SPATIAL GENETIC AND DISTRIBUTION MODELING FOR THE CONSERVATION OF HELLBENDER SALAMANDERS

by

Emily Boersma McCallen

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Forestry and Natural Resources

West Lafayette, Indiana

August 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Rod Williams, Co-chair

Department of Forestry and Natural Resources

Dr. Songlin Fei, Co-chair

Department of Forestry and Natural Resources

Dr. Jeffrey Holland

Department of Entomology

Dr. Patrick Zollner

Department of Forestry and Natural Resources

Approved by:

Dr. Robert Wagner

Head of the Graduate Program

To Aaron, the one who started this all

ACKNOWLEDGMENTS

I give special thanks to my advisors Dr. Rod Williams and Dr. Songlin Fei for their infinite patience and encouragement during this daunting process. I thank the members of my committee Dr. Jeffrey Holland and Dr. Patrick Zollner for their guidance and wisdom in the development of these projects. I thank my master's advisor Dr. Karen Gaines and my undergraduate advisor Dr. Paul Switzer for their mentorship when I was just figuring out how science works. I thank all of my amazing past and current lab mates for offering constructive feedback without judgement, but especially Dr. Obed Hernandez-Gomez, Dr. Erin Kenison, and Dr. Gabriela Nunez-Mir.

I thank all of my field and GIS technicians including Elizabeth Gilchrist, Kodiak Hengstbeck, Luke Hoehn, Kimberly Ordonez, Weston Schrank, Brian Tornabene, and Brandon Zinman for all of their heavy lifting (literally and figuratively). I thank O'Bannon Woods State Park for their hospitality, but especially Stanely Baelz and Bob Sawtelle. I thank all of our project partners for truly making this a range-wide effort including Sheena Feist, Dan Feller, Robin Foster, Michael Freake, James Godwin, John Groves, Paul Hime, Greg Lipps, Amy McMillan, Edward Thompson, Shem Unger, David Weisrock, and Lori Williams. This project could not have been possible with support provided by the Indiana Department of Natural Resources (grants T7R15 and T7R17).

I thank all of my family, but especially my mom Glory Boersma and my husband Aaron McCallen for their constant, unwavering support. I thank my son, and sunshine, Warren Wallace for giving me a reason to smile every day. Finally, I thank all of the amazing women who have cared for my baby as if he was their own so I could spend my days thinking about salamanders.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: McCallen, Emily Boersma. PhD
Institution: Purdue University
Degree Received: August 2018
Title: Spatial Genetic and Distribution Modeling for the Conservation of Hellbender Salamanders.
Major Professor: Rod Williams

Ecological data is inherently spatial; however, it is still the norm to model ecological data as spatially invariant. Failure to account for spatial structure in response variables and modeled relationships can result in inflated coefficient values, shifts in the relative importance and sign of predictors, cross-scale contradictions in relationships, and reduced predictive power due to the averaging of modeled relationships. When ecological models are used to support conservation decision-making, model error can be costly leading to both misallocation of limited resources and distrust of science-based management.

My dissertation focuses on developing methods to account for spatial structure in two models commonly used to inform conservation decisions. Both chapters focus on the imperiled hellbender salamander (*Cryptobranchus alleganiensis*) and were designed to provide guidance on the conservation and management of a species that is facing precipitous declines throughout much of its range. In chapter one, I modeled the relationship between the hellbender genome and climate and stream variables across the range of the species. I extended multiple matrix regression into a mixed modeling framework to account for strong spatial population structuring. The approach improved model fits, shrunk coefficient estimates, and increased the concordance of model results with an independent analysis of locus-specific environmental associations. The results of the model were used to forecast genomic vulnerability across the range of the species

and the resulting map suggested a potential genetic mismatch between current and future conditions in portions of the range that accommodate stable populations.

In chapter two, I developed a species distribution model to help target sampling and translocation locations for eastern hellbenders (*Cryptobranchus alleganiensis alleganiensis*). It extended presence-only modeling into a mixed modeling framework to help account for autocorrelation and nonstationarity in the intensity of hellbender occurrences and unexplained environmental heterogeneity across physiographic provinces. The spatially explicit approach improves overall model discrimination and dramatically improves model performance in regions most in need of conservation guidance. Taken together, the chapters provide flexible methods to improve the performance of common ecological models and tangible products to support hellbender conservation.

CHAPTER 1. USING A MIXED EFFECTS MODEL TO ACCOUNT FOR GENETIC STRUCTURE IMPROVES PERFORMANCE IN AN ISOLATION BY ENVIRONMENT MODEL OF A THREATENED SALAMANDER (*CRYPTOBRANCHUS ALLEGENIANSIS*)

1.1 Abstract

A common methodological approach to understand patterns of genetic variation across environmental gradients is to use multiple matrix regression (MMR) to disentangle the effects of isolation by distance (IBD) and isolation by environment (IBE) in genomic data. However, when species display hierarchical population structuring, MMR performs poorly and model results may be misleading for many taxonomic groups. I explored patterns of IBD and IBE in the genetically structured hellbender salamander (*Cryptobranchus allegeniansis*) to elucidate conservation implications across the geographic range of this threatened species. I used a mixed modeling approach to account for population structure that improved model fit (with a 55% reduction in BIC in the structured model compared to the null model) and reduced IBE coefficient inflation (with a 64% reduction in IBE coefficient intensity in the structured model compared to the null model). Accounting for genetic structure in the data also greatly improved concordance between the IBE model results and locus-specific environmental association tests. I extended the results of the analysis to forecast genomic vulnerability across species range. My results suggest a spatial mismatch between the hellbender genome and future climate conditions in portions of the species range that currently contain the only remaining stable populations. While the conservation implications of our study are specific to hellbenders, the mixed modeling methodology represents a natural extension of MMR that can be used to improve IBD and IBE model results in any taxonomic group that displays moderate to high levels of population structure.

1.2 Introduction

Exploring associations between genotypes and environmental variation is a natural extension of landscape ecology since it can shed light on underlying mechanisms of genetic change in taxonomic groups. Environmental associations may be explored in genome-wide differentiation (Wang & Bradburd, 2014) or in individual loci (Frichot & François, 2015; Günther & Coop, 2013), and the methods can be used concordantly to derive multiple lines of evidence for genome-environment relationships (Bay et al., 2018). The partial-mantel test (Smouse, Long, & Sokal, 1986), and its natural extension multiple matrix regression (MMR; Legendre, Lapointe, & Casgrain, 1994; Wang, 2013), are the most commonly used methods for detecting environmental signals in genome-wide studies (Storfer, Murphy, Spear, Holderegger, & Waits, 2010). In both methods, a pairwise genetic distance matrix is statistically related to both geographic and environmental distance matrices. The geographic distance matrix is included to account for isolation by distance (IBD; Wright, 1943) a common pattern that occurs when genomic changes due to drift accumulate faster than can be ameliorated by gene flow between populations (Rousset, 1997). The relationship between genomic variation and environmental variables after accounting for IBD is known as isolation by environment (IBE; Wang & Bradburd, 2014), and the relative strength of each process can be inferred from the coefficient values of MMR. However, these processes are inherently complicated and it can be difficult to disentangle the influence of neutral and environmental factors on the genome.

Failure to fully account for neutral genetic processes prior to exploring environmental relationships can lead to spurious associations and misleading results (Meirmans, 2012). Matrix-based approaches assume a linear IBD relationship, which is unlikely to hold true at broad scales, but this limitation may be addressed with simple transformations of distance variables (Legendre & Fortin, 2010; Wagner & Fortin, 2015). Furthermore, if migration movements are

directional, a more biologically relevant predictor such as stream distance (Mullen, Woods, Schwartz, Sepulveda, & Lowe, 2010), least-cost transect (Van Strien, Keller, & Holderegger, 2012), or resistance surface (McRae, 2006) may better explain neutral genetic patterns. However, a more substantial drawback of matrix-based models is that they are non-spatial, and do not adequately account for neutral genetic processes in species that display hierarchically structured populations. Since population structuring is common, particularly in vertebrate species (Frankham, Ballou, & Briscoe, 2010), a simple method to account for this complication could improve the robustness of IBD and IBE modeling results for many taxonomic groups.

Under hierarchical genetic structuring, populations of individuals clustered together in geographic space are more genetically similar than would be expected by IBD alone which leads to spatial autocorrelation in the genetic distance matrix (Meirmans, 2012; Wagner & Fortin, 2015). Failure to account for this inherent spatial autocorrelation in partial Mantel and MMR tests results in an inflation of coefficient values and an increased likelihood of detecting spurious relationships (Guillot & Rousset, 2013). Populations represent basic units of genetic heterogeneity on the landscape, and the strength of the IBD relationship between populations is likely to vary based on individual population demography and colonization history (Orsini, Vanoverbeke, Swillen, Mergeay, & Meester, 2013; Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998). Performing separate tests for individual populations solves the problem of spatial autocorrelation (Kuchta & Tan, 2005), but reduces the statistical power of test and breaks up the environmental variation in question. Utilizing a mixed modeling framework to account for population structure in MMR analysis may represent an effective way to model complicated IBD and IBE relationships without sacrificing sample size or reducing the range of genetic and environmental variation under examination.

Herein, I explore this approach in a long-lived, aquatic salamander species, the hellbender (*Cryptobranchus alleganiensis*). Hellbender populations are declining and threatened throughout most of their range (Burgmeier, Unger, Sutton, & Williams, 2011; Foster, McMillan, & Roblee, 2009; Wheeler, Prosen, Mathis, & Wilkinson, 2003) with the Ozark subspecies (*Cryptobranchus alleganiensis bishopi*) listed as endangered and the eastern subspecies (*Cryptobranchus alleganiensis alleganiensis*) currently under consideration for listing under the endangered species act (Gould, 2011). Threats to hellbender populations include habitat degradation disease, and limited gene flow among populations (Mayasich, Grandmaison, & Phillips, 2003). Moreover, it is uncertain how climate change may affect population persistence in this highly structured species. After determining the optimal IBE model, we extend the analysis to forecast genomic vulnerability across the range of the hellbender. Genomic vulnerability is a measure of the distance between projected future climate conditions and current local optima and may help identify locations where species are likely to face a genetic mismatch with conditions under climate change (Fitzpatrick & Keller, 2015, Bay et al., 2018). These results provide evolutionary insights into future threats for a species that is rapidly declining throughout much of its range.

1.3 Materials and Methods

Study species and sampling design

Hellbenders are the largest amphibian in North America and are distributed broadly across the eastern United States with disjunct populations occurring in Missouri and Northern Arkansas (Fig. 1.1). Besides the two distinct hellbender subspecies, range-wide genetic analyses have revealed additional moderate (Unger, Rhodes, Sutton & Williams, 2013) to high (Hime, 2017) levels of genetic structure. Analyses using the same SNP markers that were used in this

study provided evidence for four major genetic divisions in hellbenders (Fig. 1.1A; Hime personal communication).Ozark hellbenders, which occur in the Arkansas-White-Red region (Fig. 1.1A) and encompass their own subspecies, are the most genetically unique of the major demes (Hime, 2017). Eastern hellbenders in the Tennessee region (Fig. 1.1A) represent another major deme. Eastern hellbenders in Missouri are genetically similar to most hellbenders in the Ohio region (Fig. 1.1A); however, the Kanawha subregion, which is nested within the Ohio region, represents a separate genetic deme (Fig. 1.1 A). There is evidence for finer-scale genetic structure within the Tennessee deme, the Ohio deme, and the Ozark deme leading to eight distinct subpopulations (Fig. 1.1B; Hime personal communication). Within these broad genetic divisions, patterns of IBD tend to follow a dendritic pattern corresponding to stream network distances (Hime, 2017; Unger, Chapman, Regester, & Williams, 2016). The sampling sought to capture the full extent of genetic and environmental variation across the broad range of hellbenders. Since the power to detect spatial genetic patterns is more dependent on the number of locations sampled than the number of individuals sampled per location (Landguth *et al.*, 2012), the primary focus was on maximizing the number of locations sampled. In total, 194 hellbenders were sampled at 96 unique locations. Samples were obtained from all major watersheds and 14 out of 15 states known to contain hellbender populations.

Methodological overview

The primary objective of this study was to determine whether extending multiple matrix regression into a mixed modeling framework could adequately account for population structure in IBE models. I hypothesized that a random intercept model, which allows intercepts to vary between each interdeme comparison group, could help account for discrepancies in baseline genetic differences between populations. A random slope-intercept model, which allows the

slopes of the geographic distance variable and intercepts to vary between each interdeme comparison group, could help account for baseline genetic differences between populations as well as differences in the strength of the IBD relationship across the range the species. Both of these modifications may serve to more realistically capture the pattern of neutral variation across the landscape and likelihood methods can be utilized to determine the optimal model structure.

To meet the primary objective I took the following approach: I first calculated distance matrices for genetic, geographic, and environmental distance variables. Since I was uncertain of the optimal geographic distance variable, I developed models for six different IBD variables (Fig. 1.2A) across all possible levels of model structure (Fig. 1.2B) and chose the best based on mean BIC scores. I then compared model fit (BIC) and average IBE coefficient intensity across five different levels of model structure that varied in how strongly they accounted for the inherent autocorrelation in the genetic data (Fig. 1.2B). Once I determined the optimal model structure, I compared the performance of the best fitting model with the performance of the default IBE model, by comparing the IBE coefficient values with the results of locus-specific environmental association tests. Following this approach I was able to determine whether this method could improve IBE model fits, decrease IBE coefficient inflation, and result in greater concurrence with independent tests of environmental associations.

The secondary objective of our analysis was to inform hellbender conservation using the best of the developed IBE models. Once I was satisfied with the performance of the optimal model, I interpreted the IBE coefficient values and extended the analysis to forecast genomic vulnerability across the range of the species. Genomic vulnerability maps can highlight which populations are least likely to adapt quickly enough to track future climate shifts.

Distance Calculations

The genetic data for this study was collected as part of a range-wide exploration of population structure in hellbenders (Hime, 2017). Double digest restriction site-associated DNA sequencing (ddRAD; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) was used to develop a novel set of genetic markers distributed randomly throughout the hellbender genome. In total, the study produced a set of 54,532 variable loci sampled from both diploid chromosomes of individuals (for full details on marker development see Hime, 2017). To ensure none of the patterns were driven by missing values, we discarded markers missing more than 10% of data and individuals missing more than 25% of data. Since rare alleles are more likely to elicit false positives (Bay *et al*., 2018), we also removed markers with major alleles occurring in more than 90% of total markers. Individuals from the Susquehanna watershed were also removed from the analysis because of a lack of lotic connectivity with the remainder of the range. This resulted in a data set of 9125 markers for 150 individuals. The individuals occurred at 83 locations across the range of the species (Table 1.1). The locations were well distributed across genetic populations and subpopulations (Table 1.1); however, one of the Ozark hellbender subpopulations only contained a single sampling location so I combined the two subpopulations into a single Ozark group for subpopulation level analyses. All analyses were performed in R v 3.4.3 unless otherwise mentioned.

In IBE studies, genetic distance is usually represented by linearized FST values (Wang & Bradburd, 2014). However, FST values are unreliable when they are estimated with fewer than five individuals per location (Willing, Dreyer, & Van Oosterhout, 2012), and I wanted to avoid aggregation to maintain the continuous nature of the environmental and genetic variation across the landscape (Shirk, Landguth & Cushman, 2017). To accomplish this task, I estimated genetic differentiation between individuals and averaged these metrics between locations. I used the

diss.dist function (Kamvar, Tabima, and Grünwald, 2014) to measure the number of allelic differences between all individuals since this metric performs well under model selection compared to other individual-based genetic metrics (Shirk et al., 2017). I recorded the number of comparisons between locations and assigned each averaged pairwise distance value to a population and subpopulation comparison group.

I calculated six different geographic distance variables for comparison (Fig. 1.2A). I calculated Euclidean distances between locations using the point distances command in the Geospatial Modeling Environment software (Beyer, 2012). Since dendritic patterns of IBD following stream networks are common in lotic species (Hughes, Schmidt, & Finn, 2009), I also calculated stream distances between locations. I used the network analysis toolbox in ArcMap v 10.2.2 (Esri 2014) to build a stream network using all NHDPlus v2 FlowLine features (USGS 2013) east of the Mississippi River and calculate linear distances between all locations. To help linearize the IBD relationship (Legendre & Fortin, 2010), I performed square root and log transformations on both Euclidean and stream distance matrices.

I modeled IBE in both stream and climate space. Climate is often assumed to be the most important environmental driver of genomic changes; however, since hellbenders are obligate aquatic salamanders and rarely leave their natal system (Mayasich et al., 2003), I hypothesized that the stream environment would be as important as the climactic environment for shaping the genome. I measured stream variation using eight attributes associated with NHDPlus v 2 FlowLine features (USGS 2013), including stream level, stream order, upstream length, catchment area, upstream catchment area, maximum elevation, minimum elevation and slope (for a full description of the variables see Table S1.1). I used the standard 19 bioclimatic

variables (Table S1.2) derived from WorldClim version 1.4 (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) to summarize climate conditions across the hellbender range.

Since MMR analyses are sensitive to multicollinearity in predictor variables (Wagner & Fortin, 2015; Wang, 2013), I used principal components analysis (PCA) to derive orthogonal predictor variables in each environmental realm. To capture the full range of environmental variation in the range, I projected 100,000 random points within the delineated watersheds (Fig. 1.1) using the Create Random Points tool in ArcMap version 10.2.2 (Esri 2014). After snapping and joining points to the closest NHD FlowLine feature, I extracted stream and climate variables for each point. Each PCA was performed with the *princomp* function using the correlation matrix of the variables. In both PCAs, the first five axes explained ~ 95% of the variation and were retained for further analysis (Tables S1.3 & S1.4).

I interpreted the stream principal component axes as measures of elevation, stream position (upstream vs. downstream), stream level, stream size, and slope (Table S1.3). I interpreted the climate principal component axes as measures of regional temperature and precipitation patterns (generally getting colder and dryer moving from the southeast to the northwest of the range), summer temperature and precipitation patterns, temperature seasonality, precipitation seasonality, and temperature range (Table S1.4). *A*ll retained stream and climatic variables were applied and mapped across the species range (Figs. S1.1 & S1.2) including the locations of the genetic sampling. Prior to MMR analysis, pairwise Euclidean distance values between each location were calculated for each environmental variable using the *dist* function.

Multiple matrix regression

The first step of the modeling process was to determine the optimal geographic distance variable which I approached by fitting all possible geographic distance variables (Fig. 1.2A)

across all levels of model structure (Fig. 1.2B) and choosing the variable with lowest mean BIC score. For comparison purposes, I fit saturated models with all of the fixed environmental variables (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). Prior to analysis, I centered and scaled independent variables to standardize coefficients and checked for correlations among the independent variables. Since all correlation values were < 0.5, I did not remove any independent variables from the analysis. Only the upper half of distance matrices were used in the regression, but I did include zero geographic distance comparisons. All MMR models were fit using the number of comparisons at each location as a weighting variable, because it always significantly improved model fits. I used the *lmer* function to fit mixed models with maximum likelihood and specified correlated slope-intercept relationships in the slope-intercept models (Bates, Maechler, Bolker, & Walker, 2014). I fit unstructured models with the *lm* function.

Once the optimal distance variable was determined, I compared the performance of models with the best distance variable and different levels of model structure. The traditional linear MRM served as the null model comparison. I fit both random intercept and random slope-intercept models at the population and subpopulation level (Fig. 1.2B). I compared model fit by examining BIC scores and performing likelihood tests with each additional level of added model complexity. I compared IBE coefficient intensity by averaging the absolute value of all environmental coefficient values in each final model. Once I determined the optimal model structure, I made a final comparison of model performance between the null model (with the best geographic distance variable but no model structure) and the optimal model (with both the best geographic distance variable and the best model structure).

I used locus-specific association tests to serve as independent measures of model accuracy. If an IBE model is accurate, I would expect IBE coefficient intensity at a particular

environmental axis and the number of loci associated with that axis to show a strong correlation. I would also expect variables with at least one associated locus to have significant pseudo p-values so I performed permutational testing on IBE coefficients. During the permutational testing I was interested only in the significance level of the environmental variables, so I did not permute the geographic distance variable. To avoid breaking up the spatial structure in the mixed-model, I permuted environmental variables within their random variable comparison groups (Guillot & Rousset, 2013; Meirmans, 2012). I used 1000 permutations and a two-tailed test to calculate pseudo p-values for each variable.

I used latent factor mixed model analysis (LFMM; Frichot & François, 2015) to determine the number of individual loci associated with each environmental axis. LFMM is a Bayesian approach that uses latent factors to control for underlying population structure before testing for associations between allele frequencies and the environment. Following the developers guidelines, I used the function *snmf* to calculate the cross-entropy across five runs for each value of K (representing ancestral populations) and graphed the results to determine where the cross-entropy values plateaued (Frichot & François, 2015). Cross-entropy values plateaued between 11 and 15 K (Fig. S1.3), so I performed locus-specific association testing on each environmental variable at each level of K from 11 to 15 using the function *LFMM* (Frichot & François, 2015). Since the Bayesian output of the analysis can be variable, I ran each test five times. As suggested by the developer at small sample sizes, I increased the default burn-in period of 5,000 to 10,000 and the default number of iterations from 10,000 to 20,000. I took the median z-value for each locus from the series of five runs and adjusted them for multiple testing using a 10% false discovery rate (Benjamini & Hochberg, 1995). I recorded and averaged the number of associated loci over all values of K (Table S1.5). Finally, I compared the intensity and pseudo p-

value of each IBE coefficient in the null and structured models with the mean number of loci associated with each corresponding environmental axis.

Genomic Vulnerability

I used the results of the best model to develop a genomic vulnerability map. Genomic vulnerability is the distance between projected future climate conditions and the current local optima. This analysis finds the weighted (based on the model coefficient value of the environmental variable) Euclidean distance between current climate conditions and a set of future climate conditions. I took the median value from five validated global climate models (Table S1.6) for future climate forecasts. This ensemble approach improves regional climate prediction because it smooths extreme values (Pierce, Barnett, Santer, & Glecker, 2009). I used climate projections for 2050 based on a moderate emissions scenario (representative concentration pathway 4.5; Van Vuuren et al., 2011). Once values were extracted, I used the predict function to apply the principal components transformation on the new data. Climate PCAs that were retained in the final model were projected across the range of the hellbender. I calculated the weighted Euclidean distance between current conditions and projected future conditions using model environmental variable importance as a weighting factor. We visualized this raster surface of genomic vulnerability across the hellbender range.

1.4 Results

Model comparisons

Accounting for genetic structure using a mixed modeling framework improved model fits, decreased IBE coefficient inflation, and increased model agreement with locus-specific association tests. Model fits improved with increasing levels of model structure, though the greatest improvement came from adding a random intercept component at the population level (a

39% reduction in BIC score), and gains from increasing the number of divisions (a mean 10% ± 0.03 reduction in BIC score) and adding a random slope (a mean 9% ± 0.03 reduction in BIC score) were more modest (Fig. 1.3A). However, likelihood tests did suggest that the most complex model fit the data significantly better than the others, and the subpopulation level slope-intercept model was used for the final comparison against the null model. Overall, the BIC score of the most complex model was 55% lower than the null model (Fig 3A). IBE coefficient intensity showed a similar pattern (Fig. 1.3B), with the greatest decrease in the mean value occurring with the addition of a population level random intercept (a 55% reduction) and smaller decreases occurring with the addition of a random slope component (a mean 24% ± 0.07 reduction). Overall, mean coefficient intensity was 64% lower in the most complex model than the null model (Fig. 1.3B).

The structured model showed a high degree of concurrence with the locus-specific tests (Fig. 1.4A). The five environmental axes that had twoor more associated loci had higher coefficient values and pseudo p-values lower than 0.05. There was one variable that was deemed significant by the model that had no associated loci in the corresponding locus-specific test. The unstructured model showed a very low degree of concurrence with the locus-specific tests (Fig. 1.4B). The model was unable to correctly select the three variables with the highest number of associated loci. It also incorrectly selected three variables at the 0.05 alpha level that showed no associated loci in the corresponding locus-specific tests.

In the comparison of geographic distance variables, stream distances always fit better than their Euclidean distance counterparts (Table 2). The square root stream distance variable demonstrated the best fit over all levels of model structure and was retained for subsequent model comparison (Table 1.2). The strongest environmental relationship detected in the IBE

model was with temperature range (Table S1.7). Stream position, stream level, stream size all had slightly smaller coefficient intensities (Table S1.7). Temperature seasonality and precipitation seasonality displayed smaller, but still significant effects. The IBE model association with precipitation seasonality was not detected by the locus-specific association test.

Since the subpopulation level slope-intercept model had the best fit and demonstrated high concordance with locus-specific tests, we used its coefficient intensity values to develop the genomic vulnerability forecasting. Because precipitation seasonality was not detected by the locus-specific association test, we took a conservative approach and did not include it as a variable in the genomic vulnerability forecasting. The greatest levels of genomic vulnerability occurred in a patch across Ohio, Kentucky, Tennessee, and Alabama (Fig. 1.5). There is also a patch of more moderate distance values across eastern Tennessee, North Carolina, Virginia, and West Virginia (Fig. 1.5). Genomic vulnerability is minimal in Missouri, eastern Kentucky, Indiana, Mississippi, and much of Ohio, New York, and Pennsylvania (Fig. 1.5).

1.5 Discussion

My results indicate that utilizing a mixed modeling framework in MMR can be an effective way to account for hierarchical population structure in models of isolation by distance (IBD) and isolation by environment (IBE). The null model performed poorly against mixed models in likelihood tests. Average environmental coefficients were more than twice the intensity in the null model compared to the mixed models. Even with permutational testing, the null model detected spurious environmental relationships and failed to detect true environmental relationships as demonstrated by the lack of concurrence with locus-specific tests. These results further emphasize the importance of explicitly accounting for genetic structure in IBE models as failure to do so leads to inaccurate results (Guillot & Rousset, 2013; Meirmans, 2012). These

results reinforce that patterns of IBE can often be detected, but effectively pulling apart patterns of neutral and adaptive variation in genome-wide studies is difficult (Wang & Bradburd, 2014). Neutral processes are likely to explain the majority of genomic variation since environmental associations generally occur in less than 5% of tested loci (Bay et al., 2017). A thorough investigation of the appropriate model structure and IBD relationship may be required to detect subtle IBE relationships orders of magnitude weaker than neutral patterns. Failure to do so can lead to inflated IBE coefficients and misleading results.

While more sophisticated methods have been developed to account for neutral genetic structure in IBE models (Bradburd, Ralph, & Coop, 2013; Dyer, Nason, & Garrick, 2010), this method represents a natural extension of multiple matrix regression which is an intuitive and commonly used approach to model patterns of IBD and IBE (Storfer et al., 2010; Wang & Bradburd, 2014). Furthermore, mixed modeling has been embraced by the ecological community as a means of accounting for inherent variation in observational studies, and packages are already available in common statistical software to fit these models (Bolker et al., 2009). The amount of genetic structure seen in hellbenders is common in several taxonomic groups (Frankham et al., 2010), and this method may be used to account for more moderate levels of differentiation as long as population divisions are known a prioiri. An additional advantage of this approach is that likelihood tests can be used to determine what level of genetic and model structure best improves IBE model fits.

This study emphasizes that lotic species may warrant special consideration in IBE modeling, particularly in the choice of independent variables. We found that models using stream distances better accounted for patterns of IBD than Euclidean distances. We also found that variables describing stream environment were just as strongly associated with genome-wide

variation as variables describing climatic gradients. In particular, stream position, stream level, and stream size gradients all showed associations with the hellbender genome. Given the wide variety of stream conditions across the hellbender range and the ability of variation in the lotic environment to cause morphological and physiological changes in hellbenders (Kenison & Williams, 2018); it is unsurprising that stream conditions may be driving genetic differentiation. Climate conditions were also important, in particular temperature variability and annual temperature range. Temperature adaptations may be common in amphibians distributed across a latitudinal gradient (Orizaola, Quintela, & Laurila, 2010; Snyder & Weathers, 1975). Temperature may be a strong selective force for hellbenders specifically, since they rely on high levels of dissolved oxygen for cutaneous respiration (Guimond & Hutchison, 1973), and dissolved oxygen levels are directly related to stream temperatures.

The environmental associations detected in this study have implications for hellbender conservation efforts. There have already been recommendations that hellbender translocations only occur within major population boundaries, because of the strong neutral genetic structure of the species (Hime, 2017; Unger et al., 2013). I suggest that environmental matching should also be considered when moving hellbenders for conservation purposes. Breaking the affiliation between genome and the environment can reduce the probability of successful establishment (Bragg, Supple, Andrew, & Borevitz, 2015). Both lotic and climate gradients can vary widely across the geographic boundaries of a given population, and condition matching, particularly in the environmental variables that have been tested and show associations, may improve post-translocation survival rates (McKay, Christian, Harrison, & Rice, 2005). However, in genetically structured populations, new alleles represent new variation for selection to act upon so there is a tradeoff to keeping conservation actions as local as possible (Aitken & Whitlock, 2013).

Optimally, detecting patterns of IBE can act as an early step in fully understanding the genetic architecture of a species (Bay et al., 2017). Since the markers are not mapped, I cannot associate them with particular regions of the genome or understand the linkage structure among them. Linking the environmentally associated loci to phenotypes would elucidate what traits are being acted upon by stream and climate variables.

The results of the genomic vulnerability mapping are also important in the context of hellbender conservation. Hellbenders have experienced declines throughout their range (Burgmeier et al., 2011; Foster et al., 2009; Wheeler et al., 2003), but current declines have been most consistently associated with forest removal and land use change (Jachowski & Hopkins, 2018; Nickerson, Pitt, Tavano, Hecht, & Mitchell, 2017). The relatively high genomic vulnerability in regions that currently contain the only remaining stable hellbender populations (Eastern Tennessee, North Carolina, Virginia and West Virginia; Mayasich et al., 2003) is potentially concerning. Genomic vulnerability has been associated with population declines in migratory birds (Bay et al., 2018), and climate shifts appear likely to impose additional challenges to a species already stressed by anthropogenic changes. This may present a situation where we need to act before declines are detected, because the long generation time and life span of hellbenders makes them more likely to incur an extinction debt (Kuussaari et al., 2009). It should be noted that the genomic vulnerability results rely on several important assumptions including that the sampling reflects variation across important environmental gradients, that important environmental gradients are adequately measured and modeled at an appropriate spatial scale, and that the genetic data adequately captures the most important environmental associations (Bay et al., 2018). The last limitation is of particular concern since sampled markers represent a small and random sample of a large genome. While there are limitations to my

recommendations, they represent a science-based management strategy utilizing the most current genetic and analytical methods to generate conservation recommendations.

Misleading genetic and ecological models can be costly if they lead to misinformed conservation decisions and may erode the trust of stakeholders in science-based management (Addison et al., 2013). Using concurrent methods (i.e. genome wide and locus-specific association tests) to provide multiple lines of evidence can help validate model results. Several studies have demonstrated that moderate to severe hierarchical genetic structure in the genome must be adequately accounted for prior to modeling environmental associations, and the method demonstrated herein provides a simple way to do so. I suggest that IBE studies of any taxonomic group displaying hierarchical levels of genetic structure use a mixed model framework or other tested method of controlling for structure (Bradburd et al., 2013; Dyer et al., 2010) if the results are intended to draw evolutionary conclusions or help guide species conservation and management efforts.

1.6 Literature Cited

Addison, P. F. E., Rumpff, L., Bau, S. S., Carey, J. M., Chee, Y. E., Jarrad, F. C., …Burgman, M. A. (2013) Practical solutions for making models indespensible in conservation decision-making. *Diversity and Distributions 19*, 490-502. https://doi.org/10.1111/ddi.12054

Aitken S. N., Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics 44*, 367-388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package v. 1.7.

Bay, R. A., Harrigan, R. J., Le Underwood, V., Gibbs, H. L., Smith, T. B. & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science 359*, 83-86. https://doi.org/10.1126/science.aan4380

Bay, R. A., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C. K., Lasky, J. R., …Ralph, P. (2017). Predicting responses to contemporary environmental change using evolutionary response architectures. *The American Naturalist, 18*, 463-473. https://doi.org/10.1086/691233

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289-300.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*,127-135. https://doi.org/10.1016/j.tree.2008.10.008

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution 67*, 3258-3273. https://doi.org/10.1111/evo.12193

Bragg, J. G., Supple, M. A., Andrew, R. L., & Borevitz, J. O. (2015). Genomic variation across landscapes: insights and applications. *New Phytologist, 207*, 953-967. https://doi.org/10.1111/nph.13410

Burgmeier, N. G., Unger, S. D., Sutton, T. M., & Williams, R. N. (2011). Population status of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) in Indiana. *Journal of Herpetology 45*, 195-201. https://doi.org/10.1670/10-094.1

Dyer, R. J., Nason, J. D., & Garrick, R. C. (2010). Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, 19, 3746–3759. https:// doi.org/10.1111/j.1365-294X.2010.04748.x

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters 18*, 1-16. https://doi.org/10.1111/ele.12376

Foster, R. L., McMillan, A. M., & Roblee, K. J. (2009). Population status of hellbender salamanders (*Cryptobranchus alleganiensis*) in the Allegheny River drainage of New York State. *Journal of Herpetology*, *43*, 579-588. https://doi.org/10.1670/08-156.1

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). Introduction to conservation genetics (2nd ed.). Cambridge, UK: Cambridge University Press.

Frichot, E., & François, O. (2015). LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution, 6*, 925-929. https://doi.org/10.1111/2041-210X.12382

Gould, R. W. (2011). Inclusion of the hellbender, including the eastern hellbender and the Ozark hellbender, in appendix III of the convention on international trade in endangered species of wild fauna and flora (CITES). In: (Fed Register), pp. 61978-61985.

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, *4*, 336-344. https://doi.org/10.1111/2041-210x.12018

Guimond, R. W. & Hutchison, V. H. (1973). Aquatic respiration: an unusual strategy in the hellbender *Cryptobranchus alleganiensis alleganiensis* (Daudin). *Science*, *182*, 263-1265. https://doi.org/10.1126/science.182.4118.1263

Günther, T. & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics, 195*, 205-220. https://doi.org/10.1534/genetics.113.152462

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*, 1965-1978. https://doi.org/10.1002/joc.1276

Hime, P. M. (2017). *Genomic perspectives on amphibian evolution across multiple scales* (Doctoral disseration). University of Kentucky, Lexington.

Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience, 59*, 573-583. https://doi.org/10.1525/bio.2009.59.7.8

Jachowski, C. M. B., & Hopkins, W. A. (2018). Loss of catchment-wide riparian forest cover is associated with reduced recruitment in a long-lived amphibian. *Biological Conservation, 220*, 215-227. https://doi.org/10.1016/j.biocon.2018.02.012

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, e281. https://doi.org/10.7717/peerj.281

Kenison, E. K., & Williams, R. N. (2018). Rearing captive Eastern hellbenders (*Cryptobranchus a. alleganiensis*) with moving water improves swim performance. *Applied Animal Behaviour Science, 202*, 112-118. https://doi.org/10.1016/j.applanim.2018.01.013

Kuchta, S. R., & Tan, A. M. (2005). Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. *Molecular Ecology, 14*, 225-244. https://doi.org/10.1111/j.1365-294X.2004.02388.x

Kuussaari, M., Bommarco, R., Heikkinen, R. K., Helm, A., Krauss, J., Lindborg, R., …Steffan-Dewenter, I. (2009). Extinction debt: a challenge for biodiversity conservation. *Trends in Ecology & Evolution, 24*, 564-571. https://doi.org/10.1016/j.tree.2009.04.011

Landguth, E. L., Fedy, B. C., Oyler-McCance, S., Garey, A. L., Emel, S. L., Mumma, M., …Cushman, S. A. (2012). Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources, 12*, 276-284. https://doi.org/10.1111/j.1755-0998.2011.03077.x

Legendre, P., & Fortin, M. J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources, 10*, 831-844. https://doi.org/10.1111/j.1755-0998.2010.02866.x

Legendre, P., Lapointe, F.J., & Casgrain, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution, 48*, 1487-1499. https://doi.org/10.1111/j.1558-5646.1994.tb02191.x

Mayasich, J., Grandmaison, D., & Phillips, C. (2003). *Eastern hellbender status assessment report*. Duluth, MN: U.S.D.A. Forest Service.

McKay, J. K., Christian, C. E., Harrison, S., & Rice, K. J. (2005). "How local is local?"—a review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, **13**, 432-440. https://doi.org/10.1111/j.1526-100X.2005.00058.x

McRae, B. H. (2006) Isolation by resistance. *Evolution, 60*, 1551-1561. https://doi.org/10.1554/05-321.1

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology* 21, 2839-2846. https://doi.org/10.1111/mec.13243

Mullen, L. B., Woods H. A., Schwartz, M. K., Sepulveda, A. J., & Lowe, W. H. (2010). Scale-dependent genetic structure of the Idaho giant salamander (*Dicamptodon aterrimus*) in stream networks. *Molecular Ecology, 19*, 898-909. https://doi.org/10.1111/j.1365-294X.2010.04541.x

Nickerson, M. A., Pitt, A. L., Tavano, J. J., Hecht, K. A., & Mitchell, J.C. (2017). Forest removal and the cascade of effects corresponding with an Ozark hellbender populaiton decline. *Bulletin of the Florida Museum of Natural History*, *54*, 147-164.

Orizaola, G., Quintela, M., & Laurila, A. (2010). Climatic adaptation in an isolated and genetically impoverished amphibian population. *Ecography, 33,* 730-737. https://doi.org/10.1111/j.1600-0587.2009.06033.x

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology, 22*, 5983-5999. https://doi.org/10.1111/mec.12561

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one, 7,* e37135. https://doi.org/10.1371/journal.pone.0037135

Pierce, D.W., Barnett, T.P., Santer, B.D., & Gleckler, P.J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, 106, 8441-8446. https://doi.org/10.1073/pnas.0900094106

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics, 145*, 1219-1228.

Shirk, A., Landguth, E., & Cushman, S. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources, 17*, 1308-1317. https://doi.org/10.1111/1755-0998.12684

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology, 35*, 627-632. https://doi.org/10.2307/2413122

Snyder G. K. & Weathers, W. W. (1975). Temperature adaptations in amphibians. *The American Naturalist, 109*, 93-101.

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: where are we now? *Molecular Ecology, 19,* 3496-3514. https://doi.org/10.1111/j.1365-294X.2010.04691.x

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology, 7,* 453-464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Unger, S. D., Chapman, E. J., Regester, K. J., & Williams, R. N. (2016). Genetic signatures follow dendritic patterns in the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) *Herpetological Conservation and Biology,* 11, 40-51.

Unger, S. D., Rhodes, Jr., O.E., Sutton, T. M., & Williams, R. N. (2013). Population genetics of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) across multiple spatial scales. *PloS one* **8**, e74180. https://doi.org/10.1371/journal.pone.0074180

Van Strien, M. J., Keller, D., & Holderegger, R. (2012). A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. *Molecular Ecology, 21,* 4010-4023. https://doi.org/10.1111/j.1365-294X.2012.05687.x

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A. Hibbard, K., …Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change,* 109, 5-31. https://doi.org/10.1007/s10584-011-0148-z

Wagner, H. H., & Fortin, M. J. (2015). Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. In N. Balkenho, S. Cushman, A. Storfer, & L. Waits (Eds.), *Landscape genetics: Concepts, methods, applications* (pp. 77-98). Oxford, UK: Wiley Blackwell.

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution, 67*, 3403-3411. https://doi.org/10.1111/evo.12134

Wang I. J., Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology, 23*, 5649-5662. https://doi.org/10.1111/mec.12938

Wheeler, B. A., Prosen, E., Mathis, A., & Wilkinson, R. F. (2003). Population declines of a long-lived salamander: a 20+-year study of hellbenders, *Cryptobranchus alleganiensis*. *Biological Conservation, 109,* 151-156. https://doi.org/10.1016/S0006-3207(02)00136-2

Willing, E. M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PloS one, 7,* e42649.

Wright, S. (1943) Isolation by distance. *Genetics, 28*, 114-138.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R.* New York: Soringer-Verlag.

1.7 Tables

Table 1-1. The number of locations and individual hellbenders (*Cryptobranchus alleganiensis*) sampled for genetic analysis per population and sub-population.

Populations	**# of Locations**	**# of Individuals**
Kanawha	8	14
Ohio	23	48
Ozark	4	10
Tennessee	48	78
Subpopulations	**# of Locations**	**# of Individuals**
Black	3	4
Kanawha	8	14
Gasconade	4	11
Hiawassee-Ocoee	15	23
Ohio	19	37
Tennessee	25	39
Upper French Broad	8	16
White	1	6

Table 1-2. The mean BIC value over all levels of model structure (n=5) for each geographic distance variable used to model isolation by distance.

Distance	Mean BIC	SE
Euclidean	-9421.2	1097.5
Log Euclidean	-10052.8	927.9
Square Root Euclidean	-9862.0	964.2
Stream	-10247.1	721.5
Log Stream	-10202.8	866.6
Square Root Stream	-10931.3	697.5

1.8 Figures

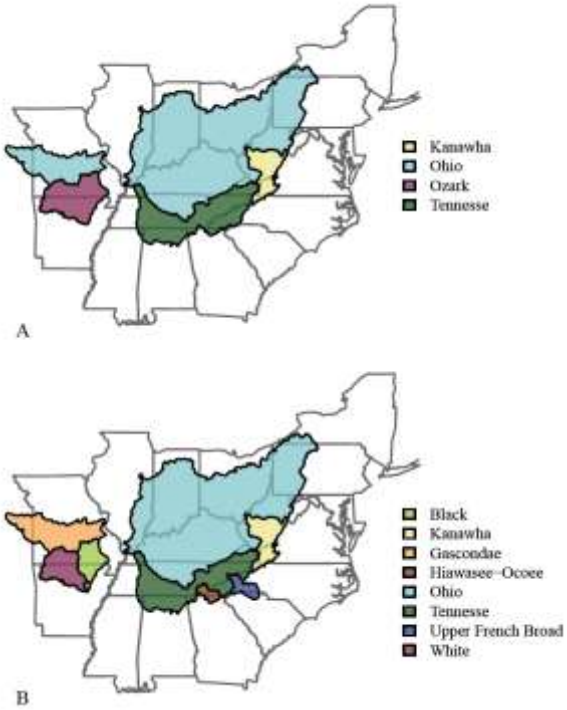

Figure 1-1. The geographic boundaries partitioning major genetic populations (A) and subpopulations (B) of hellbenders (*Cryptobranchus alleganiensis*).

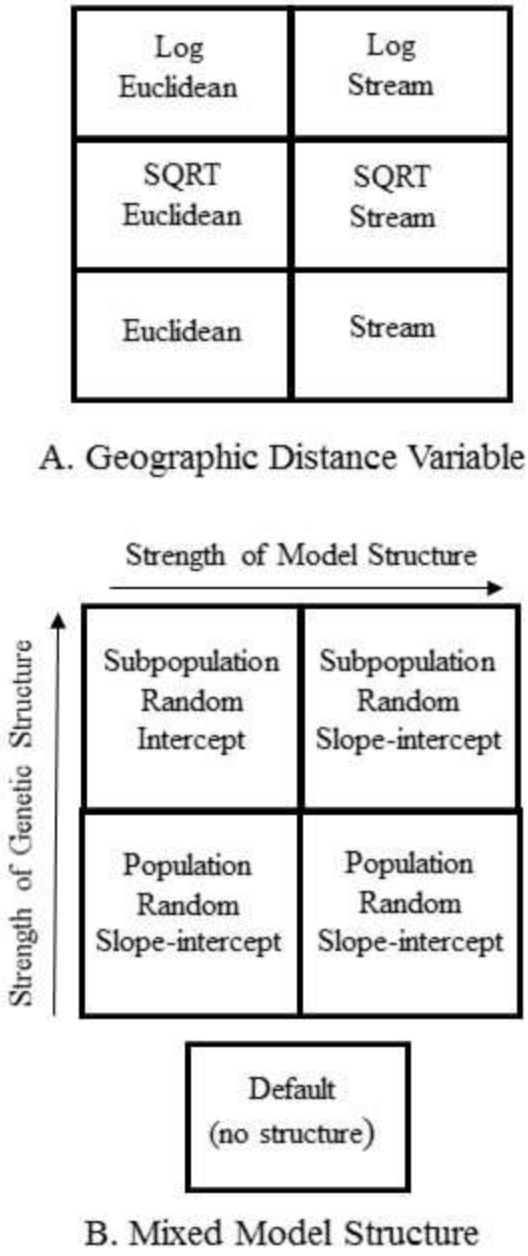

Figure 1-2. The geographic distance variables (A) and mixed model structures (B) used to model isolation by distance in the hellbender (*Cryptobranchus alleganiensis*) genome. Models were fit using each geographic distance variable and mixed model structure and compared using BIC values

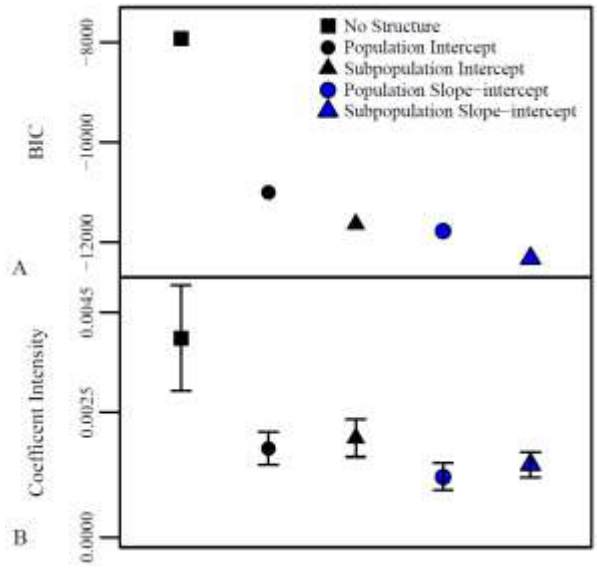

Figure 1-3. Model fit (A) and mean isolation by environment coefficient intensity (B) across five models accounting for increasingly greater amounts of genetic structure in the hellbender (*Cryptobranchus alleganiensis*) isolation by distance relationship.

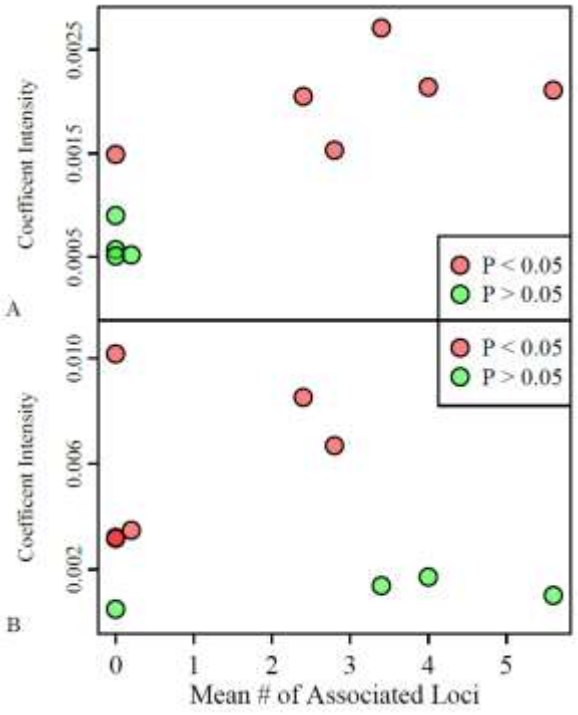

Figure 1-4. A comparison of isolation by environment (IBE) model results (coefficient intensity and p-values) and locus-specific association test results (mean number of associated loci) for each tested environmental variable for a model that accounts for genetic structure (A) and a default model that does not (B) in hellbender (*Cryptobranchus alleganiensis*) IBE models. Perfect agreement of the two methods would result in a positive linear trend line.

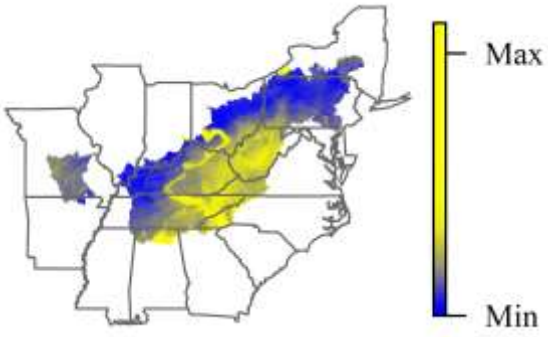

Figure 1-5. A map of genomic vulnerability across the range of the hellbender (*Cryptobranchus alleganeinsis*). Color gradient values are stretched using quantile classification with 100 classes. Higher values represent areas projected to have greater distances between current climate optima and future projected conditions.

CHAPTER 2. A REGIONAL APPROACH TO IMPROVE RANGE-WIDE SPECIES DISTRIBUTION MODELS OF IMPERILED SPECIES

2.1 Abstract

Presence-only species distribution models are important tools for conservation but lack methods to account for regional differences in relationships or spatial autocorrelation. Failure to account for these spatial complications can lead to inflation of coefficient values and poor regional model performance. Recent methodological advances have revealed the equivalency of maximum entropy presence-only modeling and appropriately weighted generalized linear logistic and Poisson models. I extended the method into a mixed modeling framework to account for environmental heterogeneity, spatial autocorrelation, and nonstationarity in a range-wide species distribution model for the imperiled, aquatic eastern hellbender salamander (*Cryptobranchus alleganiensis alleganiensis)*. I took a regional approach to develop a range-wide species distribution model. Since I knew baseline occurrence levels of hellbenders differed by physiographic province, I used the provinces to delineate data for region-specific covariates and regional performance testing. I developed three models (region-specific intercept, region-specific intercept with an autocovariate, and region-specific intercept with a region-specific autocovariate) and compared their discrimination and calibration performance using spatial-block cross validation. I also divided the testing sets into their respective physiographic provinces and tested regional model performance across the range. Modeled relationships included climate, stream, and land use variables. The model with both the region-specific intercept and region-specific autocovariate had the best discrimination and calibration performance, though there was a trade-off in increased bias. Taking a regional approach to assessment also allowed me to identify areas where the model was underperforming in order to

target future data collection. As with other aquatic species, I found that stream variables had the highest predictive power, but land use and climate variables also contributed to the model. I also saw a much higher intercept in the Blue Ridge physiographic province than others suggesting that some unexplained factor is buffering the species from declines in the region. A regional approach to model building and assessment was important, as it improved model performance in areas most in need of hellbender conservation and management.

2.2 Introduction

Species distribution models (SDMs) are a class of models that relate species occurrences to environmental predictor variables to estimate habitat suitability. SDMs can serve as a valuable tool for conservation since both the data and resources needed to make critical decisions are often limited (Leung & Steele, 2013). In the context of conservation, SDMs have been used to improve the sampling efficiency of rare species (Guisan *et al*., 2006; Williams *et al*., 2009), determine suitable sites for species reintroductions and translocations (Schadt *et al*., 2002; Hendricks et al., 2016), support resource planning and reserve selection (Ortega-Huerta & Peterson, 2004; Leathwick *et al.*, 2008; Bombi *et al.*, 2011), and assess the impact of anthropogenic stressors on populations (Junker *et al*., 2012; Radinger *et al*., 2017; Dyderski *et al.*, 2018). Given the uncertain nature of absences (Lobo *et al*, 2010), as well as the wide availability of occurrence data (Graham *et al*, 2004), presence-only SDM methods are commonly employed to develop habitat suitability predictions for conservation planning. In particular, the MAXENT program (Elith *et al*., 2006), which uses maximum entropy modeling to compare occurrence data against background points, has dominated the SDM field since its introduction (Renner *et al*., 2015). One unfortunate shortcoming of maximum entropy models is that they are spatially invariant and rarely account for regional differences in relationships or spatial

autocorrelation in the response variable (Miller, 2012). Failure to adequately account for these conditions may adversely affect predictor variable reliability and model performance. However, recent research has demonstrated the equivalence of presence-only maximum entropy modeling and appropriately weighted logistic (Fithian & Hastie, 2013) and Poisson (Renner *et al*., 2015) regression. The spatial complications of presence-only data are simpler to handle within the generalized linear modeling framework as regional random effects may be used to account for spatial heterogeneity in data and model residuals can be examined for signatures of spatial autocorrelation.

A regional approach to building range-wide SDMs is a compromise between building static models that represent all regions equally and spatially partitioning data into separate regional models. Global models can mask true species-environment relationships by averaging their effects over the entire range of the species (Osborne *et al*., 2007). Spatially partitioning data allows intercepts and species-environment relationships to vary across the range of data and often improves regional performance (El-Gabbas & Dormann, 2018; Osborne, 2002), but limits sample size and power to predict and extrapolate underlying relationships. Since species of conservation concern are often rare and data limited, building separate region-wide models can often be impractical. Moreover, if sample sizes are unequal across regions, resulting models may perform worse in low abundance and data-deficient regions that most need conservation guidance. Models that allow intercepts (Hamil *et al*., 2016) and slopes (Osborne *et al*., 2007; Miller *et. al* 2011) to vary across regions, may help account for spatial heterogeneity across the species range and improve model performance. Furthermore, taking a regional approach to model validation allows a more nuanced view of model performance. More than half of recent SDMs have relied on a single global metric to assess model performance, which can easily lead

to overconfidence in results (Fourcade *et al*., 2018). Assessing performance regionally provides a means to understand spatial variability in model results. Even within a model that accounts for regional differences, performance may be negatively affected by spatial autocorrelation.

Spatial autocorrelation in species occurrence data is common and can be problematic for model development (Legendre, 1993). Spatially invariant SDMs rely on the assumption that proper model specification will remove residual autocorrelation but rarely test this assumption (Miller, 2012). Even within a properly specified model, spatial autocorrelation may occur in the residuals if occurrences are clustered in space due to biological processes or unequal sampling effort (Segurado *et al*., 2006; Dormann *et al*., 2007). Regardless of its source, inherent autocorrelation leads to inflated coefficient estimates (Bini et al., 2009) and may even invert the slope of predicted relationships (Kühn, 2007). Filtering data to remove highly autocorrelated records has been suggested (Boria *et al.*, 2014); but, like spatial partitioning of data, is impractical when species are rare and data is limited. Using predictors to model sampling bias in data may likewise remove some autocorrelation in model residuals (El-Gabbas & Dormann, 2017); however, the method requires previous explicit knowledge and measurement of the factors driving sampling bias. Modeling autocorrelation directly by including an autocovariate within the model is a more straightforward approach to accounting for spatial autocorrelation (Dormann, 2007). Allowing the slope of the autocovariate to vary by region accounts for differences in the strength of autocorrelation across the range. Herein, I used a mixed modeling framework to account for regional differences and spatial autocorrelation in an SDM for an imperiled, aquatic salamander species (*Cryptobranchus alleganiensis*) with a clustered distribution across the landscape. I divided the range into ecologically meaningful regions and developed three models: an SDM with region-specific intercepts, an SDM with region-specific

intercepts and a spatial autocovariate, and an SDM with region-specific intercepts and a region-specific spatial autocovariate. I compared the global and regional performance of each model using spatial-block cross validation and several metrics to assess model discrimination and calibration.

2.3 Materials and Methods

Study species

Hellbenders (*Cryptobranchus alleganienesis*) are long-lived, aquatic salamanders that are threatened throughout much of their historic range (Mayasich *et al.*, 2003). Hellbenders are divided into two subspecies, with eastern hellbenders (*Cryptobranchus alleganienesis alleganienesis*) ranging from southwestern New York southward to Georgia, Alabama, and Mississippi and westward to Missouri, while Ozark hellbenders (*Cryptobranchus alleganienesis bishopi*) are restricted to watersheds in southeastern Missouri and Northeastern Arkansas (Fig. 2.1). Hellbenders once represented a considerable amount of biomass in occupied systems (Nickerson & Mays, 1973); however, populations of both subspecies have faced considerable declines throughout their range (Wheeler *et al.*, 2003; Foster *et al.*, 2009; Burgmeier *et al.*, 2011). These declines have been characterized by a lack of recruitment of younger age classes and a corresponding decrease in the body condition of remaining adults (Bothner & Gottlieb 1991; Wheeler *et al.*, 2003; Jachowski & Hopkins, 2018). The Ozark hellbender is currently protected under the Endangered Species Act, and the eastern subspecies is under consideration for listing by the United States Fish and Wildlife Service (Gould, 2011). Previous occupancy modeling has demonstrated underlying differences in hellbender occurrence rates across physiographic provinces (Jachowski *et al.*, 2016) regardless of stream conditions and surrounding land use. Given the vulnerable status of the species, a range-wide SDM can serve to

target sampling locations for exploratory population searches and identify suitable habitat for translocations.

Species data and geographic coverage

Unless otherwise noted all GIS analyses were performed in ArcMap Version 10.2.2 (Esri, 2014) and all statistical analyses were performed in R version 3.4.3 (R Core Team, 2013). Hellbender locations and sampling dates were collected from researchers throughout the range of the species. While publicly available locational databases are often used to model species distributions, this approach is unreliable in the case of hellbenders, because they are easily mistaken for other aquatic salamanders (Mayasich *et al.*, 2003). This type of misidentification error can lead to serious bias in SDM results (Lozier *et al.*, 2009). Collecting data exclusively from researchers also ensured uniformity in sampling technique since hellbender research relies on rock lifting and snorkel surveys. Because I was interested in the current distribution of hellbenders, I only used location data collected after 1990. Since I received data from multiple sources, I cross-referenced all databases to remove any repeated sampling occasions. To help normalize for sampling effort I divided the number of hellbenders discovered at any given location by the number of sampling occasions. This average number of hellbenders per sampling event served as the dependent variable.

I was unable to obtain data on Missouri hellbenders so I excluded the state from the study. I used a GIS layer of USGS HUC 6 basins to delineate the range of the species by merging all HUC 6 polygons that contained a hellbender occurrence. Two HUC 6 units in the center of the range without occurrences were also included in the delineated range to maintain spatial continuity. Since a previous hellbender study indicated that physiological provinces may play an important role in the probability of hellbender occupancy (Jachowski *et al.*, 2016), I further delineated the range using a USGS GIS layer of physiographic divisions. Any previously

selected HUC 12 subwatersheds (the smallest available hydrological unit) that fell within a physiographic division that contained occurrences were included in the final extent of the model, resulting in the inclusion of five physiographic provinces (Fig. 2.2). The majority of the hellbender occurrences (Fig. 2.3a), locations (Fig. 2.3b), and sampling occasions (Fig. 2.3c) were from rivers in the Blue Ridge province, but the number of hellbenders captured per location per sampling occasion was relatively consistent throughout the range (Fig. 2.3d). I selected background points at the midpoint of each NHDPlus version 2 FlowLine segment within the selected extent that did not contain a hellbender occurrence (Engler *et al.*, 2004). Since down-weighted Poisson regression can require a large number of background points for likelihood estimates to stabilize (Renner, 2015), I used all 220,519 background points to fit the final models.

Predictor variables

I focused on three types of predictor data for my analysis including climate, stream, and land use variables. While climate variables are often used to model species distributions at the range-wide scale, they may be an inadequate descriptor of niche space (Fourcade *et al.*, 2018). I also focused on stream variables since hellbenders are fully aquatic and likely influenced by the stream environment. I incorporated land use variables since previous studies have found associations between riparian land use and hellbender occurrence (Quinn *et al.*, 2013; Jachowski *et al.*, 2016; Pugh *et al.*, 2016). I used the standard 19 bioclimatic variables derived from WorldClim version 1.4 to explain climatic variation across the range of the species (Table S2.1). Stream variation was described using attributes associated with the NHDPlus version 2 FlowLine features including stream level, stream order, upstream length, catchment area, upstream catchment area, maximum elevation, minimum elevation and slope (for a full description of the variables see Table S2.2). I used the National Land Cover Database 2011 raster layer to define

eight distinct land cover classes (Table S2.3). Since multicollinearity in predictors can be problematic for SDMs (Graham, 2003), I performed principal components analysis to reduce the correlation and dimensionality of each set of predictor variables. To perform the analysis for climate and stream variables, I projected 100,000 random points across the hellbender range (Fig.2.1). After joining them to the closest NHD flowline feature, I extracted stream and climate variables for each point. Land cover rasters were converted to polygons and clipped to HUC 12 subwatersheds. The proportion of each land cover class was calculated for each subwatershed. I performed each PCA with the *princomp* function using the variable correlation matrix.

The first five PCA axes explained approximately 95% of the variation in the climate data and were retained as independent variables (Table S2.4). I interpreted the climate principal component axes as measures of regional temperature and precipitation patterns (generally getting colder and dryer moving from the southeast to the northwest of the range), summer temperature and precipitation patterns, temperature seasonality, precipitation seasonality, and temperature range (Table S2.4). The first five stream PCA axes also explained approximately 95% of the variation in the stream data and were retained as independent variables (Table S2.5) I interpreted the stream principal component axes as measures of elevation, stream position (upstream vs. downstream), stream level, stream size, and slope (Table S2.5). The first two land cover PCA axes explained approximately 98% of the variation in the land cover data and were retained as independent variable (Table S2.6). The first PCA represents a contrast between agricultural land use and forest cover while the second represents the amount of developed land use (Table S2.6). All retained PCA axes were projected across the entire range of the species and extracted to background and occurrence points (S2.1, S2.2, S2.3). All PCA axes showed low correlations with each other and were retained for model development after being centered and scaled.

Model development

I used the down weighted Poisson regression technique presented in Renner et al. (2015) to fit all models. The weight of background points in each physiographic province was calculated as the area of the province divided by the number of background points within the province (Table S2.7). Since physiographic province has been demonstrated to affect baseline levels of hellbender occurrence (Jachowski *et al.*, 2016), it was included as a random affect in all models. I first fit the region-specific model with all variables and examined the Pearson's residuals visually using a semivariogram (Dale & Fortin, 2014) and then confirmed the presence of spatial autocorrelation using a Moran's I test. To account for this variation, I added an autocovariate to the model. Since the traditional inverse-weighted distance technique of autocovariate calculation can lead to overcorrection (Dormann, 2007; Dormann *et al.*, 2007), I used a local autocovariate. I examined the range of autocorrelation and found that it corresponded roughly with the area of the average HUC 12 subwatershed, so I included the number of other hellbender locations within a HUC 12 watershed as an autocovariate. I also extended the model to estimate a separate autocovariate slope for each physiographic province to see if it could further improve global and regional model performance.

I used the glmnet package (Friedman, 2010) to develop Lasso regularized models for variable selection. For each of the final three models, I used 5-fold cross validation and minimum BIC score to choose an optimal lambda value (degree of regularization) and ran the Lasso model with the optimized lambda. I retained any variables that did not shrink to zero. Lasso models were fit using an equal number of presence locations and random background points. Final models were fit using *glmer* function in the lme4 package (Bates *et al.*, 2014), variables retained from the Lasso model, and all background points. In all three final models,

only climate variable two (summer temperature and precipitation patterns) was dropped from the final model.

Model validation

I validated the three final models using several measures of model performance. Since the data displayed spatial autocorrelation, cross-fold validation could lead to an overly optimistic assessment of model results (Roberts *et al.*, 2017) so I used spatial-block validation to assess model performance. Spatial blocks need to be larger than the range of residual autocorrelation yet small enough to avoid extrapolation in environmental space (Roberts *et al.,* 2017). Since watershed boundaries are hierarchically nested, and I knew autocorrelation in my model extended to the HUC 12 watershed level, I used HUC 8 subbasins (n=143) as spatial blocking units. Following the recommendation in Roberts *et al.,* 2017, I used each spatial blocking unit as an individual fold. I examined global model performance as well as regional model performance, which I averaged across all regions for comparison (individual regional results are available in table S2.8). The coastal plain province was not included in any regional analyses since it only contained three hellbender occurrences. After transforming model outputs into relative probabilities, I calculated area under the receiver operating characteristic curve as well as area under the precision-recall curve since it can be a more informative metric in imbalanced data sets (Saito & Rehmsmeier, 2015) using the PRROC package (Grau *et al.*, 2015). I also developed calibration plots and recorded the slope and R-squared values (which are indicative of model fit; Steyerberg *et al.*, 2010) and the intercept (which is indicative of bias). In addition, I used package ape to calculate the Moran's I statistic for the Pearson's residuals of the global models.

2.4 Results

A region-specific autocovariate improved model discrimination and calibration both regionally and globally; however, it resulted in increased model bias. Global AUC-ROC values were high for all three final models (ranging from 0.93 – 0.98), but regional means improved with the addition of the autocovariate (increasing 0.86 ± 0.056 to 0.93 ± 0.017) and again with addition of the region-specific autocovariate (0.97 ± 0.011; Fig. 2.4a). The pattern was similar in AUC-PR tests with values improving with the addition of the autocovariate (from 0.33 to 0.64 globally and from 0.16 ± 0.09 to 0.38 ± 0.12 regionally) and improving further with the addition of the region-specific autocovariate (0.69 globally and 0.56 ± 0.17 regionally Fig. 2.4b). Calibration slopes (Fig. 2.4c) and r-squared values (Fig. 2.4d) improved extensively with the addition of the autocovariate, but showed little additional improvement due to the addition of the region-specific autocovariate. The tradeoff for improved discrimination and calibration was increased negative bias. The calibration intercept dropped below zero both globally and regionally with the addition of the autocovariate, but it recovered slightly in the region-specific autocovariate model (Fig. 2.4e). All global and regional calibration slopes were lower than one (with a global mean of 0.71 ± 0.16 and a regional mean of 0.65 ± 0.10), suggesting under prediction in my models. The Moran's I value was high in the spatially invariant and autocovariate model (0.29 ± 0.01 and 0.27 ± 0.01 respectively), but was reduced by nearly half with the addition of nonstationarity (0.15 ±.0.01).

Unsurprisingly, the highest probabilities of the fitted model occur in the Blue Ridge province (Fig. 2.5). More disconcerting is the stark contrast of the adjacent Valley and Ridge province. This, along with the overall low discriminatory power within the Valley and Ridge province in all final models (with a mean AUC-ROC of 0.82 ± 0.07 and a mean AUC-PR of 0.10 ± 0.06; Table S1.8), suggest that there is insufficient data to adequately predict occurrences in

this region. In the random effect terms, intercepts were similarly low for all physiographic provinces (ranging from -8.76 to -6.3) except for the Blue Ridge province (ranging from -4.3 to -1.9) where it was approximately half as low as the other regions (Fig. 2.6). The slope of the autocovariate was lowest in the Blue Ridge (0.232) and Valley and Ridge (0.45) provinces, intermediate in the Coastal Plains (2.67) and Appalachian Plateaus (2.91), and highest in the Interior Low Plateaus (5.9; Fig. 2.6).

Fixed coefficient values largely agreed in the three final hellbender SDMs (Fig. 2.7). There was little consistent effect of regional climate patterns (climate variable one; 0.03 ± 0.22) and summer climate patterns were dropped from the final model. However, climate variation was more important with axes of temperature variability (climate variable three; 0.19 ± 0.11), precipitation variability (climate variable four; ± 0.13 ± 0.07), and temperature range (climate variable five; 0.22 ± 0.06) showing consistent patterns with hellbender occurrences. Stream variables had a stronger relationship with hellbender occurrences than climate variables (Fig. 2.6). Stream variable two had the strongest effect (5.14 ± 0.67) in every final model with a positive response to reduced upstream area. Stream variable three was also very influential (2.66 ± 0.28) and demonstrated a positive response to lower stream levels. I saw smaller positive responses to higher elevations (1.88 ± 0.60; stream variable 1), reduced catchment area (1.00 ± 0.11; stream variable four), and shallower slopes (1.01 ± 0.23; stream variable five). Land use effects were relatively small, but consistent, with higher levels of agriculture causing a negative response (-0.72 ±, 0.18) and lower levels of development causing a positive response (0.60 ±, 0.22).

2.5 Discussion

Accounting for inherent autocorrelation using a region-specific autocovariate improved model discrimination both globally and regionally, though it was associated with an increase in negative bias. While several studies have demonstrated improved model performance by accounting for spatial autocorrelation (Tognelli & Kelt, 2004; Bahn *et al.*, 2006) this is the first to demonstrate additional improvement by accounting for nonstationary in the autocovariate. Importantly, I did not see a decrease in residual autocorrelation until the region-specific slope was added to the model, making it a more effective method for accounting for autocorrelation than the use of an autocovariate alone. Results will likely vary depending on the nature, strength, and spatial variability of autocorrelation in a particular data set. Since the slopes of the autocovariate were highest in regions with low abundances and heavy sampling, I suspect that the autocorrelation in hellbender occurrences was related to sampling bias. As such, this method may represent a way to account for differential levels of sampling without resorting to spatial filtering (Boria *et al.*, 2014) or having to measure bias covariates (El-Gabbas & Dormann, 2017). While I used varying slopes strictly to account for regional-differences in the autocovariate, this approach could also be used to account for shifting species-environment relationships throughout the sampling range, which may also help improve model performance (Osborne *et al*., 2007). Unlike geographically weighted regression (Miller, 2012), predictions may be extrapolated outside of the modeling extent by predicting with the mean slope value.

The regional approach to modeling range-wide hellbender distribution data allowed me to account for baseline differences in occurrences across physiographic provinces. As with Jachowski et al. (2016), I found a much higher baseline level of hellbender occurrence in the Blue Ridge province compared to all other physiographic provinces. The Blue Ridge is the only province in the hellbender range that contains primarily metamorphic instead of sedimentary

bedrock. This underlying divergence may drive differences in water quality or habitat availability. The availability of shelter and nest rocks appears to limit hellbender densities (Nickerson & Mays, 1973) and geological variables were the most important predictors in a hellbender SDM of the Northeastern portion of the range (Quinn *et al*., 2013). Alternatively, land use legacy effects can be stronger predictors of stream impairment than current land use practices (Maloney *et al*., 2008; Surasinghe & Baldwin, 2014), and the Blue Ridge province was historically underutilized for agriculture compared to surrounding provinces (Price *et al*., 2006). The mixed modeling approach proved to be an effective way to account for the unexplained heterogeneity associated with physiographic province, regardless of it source.

This approach also emphasized the need to test SDMs regionally, as well as globally, as improvements in global discrimination were moderate compared to improvements in regional discrimination. Furthermore, by utilizing multiple metrics for validation I was able to gain a deeper understanding of model performance than if I had relied on the AUC-ROC metric alone since it featured a relatively small change in overall performance between models. The ROC-PR curve offered greater insights in discrimination ability since, as with most SDMs, I was working with an unbalanced data set (Saito & Rehmsmeier, 2015). The regional assessment approach also diagnosed poor performance in the Valley and Ridge province, which will allow me to focus on further data collection for the region prior to the next round of model development

Climate variables showed the lowest amount of consistent predictive power in my models, with stream variables showing the highest amount, and land use variables showing an intermediate amount. While climate variables had lower coefficient values than stream variables in these models, it is important not to underestimate their importance. The relatively low explanatory power of climate variables in the models was likely influenced by the choice of

sampling extent. I used a limited extent to draw background points since I was interested in modeling the realized distribution of the species. Had I modeled the potential distribution of the species, I would have widened the selection window which often leads to increased influence of climate variables as predictors (Guisan & Thuiller, 2005). The climate variables that were consistently reliable predictors across the range of the species are important since temperature variability (Jones, 2007), precipitation variability (Pendergrass *et al*., 2017), and temperature range (Jones, 2007) are all expected to increase under climate change. These changes could further reduce the available habitat for an already imperiled species.

The strong predictive power of upstream area is common in lotic species, acting as the strongest predictor in over half of modeled European stream fish distributions (Buisson *et al*., 2008; Logez *et al*., 2012). Its strong influence makes sense as stream habitat changes dramatically and consistently along an upstream to downstream gradient (Vannote *et al*., 1980). Downstream portions of streams tend to be warmer and deeper (Allan & Castillo, 2007). Hellbenders have long been noted for their reliance on cool, shallow habitats (Nickerson & Mays, 1973). Sedimentation also tends to increase downstream (Allan & Castillo, 2007) and high sediment loads have been shown to decrease occurrence probabilities of hellbenders in multiple regional studies (Keitzer *et al*., 2013; Quinn *et al*., 2013; Pugh *et al*., 2016). Likewise, the positive response to elevation, reduced stream order, and catchment size reflects the needs of a habitat specialist. Hellbenders only occurred in headwaters and midreaches, supporting decades of observational reports.

The predictive power of the other two stream variables is more difficult to understand. While ostensibly the inverse of stream order, stream level it is not a straightforward measure of stream size. A stream level of one will apply to both the Mississippi River and every small

stream that empties into the Atlantic Ocean. Stream PCA 2 followed a southwest to northeast gradient (Fig S2.2c) and likely helped explain the strong distributional gradient of hellbender density that ran in a similar direction. The negative response to slope seems counterintuitive, but hellbenders were never found in stream segments with very steep slopes. Furthermore, pool-riffle channels, a habitat complex nearly ubiquitous at hellbender sites (Burgmeier *et al*., 2011; Bodinof *et al*., 2012), tend to occur in low to medium gradient systems (Allan & Castillo, 2007). It is also important to note that SDMs are models of where organisms have been reported (Renner *et al*., 2015). The predictive power of upstream area and slope may be partially driven by the difficulty of sampling in deep streams and streams with steep slopes. In this context, this is an acceptable bias since the purpose of my model is to guide sampling and conservation efforts.

The negative effect of agricultural and developed land use on hellbenders is not unexpected, as several previous studies have found associations between hellbender occupancy and land use (Quinn *et al*., 2013; Jachowski *et al*., 2016; Pugh *et al*., 2016). Agricultural land use can impact streams through inputs of sediment and non-point source pollution (Allan, 2004; Dudgeon *et al*., 2006). Impervious surfaces, which are associated with developed land use, cause decreased infiltration and increased runoff in watersheds (Paul & Meyer, 2001). In associated streams, natural hydrologic variability is altered leading to increased flood magnitude and frequency, lowered base flow, and heightened erosion (Allan, 2004). However, the similar effect sizes of both land use variables suggests that the important factor may not be the type of land use, but the removal of forested areas within the catchment. Recent studies have found strong links between catchment level riparian removal, increased conductivity, and reduced hellbender occupancy (Keitzer *et al*., 2013; Pitt *et al*., 2017) and recruitment (Jachowski & Hopkins, 2018).

Regardless of the mechanism, this result stresses the importance of maintaining riparian forest cover for hellbender conservation.

The regional approach to modeling range-wide species distribution data presented herein helped account for sampling bias in occurrences and improved model performance in areas most in need of hellbender conservation. The extension of presence-only models into a mixed modeling framework represents a flexible approach to address unexplained environmental heterogeneity and nonstationarity in modeled relationships. This study demonstrates the relative ease of accounting for spatial complications when presence-only models are fit using generalized linear models instead of machine learning techniques.

2.6 Literature Cited

Allan, J.D. (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecolology, Evolution and Systematics, 35, 257-284.

Allan, J.D. & Castillo, M.M. (2007) Stream ecology: structure and function of running waters. Springer Science & Business Media, Netherlands.

Bahn, V., J O'Connor, R. & B Krohn, W. (2006) Importance of spatial autocorrelation in modeling bird distributions at a continental scale. Ecography, 29, 835-844.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014) lme4: Linear mixed-effects models using Eigen and S4.

Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araújo, M.B., Baselga, A. & Beck, J. (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography, 32, 193-204.

Bodinof, C.M., Briggler, J.T., Junge, R.E., Beringer, J., Wanner, M.D., Schuette, C.D., Ettling, J. & Millspaugh, J.J. (2012) Habitat attributes associated with short-term settlement of Ozark hellbender (*Cryptobranchus alleganiensis bishopi*) salamanders following translocation to the wild. Freshwater Biology, 57, 178-192.

Bombi, P., Luiselli, L. & D'Amen, M. (2011) When the method for mapping species matters: defining priority areas for conservation of African freshwater turtles. Diversity and Distributions, 17, 581-592.

Boria, R.A., Olson, L.E., Goodman, & and Anderson, R.P. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling, 275, 73-77.

Bothner, R.C. & Gottlieb, J.A. (1991) A study of the New York State populations of the hellbender, *Cryptobranchus alleganiensis alleganiensis.* Proceedings of the Rochester Academy of Science 17, 41-54.

Buisson, L., Blanc, L. & Grenouillet, G. (2008) Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. Ecology of Freshwater Fish, 17, 244-257.

Burgmeier, N.G., Unger, S.D., Sutton, T.M. & Williams, R.N. (2011) Population status of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) in Indiana. Journal of Herpetology, 45, 195-201.

Dale, M.R.T. & Fortin, M.J. (2014) Spatial analysis: a guide for ecologists. Cambridge University Press.

Dormann, C.F. (2007) Assessing the validity of autologistic regression. Ecological Modelling, 207, 234-242.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W. & Kissling, D.W. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30, 609-628.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D. & Stiassny, M.L.J. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews, 81, 163-182.

Dyderski, M.K., Paź, S., Frelich, L.E. & Jagodziński, A.M. (2018) How much does climate change threaten European forest tree species distributions? Global Change Biology, 24, 1150-1163.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R. & Lehmann, A. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography, 129-151.

El-Gabbas, A. and Dormann, C.F. (2017) Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. Ecography. https://doi.org/10.1111/ecog.03149

El-Gabbas, A. and Dormann, C.F. (2018) Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. Ecology and Evolution, 8, 2196-2206.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology, 41, 263-274.

Esri (2014) ArcGIS Desktop. Release 10.2.2. Redlands CA: Environmental Systems Research Institute.

Fithian, W. & Hastie, T. (2013) Finite-sample equivalence in statistical models for presence-only data. The Annals of Applied Statistics, 7, 1917-1939.

Foster, R.L., McMillan, A.M. & Roblee, K.J. (2009) Population status of hellbender salamanders (*Cryptobranchus alleganiensis*) in the Allegheny River drainage of New York State. Journal of Herpetology, 43, 579-588.

Friedman, J., Hastie, T. & Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33, 1-22.

Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography, 27, 245-256.

Gould, R.W. (2011) Inclusion of the hellbender, including the eastern hellbender and the Ozark hellbender, in appendix III of the convention on international trade in endangered species of wild fauna and flora (CITES). In: (Fed Register), pp. 61978-61985

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution, 19, 497-503.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. Ecology, 84, 2809-2815.

Grau, J., Grosse, I. & Keilwagen, J. (2015) PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics, 31, 2595-2597.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. Ecology letters, 8, 993-1009.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006) Using niche-based models to improve the sampling of rare species. Conservation Biology, 20, 501-511.

Hamil, K.-A.D., Iannone Iii, B.V., Huang, W.K., Fei, S. & Zhang, H. (2016) Cross-scale contradictions in ecological relationships. Landscape Ecology, 31, 7-18.

Hendricks, S.A., Clee, P.R.S., Harrigan, R.J., Pollinger, J.P., Freedman, A.H., Callas, R., Figura, P.J. & Wayne, R.K. (2016) Re-defining historical geographic range in species with sparse records: implications for the Mexican wolf reintroduction program. Biological Conservation, 194, 48-57.

Jachowski, B., Catherine, M., Millspaugh, J.J. & Hopkins, W.A. (2016) Current land use is a poor predictor of hellbender occurrence: why assumptions matter when predicting distributions of data-deficient species. Diversity and Distributions, 22, 865-880.

Jachowski, C.M.B. & Hopkins, W.A. (2018) Loss of catchment-wide riparian forest cover is associated with reduced recruitment in a long-lived amphibian. Biological Conservation, 220, 215-227.

Jones, P.D., Trenberth, K.E., Ambenje, P., Bojariu, R., Easterling, D., Klein, T., Parker, D., Renwick, J., Rusticucci, M. and Soden, B., 2007. Observations: surface and atmospheric climate change. In: S.D. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor & H.L. Miller (Eds.) *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change* (pp.235-336). Cambridge University Press, Cambridge, United Kingdom.

Junker, J., Blake, S., Boesch, C., Campbell, G., Toit, L.d., Duvall, C., Ekobo, A., Etoga, G., Galat-Luong, A. & Gamys, J. (2012) Recent decline in suitable environmental conditions for African great apes. Diversity and Distributions, 18, 1077-1091.

Keitzer, S.C., Pauley, T.K. & Burcher, C.L. (2013) Stream characteristics associated with site occupancy by the eastern hellbender, *Cryptobranchus alleganiensis alleganiensis*, in southern West Virginia. Northeastern Naturalist, 20, 666-677.

Kühn, I. (2007) Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distributions, 13, 66-69.

Leathwick, J., Moilanen, A., Francis, M., Elith, J., Taylor, P., Julian, K., Hastie, T. & Duffy, C. (2008) Novel methods for the design and evaluation of marine protected areas in offshore waters. Conservation Letters, 1, 91-102.

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

Leung, B. & Steele, R.J. (2013) The value of a datum–how little data do I need for a quantitative risk analysis? Diversity and Distributions, 19, 617-628.

Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33, 103-114.

Logez, M., Bady, P. & Pont, D. (2012) Modelling the habitat requirement of riverine fish species at the European scale: sensitivity to temperature and precipitation and associated uncertainty. Ecology of Freshwater Fish, 21, 266-282.

Lozier, J.D., Aniello, P. & Hickerson, M.J. (2009) Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. Journal of Biogeography, 36, 1623-1627.

Maloney, K.O., Feminella, J.W., Mitchell, R.M., Miller, S.A., Mulholland, P.J. & Houser, J.N. (2008) Landuse legacies and small streams: identifying relationships between historical land use and contemporary stream conditions. Journal of the North American Benthological Society, 27, 280-294.

Mayasich, J., Grandmaison, D. & Phillips, C. (2003) Eastern hellbender status assessment report. In. USDA Forest Service, Duluth, MN.McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. Ecography, 30, 135-151.

Miller, J.A. (2012) Species distribution models: Spatial autocorrelation and non-stationarity. Progress in Physical Geography, 36, 681-692.

Miller, J.A. (2011) Spatial nonstaionarity and the scale of species-environment relationships in the Mojave Desert, California, USA. International Journal of of Geographical Informational Science, 3, 423-438.

Nickerson, M.A. & Mays, C.E. (1973) The hellbenders: North American" giant salamanders". Milwaukee Public Museum Press Milwaukee, Wi.

Ortega-Huerta, M.A. & Peterson, A.T. (2004) Modelling spatial patterns of biodiversity for conservation prioritization in north-eastern Mexico. Diversity and Distributions, 10, 39-54.

Osborne, P.E. & Suárez-Seoane, S. (2002) Should data be partitioned spatially before building large-scale distribution models? Ecological Modelling, 157, 249-259.

Osborne, P.E., Foody, G.M. & Suárez-Seoane, S. (2007) Non-stationarity and local approaches to modelling the distributions of wildlife. Diversity and Distributions, 13, 313-323.

Paul, M.J. & Meyer, J.L. (2001) Streams in the urban landscape. Annual review of Ecology and Systematics, 32, 333-365.

Pendergrass, A.G., Knutti, R., Lehner, F., Deser, C. & Sanderson, B.M. (2017) Precipitation variability increases in a warmer climate. Scientific Reports, 7, 17966.

Pitt, A.L., Shinskie, J.L., Tavano, J.J., Hartzell, S.M., Delahunty, T. & Spear, S.F. (2017) Decline of a giant salamander assessed with historical records, environmental DNA and multi-scale habitat data. Freshwater Biology, 62, 967-976.

Price, S.J., Dorcas, M.E., Gallant, A.L., Klaver, R.W. & Willson, J.D. (2006) Three decades of urbanization: estimating the impact of land-cover change on stream salamander populations. Biological Conservation, 133, 436-441.

Pugh, M.W., Hutchins, M., Madritch, M., Siefferman, L. & Gangloff, M.M. (2016) Land-use and local physical and chemical habitat parameters predict site occupancy by hellbender salamanders. Hydrobiologia, 770, 105-116.

Quinn, S.A., Gibbs, J.P., Hall, M.H. & Petokas, P.J. (2013) Multiscale factors influencing distribution of the eastern hellbender salamander (*Cryptobranchus alleganiensis alleganiensis*) in the northern segment of its range. Journal of Herpetology, 47, 78-84.

Radinger, J., Essl, F., Hölker, F., Horký, P., Slavík, O. & Wolter, C. (2017) The future distribution of river fish: the complex interplay of climate and land use changes, species dispersal and movement barriers. Global Change Biology,

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G. & Warton, D.I. (2015) Point process models for presence-only analysis. Methods in Ecology and Evolution, 6, 366-379.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G …Warton, D.I. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40, 913-929.

Saito, T. & Rehmsmeier, M. (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10, e0118432.

Schadt, S., Revilla, E., Wiegand, T., Knauer, F., Kaczensky, P., Breitenmoser, U., Bufka, L., Červený, J., Koubek, P. & Huber, T. (2002) Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. Journal of Applied Ecology, 39, 189-203.

Segurado, P., Araujo, M.B. & Kunin, W.E. (2006) Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology, 43, 433-444.

Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J. & Kattan, M.W. (2010) Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 21, 128.

Surasinghe, T. & Baldwin, R.F. (2014) Ghost of land-use past in the context of current land cover: evidence from salamander communities in streams of Blue Ridge and Piedmont ecoregions. Canadian Journal of Zoology, 92, 527-536.

Tognelli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. Ecography, 27, 427-436.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences, 37, 130-137.

Wheeler, B.A., Prosen, E., Mathis, A. & Wilkinson, R.F. (2003) A case of amphibian decline in a long-lived salamander: a long term study of hellbender populations. Biological Conservation, 109, 151-156.

Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009) Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions, 15, 565-576.

2.7 Figures

Figure 2-1. The range of the eastern (*Cryptobranchus alleganiensis alleganiensis*) and Ozark (*Cryptobranchus alleganiensis bishopi*) hellbender subspecies.

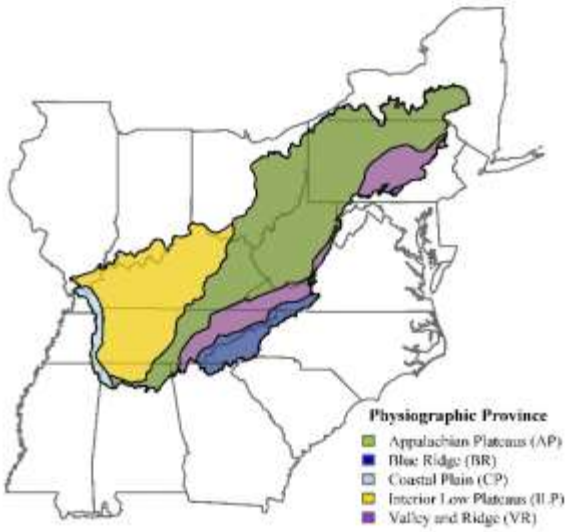

Figure 2-2. The division of physiographic provinces across the modeled extent of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) range.

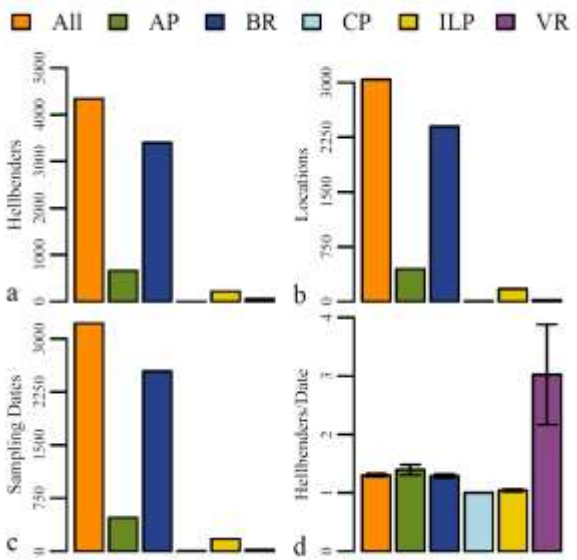

Figure 2-3. Summaries of the modeled data by physiographic province across the study range, including the number of captured *Cryptobranchus alleganiensis* (a), number of sampling locations (b), number of sampling occasions (c), and the mean number of hellbenders captured per sampling occasion (d).

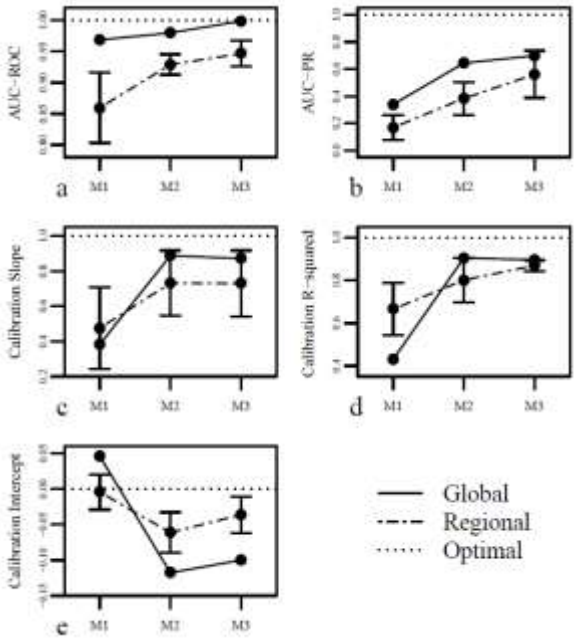

Figure 2-4. Metrics of model performance including area under the ROC curve (a), area under the PR curve (b), calibration plot slope (c), calibration plot R-squared (d), and calibration plot intercept (e) for the global dataset and averaged across all regions for three final *Cryptobranchus alleganiensis* species distribution models.

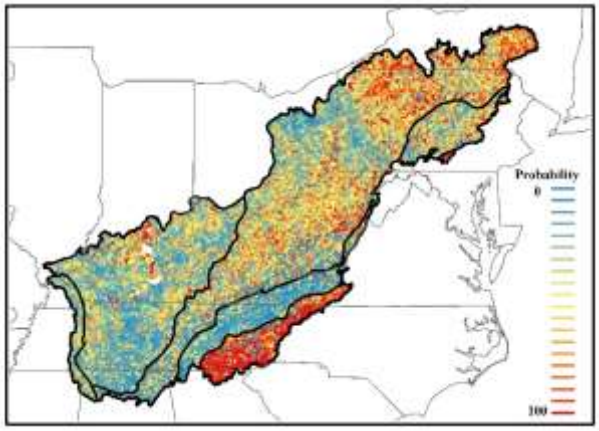

Figure 2-5. Final SDM model relative occurrence probabilities stretched using quantile classification with 20 classes.

‘

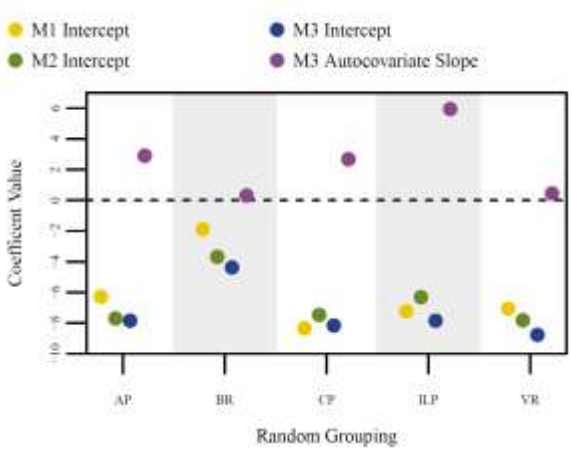

Figure 2-6. Summaries of the random effects in three final *Cryptobranchus alleganiensis* species distribution models including one with a region-specific intercept (M1), one with a region-specific intercept and an autocovariate (M2), and one with a region-specific intercept and a region-specific autocovariate (M3).

.

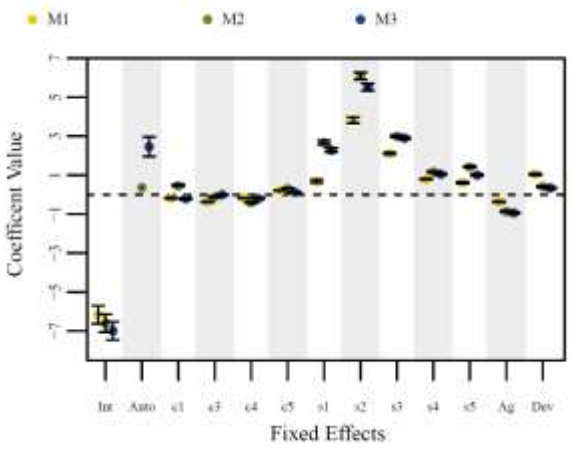

Figure 2-7. Summaries of the fixed effects in three final *Cryptobranchus alleganiensis* species distribution models including one with a region-specific intercept (M1), one with a region-specific intercept and an autocovariate (M2), and one with a region-specific intercept and a region-specific autocovariate (M3).

APPENDIX 1. CHAPTER 1 SUPPLEMENTAL MATERIAL

Table S1.1. Variables used to derive principal component axes describing stream variation across the hellbender range. All listed variables are attributes of NHDFlowline features (stream segments) in the NHDPlus Version 2 dataset.

Variable	Description
Stream Level	Reverse of stream order. Lower values represent mainstem flow lines and higher values represent tributaries.
Stream Order	Modified Strahler stream order. Headwaters receive a value of 1 and all major stream divergences add 1 to the previous value.
Upstream Length	The length (km) of all upstream portions from the downstream end of the flow line.
Catchment Area	The catchment area (km^2) of the flow line.
Upstream Catchment Area	The cumulative drainage area (km^2) at the downstream end of the flow line.
Maximum Elevation	Maximum smoothed elevation (cm) within the flow line.
Minimum Elevation	Minimum smoothed elevation (cm) within the flow line.
Slope	Slope of flow line (m/m) based on smoothed elevations

Table S1.2. Variables used to derive principal component axes describing climatic variation across the hellbender range. Bioclimatic variables were derived from WorldClim version 1.4.

Variable	Description
Annual Temperature	Annual mean temperature (°C). Derived from minimum temperature (°C) and maximum temperature (°C)
Mean Diurnal Range	Mean (24-hour period max-min) (°C). Derived from minimum temperature (°C) and maximum temperature (°C)
Isothermality	Mean Diurnal Range – Temperature Annual Range. Derived from minimum temperature (°C) and maximum temperature (°C)
Temperature Seasonality	Coefficient of variability of annual mean temperature. Derived from minimum temperature (°C) and maximum temperature (°C)
Maximum Temperature Warmest Month	Derived from maximum temperature (°C).
Minimum Temperature Coldest Month	Derived from minimum temperature (°C).
Temperature Annual Range	Maximum Temperature of Warmest Month – Minimum Temperature of Coldest Month
Mean Temperature Wettest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Mean Temperature Driest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Mean Temperature Warmest Quarter	Derived from minimum temperature (°C) and maximum temperature (°C)
Mean Temperature Coldest Quarter	Derived from minimum temperature (°C) and maximum temperature (°C)
Annual Precipitation	Derived from rainfall (mm/month)
Precipitation Wettest Month	Derived from rainfall (mm/month)
Precipitation Driest Month	Derived from rainfall (mm/month)
Precipitation Seasonality	Coefficient of variability of Annual Precipitation
Precipitation Wettest Quarter	Derived from rainfall (mm/month)
Precipitation Driest Quarter	Derived from rainfall (mm/month)
Precipitation Warmest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Precipitation Coldest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)

Table S1.3. Explained variance and variable loadings on the first five principal component axes describing stream variation across the hellbender range. Bold values represent high loadings that were used to interpret the axes.

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Variance Explained	0.363	0.250	0.138	0.111	0.082
Stream Level	0.297		**-0.533**	-0.432	**0.558**
Stream Order	-0.368	-0.207	0.312	0.389	0.405
Upstream Length	-0.344	**-0.530**	-0.243	-0.156	
Catchment Area	-0.138		**0.605**	**-0.774**	
Upstream Catchment Area	-0.342	**-0.530**	-0.246	-0.16	
Maximum Elevation	**0.451**	-0.402	0.238		0.107
Minimum Elevation	**0.413**	-0.386	0.267	0.108	0.339
Slope	0.385	-0.288			**-0.627**

Table S1.4. Explained variance and variable loadings on the first five principal component axes describing climate variation across the hellbender range. Bold values represent high loadings that were used to interpret the axes.

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Variance Explained	0.582	0.214	0.074	0.050	0.031
Annual Temperature	-0.226	**0.317**		0.121	
Mean Diurnal Range	-0.167	0.193	**0.397**	**-0.516**	0.115
Isothermality	-0.245		**0.381**	**-0.318**	-0.191
Temperature Seasonality	**0.252**	0.159	**-0.260**	0.101	**0.364**
Maximum Temperature Warmest Month	-0.138	**0.433**			0.167
Minimum Temperature Coldest Month	**-0.261**	0.199		0.205	-0.181
Temperature Annual Range	0.198	0.283	-0.112	-0.207	**0.485**
Mean Temperature Wettest Quarter	0.149		**0.578**	0.139	**0.416**
Mean Temperature Driest Quarter	**-0.254**	0.141		0.241	
Mean Temperature Warmest Quarter	-0.167	**0.398**		0.171	0.118
Mean Temperature of Coldest Quarter	**-0.262**	0.223	0.106		-0.106
Annual Precipitation	**-0.28**	-0.109	-0.182		0.168
Precipitation Wettest Month	**-0.268**		-0.231	-0.282	
Precipitaiton Driest Month	**-0.253**	-0.232		0.119	0.122
Precipitation Seasonality	0.139	0.274	-0.283	**-0.483**	-0.177
Precipitation Wettest Quarter	**-0.272**		-0.264	-0.175	0.150
Precipitation Driest Quarter	**-0.270**	-0.180			0.238
Precipitation Warmest Quarter	-0.162	**-0.344**		-0.181	**0.403**
Precipitation Coldest Quarter	**-0.287**		-0.146		

Table S1.5. The number of loci associated with each environmental axis averaged over all tested values of K (11-15). Environmental associations were tested on major alleles using latent factor mixed model analysis.

Environmental Axis	Average # of Associated Loci	Standard Error
Stream 1 – Elevation	0	0
Stream 2 – Upstream Area	2.4	0.24
Stream 3 – Stream Size	5.6	1.63
Stream 4 – Catchment Area	4	1.79
Stream 5 – Stream Level and Gradient	0	0
Climate 1 – Temperature and Precipitation	0.2	0.2
Climate 2 – Summer Temperature and Precipitation	0	0
Climate 3 – Temperature Seasonality	2.8	0.28
Climate 4 – Precipitation Seasonality	0	0
Climate 5 - Temperature Range	3.4	0.24

Table S1.6. Global climate models ensembled for future climate projections. These are the most current projections used in the Fifth Assessment IPCC report downscaled to 30 second resolution and calibrated using WorldClim 1.4 current climate data as a baseline. I used projections for 2050 based on the representative concentration pathway 4.5 in which greenhouse gas emissions peak at 2040 and then decline.

Model	Source
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory
GISS-E2-R	NASA Goddard Institute for Space Studies
INMCM4	Institute for Numerical Mathematics
IPSL-CM5A-LR	Institut Pierre-Simon Laplace
NorESM1-M	Norwegian Climate Centre

Table S1.7. Coefficient estimates for isolation by distance and isolation by environment relationships in hellbenders across models accounting for increasingly greater amounts of genetic structure.

	No Structure	**Population Intercept**	**Subpopulation Intercept**	**Population Slope**	**Subpopulation Slope**
Intercept	0.34444	0.37090	0.37980	0.34510	0.35600
Distance	0.08955	0.05661	0.05705	0.07975	0.07081
Elevation	-0.00321	-0.00093	-0.00169	-0.00091	-0.00057
Stream position	0.00852	0.00126	0.00093	0.00210	0.00205
Stream level	-0.00101	0.00317	0.00270	0.00197	0.00211
Stream size	0.00171	-0.00079	-0.00181	-0.00081	-0.00214
Slope	0.00317	0.00071	-0.00049	0.00012	-0.00051
Regional climate	0.00348	-0.00240	-0.00215	0.00019	0.00052
Summer climate	-0.01017	0.00309	0.00466	0.00041	0.00090
Temperature seasonality	0.00669	0.00060	0.00089	-0.00121	-0.00153
Precipitation seasonality	-0.00049	0.00295	0.00196	0.00204	0.00149
Temperature range	-0.00138	-0.00119	-0.00254	-0.00234	-0.00271

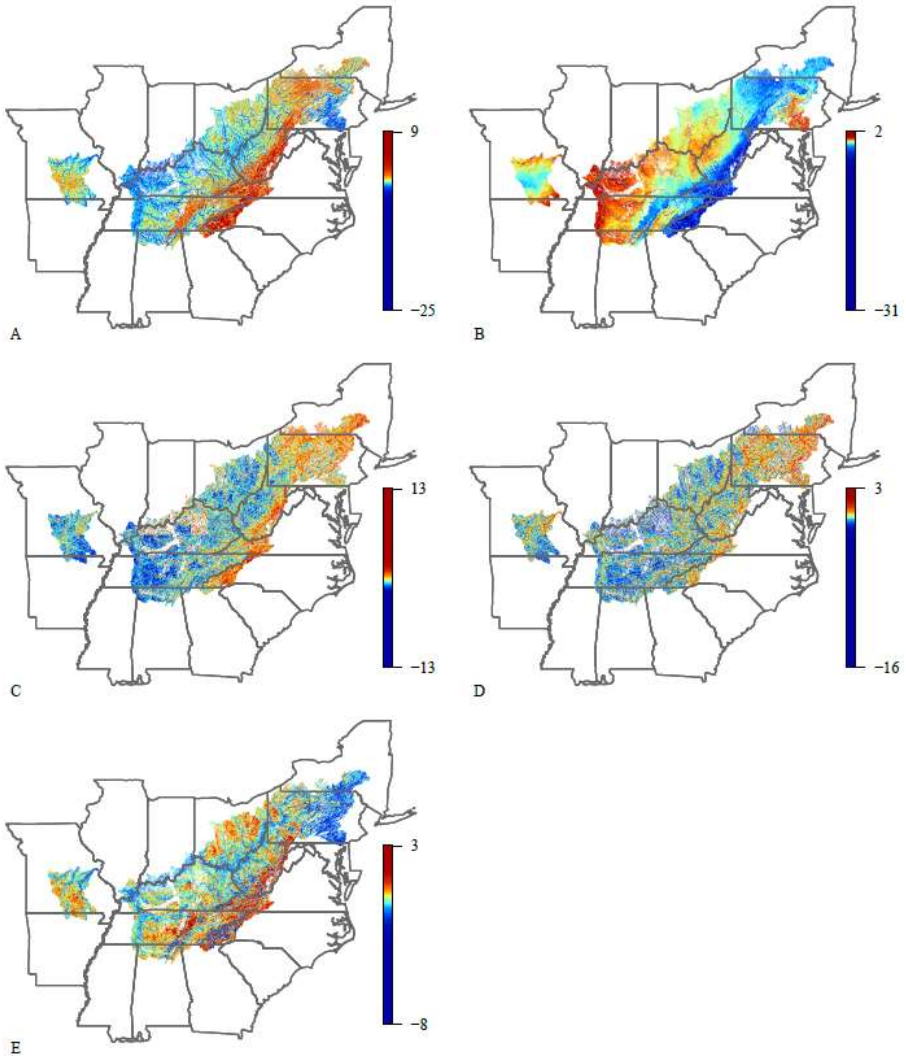

Figure S1.1. A visual representation of variation in the top five stream PCAs across the range of the hellbender. All color values are stretched using quantile classification with 100 classes. Stream PCA 1 (A) is associated with elevation with higher values representing higher elevations. Stream PCA 2 (B) is associated with upstream catchment area and upstream stream length. Higher values occur in stream segments with less upstream area. Stream PCA 3 (C) is associated with stream level and catchment area. Higher values occur in stream segments with smaller stream levels and larger catchment areas. Stream PCA 4 (D) is associated with catchment area with higher values representing stream segments with smaller catchment areas. Stream PCA 5 (E) is associated with stream level and slope. Higher values occur in stream segments with greater stream levels and flatter slopes.

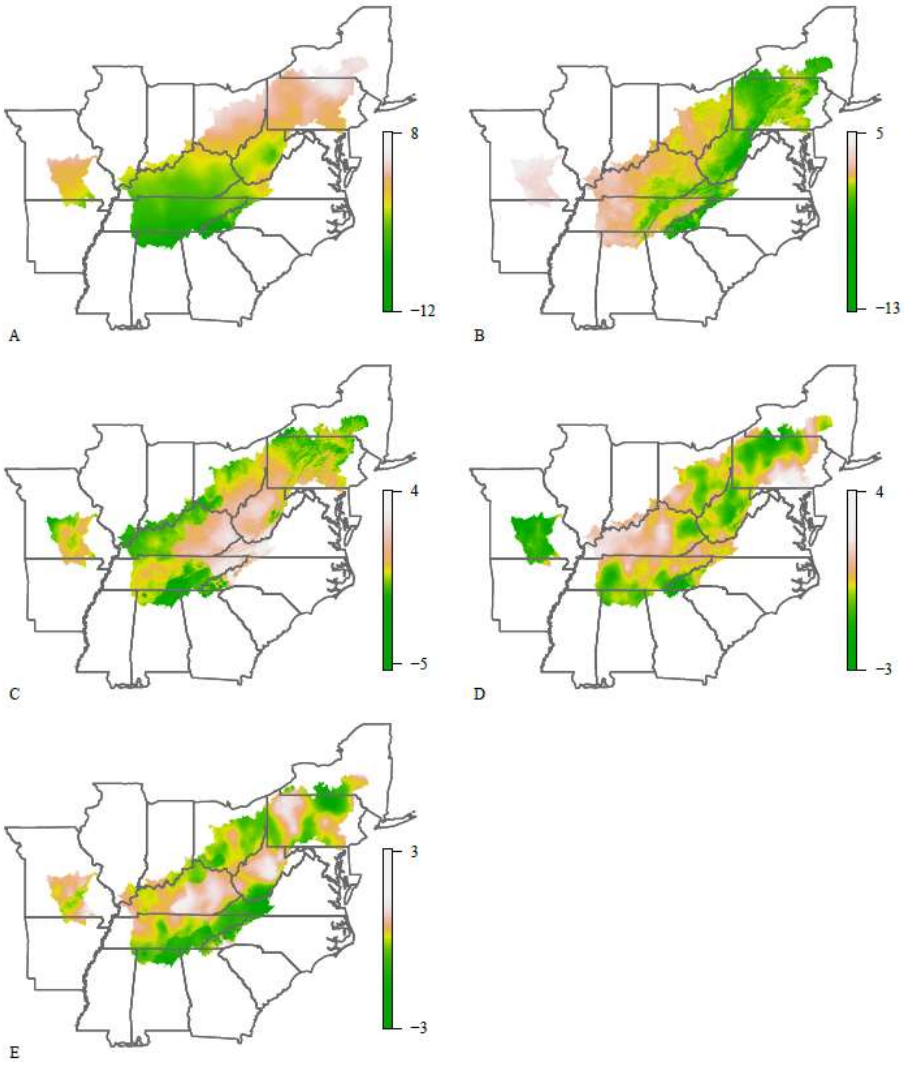

Figure S1.2. A visual representation of variation in the top five climate PCA axes across the range of the hellbender. All color values are stretched using quantile classification with 100 classes. Climate PCA 1 (A) is a general index of regional temperature and precipitation trends with both temperatures and precipitation decreasing in the higher value cells. Climate PCA 2 (B) is an index of summer climate patterns. Cells with high values have higher temperatures during the warmest periods of the year and lower precipitation during the warmest quarter. Climate PCA 3 (C) is an index of temperature variability. Cells with high values have higher daily temperature shifts, higher values of isothermality, and lower values of temperature seasonality. Cells with higher values also have higher mean temperatures during the wettest quarter. Climate PCA 4 (D) is an index of precipitation variability. Cells with higher values have lower values of precipitation seasonality and lower daily temperature shifts. Climate PCA 5 (E) is an index of temperature range. Cells with higher values have larger annual temperature ranges, greater temperature seasonality, higher mean temperatures in the wettest quarter, and higher precipitation in the warmest quarter.

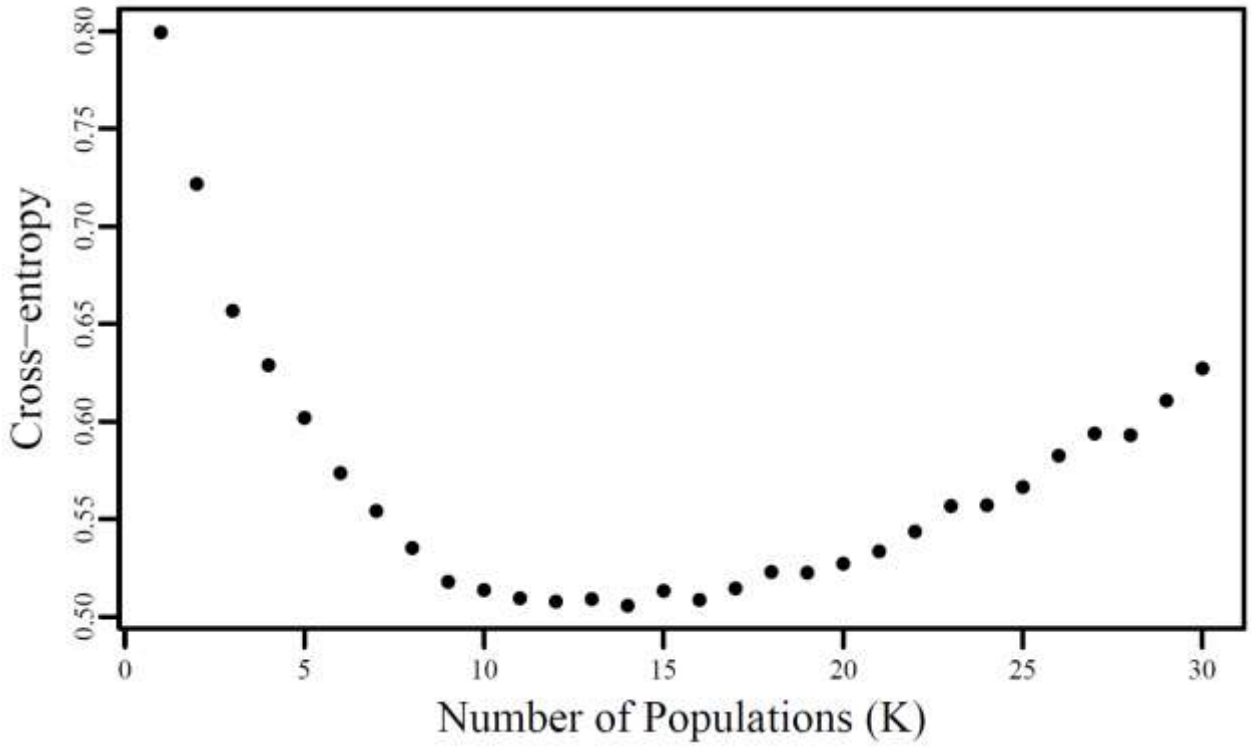

Figure S1.3. The cross-entropy value across five runs estimating ancestry coefficients at each value of K (1-30). The optimal value of K occurs when the cross-entropy is minimized. In the dataset, cross-entropy plateaus at K values of 11-15.

APPENDIX 2. CHAPTER 2 SUPPLEMENTAL MATERIAL

Table S2.1. Variables used to derive principal component axes describing climatic variation across the hellbender range. Bioclimatic variables were derived from WorldClim version 1.4.

Variable	Description
Annual Temperature	Annual mean temperature (°C). Derived from minimum temperature (°C) and maximum temperature (°C)
Mean Diurnal Range	Mean (24 hour period max-min) (°C). Derived from minimum temperature (°C) and maximum temperature (°C)
Isothermality	Mean Diurnal Range – Temperature Annual Range. Derived from minimum temperature (°C) and maximum temperature (°C)
Temperature Seasonality	Coefficient of variability of annual mean temperature. Derived from minimum temperature (°C) and maximum temperature (°C)
Maximum Temperature Warmest Month	Derived from maximum temperature (°C).
Minimum Temperature Coldest Month	Derived from minimum temperature (°C).
Temperature Annual Range	Maximum Temperature of Warmest Month – Minimum Temperature of Coldest Month
Mean Temperature Wettest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Mean Temperature Driest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Mean Temperature Warmest Quarter	Derived from minimum temperature (°C) and maximum temperature (°C)
Mean Temperature Coldest Quarter	Derived from minimum temperature (°C) and maximum temperature (°C)
Annual Precipitation	Derived from rainfall (mm/month)
Precipitation Wettest Month	Derived from rainfall (mm/month)
Precipitation Driest Month	Derived from rainfall (mm/month)
Precipitation Seasonality	Coefficient of variability of Annual Precipitation
Precipitation Wettest Quarter	Derived from rainfall (mm/month)
Precipitation Driest Quarter	Derived from rainfall (mm/month)
Precipitation Warmest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)
Precipitation Coldest Quarter	Derived from minimum temperature (°C), maximum temperature (°C), and rainfall (mm/month)

Table S2.2. Variables used to derive principal component axes describing stream variation across the hellbender range. All listed variables are attributes of NHDFlowline features (stream segments) in the NHDPlus Version 2 dataset.

Variable	Description
Stream Level	Reverse of stream order. Lower values represent mainstem flow lines and higher values represent tributaries.
Stream Order	Modified Strahler stream order. Headwaters receive a value of 1 and all major stream divergences add 1 to the previous value.
Upstream Length	The length (km) of all upstream portions from the downstream end of the flow line.
Catchment Area	The catchment area (km^2) of the flow line.
Upstream Catchment Area	The cumulative drainage area (km^2) at the downstream end of the flow line.
Maximum Elevation	Maximum smoothed elevation (cm) within the flow line.
Minimum Elevation	Minimum smoothed elevation (cm) within the flow line.
Slope	Slope of flow line (m/m) based on smoothed elevations

Table S2.3. Variables used to derive principal component axes describing land use variation within HUC 12 subwatersheds across the hellbender range. Proportions of land cover classes were calculated for each subwatershed using categories defined for the National Land Cover Database 2011 (NLDC2011) and provided by the Multi-Resolution Land Characteristics Consortium (MRLC).

Class	Raster Value	Description
Water	1	Includes open water and perennial ice and snow.
Developed	2	Includes all developed areas from open spaces to high intensity.
Barren	3	Includes any bare ground areas with less than 15% vegetation cover
Forest	4	Includes deciduous, evergreen, and mixed forests.
Shrubland	5	Includes dwarf scrub, scrub, and shrub dominated areas.
Herbaceous	7	Includes areas dominated by grasslands, sedges, lichens, and mosses.
Agriculture	8	Includes cultivated crops, hay crops, and pastures
Wetlands	9	Includes woody and emergent herbaceous wetlands

Table S2.4. Explained variance and variable loadings on the first five principal component axes describing climate variation across the hellbender range. Bold values represent high loadings that were used to interpret the axes.

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Variance Explained	0.582	0.214	0.074	0.050	0.031
Annual Temperature	-0.226	**0.317**		0.121	
Mean Diurnal Range	-0.167	0.193	**0.397**	**-0.516**	0.115
Isothermality	-0.245		**0.381**	**-0.318**	-0.191
Temperature Seasonality	**0.252**	0.159	**-0.260**	0.101	**0.364**
Maximum Temperature Warmest Month	-0.138	**0.433**			0.167
Minimum Temperature Coldest Month	**-0.261**	0.199		0.205	-0.181
Temperature Annual Range	0.198	0.283	-0.112	-0.207	**0.485**
Mean Temperature Wettest Quarter	0.149		**0.578**	0.139	**0.416**
Mean Temperature Driest Quarter	**-0.254**	0.141		0.241	
Mean Temperature Warmest Quarter	-0.167	**0.398**		0.171	0.118
Mean Temperature of Coldest Quarter	**-0.262**	0.223	0.106		-0.106
Annual Precipitation	**-0.28**	-0.109	-0.182		0.168
Precipitation Wettest Month	**-0.268**		-0.231	-0.282	
Precipitation Driest Month	**-0.253**	-0.232		0.119	0.122
Precipitation Seasonality	0.139	0.274	-0.283	**-0.483**	-0.177
Precipitation Wettest Quarter	**-0.272**		-0.264	-0.175	0.150
Precipitation Driest Quarter	**-0.270**	-0.180			0.238
Precipitation Warmest Quarter	-0.162	**-0.344**		-0.181	**0.403**
Precipitation Coldest Quarter	**-0.287**		-0.146		

Table S2.5. Explained variance and variable loadings on the first five principal component axes describing stream variation across the hellbender range. Bold values represent high loadings that were used to interpret the axes.

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Variance Explained	0.363	0.250	0.138	0.111	0.082
Stream Level	0.297		**-0.533**	-0.432	**0.558**
Stream Order	-0.368	-0.207	0.312	**0.389**	0.405
Upstream Length	-0.344	**-0.530**	-0.243	-0.156	
Catchment Area	-0.138		**0.605**	**-0.774**	
Upstream Catchment Area	-0.342	**-0.530**	-0.246	-0.16	
Maximum Elevation	**0.451**	-0.402	0.238		0.107
Minimum Elevation	**0.413**	-0.386	0.267	0.108	0.339
Slope	0.385	-0.288			**-0.627**

Table S2.6. Explained variance and variable loadings on the first two principal component axes describing land cover variation in HUC 12 subwatersheds across the hellbender range. Bold values represent high loadings that were used to interpret the axes.

	PCA 1	PCA 2
Variance Explained	0.910	0.0676
Water		-0.127
Developed		**-0.736**
Barren		
Forest	**-0.730**	**0.434**
Shrubland		
Herbaceous		
Agricultural	**0.683**	**0.501**
Wetlands		

Table S2.7. Summary of the background point model weight calculations for each physiographic province.

Physiographic Province	**Area (km^2)**	**Background Points**	**Weights**
Appalachian Plateaus	391318.3	113633	0.290
Blue Ridge	44453.0	15448	0.348
Coastal Plain	16131.5	6979	0.433
Interior Low Plateaus	208265.5	56180	0.270
Valley and Ridge	95275.2	28279	0.297

Table S2.8. Cross-validated model performance measures across the entire hellbender range and within specific physiographic provinces for the three final SDM models.

Model 1- Spatially Invariant	**AUC-ROC**	**AUC-PR**	**Calibration Intercept**	**Calibration Slope**	**Calibration R-squared**
All Provinces	0.969	0.340	0.046	0.767	0.430
Appalachian Plateaus(AP)	0.933	0.042	-0.019	0.406	0.679
Blue Ridge (BR)	0.873	0.388	0.060	0.368	0.414
Interior Low Plateaus (ILP)	0.963	0.242	-0.059	1.120	0.903
Valley and Ridge (VR)	0.700	0.001	0.001	0.004	1.000
Regional Mean	0.859	0.168	-0.004	0.475	0.665
Standard Error	0.113	0.090	0.024	0.233	0.122
Model 2- Autocovariate	**AUC-ROC**	**AUC-PR**	**Calibration Intercept**	**Calibration Slope**	**Calibration R-squared**
All Provinces	0.980	0.644	-0.118	0.888	0.871
Appalachian Plateaus(AP)	0.954	0.255	-0.112	0.893	0.566
Blue Ridge (BR)	0.938	0.741	-0.109	0.874	0.884
Interior Low Plateaus (ILP)	0.944	0.298	-0.021	0.980	0.957
Valley and Ridge (VR)	0.880	0.236	-0.003	0.180	1.000
Mean	0.929	0.383	-0.061	0.732	0.802
Standard Error	0.017	0.120	0.028	0.185	0.104
Model 3- Autocovariate and Nonstationary	**AUC-ROC**	**AUC-PR**	**Calibration Intercept**	**Calibration Slope**	**Calibration R-squared**
All Provinces	0.999	0.698	-0.100	0.871	0.897
Appalachian Plateaus(AP)	0.981	0.662	-0.003	0.897	0.899
Blue Ridge (BR)	0.934	0.746	-0.110	0.881	0.882
Interior Low Plateaus (ILP)	0.976	0.789	-0.032	0.973	0.829
Valley and Ridge (VR)	0.891	0.046	0.001	0.166	1.000
Mean	0.947	0.561	-0.036	0.729	0.869
Standard Error	0.021	0.174	0.026	0.189	0.026

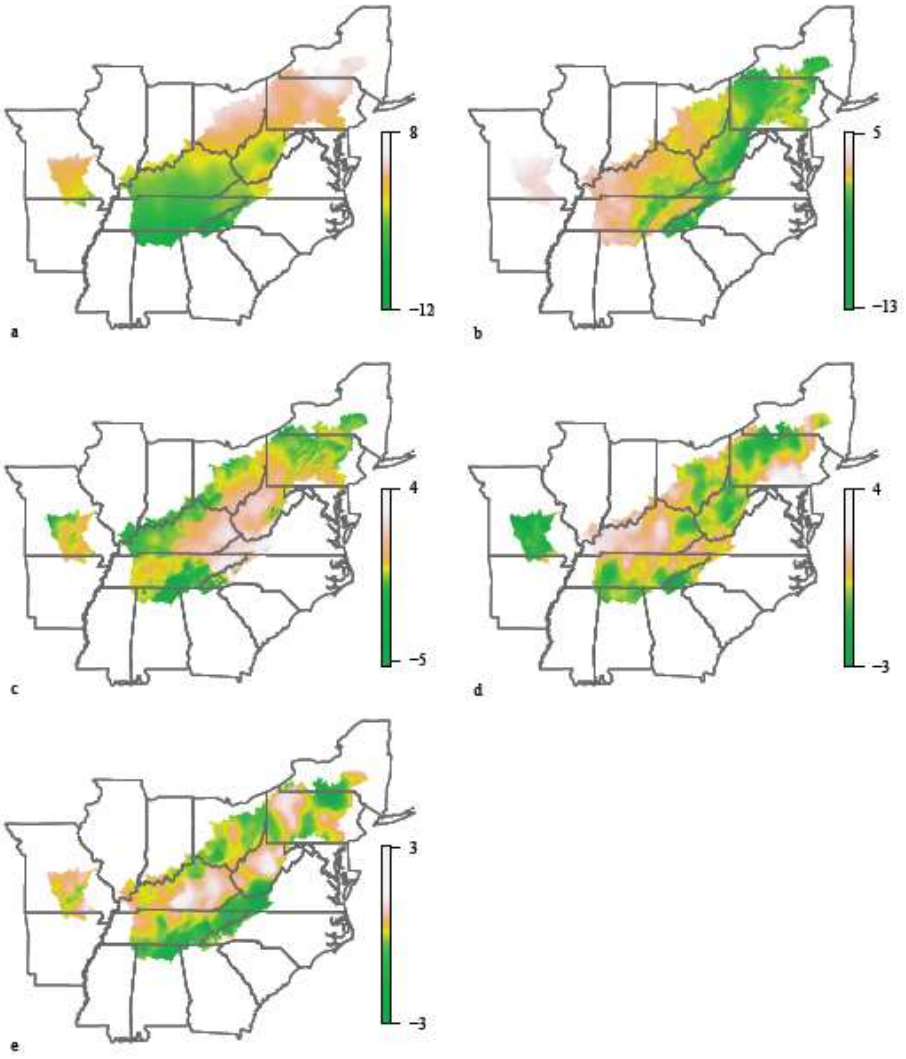

Figure S2.1. A visual representation of variation in the top five climate PCA axes across the range of the hellbender. All color values are stretched using quantile classification with 100 classes. Climate PCA 1 (a) is a general index of regional temperature and precipitation trends with both temperatures and precipitation decreasing in the higher value cells. Climate PCA 2 (b) is an index of summer climate patterns. Cells with high values have higher temperatures during the warmest periods of the year and lower precipitation during the warmest quarter. Climate PCA 3 (c) is an index of temperature variability. Cells with high values have higher daily temperature shifts, higher values of isothermality, and lower values of temperature seasonality. Cells with higher values also have higher mean temperatures during the wettest quarter. Climate PCA 4 (d) is an index of precipitation variability. Cells with higher values have lower values of precipitation seasonality and lower daily temperature shifts. Climate PCA 5 (e) is an index of temperature range. Cells with higher values have larger annual temperature ranges, greater temperature seasonality, higher mean temperatures in the wettest quarter, and higher precipitation in the warmest quarter.

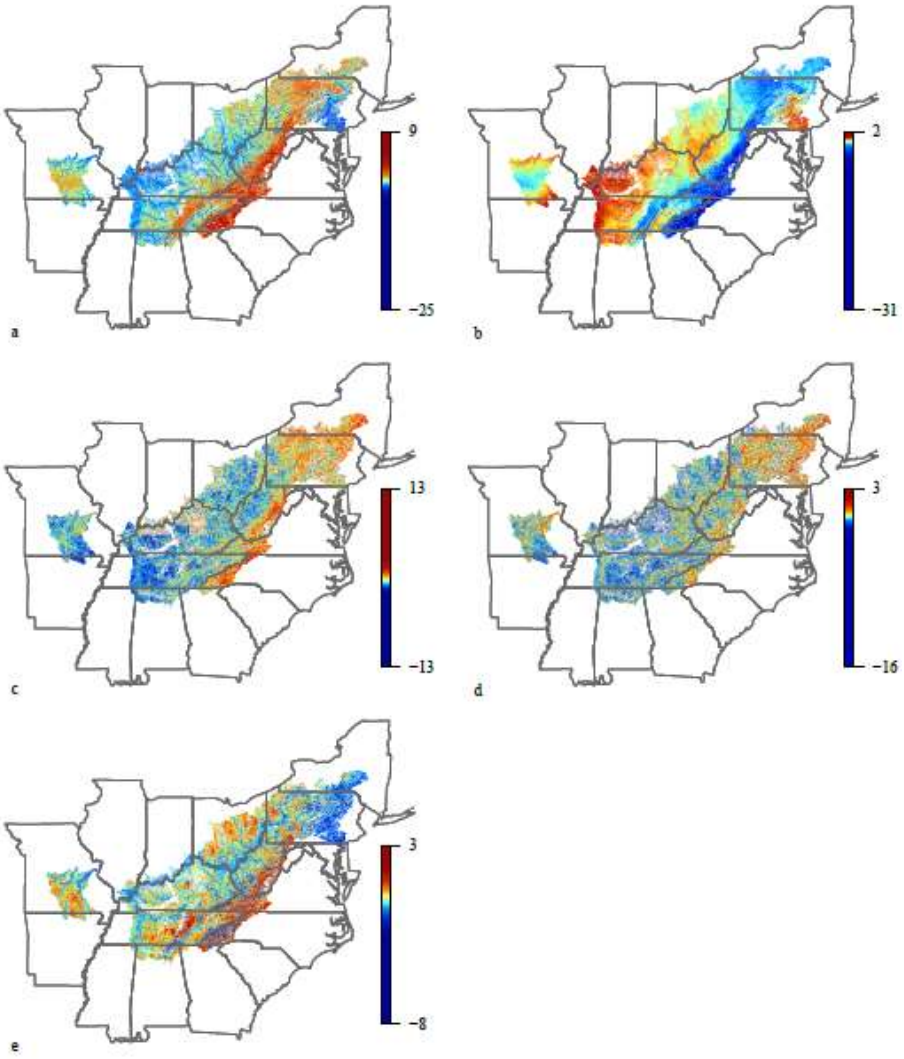

Figure S2.2. A visual representation of variation in the top five stream PCAs across the range of the hellbender. All color values are stretched using quantile classification with 100 classes. Stream PCA 1 (a) is associated with elevation with higher values representing higher elevations. Stream PCA 2 (b) is associated with upstream catchment area and upstream stream length. Higher values occur in stream segments with less upstream area. Stream PCA 3 (c) is associated with stream level and catchment area. Higher values occur in stream segments with smaller stream levels and larger catchment areas. Stream PCA 4 (d) is associated with catchment area with higher values representing stream segments with smaller catchment areas. Stream PCA 5 (e) is associated with stream level and slope. Higher values occur in stream segments with greater stream levels and flatter slopes.

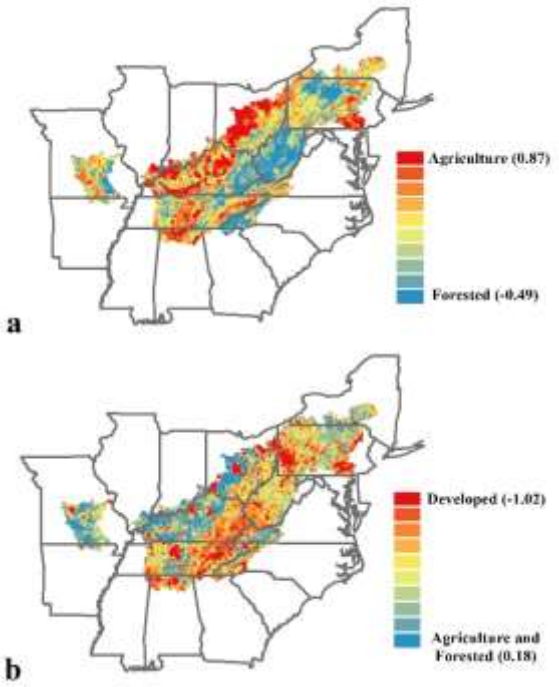

Figure S2.3. A visual representation of the variation in the top two land use PCA axes across the range of the hellbender. All color values are stretched using quantile classification with 10 classes. Land use PCA 1 (a) is a contrast between agricultural land use and forest cover. Land use PCA 2 (b) is a measure of development with negative values in more developed areas and positive values in areas with more agriculture or forested area.

VITA

Emily Boersma McCallen
715 West State Street
West Lafayette, Indiana 47907

I. Education

2018	Doctor of Science, Forestry and Natural Resources Purdue University, West Lafayette, Indiana Major Advisors: Rod N. Williams, Ph.D., Songlin Fei, Ph.D.
2013	Master of Science, Biology Eastern Illinois University, Charleston, Illinois Major Advisor: Karen F. Gaines, Ph.D.
2010	Bachelor of Science, *summa cum laude* Environmental Biology, Geographic Information Sciences Minor Eastern Illinois University, Charleston, Illinois

II. Professional Experience

08/2017	Graduate Teaching Co-instructor Introduction to R Programming Purdue University, West Lafayette, Indiana
08/2017	Graduate Teaching Co-instructor Introduction to Automated Content Analysis Purdue University, West Lafayette, Indiana
01/2017-03/2017	Graduate Teaching Laboratory Instructor Ecology and Systematics of Mammals Purdue University, West Lafayette, Indiana
01/2016-03/2016	Graduate Teaching Laboratory Instructor Ecology and Systematics of Amphibians and Reptiles Purdue University, West Lafayette, Indiana
03/2016	Guest Lecturer Ecology and Systematics of Amphibians and Reptiles Purdue University, West Lafayette, Indiana

01/2015-03/2015	Graduate Teaching Laboratory Instructor Ecology and Systematics of Amphibians and Reptiles Purdue University, West Lafayette, Indiana
01/2014-05/2014	Graduate Teaching Assistant Wildlife Investigational Techniques Purdue University, West Lafayette, Indiana
03/2014	Guest Lecturer Advanced Herpetology Purdue University, West Lafayette, Indiana
03/2014	Guest Lecturer Wildlife Investigational Techniques Purdue University, West Lafayette, Indiana
05/2013-08/2013	Graduate Research Assistant Spatial Ecology Purdue University, West Lafayette, Indiana
09/2011-12/2012	Graduate Teaching Laboratory Instructor Biological Principles and Issues Eastern Illinois University, Charleston, Illinois
01/2011-05/2011	Graduate Teaching Assistant Animal Diversity Eastern Illinois University, Charleston, Illinois
01/2011-05/2011	Graduate Teaching Assistant General Biology Eastern Illinois University, Charleston, Illinois

III. Publications

McCallen, E.B., J.A. Knott, G. Nunez-Mir, B. Taylor, J. Insu, and S. Fei. Trends in Ecology. *Frontiers in Ecology and the Environment*. Accepted.

McCallen, E.B., B.T. Kraus, S. Fei, and R.N. Williams. Movement and habitat use in eastern hellbenders (*Cryptobranchus alleganiensis alleganiensis*) following population augmentation. *Herpetologica*. Accepted.

McCallen, E.B., K.F Gaines, J.M. Novak, L.E. Ruyle, W.L. Stephens, A.L. Bryan, Jr., S.A. Blas, and. T.L. Serfass. The Development and Use of a Spatially Explicit Model for River Otters to Evaluate Environmental Hazards: A Case Study on the Department of Energy's Savannah River Site. *Environmental Monitoring and Assessment*. Accepted.

Kraus, B.T., **E.B. McCallen**, and R.N. Williams. 2017. Evaluating the survival of translocated adult and captive-reared, juvenile eastern hellbenders (*Cryptobranchus alleganiensis alleganiensis*). *Herpetologica* 73(4).

Sara, S.A., **E.B. McCallen,** and P.V. Switzer. 2013. The spatial distribution of Japanese beetles, *Popillia japonica*, in soybean fields. *Journal of Insect Science* 13.

IV. Presentations

McCallen, E.B., O. Hernandez-Gomez, N. Burgmeier, V. Yager, A. Wang, and R.N. Williams. A community approach to hellbender reintroduction site selection. Technical Advisory Committee Amphibians and Reptiles Meeting. Indianapolis, Indiana. October 2017.

McCallen, E.B., O. Hernandez-Gomez, N. Burgmeier, V. Yager, A. Wang, and R.N. Williams. A combined abiotic and biotic approach to hellbender reintroduction site selection at multiple scales. Indiana Hellbenders Partners Meeting. West Lafayette, Indiana. October 2017.

McCallen, E.B., B.T. Kraus, N.G. Burgmeier, S. Fei, and R.N. Williams. Movement and habitat use in eastern hellbenders following population augmentation. Joint Meeting of Ichthyologists and Herpetologists. Austin, TX. July 2017.

McCallen, E.B., B.T. Kraus, N.G. Burgmeier, S. Fei, and R.N. William. Movement and habitat use in eastern hellbenders following population augmentation. 8th Biennial Hellbender Symposium. Jackson, Mississippi. June 2017.

McCallen, E.B., and E.K. Kenison. The multiple phases of hellbender conservation in Indiana: research at work. The Wildlife Society North Central Section Student Conclave. Brookston, Indiana. March 2017.

McCallen, E.B., B.T. Kraus, N.G. Burgmeier, and R.N. Williams. The effects of management on hellbender movement and habitat use in Indiana. Indiana Hellbenders Partners Meeting. West Lafayette, Indiana. October 2016.

McCallen, E.B. Spatial modeling for hellbender conservation in Indiana flash talk. Purdue Forestry and Natural Resources Seminar Series. West Lafayette, Indiana. March 2016.

McCallen, E.B. Species distribution models for hellbenders in Indiana. Indiana Hellbenders Partners Meeting. West Lafayette, Indiana. November 2015.

McCallen, E.B. Spatial modeling for hellbender conservation in Indiana. Indiana Technical Advisory Committee Amphibians and Reptiles Meeting. Bloomington, Indiana. October 2015.

McCallen, E.B., L. E. Ruyle, J. M. Novak, and K. F. Gaines. A resource selection model for river otters on the department of energy's Savannah River Site. Wildlife Society Annual Conference. Milwaukee, Wisconsin. October, 2013.

McCallen, E.B., S. A. Sara, and P. V. Switzer. The effect of land use type and soil variables on Japanese beetle population attributes in Central Illinois soybean fields. Midwest Ecology and Evolution Conference. Cincinnati, Ohio. April 2012.

McCallen, E.B. Lessons on the evolutionary mechanisms of domestication from the domestication of the silver fox (*Vulpes vulpes*). The Behavioral Ecology Graduate Seminar Series. Charleston, Illinois. May 2011.

V. Poster Presentations

Yager, V., N. Burgmeier, **E.B. McCallen,** S. Unger, and R.N. Williams. Hellbender reintroduction in Indiana. 8th Biennial Hellbender Symposium. Jackson, Mississippi. June 2017.

Ordonez, K.A., **E.B. McCallen,** S. Fei, and R.N. Williams. Innovative object-based classification of river substrates utilizing sonar imaging data for hellbender conservation. 8th Biennial Hellbender Symposium. Jackson, Mississippi. June 2017.

Knott, J. **E.B. McCallen,** G. Nunez-Muir, B. Taylor, I. Jo, L. Pataro, and S. Fei. What's trending in ecology? An automated content analysis of the top concepts in ecology. Office of Interdisciplinary Graduate Programs Spring Reception. West Lafayette, Indiana. May 2017.

McCallen, E.B., B.T. Kraus, N.G. Burgmeier, and R.N. Williams. The Impact of management on the movement and home range size of Indiana's eastern hellbender salamanders. Poster. Purdue GIS Day. West Lafayette, Indiana. November 2016.

McCallen, E.B., R.N. Williams, and S. Fei. A multispecies distribution model for the conservation and management of Indiana's eastern hellbenders. 7th Biennial Hellbender Symposium. Saint Louis, Missouri. June 2015.

McCallen, E.B., R.N. Williams, and S. Fei. A multispecies distribution model for the conservation of *Cryptobranchus alleganiensis alleganiensis* in Indiana. Midwest Fish and Wildlife Conference. Indianapolis, Indiana. February 2015.

McCallen, E. B., S. Fei, and R. N. Williams. Ecological niche modeling of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) in Indiana. Purdue GIS Day. West Lafayette, Indiana. October 2014.

McCallen, E.B., L. E. Ruyle, J. M. Novak, and K. F. Gaines. A spatially explicit resource slection model for river otters on the department of energy's Savannah River Site. Wildlife Society Annual Conference. Milwaukee, Wisconsin. October, 2013.

McCallen, E. B., S. Fei, and R. N. Williams. Ecological niche modeling of the eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) in Indiana. 6th Biennial Hellbender Symposium. Chattanooga, Tennessee. June 2013.

McCallen, E.B., K.F. Gaines, L.E. Ruyle, and J.M. Novak. A spatially explicit model to estimate contaminant burdens in tiver oters (*Lontra Canadensis*) on the Department of Energy's Savannah River Site. Eastern Illinois University's College of Sciences Sciencefest. Charleston, Illinois. April 2013.

McCallen, E.B., S.A. Sara, and P.V. Switzer. The effect of land use type and soil variables on Japanese beetle distribution in central Illinois soybean fields. Eastern Illinois University's College of Sciences Sciencefest . Charleston, Illinois. April 2012.

Sara, S.A., **E.B. McCallen,** and P.V. Switzer. Japanese Beetle Distributions in Soybean Fields. Graduate Exposition Poster Presentations. Charleston, Illinois. April 2010.

VI. Honors and Awards

Graduate School Scholarship 2017
College of Agriculture, Purdue University

GIS Day Poster Competition Award 2016
Second Place for Graduate and Post-Doc Posters
Purdue University Libraries

GIS Day Poster Competition Award 2014
Third Place for Graduate and Post-Doc Posters
Purdue University Libraries

Frederick N. Andrews Fellowship 2013
The Graduate School, Purdue University

Teaching Award of Excellence for Graduate Assistants 2012
Eastern Illinois University and Midwestern Association of Graduate Schools

Research/Creative Activity Grant 2012
The Graduate School, Eastern Illinois University

Williams Travel Grant 2012
The Graduate School, Eastern Illinois University

Graduate Student Investigator Award 2012
College of Sciences, Eastern Illinois University

Stephen J. Gould Award 2012
Department of Biological Sciences, Eastern Illinois University

VII. Outreach and Extension

04/2017	Wildlife Techniques and Awareness Day Presentation for FNR 59800 Purdue University, West Lafayette, Indiana
03/2017	Graduate School Question and Answer Panel The Wildlife Society North Central Section Student Conclave Brookston, Indiana. March 2017
02/2017	Strawberry DNA Extraction Presentation for 4th Grade Students Hershey Elementary School, Lafayette, Indiana
12/2016	Exploring Amphibians and Reptiles Presentation for Preschool Students LLL Childcare, Lafayette, Indiana
11/2016	Strawberry DNA Extraction Presentation for 7th Grade Students Tecumseh Junior High School, Lafayette, Indiana
10/2016	Hell-bent on Integrating Extension and Research Presentation for FNR Wildlife Potluck Purdue University, West Lafayette, Indiana
08/2016	Help the Hellbender Day Public Education Volunteer Columbian Park Zoo, Lafayette, Indiana
04/2016	College of Agriculture Spring Fest Public Education Volunteer Purdue University, West Lafayette, Indiana
04/2016	Hellbender Hustle Public Education Volunteer O'Bannon Woods State Park, Corydon, Indiana
08/2015	Help the Hellbender Day Public Education Volunteer

Columbian Park Zoo, Lafayette, Indiana

04/2015 Feast Like a Hellbender
Public Education Volunteer
O'Bannon Woods State Park, Corydon, Indiana

02/2015 Giant Salamanders Part 1: Climate Change and Immune Function
Podcast Host
Forestry and Natural Resources, Purdue University

02/2015 Strawberry DNA Extraction
Presentation for 4th Grade Students
Hershey Elementary School, Lafayette, Indiana

11/2014 High School GPS Data Gathering Event
GIS Day 2014 Volunteer
Purdue University, West Lafayette. Indiana

06/2014 Video and PDF Files for Department Website
Student Example Extension Material
Forestry and Natural Resources, Purdue University

05/2014 The Use of Statistics in Ecological Research
Presentation for 11th – 12th Grade Students
Southmont Senior High School, Crawfordsville, Indiana

05/2014 The Eastern Hellbender in Indiana
Presentation for 8th Grade Students
Southmont Junior High School, Crawfordsville, Indiana

04/2014 Hellbender Hustle
Public Education Volunteer
O'Bannon Woods State Park, Corydon, Indiana

03/2014 Exploring Indiana's Amphibians and Reptiles
Presentation for 8th Grade Students
Southmont Junior High School, Crawfordsville, Indiana

07/2013 National Wildlife Habitat Education Program
Indiana Committee Member

Trafalgar, Indiana

04/2013 Feast Like a Hellbender
Public Education Volunteer
Purdue Extension Office, Corydon, Indiana

05/2012 -08/2012 Summer Science Program
Coordinator/Director
Weekly Outreach for 2nd – 5th Grade Students
Monroe Elementary School, Casey, Illinois