminous United States. U.S. Geological Survey. Accessed November 17, 2011 at URL http://water.usgs.gov/GIS/metadata /usgswrd/XML/rech48grd.xml.

Wolock, D.M. 2003b. Infiltration-excess overland flow estimated by TOPMODEL for the conterminous United States. U.S. Geological Survey. Accessed on July 18, 2012 at URL http://water.usgs.gov./lookup/getspatial?ieof48.