Integrating species distribution and occupancy modeling to study hellbender (*Cryptobranchus alleganiensis*) occurrence based on eDNA surveys

Jeronimo G. Da Silva Neto[a], William B. Sutton[a,*], Stephen F. Spear[b], Michael J. Freake[c], Marc Kéry[d], Benedikt R. Schmidt[e,f]

[a] *Wildlife Ecology Laboratory, Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, TN 37209, USA*
[b] *The Wilds, Cumberland, OH 43732, USA*
[c] *Department of Natural Sciences and Mathematics, Lee University, Cleveland, TN 37311, USA*
[d] *Swiss Ornithological Institute, 6204 Sempach, Switzerland*
[e] *Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*
[f] *Info Fauna Karch, UniMail, Bâtiment G, Bellevaux 51, 2000 Neuchâtel, Switzerland*

ABSTRACT

Managers often rely on species surveys and distribution models to evaluate species occurrence and develop management and conservation plans. However, these tools are rarely used in concert. We used a three-step framework to evaluate the distribution of a declining and elusive freshwater amphibian species, the hellbender salamander (*Cryptobranchus alleganiensis*). We used the Maximum Entropy (MaxEnt) algorithm to develop a habitat suitability model to guide sampling, environmental DNA (eDNA) surveys to ground truth the habitat suitability model, and multi-level occupancy modeling to assess species presence, while accounting for eDNA detection errors. Our suitability model (AUC = 0.941, True Skill Statistic = 0.7, sensitivity and specificity = 0.86) identified the greatest amount of high and very highly suitable habitat in the Interior Plateau and Blue Ridge ecoregions of the study area. We used eDNA survey results (n = 284 sites) to evaluate model fit and detected the species at 65 sites. Detection probability (p) was 0.692 (95% CRI: 0.547, 0.818) at the site level and 0.674 (95% CRI: 0.621, 0.721) at the quantitative PCR level. Ecoregion was the primary covariate that explained occupancy, with greatest estimates in the Blue Ridge ecoregion 0.695 (95% CRI, 0.390, 0.925). Although the MaxEnt output was not significantly correlated with occupancy probability, we established a relationship between habitat quality and the number of eDNA detections. This study highlights the use of a multi-level framework to optimize sampling, assess model fit, account for imperfect detection, and evaluate the distribution of rare species that have limited occurrence data available.

1. Introduction

Efforts to inform conservation management and remediate causes of species declines are often impaired by the lack of baseline distribution data (Collins and Crump, 2009). True absence data are rare, and existing species occurrence datasets can be unrepresentative of true species occurrence due to imperfect detection, sampling bias, and inconsistent sampling approaches (Newbold, 2010; Kéry and Royle, 2016, 2020). Lack of distribution data can be especially challenging when developing range-wide conservation plans for elusive, rare, or declining species (Collins and Crump, 2009). Large scale surveys for rare species can be costly, time consuming, and random sampling is often unlikely to result in acquisition of numerous species detections (Buckley and Beebee, 2004; Beebee and Griffiths, 2005; Goldberg et al., 2011).

An effective method to inform range-wide management decisions is to develop species distribution models (SDM) that permit managers to optimize their conservation funding through prioritization of survey efforts. As presence data may be the only occurrence data available for rare species, presence-only SDM approaches can be used to create habitat suitability models (Newbold, 2010; Elith et al., 2011). In general, SDMs estimate the potential distribution of target species across the landscape by relating species distribution data (e.g., presence-only data and pseudo-absences), with known or suggested environmental factors required to maintain viable populations of the target species (Elith and Leathwick, 2009; Phillips et al., 2006). Although habitat suitability models can be an effective tool to evaluate potential species distribution and identify priority conservation areas, models are often not validated through field surveys due to cost limitations or time restrictions, which
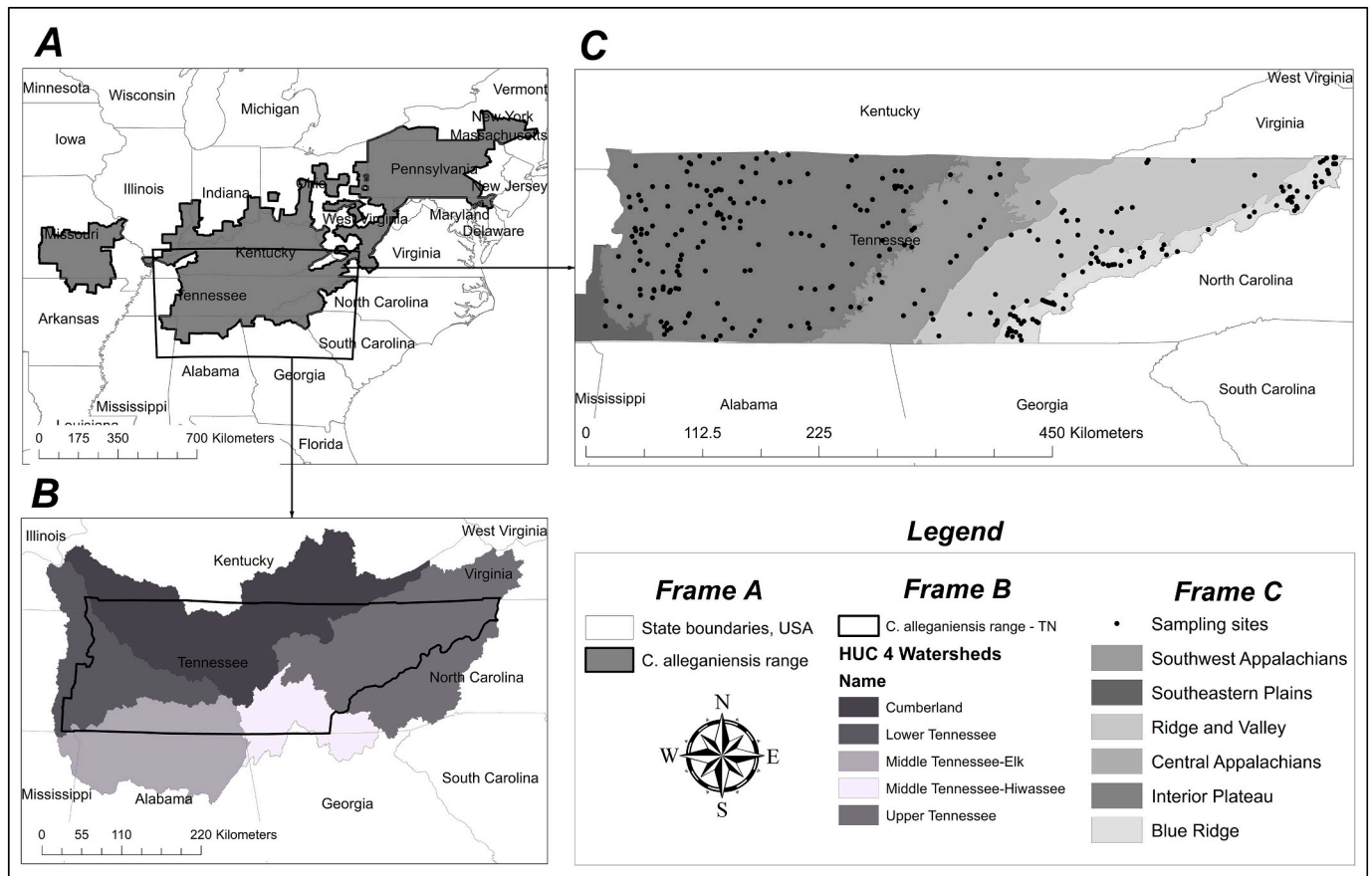
---

* Corresponding author.
E-mail address: wsutton@tnstate.edu (W.B. Sutton).

Fig. 1. Maps representing the *C. alleganiensis* total geographic range that encompasses 15 States (Frame A; Hammerson and Phillips, 2004), the main river basins (Frame B) and ecoregions (Frame C) found within the study area. Frame C also displays all 284 sampling sites surveyed for eDNA between 2012 and 2016 in the current study.

limits their predictive performance (Guisan et al., 2013; Riaz et al., 2019). As a result, advancement of conservation efforts may be jeopardized if decisions are based on inaccurate models (Loiselle et al., 2003).

Ground-truthing (i.e., evidence of species presence confirmed through surveys) is an important validation step that can be used to assess the performance of a habitat suitability model, strengthen model inference, and better inform occurrence-based management decisions (Greaves et al., 2006). Environmental DNA (eDNA, hereafter) sampling, for example, is a non-invasive method for monitoring species in both terrestrial and aquatic systems (Rees et al., 2014). Recent research on amphibians, reptiles, and fish indicate that eDNA can provide a more rapid, cost-effective, and less labor-intensive survey method compared to traditional field surveys (Goldberg et al., 2011; Spear et al., 2015). Even though eDNA has shown to be a reliable alternative to traditional field survey methods, species detection during eDNA surveys is not perfect and can be impacted by a variety of factors, including target animal abundance, weather conditions, and water and soil chemistry (Stewart, 2019). Occupancy modeling (MacKenzie et al., 2003) represents one SDM method that permits prediction of a species distribution, while accounting for factors that impact detection. By accounting for imperfect detection, researchers can assess survey effectiveness and accuracy of species distribution predictions (Kéry et al., 2013; Peterman et al., 2013). These data in turn can be used to assess accuracy of a habitat suitability model used to guide sampling efforts.

Field surveys and SDMs are commonly used to effectively develop and manage species-specific conservation plans but are rarely used in concert (but see Riaz et al., 2019 and Peterman et al., 2013). In this

study we demonstrate the development and implementation of a robust multi-level approach to evaluate the distribution of a declining and elusive freshwater amphibian species. Our approach consisted of three main phases, including 1) development of a habitat suitability model, 2) ground-truthing of the habitat suitability model via eDNA surveys, and 3) evaluation of survey data using multi-level occupancy modeling to account for imperfect detection. Through this three-tier approach, we demonstrate how researchers can enhance distribution data for species by optimizing sampling efforts and accounting for imperfect detection. The focal species for this study was the hellbender salamander (*Cryptobranchus alleganiensis*), which was chosen as a model species due to rapid range-wide declines, lack of range-wide distribution data, and unique life history characteristics (e.g., strong site fidelity, fully aquatic dependence, elusive behaviors). Our objectives were to 1) model and quantify suitable habitat for hellbenders within our study area, 2) use eDNA to assess hellbender distribution and validate the habitat suitability model, 3) use occupancy modeling to evaluate field survey results, and 4) identify covariates that describe occupancy of *C. alleganiensis*. These four objectives, accomplished in their respective order, provide a robust species distribution evaluation approach that can be applied to a variety of other elusive, threatened, and/or endangered species.

2. Materials and methods

2.1. Predictive habitat suitability model

2.1.1. Model species and study area

Hellbender salamanders, including the Eastern (*C. alleganiensis*

alleganiensis) and Ozark (*C. alleganiensis bishopi*) subspecies, are fully aquatic, benthic inhabitants of streams and rivers in the eastern and midwestern United States of America (Petranka, 1998). Both subspecies have experienced range-wide population declines due to a variety of stressors, including habitat fragmentation, accelerated sedimentation in waterways, aquatic contaminants, infectious diseases, and poaching (Nickerson and Mays, 1973; Wheeler et al., 2003; Briggler et al., 2007; Burgmeier et al., 2011; Freake and DePerno, 2017). Despite research completed over the past four decades, challenges remain in the field of *C. alleganiensis* conservation and standardized population assessments throughout the geographic range have often been limited to rivers and streams with known populations.

Our study area included the known range of *C. alleganiensis* in Tennessee, USA (Petranka, 1998; Fig. 1), which includes the Upper, Middle, and Lower Tennessee River, and Cumberland River basins. A considerable portion of the *C. alleganiensis* range in Tennessee occurs within the Tennessee River Basin, which currently holds the greatest number of hellbender populations across the historical range (USFWS, 2018). Therefore, our study area constitutes a significant portion of the species total geographic range.

2.1.2. Step 1: predicting habitat suitability with MaxEnt

We used the MaxEnt algorithm (Phillips et al., 2006, 2009) to develop a habitat suitability model for *C. alleganiensis* within the study area (Fig. 2 – Step 1). We acquired a total of 153 occurrence data points from seven main sources, including the Biodiversity Information Serving Our Nation (BISON), Global Biodiversity Information Facility (GBIF), Conservation Fisheries, Tennessee Species Wildlife Action Plan (TNSWAP), Tennessee Herpetofaunal Atlas (Austin Peay University), VertNet (www.vertnet.org), and the Tennessee Valley Authority (TVA) databases. We selected environmental variables based on our biological knowledge of the target species, and from studies that evaluated habitat parameters associated with the presence of *C. alleganiensis* and other sensitive aquatic taxa (Heino et al., 2003; Messerman, 2014; Pugh et al., 2016). We used multiple environmental variables, including elevation, stream flow direction, geology, hillshade, land use, mean annual precipitation, slope, Strahler stream order, and temperature, with a minimum data grain of 30 m (Appendix A - Table A1). We developed a background points layer (i.e., pseudo-absences) by allocating 10,000 randomly generated points among modeled streams and rivers throughout the study area. For more information on environmental variables, see Appendix A – Table A1.

We used a ten-fold cross-validation approach to estimate habitat suitability across the study area (Messerman, 2014; Sutton et al., 2015).

The initial model layer represented all streams within the study area with raster cells classified from least suitable (0.0001) to highly suitable (0.9999) on a logistic scale. We applied a 10-percentile training presence logistic threshold (0.141; hereafter referred as F10 threshold) to the resulting model layer to generate a binary map (i.e., suitable vs. not suitable), where cells were considered "suitable" if the suitability score was greater than the F10 threshold (Freeman and Moisen, 2008). Although threshold selection is often arbitrary, the F10 threshold provides a more conservative approach and avoids overfitting compared to other thresholds such as the minimum training presence (Tinoco et al., 2009; Jarnevich and Reynolds, 2011). We used Jenk's natural breaks classification (Chen et al., 2013) in ArcGIS v.10.3 to further classify logistic suitability scores above the F10 threshold score into the four following habitat suitability categories: Category 1 (low suitability = 0.141–0.348), Category 2 (medium suitability = 0.348–0.555), Category 3 (high suitability = 0.555–0.761), and Category 4 (very high suitability = 0.761–1.0). By applying this two-step process (i.e., applying a threshold and using natural breaks), we developed a raster layer of waterways within our study area divided into four suitability classes to aid in site selection for eDNA sampling.

We evaluated fit of the resulting habitat suitability model using the area under the Receiver Operating Characteristic Curve (AUC; Appendix C - Fig. C1) and True Skill Statistic (TSS). We considered the model informative when the AUC value was above 0.75 (i.e., 75% of the time the model correctly differentiates a test point from a random point; Fielding and Bell, 1997; Phillips, 2006; Sobek-Swant et al., 2012). The average TSS was calculated based on recommendations in Allouche et al. (2006) and Monserud and Leemans (1992) and is evaluated on a scale from −1 to +1, where values below 0.2 represent poor/as good as random model performance, and values above 0.7 represent very good/excellent model performance. We used both the AUC and TSS to evaluate model performance because AUC evaluates differences in distribution between random and occurrence points independent of thresholds, whereas TSS accounts for omission and commission errors based on a threshold and is not as sensitive to size of the modeled area (Lobo et al., 2008; Phillips and Elith, 2010; Warren and Seifert, 2011).

We evaluated the MaxEnt output via response curves and Jackknife tests, which we used to assess how each variable contributed to the model predictions. Output response curves showed the mean effect of individual variables on model predictions over 10 replicates, whereas Jackknife tests assessed differences between variables and individual variable contribution.
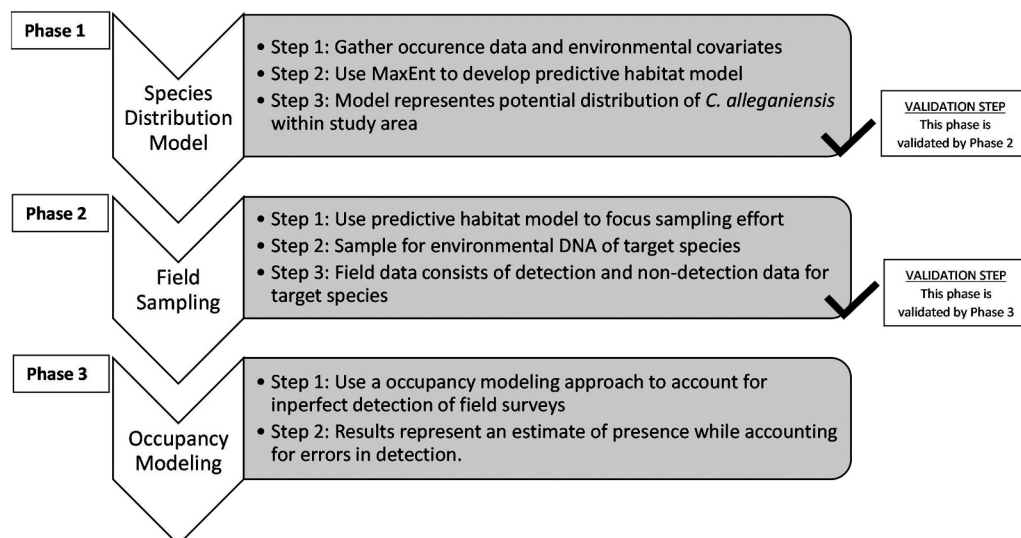

Phase 1

Species Distribution Model
- Step 1: Gather occurence data and environmental covariates
- Step 2: Use MaxEnt to develop predictive habitat model
- Step 3: Model representes potential distribution of *C. alleganiensis* within study area

VALIDATION STEP
This phase is validated by Phase 2

Phase 2

Field Sampling
- Step 1: Use predictive habitat model to focus sampling effort
- Step 2: Sample for environmental DNA of target species
- Step 3: Field data consists of detection and non-detection data for target species

VALIDATION STEP
This phase is validated by Phase 3

Phase 3

Occupancy Modeling
- Step 1: Use a occupancy modeling approach to account for inperfect detection of field surveys
- Step 2: Results represent an estimate of presence while accounting for errors in detection.

Fig. 2. We used a three-phase process to assess the distribution and occupancy of *C. alleganiensis* within our study area. Briefly, these phases consisted of a presence-only habitat suitability model (Phase 1), range-wide ground-truthing of the habitat suitability model using eDNA data (Phase 2), and final assessment of occupancy and detection probabilities based on ground-truthing data (Phase 3). Each main step was composed of two or three steps completed in order. The "Validation step" checkmarks demonstrate that one next step was used to validate the results from the previous step.

2.2. Step 2: validation of habitat suitability model through environmental DNA sampling

2.2.1. Field surveys

Between 2012 and 2016, we collected and filtered water samples from 284 sites to assess presence of *C. alleganiensis* DNA (Fig. 2 – Step 2) using the collection protocol described by Spear et al. (2015). Between 2012 and 2015, we sampled 84 sites once based on goals of a separate ongoing study. In 2016, we allocated and sampled 200 sites based on our habitat suitability model. Of the 200 sites, we sampled 170 sites once and 30 sites three times (i.e., each of the 30 sites was sampled on June, July, and August 2016). We selected the 170 sampling sites by allocating 1000 random points across modeled waterways in the study area and selected approximately 42 survey points within each habitat suitability category based on accessibility. Thirty out of 200 sampling sites were not randomly selected; instead, we selected them based on previous data suggesting that 5 sites were negative, and 25 sites were positive for *C. alleganiensis* eDNA. These 30 sites were equally stratified among habitat suitability categories based on the MaxEnt suitability model and Level III ecoregions throughout the range of *C. alleganiensis* within the study area to control for potential influences of habitat quality and ecoregion effects on detectability. We ensured a minimum sampling distance of 5 km between all sites based on both home-range estimates for *C. alleganiensis* and uncertainties associated with eDNA sampling techniques to increase site independence (Peterson and Wilkinson, 1996; Deiner and Altermatt, 2014). For detailed field survey and sample collection methods, refer to Appendix B.

2.2.2. Laboratory methods

We used modified protocols reported in Goldberg et al. (2011) and Spear et al. (2015) to extract, purify, and amplify DNA from filters utilized to collect eDNA from water samples. We used DNeasy Blood and Tissue Kits (Qiagen, Inc.) and followed the standard protocols except we used a Qiashredder (Qiagen, Inc.) spin column after the lysis buffer step. We amplified DNA via quantitative PCR (qPCR) with reverse and forward primers and probe that amplified a 104 bp region of mitochondrial *cytochrome-B* as recommended in Spear et al. (2015) These primers and probes were developed based on DNA from individuals collected across the entire *C. alleganiensis* range as described by Sabatino and Routman (2009). We considered a sample positive for *C. alleganiensis* DNA if 2/3 or 3/3 replicates amplified. We considered samples that had 1/3 replicates amplify as questionable positives. We amplified all questionable positive samples a second time and considered these samples positive if 1/3, 2/3, or 3/3 replicates amplified. We considered a sample negative for eDNA if 0/3 replicates amplified, or if a questionable sample had 0/3 replicates amplify the second round. For a detailed description of laboratory methods, refer to Appendix B.

2.3. Step 3: occupancy modeling

We used the multilevel occupancy model described by Nichols et al. (2008), Mordecai et al. (2011), Schmidt et al. (2013), and Kéry and Royle (2016, Section 10.10 which includes R and JAGS code to fit the model to the data) to evaluate potential covariates that describe occupancy of *C. alleganiensis* while accounting for imperfect detection of the eDNA sampling method (Fig. 2 – Step 3). We used eDNA detection/non-detection data based on repeated site visits and triplicate PCR from 284 sites sampled during seven sampling events: 2012, 2013, 2014, 2015, 2016A (May – July), 2016B (July – August), 2016C (August – October) to assemble a detection matrix. The model estimates an occupancy probability and detectability at two scales: detectability in the water sample (given presence at the sampling site; "availability" in Schmidt et al., 2013) and detectability in the PCR, given that eDNA is in the water sample. Using uniform priors as in Schmidt et al. (2013), we fitted a single model with all covariates of interest to the data using

JAGS and R (package "jagsUI"; Plummer, 2003; Kéry and Schaub, 2012; Kellner, 2018). For all probabilities, we used uniform priors in the interval (0,1). For regression coefficients, we used uniform priors on the interval (−10,10). All models were run using a burn-in of 2000, 3 chains, and thinning 1 in 10, with 20,000 iterations. Convergence was assessed using the Brooks-Gelman-Rubin statistic (Kéry and Schaub, 2012). Convergence was satisfactory when R-hat values were smaller than 1.1

We evaluated how linear and quadratic sampling date patterns may impact detection probability in the water sample while detectability was held constant at the qPCR (as in Schmidt et al., 2013). We evaluated the influence of sampling day because previous studies of aquatic salamanders found a quadratic effect of date on detectability (e.g., de Souza et al., 2016). For occupancy, we used five covariates that we hypothesized might affect *C. alleganiensis* occupancy across the study area, including elevation, Strahler order (as a continuous variable), agricultural and forest land cover, and ecoregion (i.e., geographical units defined by similar ecosystems, physiography, geology; Wineland et al., 2019). The effect of land cover was measured at the HUC 12 scale (i.e., hydrological units that capture tributary systems). We used the Geospatial Modelling Environment software (Spatial Ecology LLC 2015) to calculate the proportion of land cover categories within each HUC 12 watershed across the study area and used forest and agriculture land use categories in our analysis. Because different sites were sampled in different years, we used a year-specific intercept which was defined as a random effect. To further assess the fit of our MaxEnt model, we compared the MaxEnt logistic output with occupancy model results.

3. Results

3.1. MaxEnt habitat suitability model

Our habitat suitability model performed considerably better than random as indicated by an AUC value of 0.941 (STD = 0.023; Appendix C - Fig. C1), along with a TSS value of 0.77, and sensitivity and specificity values of 0.86. According to our analysis of variable contribution, the Strahler stream order variable contributed greatest to the suitability model (85.4%), followed by geology (7.4%), mean annual precipitation (2%), mean annual temperature (1.5%), land use and elevation (0.9% each), slope (0.8%), hillshade (0.6%), and flow accumulation (0.5%; Appendix C – Table C1). Strahler stream order was also the most informative variable (i.e., contributed the most amount of unique information to the model), followed by geology and land use (Appendix C - Fig. C2).

After application of the F10 threshold, the total suitable stream area (i.e., stream suitability Categories 1–4) for *C. alleganiensis* within the study area represented 12.2% (~422 km^2) of the total modeled stream area (~3463 km^2 before F10 threshold was applied; Table 1 - Frame A). The total suitable stream area was comprised of 48.6% (~205 km^2) Category 1 streams, 24.4% (~103 km^2) Category 2 streams, 24.6% (~104 km^2) Category 3 streams, and 2.4% (~10 km^2) Category 4 streams (Table 1 - Frame A). The Interior Plateau ecoregion contained the greatest percent of all suitable (Category 1 or greater) streams (53%, ~225 km^2), followed by the Ridge and Valley (23%, ~97 km^2), Blue Ridge (13%, ~55 km^2), Southwest Appalachian (8%, ~33 km^2), Southeastern Plains (2%, ~8 km^2), and Central Appalachian ecoregions (1%, ~5 km^2; Table 1 - Frame A). After controlling for ecoregion area, the Interior Plateau had the greatest proportion (~31%, ~70 km^2) of Category 3 streams, followed by the Ridge and Valley (~23%, ~23 km^2) and Blue Ridge (~14%, ~7 km^2), whereas the Central Appalachian ecoregion had the least (~3%, ~0.1 km^2; Table 1 – Frame B). When considering only the highest suitability category, the Southeastern Plains had the greatest proportion (~10%, ~0.7 km^2) of Category 4 streams followed by the Blue Ridge (~9%, ~5 km^2), whereas the Central Plains and Southwest Appalachian ecoregions had no or very few streams in Category 4 (Table 1 – Frame B).

Table 1
Frame A) Suitable stream area (km^2) per suitability category (i.e., Category 1 = low suitability, Category 2 = medium suitability, Category 3 = high suitability, and Category 4 = very high suitability) within each ecoregion, and B) respective percent of each suitability category per ecoregion after accounting for ecoregion area. The "Total" column represents total suitable stream area per ecoregion (km^2). The "Total" row represents the total stream area per suitability category and the total area (km^2) of suitable stream within the study area (422.39 km^2).

A

Ecoregion	Suitable stream area per suitability category (km^2)				Total (km^2)
Suitability category	1	2	3	4	
Central Appalachians	3.09	1.74	0.13	0.00	4.96
Blue Ridge	29.72	12.07	7.80	5.02	54.60
Southwest Appalachians	19.33	11.27	2.75	0.13	33.48
Ridge and Valley	46.75	25.59	22.63	1.91	96.89
Southeastern Plains	3.59	2.24	0.99	0.76	7.58
Interior Plateau	102.19	50.64	69.94	2.11	224.88
Total (km^2)	204.68	103.54	104.25	9.93	422.39

B

Ecoregion	Percent of each suitability category per ecoregion (%)			
Suitability category	1	2	3	4
Central Appalachians	62.39	35.03	2.58	0.00
Blue Ridge	54.43	22.10	14.29	9.19
Southwest Appalachians	57.73	33.65	8.22	0.39
Ridge and Valley	48.26	26.41	23.36	1.97
Southeastern Plains	47.36	29.51	13.09	10.04
Interior Plateau	45.44	22.52	31.10	0.94

3.2. Environmental DNA field sampling

Overall, from 2012 to 2016, we sampled 284 sites and detected *C. alleganiensis* eDNA at 65 individual sites (Table 2). Category 1 streams had the fewest proportion of detections (13.2%) and greatest proportion of non-detections (86.8%), whereas Category 4 streams had the greatest proportion of detections (46.3%) and fewest proportion of non-detections (53.7%; Table 2). The Blue Ridge ecoregion had the greatest number of detections (34), followed by the Interior Plateau (17), Ridge and Valley (10), and Southwest Appalachian ecoregions (4), for a total of 65 detections across the study area (Table 2). All field and lab controls were negative, which indicated that there was no contamination during field sampling or lab procedures. Although we were able to consistently amplify standards, our 10^{-6} μL standard failed to amplify completely (all three wells did not amplify) in 9 out of 19 plates and failed to amplify at least once (one or two wells did not amplify) in 5 additional plates. Therefore, we considered our limit of detection to be

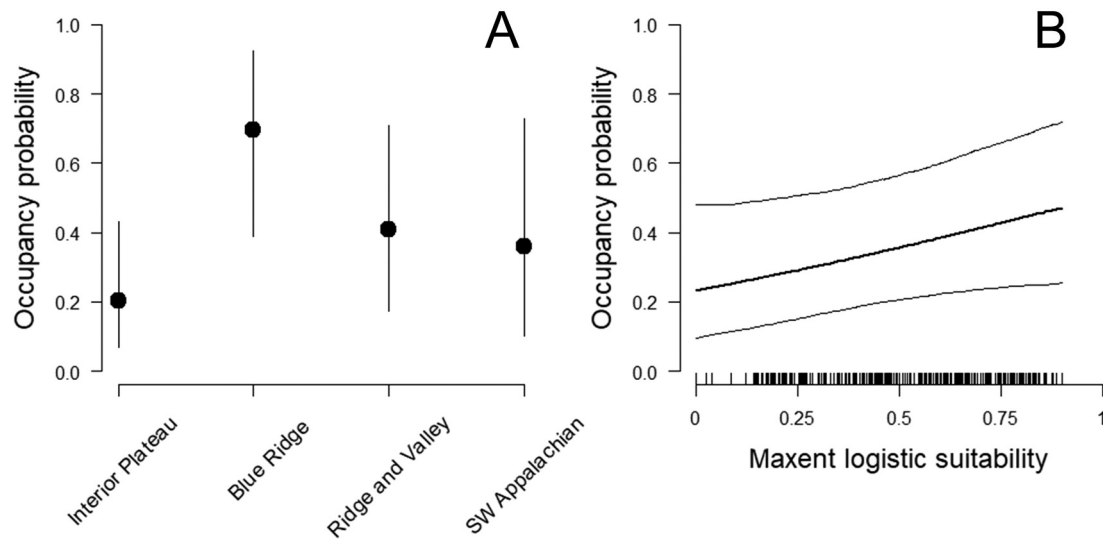0.000524 ng/μL (10^{-5}).

3.3. Occupancy model

The occupancy model showed that detection probability at the PCR level was 0.674 (95% CRI: 0.621, 0.721). Detection probability at the field level was on average 0.692 (95% CRI: 0.547, 0.818) and did not depend on sampling date, but we note that the extreme uncertainty in both the linear and the quadratic term of Julian date makes this result highly inconclusive. Mean occupancy across the study area was 0.436 (95% CRI: 0.073, 0.858). Parameter estimates of the variables describing occupancy are shown in Appendix A - Table A2. Occupancy varied among ecoregions. Estimated occupancy (and 95% CRI) for the ecoregions were 0.203 (0.070, 0.432) for the Interior Plateau, 0.695 (0.390, 0.925) for the Blue Ridge, 0.408 (0.172, 0.710) for the Ridge and Valley, and 0.360 (0.165, 0.729) for the Southwestern Appalachians (Fig. 3 – Frame A). Other covariates had no strong effects and credible intervals of estimated coefficients overlapped zero.

In our analysis of the relationship between MaxEnt logistic output and occupancy modeling results, MaxEnt logistic output positively influenced occupancy probability, but the credible intervals overlapped zero (coefficient: 1.123, 95% CRI: −0.268, 2.558, probability that the coefficient was greater than zero = 0.943; Fig. 3 – Frame B).

4. Discussion

We quantified the predicted suitable habitat area for *C. alleganiensis* to be only approximately 422 km^2 within the study area. A large portion of Category 3 and 4 streams were located within ecoregions that were, and will most likely continue to be, impacted by anthropogenic factors (Green et al., 2003; Drummond and Loveland, 2010). With limited availability of suitable habitat for *C. alleganiensis* within the study area, and the growing threats to stream habitat, the need is apparent for additional biological information on the species (Alig et al., 2003). The MaxEnt suitability model indicated that Strahler stream order (stream orders 4–6) and geology (granitic and sedimentary substrates) were important contributors to the habitat suitability model, while our occupancy modeling results show that ecoregion explained *C. alleganiensis* occupancy. These factors may influence habitat suitability and site occupancy because the species has a strong relationship with substrate composition and stream hydrological patterns, which are determined by ecoregion-specific geomorphology and topography (Platts, 1979; Quinn et al., 2013).

The effect of ecoregion on occupancy can serve as a proxy for the effect of land use practices on habitat quality. Bodinof Jachowski et al. (2016) suggested that even though *C. alleganiensis* occupancy was explained mainly by geology and physiography, agriculture and development could have a negative effect on species occurrence. Urbanization and agricultural practices affect water quality, increase sediment input into waterways, and significantly disrupt hydrological patterns (Malmqvist and Rundle, 2002). Our occupancy model results support

Table 2
C. alleganiensis detection and non-detections within each stream suitability category (i.e., Category 1 = low suitability, Category 2 = medium suitability, Category 3 = high suitability, and Category 4 = very high suitability), and ecoregion (BR = Blue Ridge, IP = Interior Plateau, RV = Ridge and Valley, SA = Southwest Appalachians, CA = Central Appalachians, SP = Southeastern Plains). The "Total" column represents the total number of sites sampled at each ecoregion. The lower portion of the table represents the naïve occupancy per category.

	MaxEnt suitability category					Ecoregions						
	1	2	3	4	Total	BR	IP	RV	SA	CA	SP	Total
Detection (#)	9	12	19	25	65	34	17	10	4	0	0	65
Non-detection (#)	59	66	65	29	219	36	139	29	12	1	2	219
Total (#)	68	78	84	54	284	70	156	39	16	1	2	284
Detection (%)	13.2	15.4	22.6	46.3	–	48.6	10.9	25.6	25	0	0	–
Non-detection (%)	86.8	84.6	77.4	53.7	–	51.4	89.1	74.4	75	100	100	–

Fig. 3. Frame A shows the relationship between occupancy probability and ecoregion. Symbols are posterior means and 95% CRI. The estimates are for the first year (the model has a year-specific intercept). Other predictor values in the model were fixed at their mean value. Frame B shows the relationship between the MaxEnt logistic suitability and occupancy probability. The thick line shows the posterior mean and the thin lines the 95% CRI. The "rug" on the x-axis shows predicted MaxEnt logistic suitability values.

the hypothesis that *C. alleganiensis* occupancy was greater within ecoregions currently with lower (and likely lower historical) levels of anthropogenic land use change (i.e., conversion of forest to agriculture and urban areas). Congruent with results presented by Bodinof Jachowski et al. (2016), the Blue Ridge ecoregion within the study area had the greatest occupancy estimates compared to the Ridge and Valley, Southwestern Appalachian, and Interior Plateau ecoregions. The Blue Ridge ecoregion has historically experienced the smallest net percent change in forest cover between 1973 and 2000, whereas the Interior Plateau and Ridge and Valley experienced greater net percent change in forest cover, mainly due to agriculture and mechanical forest disturbance (Drummond and Loveland, 2010). This historical land use pattern may also explain why the Blue Ridge ecoregion contained the greatest proportion of very high suitability streams and the greatest occupancy estimate compared to all other ecoregions. Although our habitat suitability model shows that the Southeastern Plains ecoregion had a relatively greater proportion of very high suitability streams than the Blue Ridge ecoregion; only a small portion of the Southeastern Plains ecoregion is located within the study area. Therefore, we believe that these results do not represent the proportion of suitability streams within the entire Southeastern Plains ecoregion.

Compared to results from Franklin (2016), where detection probabilities for *C. alleganiensis* via eDNA in North Carolina were approximately 90%, we reported considerably lower detection estimates of approximately 67%. Since eDNA concentration is also a function of species abundance, density, and biomass, lower quality sites with lower species abundance would consequently have lower eDNA available for detection (Pilliod et al., 2013). We believe the difference in detection estimates between studies can be explained by the large disparities in species range size, habitat quality, and species abundance sampled in each study. We recommend that future studies conduct abundance surveys at a sub-sample of sites to account for errors in detection due to abundance-based heterogeneity.

Although our model sensitivity and specificity were greater than 85%, we acknowledge that our MaxEnt habitat suitability model has limitations. Because we created our habitat suitability model using *C. alleganiensis* occurrence data (1950–2016) from a different period than some of the environmental covariates (e.g., land use cover 2001–2010), the effect of these covariates on the target species may not be fully represented. In addition, our MaxEnt model does not take into consideration small-scale biotic, abiotic, and anthropogenic factors that are

important determinants of species distribution patterns (Barve et al., 2011). For example, our habitat suitability model does not take into consideration inter and intra-specific competition or lack of micro-habitat characteristics such as cover rock presence or low levels of substrate embeddedness (e.g., Da Silva Neto et al., 2019). The lack of a significant relationship between the MaxEnt logistic output and occupancy probability highlights the predictive limitations of our model. However, there were several benefits of using a habitat suitability model in this study. Given the status of the species and its rapid decline, random sampling within the thousands of stream miles within the study area would likely be ineffective for providing informative distribution data. In addition, the resulting MaxEnt model is a tool that can guide future local and regional conservation efforts such as watershed level habitat mitigation, implementation of best management practices, and microhabitat surveys. Our overarching goal was to provide managers with a robust, but adaptable, framework to evaluate *C. alleganiensis* distribution across the study area. Following the principles of adaptive management, the model can be updated based on the acquisition of new locality data, which will aid in the development of management and conservation strategies that will contribute to the long-term conservation of rare and elusive species such as *C. alleganiensis*.

5. Conclusions

Our study demonstrates that by using a structured analysis framework that combines species distribution modeling, ground-truthing (via eDNA surveys), and multi-level occupancy modeling, we were able to make robust inferences on baseline site occupancy for *C. alleganiensis* across the study area, evaluate factors that influence occupancy, and identify areas of high conservation concern. Further, we were able to estimate the area of suitable streams predicted to be available for the target species. A large portion of suitable streams located within ecoregions under high levels of development will most likely continue to be affected by anthropogenic factors. Furthermore, our study provides detection and non-detection data for *C. alleganiensis* at 284 sites that were previously not sampled. This information can contribute to future species status assessments and the development of species-specific management plans. Overall, this multi-level distribution analysis approach can be applied to many rare, threatened, and endangered species with minimal occurrence and/or abundance data.

It is important to note that our habitat suitability model may not

represent habitat requirements for long-term persistence of the target species. Therefore, to increase the inference of this study, future research should focus on selecting high quality streams with positive detections within each ecoregion and conduct physical surveys to assess the habitat associated with *C. alleganiensis* abundance. This information will not only help further ground-truth model results, but will also provide baseline population demographics that will facilitate long-term assessment of the effects of land use change on *C. alleganiensis* populations within different ecoregions. Future conservation of the species may rely on prioritizing populations and areas of highest conservation need. Therefore, demographic and distribution data will be essential if future conservation strategies include targeted stream restoration, enhanced watershed protection, population augmentation, and repatriation of individuals into priority conservation areas.

CRediT authorship contribution statement

J. G. da Silva Neto: Conceptualization, Methodology, Formal Analysis, Visualization, Investigation, Writing – Original draft, Writing – Review & Editing; **W.B. Sutton**: Conceptualization, Methodology, Resources, Formal Analysis, Writing – Original draft, Writing – Review & Editing, Supervision, Project administration; **S. F. Spear**: Conceptualization, Investigation, Resources, Supervision; **M. J. Freake**: Conceptualization, Investigation, Resources; **M. Kéry**: Formal Analysis, Writing – Original draft, Writing – Review & Editing; **B. R. Schmidt**: Formal Analysis, Writing – Original draft, Writing – Review & Editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2020.108787.

References

Alig, R.J., Plantinga, A.J., Ahn, S., Kline, J.D., 2003. Land Use Changes Involving Forestry in the United States: 1952 to 1997, With Projections to 2050. US Department of Agriculture, Forest Service.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberon, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol. Model. 222, 1810–1819.

Beebee, T.J.C., Griffiths, R.A., 2005. The amphibian decline crisis: a watershed for conservation biology? Biol. Conserv. 125, 271–285.

Bodinof Jachowski, C.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. Divers. Distrib. 22, 865–880.

Briggler, J.T., Ettling, J., Wanner, M., Schuette, C., Ducan, M., 2007. Cryptobranchus alleganiensis chytrid fungus. Herpetol. Rev. 38.

Buckley, J., Beebee, T.J., 2004. Monitoring the conservation status of an endangered amphibian: the natterjack toad Bufo calamita in Britain. Animal Conservation forum. 3, 221–228.

Burgmeier, N.G., Unger, S.D., Sutton, T.M., Williams, R.N., 2011. Population status of the eastern hellbender (Cryptobranchus alleganiensis alleganiensis) in Indiana. J. Herpetol. 45, 195–201.

Chen, J., Yang, S., Li, H., Zhang, B., Lv, J., 2013. Research on geographical environment unit division based on the method of natural breaks (Jenks). Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 3, 47–50.

Collins, J.P., Crump, M.L., 2009. Introduced species, commerce, and land use change. pp. 42–58 in Extinction in Our Times: Global Amphibian Decline. Oxford University Press, UK.

Da Silva Neto, J.G., Sutton, W.B., Freake, M., 2019. Life-stage differences in microhabitat use by hellbenders (Cryptobranchus alleganiensis). Herpetologica 75, 21–29.

de Souza, L.S., Godwin, J.C., Renshaw, M.A., Larson, E., 2016. Environmental DNA (eDNA) detection probability is influenced by seasonal activity of organisms. PLoS One 11, e0165273.

Deiner, K., Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river. PLoS One 9, e88786.

Drummond, M.A., Loveland, T.R., 2010. Land-use pressure and a transition to forest-cover loss in the eastern United States. BioScience 60, 286–298.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

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

Franklin, T.W., 2016. Estimates of Eastern Hellbenders (*Cryptobranchus alleganiensis alleganiensis*) Occupancy and Detection Using Two Sampling Methods. Doctoral dissertation, Appalachian State University.

Freake, M.J., DePerno, C.C., 2017. Importance of demographic surveys and public lands for the conservation of eastern hellbenders Cryptobranchus alleganiensis alleganiensis in southeast USA. PLoS One 12, e0179153.

Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. 217, 48–58.

Goldberg, C.S., Pilliod, D.S., Arkle, R.S., Waits, L.P., 2011. Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS One 6, e22746.

Greaves, G.J., Mathieu, R., Seddon, P.J., 2006. Predictive modelling and ground validation of the spatial distribution of the New Zealand long-tailed bat (Chalinolobus tuberculatus). Biol. Conserv. 132, 211–221.

Green, H.A., Murray, M.N., Marshall, J.L., Couch, S.E., Ransom, A., McLeod, Belliveau, 2003. Population Projections for the State of Tennessee 2005 to 2025. Tennessee Advisory Committee on Intergovernmental Relations and the University of Tennessee Center for Business and Economic Outreach, Nashville, TN.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Martin, T.G., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16, 1424–1435.

Hammerson, G., Phillips, C., 2004. Cryptobranchus alleganiensis. The IUCN Red List of Threatened Species 2004. e.T59077A11879843.

Heino, J., Muotka, T., Paavola, R., 2003. Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. J. Anim. Ecol. 72, 425–434.

Jarnevich, C.S., Reynolds, L.V., 2011. Challenges of predicting the potential distribution of a slow-spreading invader: a habitat suitability map for an invasive riparian tree. Biol. Invasions 13, 153–163.

Kellner, K.F., 2018. jagsUI: a wrapper around 'rjags' to streamline 'JAGS' analyses. R package version 1.5.0. https://CRAN.R-project.org/package=jagsUI.

Kéry, M., Royle, J.A., 2016. Applied hierarchical modeling in ecology—modeling distribution, abundance and species richness using R and BUGS. In: Volume 1: Prelude and Static Models. Elsevier/Academic Press.

Kéry, M., Royle, J.A., 2020. Applied hierarchical modeling in ecology—modeling distribution, abundance and species richness using R and BUGS. In: Volume 2: Dynamic and Advanced Models. Elsevier/Academic Press.

Kéry, M., Schaub, M., 2012. Bayesian Population Analysis Using WinBUGS. Academic Press.

Kéry, M., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2013. Analyzing and mapping species range dynamics using dynamic occupancy models. J. Biogeogr. 40, 1463–1474.

Lobo, J.M., Jimenez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecol. Biogeography. 17, 145–151.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H., 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conserv. Biol. 17, 1591–1600.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., Franklin, A.B., 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology. 84, 2200–2207.

Malmqvist, B., Rundle, S., 2002. Threats to the running water ecosystems of the world. Environ. Conserv. 29, 134–153.

Messerman, A., 2014. The Use of Nest Boxes by the Hellbender Salamander in Western North Carolina. M.S. Thesis. Duke University.

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistics. Ecol. Model. 62, 275–293.

Mordecai, R.S., Mattsson, B.J., Tzilkowski, C.J., Cooper, R.J., 2011. Addressing challenges when studying mobile or episodic species: hierarchical Bayes estimation of occupancy and use. J. Appl. Ecol. 48, 56–66.

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Prog. Phys. Geogr. 34, 3–22.

Nichols, J.D., Bailey, L.L., O'Connell Jr., A.F., Talancy, N.W., Campbell Grant, E.H., Gilbert, A.T., Annand, E.M., Husband, T.P., Hines, J.E., 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. J. Appl. Ecol. 45, 1321–1329.

Nickerson, M.A., Mays, C.E., 1973. The Hellbenders: North American Giant Salamanders. No. 1. Milwaukee Public Museum Publications in Biology and Geology.

Peterman, W.E., Crawford, J.A., Kuhns, A.R., 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. J. Nat. Conserv. 21, 114–121.

Peterson, C.L., Wilkinson, R.F., 1996. Home range size of the hellbender (*Cryptobranchus alleganiensis*) in Missouri. Herpetological Review. 27, 126–127.

Petranka, J.W., 1998. Salamanders of the United States and Canada. Smithsonian Institution Press, Washington, DC.

Phillips, S.J., 2006. A brief tutorial on Maxent. AT&T Research. Available from url. http://biodiversityinformatics.amnh.org/open_source/maxent/.

Phillips, S.J., Elith, J., 2010. POC plots: calibrating species distribution models with presence-only data. Ecology. 91, 2476–2484.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197.

Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P., 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. Can. J. Fish. Aquat. Sci. 70, 1123–1130.

Platts, W.S., 1979. Relationships among stream order, fish populations, and aquatic geomorphology in an Idaho river drainage. Fisheries. 4, 5–9.

Plummer, M., 2003. JAGS: a program for analysis of bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. 124.

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. J. Herpetol. 47, 78–84.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R., Gough, K.C., 2014. The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. J. Appl. Ecol. 51, 1450–1459.

Riaz, M., Kuemmerlen, M., Wittwer, C., Cocchiararo, B., Khaliq, I., Pfenninger, M., Nowak, C., 2019. Combining environmental DNA and species distribution modeling to evaluate reintroduction success of a freshwater fish. Ecological Applications 30, e02034.

Sabatino, S.J., Routman, E.J., 2009. Phylogeography and conservation genetics of the hellbender salamander (*Cryptobranchus alleganiensis*). Conserv. Genet. 10, 1235.

Schmidt, B.R., Kéry, M., Ursenbacher, S., Hyman, O.J., Collins, J.P., 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. Methods Ecol. Evol. 4, 646–653.

Sobek-Swant, S., Kluza, D.A., Cuddington, K., Lyons, D.B., 2012. Potential distribution of emerald ash borer: what can we learn from ecological niche models using MaxEnt and GARP? For. Ecol. Manag. 281, 23–31.

Spear, S.F., Groves, J.D., Williams, L.A., Waits, L.P., 2015. Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. Biol. Conserv. 183, 38–45.

Stewart, K.A., 2019. Understanding the effects of biotic and abiotic factors on sources of aquatic environmental DNA. Biodivers. Conserv. 28, 983–1001.

Sutton, W., Barrett, K., Moody, A., Loftin, C., deMaynadier, P., Nanjappa, P., 2015. Predicted changes in climatic niche and climate refugia of conservation priority salamander species in the Northeastern United States. Forests. 6, 1–26.

Tinoco, B.A., Astudillo, P.X., Latta, S.C., Graham, G.H., 2009. Distribution, ecology and conservation of an endangered Andean hummingbird: the violet-throated Metaltail (*Metallura baroni*). Bird Conservation International. 9, 63–76.

USFWS, 2018. Species Status Assessment Report for the Eastern Hellbender (*Cryptobranchus alleganiensis alleganiensis*). 104pp.

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in MaxEnt: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

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.

Wineland, S.M., Welch, S.M., Pauley, T.K., Apodaca, J.J., Olszack, M., Mosher, J.J., Waldron, J.L., 2019. Using environmental DNA and occupancy modelling to identify drivers of eastern hellbender (*Cryptobranchus alleganiensis alleganiensis*) extirpation. Freshw. Biol. 64, 208–221.