Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists

Cody M Rhoden [Corresp., 1] , **William E Peterman** [2] , **Christopher A Taylor** [1]

[1] Illinois Natural History Survey, University of Illinois Urbana–Champaign, Champaign, Illinois, United States

[2] School of Environment and Natural Resources, Ohio State University, Columbus, Ohio, United States

Corresponding Author: Cody M Rhoden
Email address: codyrhoden@gmail.com

Background. Rare or narrowly endemic organisms are difficult to monitor and conserve when their total distribution and habitat preferences are incompletely known. One method employed in determining distributions of these organisms is species distribution modeling (SDM).

Methods. Using two species of narrowly endemic burrowing crayfish species as our study organisms, we sought to ground validate Maxent, a commonly used program to conduct SDMs. We used fine scale (30 m) resolution rasters of pertinent habitat variables collected from historical museum records in 2014. We then ground validated the Maxent model in 2015 by randomly and equally sampling the output from the model.

Results. The Maxent models for both species of crayfish showed positive relationships between predicted relative occurrence rate and crayfish burrow abundance in both a Receiver Operating Characteristic and generalized linear model approach. The ground validation of Maxent led us to new populations and range extensions of both species of crayfish.

Discussion. We conclude that Maxent is a suitable tool for the discovery of new populations of narrowly endemic, rare habitat specialists and our technique may be used for other rare, endemic organisms.

1   **Maxent-directed field surveys identify new populations of**

2   **narrowly endemic habitat specialists**

3

4   Cody M. Rhoden[1], William E. Peterman[2], Christopher A. Taylor[1]

5

6   [1]Illinois Natural History Survey, University of Illinois Urbana–Champaign,

7   Champaign, Illinois, United States of America

8   [2]Ohio State University, Columbus, Ohio, United States of America

9

10 Corresponding author:

11 Cody Rhoden

12

13 E-mail: codyrhoden@gmail.com

Abstract

14

15   **Background.** Rare or narrowly endemic organisms are difficult to monitor and conserve
16   when their total distribution and habitat preferences are incompletely known. One method
17   employed in determining distributions of these organisms is species distribution modeling
18   (SDM).

19   **Methods.** Using two species of narrowly endemic burrowing crayfish species as our
20   study organisms, we sought to ground validate Maxent, a commonly used program to conduct
21   SDMs. We used fine scale (30 m) resolution rasters of pertinent habitat variables collected from
22   historical museum records in 2014. We then ground validated the Maxent model in 2015 by
23   randomly and equally sampling the output from the model.

24   **Results.** The Maxent models for both species of crayfish showed positive relationships
25   between predicted relative occurrence rate and crayfish burrow abundance in both a Receiver
26   Operating Characteristic and generalized linear model approach. The ground validation of
27   Maxent led us to new populations and range extensions of both species of crayfish.

28   **Discussion.** We conclude that Maxent is a suitable tool for the discovery of new
29   populations of narrowly endemic, rare habitat specialists and our technique may be used for
30   other rare, endemic organisms.

Introduction

32   Understanding the factors influencing species distributions and habitat selection are

33   critical to researchers (Baldwin, 2009) because rare species or those with small native ranges

34   (defined herein as those occurring in a single river drainage or a 1000 sq. km area), are difficult

35   to monitor and conserve when their total distribution and habitat preferences are not completely

36   known. These problems can be addressed using species distribution models (SDMs), which are

37   correlative models using environmental and/or geographic information to explain observed

38   patterns of species occurrences (Elith & Graham, 2009). SDMs can provide useful information

39   for exploring and predicting species distributions across the landscape (Elith et al., 2011).

40   Models estimated from species observations can also be applied to produce measures of habitat

41   suitability (Franklin, 2013). This information can be useful for detecting unknown populations of

42   rare, endemic, or threatened species (e.g. Williams et al., 2009; Rebelo & Jones, 2010; Peterman,

43   Crawford & Kuhns, 2013; Searcy & Shaffer, 2014; Fois et al., 2015). SDMs can also limit search

44   efforts by selecting suitable sampling areas *a priori*, leading to a cost-effective and efficient use

45   of sampling effort (Fois et al., 2015).

46   One of the most widely used SDMs in recent years is the program Maxent (Kramer-

47   Schadt et al., 2013). Maxent is a presence-only modeling algorithm using predictor variables

48   such as climatic and remotely sensed variables (Phillips, Anderson & Schapire, 2006; Phillips &

49   Dudík, 2008). These data are used to predict the relative occurrence rate (ROR) of a focal species

50   across a predefined landscape (Fithian & Hastie, 2013). Recent studies focusing on the

51   performance of Maxent have revealed it to perform well in comparison to other SDMs (Elith et

52   al., 2006). Maxent also performs well with small sample sizes (Pearson et al., 2007; Wisz et al.,

53   2008), rare species (Williams et al., 2009; Rebelo & Jones, 2010), narrowly endemic species

54   (Rinnhofer et al., 2012), and when used as a habitat suitability index (Latif et al., 2015).

55   However, the potential for the inaccurate execution and interpretation of an SDM is well

56   documented (Baldwin, 2009; Syfert, Smith & Coomes, 2013; Fourcade et al., 2014; Guillera-

57   Arroita, Lahoz-Monfort & Elith, 2014). Specific issues surrounding the interpretation of Maxent

58   analyses include sampling bias (Phillips & Dudík, 2008; Syfert, Smith & Coomes, 2013;

59   Fourcade et al., 2014), the lack of techniques to assess model quality (Hijmans, 2012),

60   overfitting of model predictions (Elith, Kearney & Phillips, 2010; Warren & Seifert, 2011), or

61   assessment of detection probabilities (Lahoz-Monfort, Guillera-Arroita & Wintle, 2014).

62   Researchers have sought to solve the aforementioned issues by reducing sampling bias through

63   spatial filtering (Boria et al., 2014), assessing model quality with a null model approach (Raes &

64   Ter Steege, 2007), utilizing the R package ENMeval (Muscarella et al., 2014) to balance

65   goodness-of-fit and model complexity, and collecting data informative about imperfect

66   detectability (Lahoz-Monfort, Guillera-Arroita & Wintle, 2014). The utility of Maxent has also

67   been burdened with issues of model validation (Hijmans, 2012). Most model validation methods

68   involve subsets of the input data with the predictions generated by the models (Rebelo & Jones,

69   2010). Historically, validation of Maxent predictions has lacked an independent assessment of

70   model performance (Greaves, Mathieu & Seddon, 2006), such as a novel set of presence

71   locations. Recent studies have found ground validation of Maxent has been a suitable method to

72   determine the accuracy of predictions (Stirling et al., 2016). The need for independent validation

73   is especially important for rare species exhibiting a wider knowledge gap in distribution than

74   more common species (Rebelo & Jones, 2010). For example, North American primary

75   burrowing crayfishes are a poorly understood and understudied taxon for which SDMs could

76   provide novel insight into distributions and habitat relationships and thus provide an excellent

77   case study for validation of SDMs.

78   North America has the highest diversity of crayfishes worldwide (Taylor et al., 2007).

79   Within North America, 22% of the species listed as endangered or threatened in a conservation

80   review of crayfishes were primary burrowing crayfishes (Taylor et al., 2007). It is hypothesized

81   all crayfishes have the ability to construct refugia by way of burrowing down into the soil or

82   substrate (Hobbs Jr., 1981; Berrill & Chenoworth, 1982). Primary burrowing crayfishes differ

83   from stream dwelling crayfishes in their life history traits, they spend most of their life cycle

84   underground, leaving their burrows only to forage and find a mate (Hobbs Jr., 1981). This

85   difference in life history traits allow primary burrowing crayfish to persist in areas that are not

86   connected to above-ground sources of water. This persistence allows primary burrowing crayfish

87   to use habitats such as seeps, perched wetlands, and even roadside ditches.

88   Amongst the three types of burrowers, the least is known regarding the natural history of

89   primary burrowing crayfishes (Taylor et al., 2007; Moore, DiStefano & Larson, 2013) due to the

90   challenges in sampling these largely fossorial animals (Larson & Olden, 2010). However, the

91   narrowly endemic nature of North American crayfishes is well documented (Page, 1985; Taylor

92   et al., 2007; Simmons & Fraley, 2010; Morehouse & Tobler, 2013). Primary burrowing

93   crayfishes in Arkansas are no exception (Robison et al., 2008). Of the 12 species of primary

94   burrowers in Arkansas (*Fallicambarus dissitus, F. fodiens, F. gilpini, F. harpi, F. jeanae, F.*

95   *petilicarpus, F. strawni, Procambarus curdi, P. liberorum, P. parasimulans, P. regalis,* and *P.*

96   *reimeri*) six (50%) are known from only one ecoregion. The limited geographic distribution of

97   any taxa makes them more vulnerable to localized extirpation. Because these animals occur at

98   such a constrained geographic scale, it is important to understand and document their existing

99   distribution to manage and preserve current populations.

100   The rarity of and difficulties surrounding the collection of natural history information,

101   specifically habitat suitability, make primary burrowing crayfishes ideal candidates for SDMs.

102   To test the ability of Maxent to predict the distribution of suitable habitat for two narrowly

103   endemic habitat specialists, we constructed SDMs for *Fallicambarus harpi* and *Procambarus*

104   *reimeri* and validated the models using sampling data collected after completion of each SDM.

105   These species are vulnerable to population declines and are currently recorded under the

106   Endangered (*P. reimeri*) and Vulnerable (*F. harpi*) conservation status categories (Taylor et al.,

107   2007) based on modifications to or reductions of habitat in their already restricted ranges. Both

108   crayfishes are endemic to the Ouachita Mountains Ecoregion (OME; Woods et al., 2004), which

109   is characterized by remnant pine-bluestem (*Pinus-Schizachyrium*) communities and silty loam

110   soil (Hlass, Fisher & Turton, 1998). We used these two narrowly endemic species to reinforce

111   the performance of Maxent with small sample sizes and rare species along with addressing

112   problems associated with Maxent to maximize the accuracy of our predictions. We also sought to

113   investigate the use of Maxent to identify suitable habitat and locate new occurrences of both

114   species of crayfish.

Materials and Methods

115

Presence data and environmental variables

116

117   To determine habitat requirements of *F. harpi* and *P. reimeri*, we queried natural history

118   museums or databases (Illinois Natural History Survey Crustacean Collection, the National

119   Museum of Natural History Smithsonian Institution, and the Arkansas Department of Natural

120   Heritage) for historic locations of both species, and a subset of those locations were visited in

121   2014 (Rhoden, Taylor & Peterman, 2016). We sampled only those sites with confirmed

122   presences in the past 20 years that were not based on obvious misidentifications occurring well

123   outside of the known range of each species known in 2014 and were able to be located given

124   historical information associated with respective collection events. At each location, we

125   measured habitat variables hypothesized to determine burrow placement: percent tree canopy

126   cover, percent herbaceous ground cover, stem density, the number of burrows, presence of

127   standing water at the site, remotely sensed variables and the presence or absence of hydrophilic

128   sedges. We found canopy cover and the presence of hydrophilic sedges were the most important

129   factors in predicting crayfish abundance (Rhoden, Taylor & Peterman, 2016).

130   The presence locations used for the Maxent analysis, based on the field surveys of 2014,

131   consisted of 58 locations for *F. harpi* (of which 56 were used for the SDM analysis) and 53

132   locations for *P. reimeri* (of which 50 were used for the SDM analysis). To minimize spatial

133   autocorrelation, a subset of the original presence data was used. All duplicate presence locations

134   falling within the same cell of a 30 m resolution raster were removed before the SDM analysis.

135   The selected presence locations used for the SDM analysis were near (<90 m) primary,

136   secondary, and tertiary roadways. The environmental variables used for the SDM analysis

137   consisted of canopy cover, elevation, distance to nearest waterbody, compound topographic

138   index value (CTI), and solar radiation value (Table 1). These habitat variables reflect habitat

139   characteristics associated with *F. harpi* (Robison & Crump, 2004) and *P. reimeri* (Robison,

140   2008), and other primary burrowing crayfish species (Hobbs Jr., 1981; Welch & Eversole, 2006;

141   Loughman, Simon & Welsh, 2012). Canopy cover was estimated using a United States Forest

142   Service percent canopy raster (National Land Cover Database 2011; 30 m). Elevation was

143   estimated using a United States Geological Survey digital elevation map (DEM; 10 m). Distance

144   to nearest water body was estimated by constructing a raster of the Euclidean distance from all

145   permanent waterbodies. Compound topographic index values were determined using the

146   Geomorphometry and Gradient Metrics (version a1.0; Evans et al., 2010) toolbox; this metric is

147   a representation of surface wetness across the landscape (Evans et al., 2010). CTI is a steady

148   state wetness index, where a larger CTI value represents areas topographically suitable for water

149   accumulation. We measured solar radiation by calculating the watt-hour/m^2 of the delineated

150   sampling area using the Area Solar Radiation tool in ArcMap (Table 1). These values were

151   calculated using digital elevation maps (National Elevation Dataset http://ned.usgs.gov/ accessed

152   07/21/2014) and surface water maps (National Hydrography Dataset

153   http://nhd.usgs.gov/index.html accessed 07/21/2014). The entire OME was used as a delineation

154   for both species of crayfish in the SDM analysis. Each surface was resampled to a common

155   resolution of 30 m to match the resolution of the canopy surface.

156   Table 1. Environmental Variables.
157   Description, origin, resolution, general statistics, and units of environmental variables used in the
158   Maxent analysis of two primary burrowing crayfish species (*Fallicambarus harpi* and
159   *Procambarus reimeri*) in western Arkansas.

Variable	Description	Source	Resolution	Min/max(unit)	μ (sd)
Canopy cover	Percent tree canopy cover	National Land Cover Database 2011 USFS	30 m	0/100 (% cover)	52.23(448)
Elevation	Digital elevation model of the study site	USGS National Elevation Dataset	10 m	50.50/818.96 (m)	229(100.85)
Distance to nearest waterbody	Euclidean distance to nearest	ESRI Spatial Analyst Tools; National	10 m	0/1740.26 (m)	187.48(161.04)

	permanent waterbody across the study site	Hydrology Dataset composed of stream segments of study site			
Compound topographic index	A function of slope and the upstream contributing area per unit width orthogonal to the flow direction (Evans et al., 2010)	ArcGIS Geomorphometry and Gradient Metrics Toolbox 2.0 (Evans et. al., 2010); National Elevation Dataset	10 m	2.67/27.58 (index value)	7.58(1.94)
Solar radiation	Incoming solar radiation value (watt hours per m^2) based on direct and diffuse insolation from the unobstructed sky directions	ESRI Spatial Analyst Tools; National Elevation Dataset	10 m	3542.41/64134 (watthours/m^2)	5955.14(116.01)

160

Maxent analysis

162   We created species-specific distribution models using Maxent (version 3.3k; Phillips,

163   Anderson & Schapire, 2006). For each species, we generated 2500 random background points

164   within 10 km^2 polygons that were situated around the historic museum localities for which we

165   confirmed species presence in the field in 2014. This approach follows Peterman et al. 2013 and

166   was implemented to reduce model bias described by Phillips 2008. We fit a full model for each

167   species, and used the ENMeval package (Muscarella et al., 2014) in program R (R 3.1.1; R

168   Development Core Team, 2014) to tune the Maxent model parameter settings minimizing the

169   SDM model AICc. ENMeval automatically executes Maxent across a range of settings and

170   outputs evaluation metrics to aid in identifying settings balancing model fit and predictive ability

171   (Muscarella et al., 2014). Using jackknife and the standard settings, this analysis suggested the *F.*

172   *harpi* model should be fit with a betamultiplier of 2.5 and linear, quadratic, and hinge features,

173   and the *P. reimeri* model should be fit with a betamultiplier of 1.5 and linear, quadratic, and

174   hinge features to provide the most parsimonious fit to our data. We then re-ran each species'

175 model using the refined regularization multiplier and feature classes to increase the rigor in
176 building and evaluating the distribution model for each species based on presence only data.

177 We assessed the performance of the tuned models using the null model approach of Raes
178 and ter Steege (2007) with package ENMtools (Warren, Glor & Turelli, 2010). We generated
179 two groups of 999 random data sets containing 56 and 50 samples, which correspond to the
180 number of presence locations used for *F. harpi* and *P. reimeri* (respectively) in the initial model.
181 These points were drawn without replacement from the OME delineation used in the initial
182 model. Both model Area Under the Curve (AUC) values were compared to the 95th percentile of
183 the null AUC frequency distribution.

184 The final Maxent models were calculated with the maximum number of iterations set to
185 5000 and the analysis of variable importance was measured by jackknife and response curves.
186 The form of replication used was bootstrap. These settings, the refined regularization multiplier
187 and feature classes, and the recommended default values were used for our final Maxent model
188 runs. Due to the endemic nature of both species and the small amount of presence locations in
189 the initial model, we did not include a bias file or spatial filtering.

Field sampling and validation

191 The refined Maxent models (one for each species) were used to select 80 semi-random
192 sampling sites for each species within the OME. These sites were semi-random because we
193 restricted our sampling to areas of public access (roadside ditches). We sampled thirteen counties
194 encompassing the known range of both species of primary burrowing crayfish: from east to west,
195 those counties were Pulaski, Saline, Perry, Garland, Hot Spring, Clark, Yell, Montgomery, Pike,
196 Scott, Howard, Polk, and Sevier (Fig 1). The Maxent output for both species was discretized into
197 four categories based on the relative occurrence rate (ROR; Fithian and Hastie, 2013). The
198 Maxent output is considered a relative occurrence rate because the presence data are proportional
199 but not equal to occurrence. The first category ranged from an ROR of 0 to the lowest presence
200 threshold (LPT =minimum training presence threshold of Maxent software; Wisz et al., 2008) of
201 each species. The LPT is the smallest logistic value associated with one of the observed species
202 localities. The second class ranged from the LPT to 50% of the maximum ROR of each species.
203 The third category ranged from 50% of the maximum ROR to 75% of the maximum ROR of

204   each species. The fourth category ranged from 75% of the maximum ROR to the maximum ROR

205   of each species.

206

207   Figure 1: Sampling Sites. Map depicting the location of sites sampled in western Arkansas in the
208   spring of 2015 based on the predictions from a Maxent analysis of two primary burrowing
209   crayfish species (*Fallicambarus harpi* and *Procambarus reimeri*).
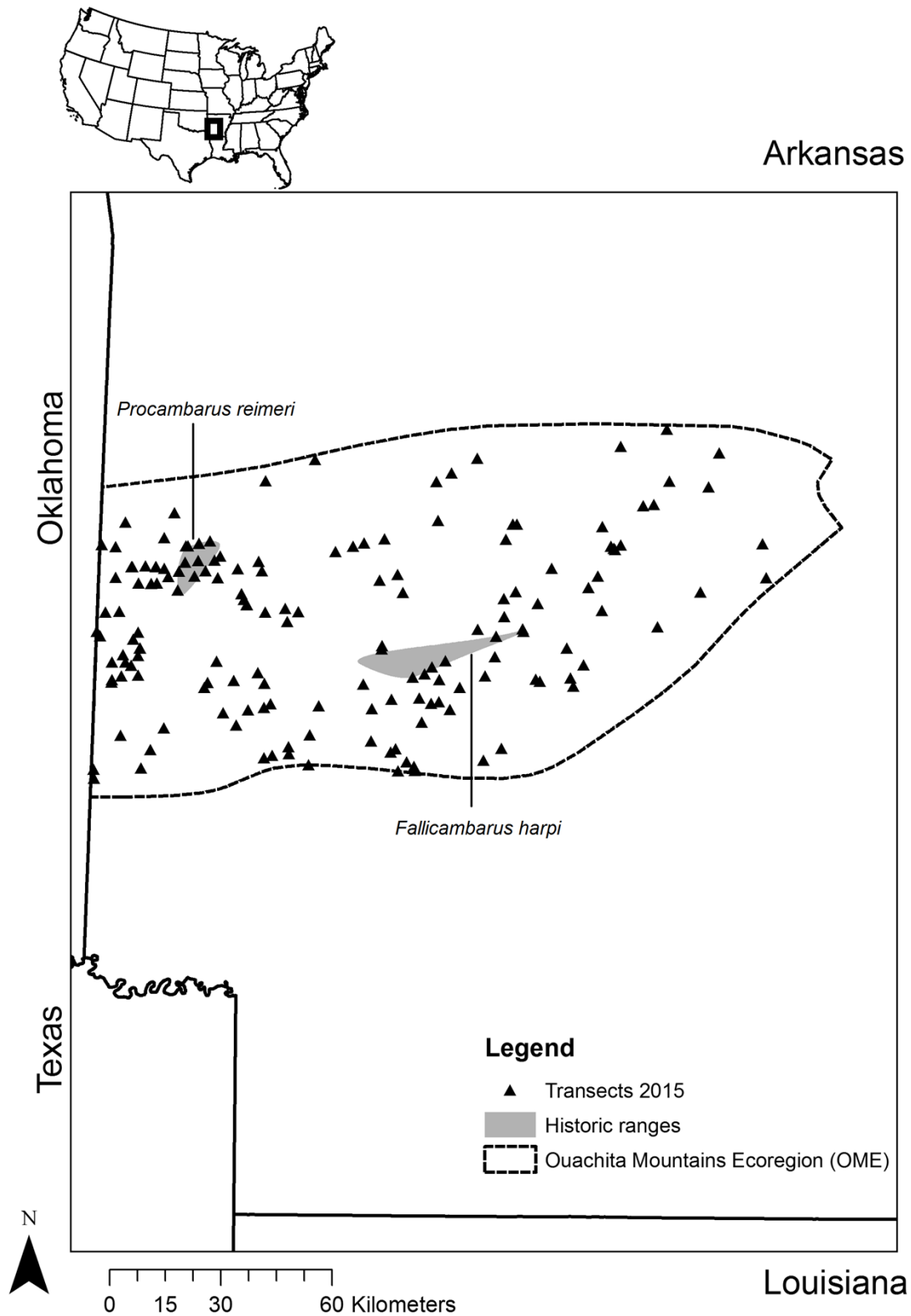
211   The final Maxent model outputs for both species were placed into the described

212   categories in ArcMap. The projection of the Maxent model onto the environmental variables was

213   converted into polygons in ArcMap, which represented each category. Any polygon representing

214   a single pixel or island (one 30 m x 30 m area in original output raster) was removed. All

215   category polygons were then overlaid with a layer representing public right of ways and other

216   public areas (state parks, natural areas, etc.).

217   We generated 40 random points in each category polygon using the final polygon layer.

218   All points within each category polygon had a spatial buffer of 2 km and were checked before

219   sampling to ensure accessibility. If a point was inaccessible in the field, the next closest

220   accessible point within the respected category was chosen and sampled. To assess the accuracy

221   of the Maxent predictions, we calculated the receiver operating characteristic (ROC) and the

222   AUC for the average ROR of occupied transects vs. the average ROR of unoccupied transects

223   (Fawcett, 2006) with the pROC package in program R (Robin et al., 2011). A ROC graph is a

224   technique for visualizing, organizing and selecting classifiers based on their performance

225   (Fawcett, 2006), and displays the performance of a binary classification method

226   (presence/absence) with a continuous (Maxent prediction) ordinal output (Robin et al., 2011).

227   Furthermore, the ROC plot shows the sensitivity (proportion of correctly classified positive

228   observations) and specificity (the proportion of correctly classified negative observations) as the

229   output threshold is moved over the range of all possible values (Robin et al., 2011).

230   To assess how the number of burrows encountered on a transect related to habitat

231   variables and ROR, we fit zero-inflated (package:pscl; Zeileis, Kleiber, and Jackman, 2008) and

232   negative binomial generalized linear models (package:MASS; Venables and Ripley, 2002) for *F.*

233   *harpi* and *P. reimeri*, respectively. Zero-inflated models were fit for both species, however

234   model selection suggested that the negative binomial was a better fit for the *P. reimeri* data. The

235   response variable was the number of burrows in each transect for the *F. harpi* and *P. reimeri*

236   models. We modeled excess zeros in *F. harpi* data by including "sedge" as a predictor in the

237   zero-inflation logit model (Table 2). The sedge variable indicated the number of quadrats in a

238   transect that contained sedges. Sedge was modeled in this manner due to its significant

239   relationship with the presence of both crayfish species across the landscape (see Rhoden, Taylor,

240   & Peterman 2016), as well as our inability to accurately identify sites with sedges from spatial

241 GIS data. The predictor variables for both analyses were based on averages of habitat data

242 collected at each transect during the search of burrows in each quadrat (Table 2).

243   We assessed model convergence and fit and then adjusted the optimization algorithm as

244 needed. The full candidate model set is shown in Table 3. We compared candidate models with

245 Akaike Information Criterion corrected for small sample sizes (AICc; Akaike 1974) with the

246 package MuMIn (Barton 2014) by means of model selection and averaging described by

247 Burnham and Anderson (2002) and Luckacs et al. (2009).

248   Field sampling occurred in March and April of 2015, the period of peak activity for both

249 *F. harpi* and *P. reimeri* (Robison & Crump, 2004; Robison, 2008). Field sampling was

250 conducted under funding agency Scientific Collection Permit number 030620151. At each

251 sampling point, one 50-m linear transect was searched for the presence of burrows in six 1-m^2

252 quadrats placed at 10 m intervals along each transect. Within a sampling polygon, the area

253 surrounding the transect was also thoroughly searched for burrows. If burrows were present

254 along the transect, quadrat, or within the vicinity of the transect, animals were captured with

255 hand excavation by using a hand shovel to slowly dig around the burrow entrance and inserting

256 one's arm into the burrow feeling for the crayfish. This method was chosen over other methods

257 due to the success rate and limited amount of time spent at each burrow location (Ridge et al.,

258 2008).

259 Table 2. Model Variables
260 Variables and their descriptions for generalized linear model analysis of two primary burrowing
261 crayfishes in Arkansas (*Fallicambarus harpi* and *Procambarus reimeri*). Quadrats were 1 m^2 and
262 transects were 50 m in length.

Variable	Description
Sedge	Presence of hydrophilic sedge in transect (binary: yes/no)
Herb	% herbaceous ground cover measured in each quadrat, averaged across each transect
Solar	Incoming solar radiation value (watt-hour/m^2) averaged across each transect location based on direct and diffuse insolation from the unobstructed sky directions (ArcGIS, Environmental System Research Institute, Redlands, California)
Water_dist	Euclidean distance to nearest waterbody calculated at central point (25 m) of each transect location (National Hydrography Dataset; http://nhd.usgs.gov/index.html)

CTI	Average compound topographic index value calculated for each transect location (Evans et al. 2010)
Soil1, Soil2	Transformed soil composition (% sand, silt, clay) value calculated for each soil sample averaged across each transect (van den Boogaart et al. 2014)
Mxnt	Average relative occurrence rate (ROR) calculated for each transect

263

264

265 Table 3. Candidate Models
266 Candidate models in the generalized linear model analysis for *Fallicambarus harpi* and
267 *Procambarus reimeri* in Arkansas. The response variable used in each model was burrow
268 abundance/presence in each 50 m transect. See Table 2 for variable names.

Model name:	Variables
Mod 1(global)	herb + solar + water_dist + cti + soil1 + soil2 + mxnt
Mod 2	mxnt+ soil2
Mod 3	soil2 + soil1
Mod 4	solar + water_dist
Mod 5	cti + mxnt
Mod 6	herb + water_dist + cti
Mod 7	mxnt
Mod 8	mxnt + soil1

269

Results

Maxent analysis

272   The AUC converged to 0.959 and 0.976 for the final *F. harpi* and *P. reimeri* models,

273 respectively. The model for *F. harpi* converged after 520 iterations and the model for *P. reimeri*

274 converged after 420 iterations. Both models were significantly better than the random AUC

275 estimations from the null models ($p<0.01$). Of the parameters included in the model, canopy

276 cover was the variable with the highest percent contribution for both species (48.8% and 47.2%

277 *F. harpi* and *P. reimeri*, respectively; Table 4). Both species showed a steady decline in the

278 probability of presence as canopy cover increased. The variable with the highest gain when used

279 in isolation was elevation for both species (Table 4). An elevation between 150 m and 200 m was

280 most suitable for *F. harpi* and between 300 m and 350 m was most suitable for *P. reimeri*. The

281 concentration of the highest ROR was centered around the presence locations for both species

282 (Fig 2**)**. The LPT was 0.07 for *F. harpi* and 0.26 for *P. reimeri*. In the *F. harpi* model, 10% of the

283 area in the OME was predicted to be above the LPT. In the *P. reimeri* model, 2% of the OME

284 was predicted to be above the LPT (Table 5).

285 Table 4: Maxent Results.
286 Percent contribution and permutation importance of each environmental variable analyzed in the
287 final Maxent models for two primary burrowing crayfish species (*Fallicambarus harpi* and
288 *Procambarus reimeri*) in western Arkansas.

Fallicambarus harpi		
Variable	Percent contribution	Permutation importance
Canopy	48.8	15.7
Elevation	37.9	56.3
CTI	7.6	2
Solar	4	25.4
Distance to nearest waterbody	1.8	0.6
Procambarus reimeri		
Variable	Percent contribution	Permutation importance
Canopy	47.2	27.5
Elevation	39.8	41.9
Distance to nearest waterbody	7.1	16.8
CTI	5.5	9
Solar	0.4	4.9

289
290
291 Table 5: Ground Validation Statistics.
292 (A) Threshold values (relative occurrence rate; ROR), land area (ha), and percentage of Ouachita
293 Mountains Ecoregion (OME) of each relative occurrence category; and (B) number of presence
294 and absence quadrats, average canopy cover (%) of quadrats sampled in each relative occurrence
295 category, and percentage of quadrats in each relative occurrence category with sedges present
296 from the field sampling based on Maxent models for two primary burrowing crayfish species
297 (*Fallicambarus harpi* and *Procambarus reimeri*) in western Arkansas.

<div align="center">A.</div>

Species	Category 1	Category 2	Category 3	Category 4
	Thresholds (ROR)			
Fallicambarus harpi	0.00 – 0.07	0.07 – 0.44	0.44 – 0.66	0.66 – 0.88
Procambarus reimeri	0.00 – 0.26	0.26 – 0.42	0.42 – 0.64	0.64 – 0.85
	Land area (ha)			
Fallicambarus harpi	1374105	143139	21894	4996
Procambarus reimeri	1515441	15209	9756	3728

	Percentage of OME			
Fallicambarus harpi	89	9	1	<1
Procambarus reimeri	98	1	1	<1
	B.			
Variable	Category 1	Category 2	Category 3	Category 4
	Fallicambarus harpi			
Present	0	0	5	14
Absent	121	120	115	105
Average Canopy Cover	34	26	16	7
Percent Quad w/ Sedge	27	73	44	46
	Procambarus reimeri			
Present	0	12	14	15
Absent	122	106	106	105
Average Canopy Cover	38	18	19	15
Percent Quad w/ Sedge	51	55	63	45

298
299

300 Figure 2: Projection of Maxent Results. Projection of the Maxent models for (A) *Procambarus*
301 *reimeri* and (B) *Fallicambarus harpi* onto the environmental variables (Table 1) used for
302 analysis in western Arkansas. The total shaded area represents the Ouachita Mountains
303 Ecoregion (OME). Cooler colors show areas with better predicted conditions (relative occurrence
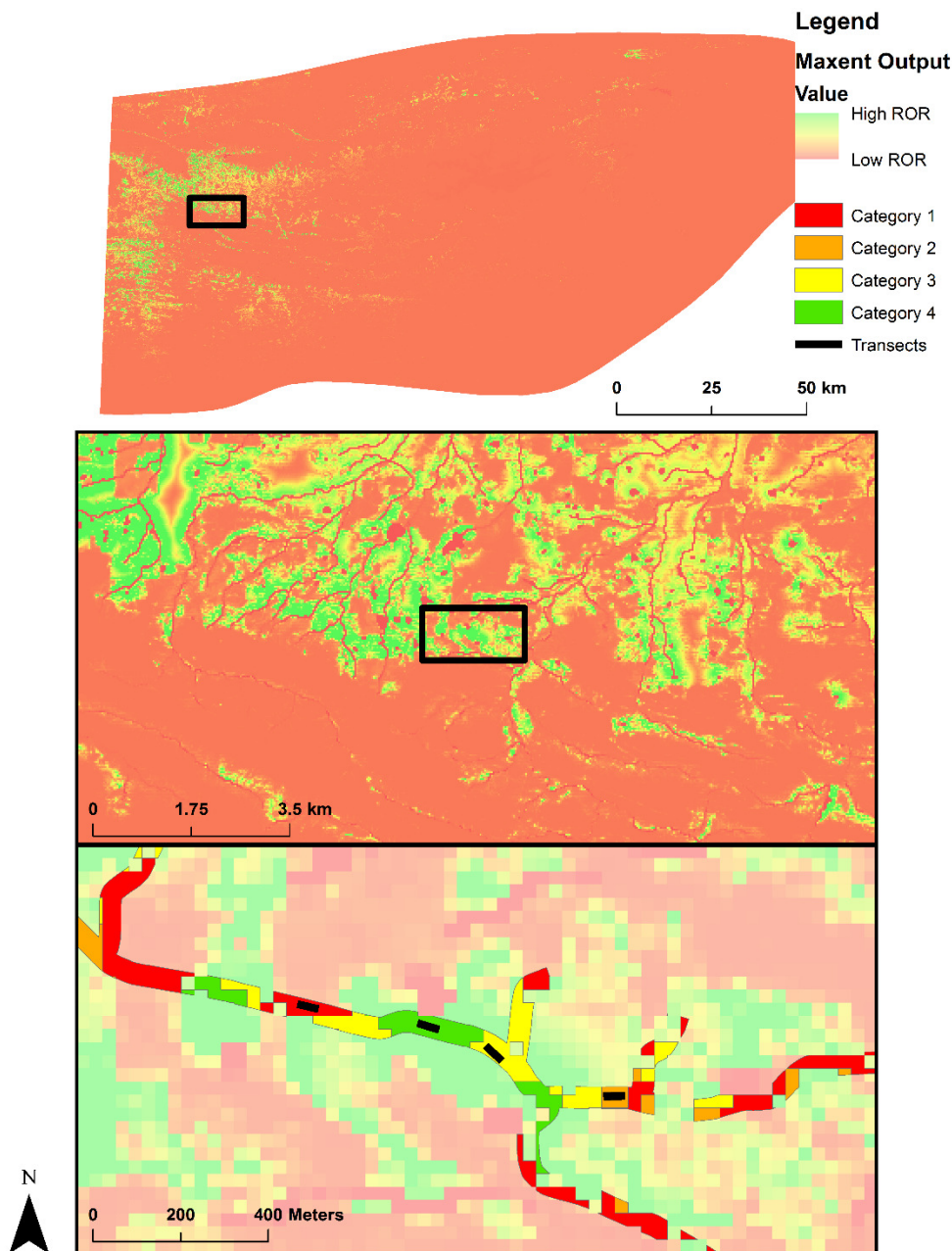304 rates [ROR]).

305

306
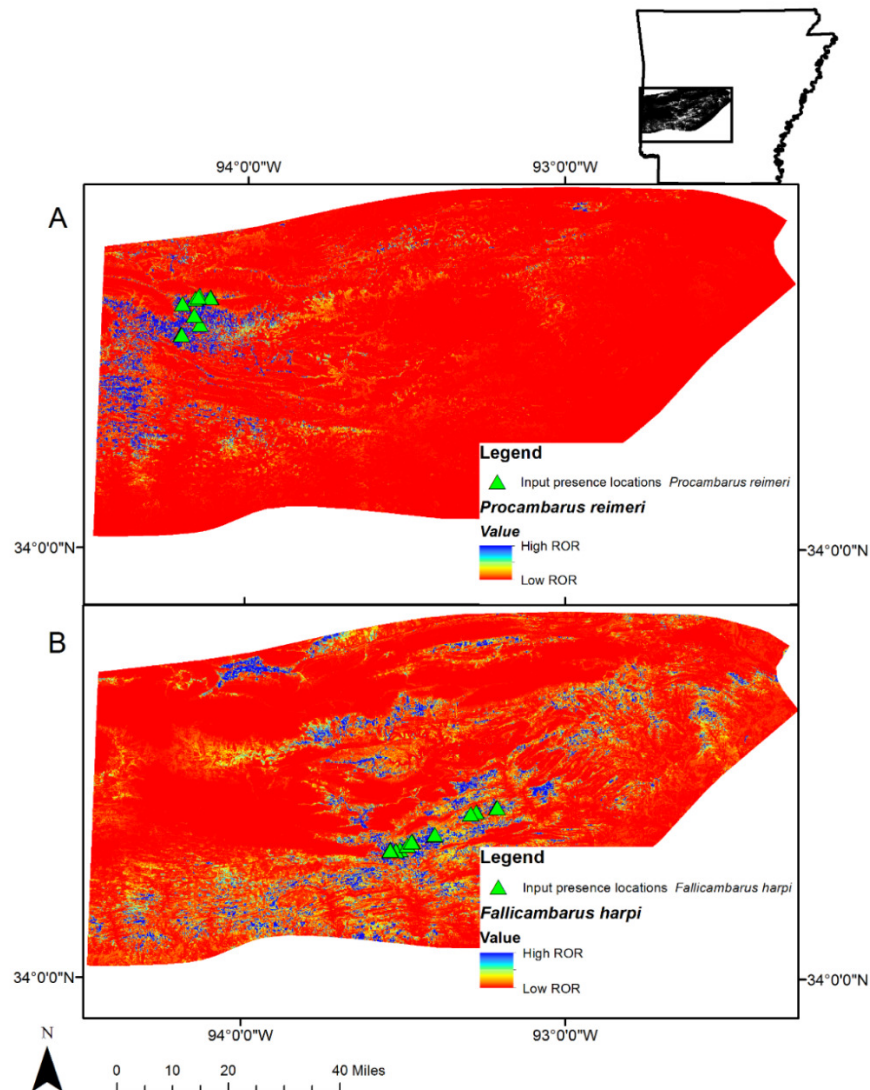
Field sampling and validation

308   All sites were sampled in the right of way of primary, secondary, and tertiary roadways

309   (Fig 3). Most (89% for *F. harpi* and 98% for *P. reimeri*) of the land area in the OME was in the

310   first (lowest ROR) category (Table 5). No individuals of either species were caught in areas

311   predicted below the LPT (category 1). Most (74%) of the presence locations for *F. harpi* were in

312   category 4 (Table 5). The presence locations for *P. reimeri* were more evenly distributed

313   between categories 2, 3, and 4 (Table 5). *Fallicambarus harpi* was captured in 19 of 480

314   quadrats within 5 of the 80 transects surveyed for the species. *Procambarus reimeri* was

315   captured in 41 of 480 quadrats within 15 of the 80 transects surveyed for the species. We counted

316 70 burrows each for *F. harpi* and *P. reimeri*. The updated range of *F. harpi* extends 2.8 km to the

317 north and 2 km to the south of its historical range while the updated range of *P. reimeri* extends

318 51.6 km to the east, 12.1 km to the south, and 19.2 km to the west of its historical range. Thus,

319 the total range for both species was approximately 265 km^2 for *F. harpi* and 1467 km^2 for *P.*

320 *reimeri* using a minimum convex polygon approach in ArcGIS encompassing all known capture

321 localities from both years and historic museum data.

322 Figure 3: Map of Ground Validation**.**
323 Map representing the sampling scheme based on the predictions from a Maxent analysis of two
324 primary burrowing crayfish species (*Fallicambarus harpi* and *Procambarus reimeri*) in western
325 Arkansas in the spring of 2015. Each color represents a relative occurrence category upon which
326 the field validation sampling procedure was based. The black lines in the lower graphic depict
327 50-m transects used to assess presence or absence of the target species at each site. The linear,
328 focused colors in the bottom graphic represent the accessible polygons in which the transect
329 sampling was carried out.

Legend

Maxent Output

Value

High ROR

Low ROR

Category 1

Category 2

Category 3

Category 4

Transects

331   The AUC for the *F. harpi* field validation was 78.67 (63.29-94.04). The AUC for the *P.*

332   *reimeri* field validation was 69.54 (56.16 – 82.92). The threshold values (prediction with the

333   highest specificity and sensitivity) were 0.48 and 0.29 for *F. harpi* and *P. reimeri*, respectively.

334   In both the *F. harpi* and *P. reimeri* models, the variable of ROR (Mxnt; Table 3) was in the top

335   (AICc ≤4) models and was shown to have a positive relationship with the abundance of crayfish

336   burrows in each transect (Table 6). Sedge was an important predictor of excess zeros in our *F.*

337   *harpi* data (Table 7). As the number quadrats in a transect containing sedges increased, the

338   likelihood of an excess zero decreased.

339   Table 6. Candidate generalized linear model results.
340   Model name, number of model parameters (K), Akaike's Information Criterion adjusted for
341   small sample size (AICc), difference in AICc (ΔAICc), Akaike cumulative weights (w_c), and log
342   liklihood (LL) for models from a suite of variables modeled with a generalized linear model
343   analysis for 2 primary burrowing crayfish species, *Fallicambarus harpi* ($n = 80$ transects) and
344   *Procambarus reimeri* ($n = 80$ transects) in Arkansas. See Tables 2 and 3 for a description of each
345   model and the variables included. Models marked with * represent inclusion of ROR parameter
346   (mxnt).

Model	K	AICc	ΔAICc	w_c	LL
Fallicambarus harpi					
Mod 5*	6	65.5	0	0.69	-26.19
Mod 1*	11	68.9	3.35	0.82	-21.50
Mod 6	7	69.3	3.78	0.92	-26.88
Mod 8*	6	71.1	5.6	0.96	-29
Mod 7*	5	71.9	6.4	0.99	-30.6
Mod 2*	6	74.3	8.8	1	-30.6
Mod 3	6	77.1	11.5	1	-32.0
Mod 4	6	80.8	15.3	1	-33.8
Procambarus reimeri					
Mod 7*	3	137.7	0	0.40	-65.68
Mod 4	4	139.3	1.67	0.57	-65.41
Mod 2*	4	139.8	2.12	0.71	-65.63
Mod 5*	4	139.8	2.17	0.85	-65.66
Mod 8*	4	139.9	2.22	0.98	-65.68
Mod 3	4	144.2	6.56	1	-67.85
Mod 6	5	145.5	7.84	1	-67.35
Mod 1*	9	150.5	12.86	1	-64.99

347
348
349   Table 7. Parameter estimates of generalized linear model analysis.
350   Conditional model-averaged parameter estimates of the full candidate models (Table 6) for two
351   primary burrowing crayfishes species (*Fallicambarus harpi* and *Procambarus reimeri*) in
352   Arkansas. See Table 2 for a description of the variables included.

Species and variable	Model-averaged estimate (SE)	$p > \lvert z \rvert$
Fallicambarus harpi		
Herb	-0.75 (1.55)	0.63
Solar	1.45 (5.14)	0.78
Water_dist	-1.93 (1.75)	0.23
CTI	-0.98 (0.38)	0.01
Soil1	-5.77 (5.43)	0.29

Soil2	2.08 (1.68)	0.21
mxnt	3.96 (1.70)	0.02
Sedge (zero-infl)	-0.50 (0.30)	0.10
Count Intercept	-1.16 (2.38)	0.63
Zero-infl Intercept	3.51 (1.28)	0.01
Procambarus reimeri		
Solar	1.70 (0.69)	0.02
Water_dist	0.11 (0.47)	0.80
Cti	0.11 (0.50)	0.82
Soil1	-0.07 (0.55)	0.90
Soil2	-0.16 (0.56)	0.78
mxnt	1.25 (0.49)	0.01
Intercept	-0.61 (0.47)	0.20

353

Discussion

355   We demonstrate that Maxent is a useful tool to predict new occurrences and the

356 distribution of suitable habitats for two narrowly endemic, rare species with unique natural

357 histories that span both terrestrial and aquatic life styles. Our models were successful in directing

358 us to new populations of both species. We used a suite of functions to assess model fit and

359 safeguard against potential pitfalls associated with the Maxent program (Phillips & Dudík, 2008;

360 Warren & Seifert, 2011; Elith et al., 2011; Hijmans, 2012; Lahoz-Monfort, Guillera-Arroita &

361 Wintle, 2014). We also used biologically relevant habitat information at a constrained

362 geographic scale to increase the accuracy of our predictions (Guisan & Thuiller, 2005). These

363 habitat variables and the scale at which we delineated them were a result of previous field

364 sampling and analysis of habitat preference of both species (Rhoden, Taylor & Peterman, 2016),

365 which revealed both crayfish to be microhabitat specialists; using open, low-herbaceous

366 microhabitats. We validated the models through a stratified sampling of our Maxent model

367 predictions based on the LPT and the maximum ROR. We then equally sampled each category

368 across the entire OME. Both models performed well in the ROC analysis and subsequent

369 generalized linear models. The ROR was shown to be positively associated with the number of

370 burrows in a transect (represented as "mxnt" in Table 7) during the generalized linear models for

371 each species. This analysis revealed that regions with higher estimated ROR not only are more

372 likely to be occupied, but will harbor more individuals. This validation resulted in an expansion

373   of both species' known ranges and the discovery of new populations. The models performed well

374   by directing sampling efforts to treeless areas on the landscape that tended to have greater

375   predicted probabilities of occurrence. However, the models did a poor job of identifying the wet,

376   low-herbaceous microhabitats most frequently associated with occurrence in the field and

377   previous studies (Robison & Crump, 2004; Robison, 2008; Rhoden, Taylor & Peterman, 2016).

378   The habitat attributes of sites in which animals were present consisted of treeless, wet,

379   low-herbaceous microhabitats. The average canopy cover for the categories above the LPT

380   (category 2, 3, and 4) was 17% for both species. Quadrats where we detected our focal crayfish

381   species had an average canopy cover of 5%. Hydrophilic sedges were present in over 90% of the

382   quadrats having *F. harpi* and *P. reimeri* but were present in less than half of the quadrats

383   predicted above the LPT (categories 2, 3, and 4). The sites recorded as being above the LPT

384   (categories 2, 3, and 4) not having the target species were treeless for the most part, but those

385   sites did not exhibit a moist microhabitat. The Maxent models thus did not capture the perched

386   water table observed across the landscape associated with other primary burrowing crayfishes

387   (Welch, Eversole & Riley, 2007). It is likely the model did not capture these moist, low

388   herbaceous habitats due to the spatial resolution and variables chosen for the Maxent analysis

389   (canopy cover, CTI, elevation, solar radiation, and distance to nearest waterbody). Future studies

390   could incorporate remotely sensed data to better identify these unique habitats.

391   The use of the LPT to determine the threshold between the probability of presence or

392   absence at any given predicted output location (Pearson et al., 2007) is well documented

393   (Rinnhofer et al., 2012; Boria et al., 2014; Fois et al., 2015). We successfully used this value in

394   our field validation techniques: no animal was captured in an area predicted below the LPT

395   (Table 3). The land area above the LPT for the *F. harpi* model comprised 10% of the OME and

396   2% for the *P. reimeri* model in Arkansas. The ROC analysis identified threshold values of 0.48

397   and 0.29 for the *F. harpi* and *P. reimeri* models, respectively, which optimized the sensitivity

398   (100 for both *F. harpi* and *P. reimeri*) and specificity (58.7 and 38.5 for *F. harpi* and *P. reimeri*,

399   respectively) of our model (Robin et al., 2011). These values are far more conservative than the

400   LPT and are based on the field validation results from both species. Using these threshold

401   metrics, the area predicted as suitable habitat for *F. harpi* and *P. reimeri* is less than 1% of the

402   OME. We recommend the use of this threshold based on the ROC analysis for a more fine-tuned

403   sampling effort for high-quality habitat for both species in the future.

404   Our SDMs used fine-scale (30 m) rasters of biological variables relevant to our two study

405   species (canopy cover, CTI, solar radiation, elevation, and distance to waterbody). In the past, it

406   has been common to use coarse (≥1 km) climatic data to construct models (e.g. Peterson, 2001;

407   Welch, Eversole & Riley, 2007; Chunco et al., 2013). The use of coarse-scale habitat variables in

408   Maxent has been addressed in previous studies (Araújo & Guisan, 2006; Jiménez-Alfaro, Draper

409   & Nogués-Bravo, 2012). Others using fine-scale inputs have found new populations of other rare

410   species such as the discovery of new breeding ponds for a salamander species in east central

411   Illinois (*Ambystoma jeffersonianum;* Peterman, Crawford & Kuhns, 2013). Using fine-scale

412   spatial surfaces of specific habitat variables for narrowly endemic habitat specialists was more

413   appropriate than the more general approach of coarse-scale climatic data due to the resolution

414   one gains with specific habitat information and fine-scale inputs. This fine-scale resolution was

415   necessary to capture elements of the microhabitat the crayfishes prefer by differentiating between

416   suitable and unsuitable habitat within anthropogenically altered habitat situated in natural

417   landscapes (e.g. roadside ditches). However, we note that the specific surfaces or resolution in

418   our study still failed to completely capture essential habitat features or indicators of preferred

419   habitat, such as sedges.

420   Conservation efforts for rare species benefit by narrowing the knowledge gap in

421   distribution information, adding localities for monitoring persistence in roadside ditches, and

422   providing habitat preference information. Our study shows Maxent was an appropriate tool to

423   analyze habitat suitability and discover populations of narrowly endemic, rare species. Our

424   method of reinvestigating museum localities, verifying species persistence, and collecting habitat

425   data from verified locations added precision to the presence locations we used for analysis. Our

426   initial surveys also added valuable information regarding the habitat preferences of both *F. harpi*

427   and *P. reimeri*, which in turn guided the selection of our habitat variables for both models. Our

428   concentrated search efforts resulted in the discovery of five new populations of *F. harpi* and 16

429   new populations of *P. reimeri* and known range expansions of approximately 91 km^2 and 1404

430   km^2, respectively.

431   # Conclusion

432   Recent studies have found ground validation of Maxent has been a suitable method to

433   determine the accuracy of predictions (Stirling et al., 2016). Our study supports this conclusion

434   and offers a unique method, incorporating historic museum localities to inform an SDM of

435   pertinent habitat variables and validating the localities before conducting the SDM. We have also

436   shown Maxent works well with narrowly endemic, rare habitat specialists and fine scale (30 m)

437   raster inputs. Constructing models followed by ground validation has added valuable habitat

438   information to two spatially restricted, understudied species and illustrates the potential

439   effectiveness of such a strategy for other rare habitat specialists.

440   ## Acknowledgements

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19:716–723.

Araújo MB., Guisan A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677–1688. DOI: 10.1111/j.1365-2699.2006.01584.x.

Baldwin RA. 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11:854–866. DOI: 10.3390/e11040854.

Berrill M., Chenoworth B. 1982. The Burrowing Ability of Nonburrowing Crayfish. *American Midland Naturalist* 108:199–201.

Boria RA., Olson LE., Goodman SM., Anderson RP. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275:73–77. DOI: 10.1016/j.ecolmodel.2013.12.012.

Chunco AJ., Phimmachak S., Sivongxay N., Stuart BL. 2013. Predicting Environmental Suitability for a Rare and Threatened Species (Lao Newt, Laotriton laoensis) Using Validated Species Distribution Models. *PLoS ONE* 8. DOI: 10.1371/journal.pone.0059853.

Elith J., Graham CH. 2009. Do they? How do they? Why do they differ? on finding reasons for differing performances of species distribution models. *Ecography* 32:66–77. DOI: 10.1111/j.1600-0587.2008.05505.x.

Elith J., Graham CH., Anderson RP., Dudık M., Ferrier S., Guisan A., Hijmans RJ., Huettmann F., Leathwick JR., Lehmann A., Li J., Lohmann LG., Loiselle BA., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton JM., Peterson AT., Phillips SJ., Richardson K., Scachetti-Pereira, Ricardo S, Schapire RE., Soberon J., Williams S., Wisz MS., Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.

Elith J., Kearney M., Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1:330–342. DOI: 10.1111/j.2041-210X.2010.00036.x.

Elith J., Phillips SJ., Hastie T., Dudík M., Chee YE., Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17:43–57. DOI: 10.1111/j.1472-4642.2010.00725.x.

Evans J., Oakleaf J., Cushman S., Theobald D. 2010. An ArcGIS Toolbox for Surface Gradient and Geomorphic Modeling, version a1.0.

Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27:861–874. DOI: 10.1016/j.patrec.2005.10.010.

Fithian W., Hastie T. 2013. Finite-sample equivalence in statistical models for presence-absence only data. *Annals of Applied Statistics* 7:1917–1939. DOI: 10.1214/13-AOAS667.Finite-Sample.

Fois M., Fenu G., Cuena Lombraña A., Cogoni D., Bacchetta G. 2015. A practical method to speed up the discovery of unknown populations using Species Distribution Models. *Journal*

482   *for Nature Conservation* 24:42–48. DOI: 10.1016/j.jnc.2015.02.001.

483   Fourcade Y., Engler JO., Rödder D., Secondi J. 2014. Mapping species distributions with
484   MAXENT using a geographically biased sample of presence data: a performance
485   assessment of methods for correcting sampling bias. *PloS one* 9:e97122. DOI:
486   10.1371/journal.pone.0097122.

487   Franklin J. 2013. Species distribution models in conservation biogeography: developments and
488   challenges. *Diversity and Distributions* 19:1217–1223. DOI: 10.1111/ddi.12125.

489   Greaves GJ., Mathieu R., Seddon PJ. 2006. Predictive modelling and ground validation of the
490   spatial distribution of the New Zealand long-tailed bat (Chalinolobus tuberculatus).
491   *Biological Conservation* 132:211–221. DOI: 10.1016/j.biocon.2006.04.016.

492   Guillera-Arroita G., Lahoz-Monfort JJ., Elith J. 2014. Maxent is not a presence-absence method:
493   a comment on Thibaud et al. *Methods in Ecology and Evolution* 5:1192–1197. DOI:
494   10.1111/2041-210X.12252.

495   Guisan A., Thuiller W. 2005. Predicting species distribution: Offering more than simple habitat
496   models. *Ecology Letters* 8:993–1009. DOI: 10.1111/j.1461-0248.2005.00792.x.

497   Hijmans RJ. 2012. Cross-validation of species distribution models: Removing spatial sorting bias
498   and calibration with a null model. *Ecology* 93:679–688. DOI: 10.1890/11-0826.1.

499   Hlass LJ., Fisher WL., Turton DJ. 1998. Use of the Index of Biotic Integrity to Assess Water
500   Quality in Forested Streams of the Ouachita Mountains Ecoregion, Arkansas. *Journal of*
501   *Freshwater Ecology* 13:181–192. DOI: 10.1080/02705060.1998.9663606.

502   Hobbs Jr. HH. 1981. *The Crayfishes of Georgia*. City of Washington: Smithsonian Contributions
503   to Zoology.

504   Jiménez-Alfaro B., Draper D., Nogués-Bravo D. 2012. Modeling the potential area of occupancy
505   at fine resolution may reduce uncertainty in species range estimates. *Biological*
506   *Conservation* 147:190–196. DOI: 10.1016/j.biocon.2011.12.030.

507   Kramer-Schadt S., Niedballa J., Pilgrim JD., Schröder B., Lindenborn J., Reinfelder V., Stillfried
508   M., Heckmann I., Scharf AK., Augeri DM., Cheyne SM., Hearn AJ., Ross J., Macdonald
509   DW., Mathai J., Eaton J., Marshall AJ., Semiadi G., Rustam R., Bernard H., Alfred R.,
510   Samejima H., Duckworth JW., Breitenmoser-Wuersten C., Belant JL., Hofer H., Wilting A.
511   2013. The importance of correcting for sampling bias in MaxEnt species distribution
512   models. *Diversity and Distributions* 19:1366–1379. DOI: 10.1111/ddi.12096.

513   Lahoz-Monfort JJ., Guillera-Arroita G., Wintle BA. 2014. Imperfect detection impacts the
514   performance of species distribution models. *Global Ecology and Biogeography* 23:504–
515   515. DOI: 10.1111/geb.12138.

516   Larson ER., Olden JD. 2010. Latent extinction and invasion risk of crayfishes in the southeastern
517   United States. *Conservation Biology* 24:1099–1110. DOI: 10.1111/j.1523-
518   1739.2010.01462.x.

519   Latif QS., Saab VA., Mellen-Mclean K., Dudley JG. 2015. Evaluating habitat suitability models
520   for nesting white-headed woodpeckers in unburned forest. *The Journal of Wildlife*

521   *Management* 79:263–273. DOI: 10.1002/jwmg.842.

522   Loughman ZJ., Simon TP., Welsh SA. 2012. Occupancy rates of primary burrowing crayfish in
523   natural and disturbed large river bottomlands. *Journal of Crustacean Biology* 32:557–564.
524   DOI: 10.1163/193724012X637339.

525   Moore MJ., DiStefano RJ., Larson ER. 2013. An assessment of life-history studies for USA and
526   Canadian crayfishes: identifying biases and knowledge gaps to improve conservation and
527   management. *Freshwater Science* 32:1276–1287. DOI: 10.1899/12-158.1.

528   Morehouse RL., Tobler M. 2013. Crayfishes (Decapoda : Cambaridae) of Oklahoma:
529   identification, distributions, and natural history. *Zootaxa* 3717:101–157. DOI:
530   10.11646/zootaxa.3717.2.1.

531   Muscarella R., Galante PJ., Soley-Guardia M., Boria R a., Kass JM., Uriarte M., Anderson RP.
532   2014. ENMeval: An R package for conducting spatially independent evaluations and
533   estimating optimal model complexity for Maxent ecological niche models. *Methods in*
534   *Ecology and Evolution* 5:1198–1205. DOI: 10.1111/2041-210X.12261.

535   Page L. 1985. The crayfishes and shrimps (Decapoda) of Illinois. *Illinois Natural History Survey*
536   *Bulletin* 33.

537   Pearson RG., Raxworthy CJ., Nakamura M., Townsend Peterson A. 2007. Predicting species
538   distributions from small numbers of occurrence records: a test case using cryptic geckos in
539   Madagascar. *Journal of Biogeography* 34:102–117. DOI: 10.1111/j.1365-
540   2699.2006.01594.x.

541   Peterman WE., Crawford JA., Kuhns AR. 2013. Using species distribution and occupancy
542   modeling to guide survey efforts and assess species status. *Journal for Nature Conservation*
543   21:114–121. DOI: 10.1016/j.jnc.2012.11.005.

544   Peterson AT. 2001. Predicting Species ℠ Geographic Distributions Based on Ecological
545   Niche Modeling. *The Condor* 103:599–605. DOI: 10.1650/0010-5422(2001)103.

546   Phillips SJ., Anderson RP., Schapire RE. 2006. Maximum entropy modeling of species
547   geographic distributions. *Ecological Modelling* 190:231–259. DOI:
548   10.1016/j.ecolmodel.2005.03.026.

549   Phillips SJ., Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and
550   a comprehensive evaluation. *Ecography* 31:161–175. DOI: 10.1111/j.2007.0906-
551   7590.05203.x.

552   R Development Core Team. 2014. R: A Language and Environment for Statistical Computing.

553   Raes N., Ter Steege H. 2007. A null-model for significance testing of presence-only species
554   distribution models. *Ecography* 30:727–736. DOI: 10.1111/j.2007.0906-7590.05041.x.

555   Rebelo H., Jones G. 2010. Ground validation of presence-only modelling with rare species: A
556   case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). *Journal*
557   *of Applied Ecology* 47:410–420. DOI: 10.1111/j.1365-2664.2009.01765.x.

558   Rhoden CM., Taylor CA., Peterman WE. 2016. Highway to heaven? Roadsides as preferred

559   habitat for two narrowly endemic crayfish. *Freshwater Science* 35. DOI: 10.1086/686919.

560 Ridge J., Simon TP., Karns D., Robb J. 2008. Comparison of Three Burrowing Crayfish Capture
561   Methods Based on Relationships with Species Morphology, Seasonality, and Habitat
562   Quality. *Journal of Crustacean Biology* 28:466–472.

563 Rinnhofer LJ., Roura-Pascual N., Arthofer W., Dejaco T., Thaler-Knoflach B., Wachter GA.,
564   Christian E., Steiner FM., Schlick-Steiner BC. 2012. Iterative species distribution modelling
565   and ground validation in endemism research: an Alpine jumping bristletail example.
566   *Biodiversity and Conservation* 21:2845–2863. DOI: 10.1007/s10531-012-0341-z.

567 Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J-C., Müller M. 2011. pROC: an
568   open-source package for R and S+ to analyze and compare ROC curves. *BMC*
569   *bioinformatics* 12:77. DOI: 10.1186/1471-2105-12-77.

570 Robison HW. 2008. *Distribution, life history aspects, and conservation status of three Ouachita*
571   *Mountain crayfishes: Procambarus tenuis, P. reimeri, and Orconectes menae*. Hot Springs,
572   AR.

573 Robison HW., Crump B. 2004. Distribution, natural history aspects, and status of the Arkansas
574   endemic crayfish, Fallicambarus harpi Hobbs and Robison, 1985. *Journal of the Arkansas*
575   *Academy of Science* 58:91–94.

576 Robison HW., McAllister C., Carlton C., Tucker G. 2008. The Arkansas endemic biota: An
577   update with additions and deletions. *Journal of the Arkansas Academy of Science* 62.

578 Searcy CA., Shaffer HB. 2014. Field validation supports novel niche modeling strategies in a
579   cryptic endangered amphibian. *Ecography* EV:EV1-10. DOI: 10.1111/ecog.00733.

580 Simmons JW., Fraley SJ. 2010. Distribution, Status, and Life-History Observations of Crayfishes
581   in Western North Carolina. *Southeastern Naturalist* 9:79–126. DOI: 10.1656/058.009.s316.

582 Stirling DA., Boulcott P., Scott BE., Wright PJ. 2016. Using verified species distribution models
583   to inform the conservation of a rare marine species. *Diversity and Distributions* 22:808–
584   822. DOI: 10.1111/ddi.12447.

585 Syfert MM., Smith MJ., Coomes DA. 2013. The effects of sampling bias and model complexity
586   on the predictive performance of MaxEnt species distribution models. *PloS one* 8. DOI:
587   10.1371/journal.pone.0055158.

588 Taylor CA., Schuster GA., Cooper JE., DiStefano RJ., Eversole AG., Hamr P., Hobbs III HH.,
589   Robison HW., Skelton CE., Thoma RF. 2007. A reassessment of the conservation status of
590   crayfishes of the United States and Canada after 10+ years of increased awareness.
591   *Fisheries* 32:372–389. DOI: 10.1577/1548-8446(2007)32.

592 Warren DL., Glor RE., Turelli M. 2010. ENMTools: A toolbox for comparative studies of
593   environmental niche models. *Ecography* 33:607–611. DOI: 10.1111/j.1600-
594   0587.2009.06142.x.

595 Warren DL., Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model
596   complexity and the performance of model selection criteria. *Ecological Applications*
597   21:335–342. DOI: 10.1890/10-1171.1.

598   Welch SM., Eversole AG. 2006. The occurrence of primary burrowing crayfish in terrestrial
599   habitat. *Biological Conservation* 130:458–464. DOI: 10.1016/j.biocon.2006.01.007.

600   Welch SM., Eversole AG., Riley J. 2007. Using the spatial information implicit in the habitat
601   specificity of the burrowing crayfish Distocambarus crockeri to identify a lost landscape
602   component. *Ecography* 30:349–358. DOI: 10.1111/j.2007.0906-7590.04815.x.

603   Williams JN., Seo C., Thorne J., Nelson JK., Erwin S., O'Brien JM., Schwartz MW. 2009. Using
604   species distribution models to predict new occurrences for rare plants. *Diversity and*
605   *Distributions* 15:565–576. DOI: 10.1111/j.1472-4642.2009.00567.x.

606   Wisz MS., Hijmans RJ., Li J., Peterson AT., Graham CH., Guisan A. 2008. Effects of sample
607   size on the performance of species distribution models. *Diversity and Distributions* 14:763–
608   773. DOI: 10.1111/j.1472-4642.2008.00482.x.

609   Woods AJ., Foti TL., Chapman SS., Omernik JM., Wise JA., Murray EO., Prior WL., Pagan
610   JBJ., Comstock JA., Radford M. 2004. Ecoregions of Arkansas (color poster with map,
611   descriptive text, summary tables, and photographs). *US Geological Survey*:map scale
612   1:1,000,000.

613