

3-2015

Development and Evaluation of Species Distribution Models for Fourteen Native Central U.S. Fish Species

Kristen L. Bouska

Gregory Whitledge

Southern Illinois University Carbondale, gwhit@siu.edu

Christopher Lant

Follow this and additional works at: http://opensiuc.lib.siu.edu/fiaq_pubs

Recommended Citation

Bouska, Kristen L., Whitledge, Gregory and Lant, Christopher. "Development and Evaluation of Species Distribution Models for Fourteen Native Central U.S. Fish Species." *Hydrobiologia* 747, No. 1 (Mar 2015): 159-176. doi:10.1007/s10750-014-2134-8.

This Article is brought to you for free and open access by the Fisheries and Illinois Aquaculture Center at OpenSIUC. It has been accepted for inclusion in Publications by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

Title: Development and evaluation of species distribution models for fourteen native Central U.S. fish species

Authors:

Kristen L. Bouska, Department of Fisheries and Wildlife Sciences, University of Missouri, Columbia, MO 65211 bouskak@missouri.edu

Gregory W. Whitledge, Center for Fisheries, Aquaculture and Aquatic Sciences and Department of Zoology, Southern Illinois University, Carbondale, IL 62901, gwhit@siu.edu

Christopher Lant, Department of Environment and Society, Quinney College of Natural Resources, Utah State University, Logan, UT 84321, chris.lant@usu.edu

Corresponding Author: Kristen L. Bouska, Department of Fisheries and Wildlife Sciences, University of Missouri, 302 Anheuser-Busch Natural Resources Building, 1105 E. Rollins St. Columbia, MO 65211. Phone: (618) 924-2592, email: bouskak@missouri.edu

Key Words: ensemble model; fish distributions; model performance; warmwater fishes; range projections

1 **Abstract**

2 Environmental change has and will continue to adversely influence aquatic communities. Efforts
3 to model impacts of environmental change on fisheries have largely focused on cold-water,
4 commercial, and recreationally-valued species, even though warmwater, non-game species have
5 important roles in ecosystem services and processes. We developed species distribution models
6 for fourteen warmwater fish species native to the Central United States and evaluated
7 environmental drivers and predictive performance. We used an ensemble model approach
8 produced by combining forecasts of five single-model techniques. Response plots and variable
9 importance calculations were used to evaluate the influence of individual variables. The
10 predictive performance of the ensemble models was assessed using area under the curve (AUC)
11 of the receiver-operating characteristic plot. Ensemble model AUC values generally performed
12 better than single-model types, suggesting ensemble models are more reliable and applicable for
13 management purposes than single models. Most models were influenced by a mix of climate,
14 land use and geophysical variables; however, climate variables were the dominant environmental
15 drivers across models. Given the high sensitivity of models to climate and land use, we expect
16 future climate and land use changes to influence distributions.

17

18 **Introduction**

19 Freshwater communities are facing losses in biodiversity far greater than terrestrial systems as a
20 result of habitat loss and fragmentation, pollution, hydrologic alteration, invasive species,
21 climate change, and overexploitation (Dudgeon *et al.*, 2006; Heino *et al.*, 2009). As
22 anthropogenic activities continue to degrade freshwater systems, it will become increasingly
23 important to develop reliable conservation planning tools that can be used to evaluate the
24 tradeoffs of management and conservation strategies under future scenarios of environmental
25 change. Substantial efforts to assess environmental change impacts, particularly climate change,
26 in fisheries have been targeted at cold-water species (upper thermal tolerance $<26^{\circ}$ C; Hokanson
27 1977), such as salmonids (Comte *et al.*, 2013), that are unable to tolerate warm water
28 temperatures. Further, fish species with little commercial or recreational value are not well-
29 represented in risk assessment and impacts modeling literature (Comte *et al.*, 2013), despite their
30 important role in ecosystem services and processes (Vanni, 2002; Dudgeon *et al.*, 2006).

31 Species' distribution models have been utilized for a number of reasons, including testing
32 ecological hypotheses, conservation planning, impact assessment, resource management,
33 ecosystem restoration, and invasive species risk assessment (Manel *et al.*, 2001; Townsend
34 Peterson, 2006; Franklin, 2009). Although there remain limitations, species distribution models
35 are one of few practical tools to assess impacts of environment change on species (Araujo &
36 Guisan 2006; Thuiller, 2007; Franklin, 2009; Morin & Thuiller, 2009; Bellard *et al.*, 2012). Due
37 to the wide application of species distribution models and access to digital databases, use of
38 species distribution modeling has greatly increased and led to the advancement of modeling
39 techniques (Guisan & Thuiller, 2005; Franklin, 2009; Comte *et al.*, 2013). There are numerous
40 modeling methods used to describe relationships between environmental predictors and a

41 species' distribution, each with different tradeoffs (Austin, 2002; Segurado & Araujo, 2004; Elith
42 *et al.*, 2006; Franklin, 2009; Aguirre-Gutierrez *et al.*, 2013). As techniques have moved from
43 climate envelope models to generalized linear models (GLM) to multivariate adaptive regression
44 splines (MARS), models have become more reflective of realistic ecological relationships
45 (Leathwick *et al.*, 2006; Elith & Leathwick, 2009). Machine learning approaches such as
46 classification and regression trees (CTA or CART), artificial neural networks (ANN), random
47 forest (RF), generalized boosted regression (GBM), and maximum entropy (MAXENT) models
48 have become common analytical tools (Elith & Leathwick, 2009; Franklin, 2009). While
49 machine learning techniques tend to have better performance, modern regression techniques
50 often allow the modeler to more clearly interpret response curves and have more control over
51 model-fitting (Elith & Leathwick, 2009; Franklin, 2009; Cianfrani *et al.*, 2011). Model selection
52 should be considered carefully as it is the major source of uncertainty in species distribution
53 modeling (Pearson *et al.*, 2006; Diniz-Filho *et al.*, 2009; Buisson *et al.*, 2010).

54 Recently, there has been a move towards ensemble models, which combine projections from
55 different model types to overcome inter-model variability (Marmion *et al.*, 2009; Thuiller *et al.*,
56 2009). The thought behind model ensembles is that a combination of unbiased model outputs,
57 each limited by its own assumptions and algorithms, will result in a more accurate prediction.
58 Multi-model ensembles can utilize different initial conditions, model classes, model parameters
59 and boundary conditions in order to provide more robust projections (Araujo & New, 2006;
60 Grenouillet *et al.*, 2011). Providing reliable projections is essential for natural resource agencies,
61 which are often tasked with difficult management decisions regarding species and habitat
62 (Mawdsley *et al.*, 2009; Cianfrani *et al.*, 2011; Dawson *et al.*, 2011; Kostyack *et al.*, 2011;
63 Lawler & Olden, 2011; Olden *et al.*, 2011).

64 A variety of species distribution modeling approaches have been used to project fish
65 distributions, including GLM (Lassalle *et al.*, 2008; Wenger *et al.*, 2011), GAM (Brosse & Lek,
66 2000; Buisson *et al.*, 2008; Lassalle *et al.*, 2008), CTA (Lyons *et al.*, 2010; Steen *et al.*, 2010),
67 and MAXENT (Labay *et al.*, 2011). The ensemble modeling approach, however, has not been
68 widely applied to fish species (but see Buisson & Grenouillet, 2009; Grenouillet *et al.*, 2011;
69 Poulos *et al.*, 2012). Within the United States, fish distribution models have largely been limited
70 to small watersheds (Steen *et al.*, 2010; Labay *et al.*, 2011), single species (Brewer *et al.*, 2007),
71 or impacts of non-native species (Sharma *et al.*, 2011; Wenger *et al.*, 2011; Poulos *et al.*, 2012).
72 Recent state-wide modeling efforts in Wisconsin (Lyons *et al.*, 2010), Maryland (Maloney *et al.*,
73 2013), and other states have resulted in the development of several distribution models for
74 common stream fishes. However, developing models at larger spatial scales that incorporate the
75 full range of natural variability associated with species' occurrence may produce different results
76 than models developed for sub-sections of a species' range (Babet-Massin *et al.*, 2010).

77 The primary objective of this research was to explore the utility of an ensemble modeling
78 approach for characterizing distributions of 14 fish species native to the Central United States, of
79 which the majority are warmwater (upper thermal tolerance $>34^{\circ}$ C; Hokanson 1977), non-game
80 species. In the development of these models we were interested in the environmental drivers of
81 each model and whether the response curves made logical sense given the known ecology of
82 each species. In the evaluation phase, we were interested in the predictive performance of
83 individual models as well as the model ensemble approach, given the large-scale environmental
84 variables selected. Finally, model performance is often explained to be an artifact of species
85 ecology in that habitats associated with a specialist species are thought to be easier to
86 discriminate than habitats for a more ubiquitous species (Segurado & Araujo, 2004; Allouche *et*

87 *al.*, 2006). Therefore, we were also interested if there were trends in model performance with
88 respect to prevalence rates.

89

90 **Materials and Methods**

91 *Data Collection*

92 Natural resource agencies from twelve states representing the central United States (AR, IL, IA,
93 KS, LA, MN, MS, MO, NE, ND, SD, and WI) were sent data requests for representative fish
94 stream sampling events between the years 2000 and 2010. Ten states returned presence/absence
95 fish data for a combined total of 15,710 unique fish sampling events for the time period, with
96 data for 218 species. Due to the high number of sampling events in Wisconsin compared to the
97 remainder of the study area, sampling events were randomly sub-sampled to provide a more even
98 spatial coverage of sampling events across the entire study area (Figure I). The list of candidate
99 fish species was narrowed down through a number of different methods. First, native range and
100 study area overlap was examined for each species. Modeling of species from a restricted range of
101 occurrence can lead to misleading conclusions (Babet-Massin *et al.*, 2010); therefore, species
102 whose known distribution was predominantly or completely contained within the study area were
103 initially selected for model development. Second, records for each species were examined to
104 make sure observations are representative of the environmental space in which it occupies
105 (Kadmon *et al.*, 2003). Although some successful models have been developed for species with
106 small sample sizes (Hernandez *et al.* 2006, Franklin *et al.* 2009), species with low recorded
107 occurrences (<30) were removed (Guisan *et al.*, 2007; Wisz *et al.*, 2008). The selection process

108 reduced the number of species down to fourteen, listed in Table I with their scientific name and
109 nomenclatural authority.

110 Environmental attributes, including geology, land use, discharge, and climate variables, were
111 quantified for each stream segment containing a fish sampling event. Physical variables included
112 elevation, slope, drainage area, and stream order from the National Hydrography Dataset Plus
113 (<http://www.horizon-systems.com/nhdplus/>), as well as number of dams in the watershed
114 (Esselman *et al.*, 2011). Geological variables included presence of karst (Tobin & Weary, 2004)
115 and 18 surficial geology variables from the USGS Surficial Geology by Major River Basins
116 dataset (http://water.usgs.gov/GIS/metadata/usgswrd/XML/mrb_e2rf1_sgeol.xml). Land use
117 variables included row crops, wheat, forest, grassland, developed, and water for the year in
118 which each fish sampling event took place and originated from the National Agricultural
119 Statistics Service's Cropland Data Layer (<http://nassgeodata.gmu.edu/CropScape/>). Land use and
120 surficial geology classes were calculated as a percentage within a stream segment's watershed.

121 Discharge variables included the 10th, 50th and 90th percentile of flow, and a flow variability
122 index (90th/10th percentiles). Discharge variables were calculated using regression models
123 developed and evaluated using USGS gage data (<http://waterdata.usgs.gov/nwis>) for the study
124 area. Climate variables were obtained through the National Climate Data Center
125 (<http://www.ncdc.noaa.gov/>) and included mean seasonal and annual precipitation, mean annual
126 maximum air temperature and mean annual minimum air temperature. Climate variables were
127 calculated as ten-year averages with the year of sampling being the tenth chronological year used
128 in the calculation. All environmental variables were calculated for each stream segment within
129 the study area (roughly 700,000 stream segments) for the year 2011 in order to project the
130 current distribution of each species.

131

132 ***Modeling approach***

133 Five different individual model types were used to model the presence and absence of species:
134 GLM, GBM, CTA, RF, and MARS. These five models were selected based on computation
135 requirements and ability to evaluate response curves. The models were developed within the
136 BIOMOD2 package in R (<http://cran.r-project.org/web/packages/biomod2/index.html>). For each
137 species, the data were split into training and testing subsets to develop and then evaluate the
138 model's performance (Guisan & Zimmerman, 2000). Splitting of the original dataset into
139 training and testing subsets was done using a stratified random sampling design. First, a 100x100
140 km grid was overlain on the study area. Presences within each stratum were randomly assigned
141 to calibration (also called training) and testing datasets, and this process was repeated for
142 absences. Approximately 75 percent of the data were allocated to the calibration subset and 25
143 percent allocated to the testing subset (Franklin, 2009). The calibration subset was randomly split
144 using 75 percent for calibrating the models and the remaining 25 percent for evaluating single
145 models. When a final set of variables were identified, a ten-fold cross validation resulted in 50
146 single models for each species. Each of the ten model runs for cross validation was developed
147 using a randomized selection of 75% of the calibration dataset. The initial testing subset was
148 used to calculate model evaluation scores for model ensembles using the optimal threshold
149 identified (Figure II).

150 During the calibration process, a variable selection process reduced multi-collinearity through
151 removal of correlated variables. Response plots and variable importance assessments were useful
152 in identifying the best suite of variables for the final model. Response plots for individual models

153 were generated by holding N-1 variables constant at their mean value while the variable of
154 interest was plotted across its range. Variable importance was evaluated in each model using a
155 permutation procedure that compared projections made by a trained model with projections made
156 by the same model, but with one variable randomized. This was done for each variable
157 independently and projections were evaluated using correlation. A high correlation between
158 projections suggested the randomized variable had little influence on the model and a low
159 correlation suggested the variable to be more important to the model. The variable importance
160 metric was calculated in BIOMOD2 and is equal to one minus the correlation between the
161 trained and randomized projection. Of correlated variables ($r \geq 0.60$), the variable with the highest
162 variable importance was included in the final model. Variables accounting for <5% of average
163 model variable importance were assumed to have little influence on model response and were
164 removed.

165 A model ensemble was produced using a weighted average of the individual-model evaluation
166 values (Marmion *et al.*, 2009). The primary evaluation metric used was the area under the curve
167 of the receiver-operating characteristic plot (AUC), which measures the ability of a model to
168 discriminate between sites where a species is present, versus those where it is absent. AUC
169 values of 1 represent a perfect model and values of 0.5 reflect a model as good as random
170 assignment of presence or absence (Pearce & Ferrier, 2000). Model ensemble probability
171 projections were converted to binary projections using thresholds identified by the model
172 ensemble AUC that maximizes sensitivity (true positive rate) and specificity (true negative rate).
173 Sensitivity and specificity were also used to evaluate model ensemble performance. We
174 evaluated the influence of prevalence rates on model performance metrics using Pearson

175 correlations. Model ensembles were used to project current distributions and were evaluated
176 spatially using ArcGIS.

177

178 **Results**

179 *Model performance*

180 Species distribution models performed well, with individual models producing AUC values
181 between 0.65 and 0.99 and ensemble models producing AUC values between 0.89 and 0.99
182 (Figure III). Sensitivity across species' ensemble models ranged from 80% to 100% (average of
183 92.3%) and specificity ranged from 73% to 97% (average of 85.9%) suggesting models could
184 predict both presences and absences well. Ensemble model AUC values generally performed
185 better than individual model types. Although no single model type consistently out-performed
186 other model types across species, RF and GBM models performed better than GLM, CTA, and
187 MARS for almost all species. Variability of AUC scores within an individual model type varied
188 across species. For example, GLM runs produced consistent AUC values for orangethroat darter
189 (*Etheostoma spectabile*), but were highly variable for shortnose gar (*Lepisosteus platostomus*).
190 Variability of AUC values across model types within a species also varied, with some species
191 showing fairly consistent values, like river carpsucker (*Carpionodes carpio*), and other species, like
192 smallmouth bass (*Micropterus dolomieu*), showing more variability. Species prevalence was
193 negatively correlated with ensemble model AUC values ($r = -0.64$, $p < 0.01$), sensitivity ($r = -0.60$,
194 $p = 0.02$), and specificity ($r = -0.55$, $p = 0.04$).

195

196 ***Environmental drivers***

197 The final set of models included between 8 and 18 environmental variables per species. Most
198 models were influenced by a mix of climate, land use and geophysical variables; however,
199 average variable importance was highest for climate variables across models (Figure IV). For
200 most species, there was fairly high variability in variable importance for environmental drivers
201 across individual model types and model runs. For example, variable importance for mean
202 annual minimum air temperature for bigeye shiner (*Notropis boops*) and total annual
203 precipitation for plains topminnow (*Fundulus sciadicus*) averaged 0.45, but ranged across
204 models and model runs from 0.1 to 0.8. Response curves for mean annual maximum temperature
205 typically resembled unimodal or threshold responses for most species. For example, response
206 curves for mean annual maximum temperatures suggest smallmouth bass occurrence is
207 associated with temperatures less than 31°C, river carpsucker occurrence is associated with
208 temperatures greater than 28°C, and slender madtom (*Noturus exilis*) is associated with
209 temperatures between 28-33°C (Figure V). Response curves for flow variability varied across
210 species, with presence of smallmouth bass, bigmouth shiner (*Hybopsis dorsalis*), and
211 slenderhead darter (*Percina phoxocephala*) associated with low flow variability, presence of
212 plains topminnow and freckled madtom (*Noturus nocturnus*) associated with low to medium
213 flow variability, and presence of suckermouth minnow (*Phenacobius mirabilis*) and orangethroat
214 darter associated with medium to high flow variability.

215 While there was some variability in response curves across model types, there were consistent
216 relationships for many species. Occurrence of bigeye shiner and slender madtom was associated
217 with mean annual minimum air temperatures greater than -5°C, while occurrence of suckermouth
218 minnow was associated with mean annual minimum air temperatures between -10° and 0°C.

219 Occurrence of many species were associated with specific ranges of precipitation. For example,
220 presence of plains topminnow was associated with total annual precipitation less than
221 approximately 700 mm and presence of bigeye shiner was associated with total annual
222 precipitation between 700 and 1300 mm.

223 At least one land use variable remained in the final suite of variables for each species model.
224 Percent row crops, water and forest were the variables most commonly found across all species.
225 A few models were strongly influenced by percent row crops, including bigmouth shiner,
226 suckermouth minnow, smallmouth bass, and slenderhead darter. Response plots suggest
227 occurrence of all four of these species to be positively associated with row crops. The plains
228 minnow (*Hybognathus placitus*) model indicated that this species was associated with less than
229 five percent of developed land in a watershed.

230 Of the geophysical variables, elevation was a particularly strong driver of occurrence of
231 shortnose gar and river carpsucker. Presence of river carpsucker was associated with elevation
232 less than 180 meters, while presence of shortnose gar was associated with elevation less than 60
233 meters. Though generally less sensitive, surficial geology variables were present in every
234 species' model except suckermouth minnow.

235

236 ***Current range projections***

237 The current range projections of the models closely matched known distributions (NatureServe,
238 2010) for all species (Figure VI – Figure IX). Although the known distribution outlines were
239 produced using reputable sources, their accuracy is not guaranteed. Models that did considerably
240 well without over-projecting into areas where the species is not known to exist, or under-

241 projecting within the known range, included smallmouth bass, suckermouth minnow, largescale
242 stoneroller (*Campostoma oligolepis*), plains minnow, river carpsucker, and slender madtom. The
243 coefficient of variation of the five model types was quantified for each stream segment to show
244 where model agreement and disagreement occurred for each species (Table II). Generally, the
245 models agreed (had lower coefficient of variation values) where each species was projected to be
246 present by the ensemble models, but had greater disagreement (higher coefficient of variation
247 values) where the ensemble model projected the species' to be absent (Figure VI - Figure IX).

248 High densities of certain species often resulted in higher probabilities of occurrence in those
249 areas. The southeast corner of Kansas had relatively high frequencies of freckled madtom,
250 orangespotted sunfish (*Lepomis humilis*), and slenderhead darter occurrences and therefore the
251 models projected higher probabilities of occurrence for these species in that general area. Some
252 models failed to project where there were relatively high densities of occurrence records. For
253 example, the bigmouth shiner model failed to project distributions in western Nebraska despite
254 having numerous occurrence records located in that region. The bigeye shiner model also failed
255 to project where there was presence data in southeastern Kansas.

256 In areas where occurrence data were limited, the models were often unable to project
257 distribution. For example, the plains topminnow model failed to project high probabilities of
258 occurrence within its known range in central Missouri due to limited occurrence records.

259 Freckled madtom also had limited occurrences in the southern extent of its range that the model
260 was unable to capture. Although there were only 69 occurrence events for plains minnow, the
261 model captured the known range well.

262 Some models over-projected species range well outside known ranges. Over-projections were
263 most commonly located north and west of the known ranges. For example, the orangespotted
264 sunfish and slenderhead darter models projected occurrences to extend north and west of their
265 current known ranges. Although the known range overlay makes it appear as though river
266 carpsucker is over-projected into Nebraska and southern Minnesota, there are numerous
267 occurrence records from these regions. Additionally, there are records of river carpsuckers
268 present throughout the Platte River watershed in Nebraska dating back to the 1950's (University
269 of Kansas Biodiversity Institute, Cornell University Museum of Vertebrates [Accessed through
270 Biodiversity Information Serving Our Nation (BISON), bison.usgs.ornl.gov, on 2014-04-09]).
271 Similarly, although the known range overlay suggests there are overprojections in the Ozarks of
272 southern Missouri and northern Arkansas, there are occurrence records in that region.

273

274 **Discussion**

275 *Model development and evaluation*

276 We developed species distribution models for fourteen fish species that were able to discriminate
277 suitable habitat for these species quite well. We used existing state fisheries data, suggesting
278 these datasets are valuable resources for modeling species distributions. Differences in sampling
279 practices and database management, however, may influence model performance. For example,
280 prior to sub-sampling data from Wisconsin, models were developed using all fish data and
281 projections were skewed due to the high density of samples. Others have shown that well-
282 distributed data improve model performance (Kadmon *et al.*, 2003). While requesting fisheries
283 data from various state agencies, numerous personnel made clear the potential for errors due to

284 misidentification of a species or data entry mistakes. Additionally, some states lumped all types
285 of sampling into a single database (e.g., fish kills, standardized basin sampling, targeted
286 sampling), which made it difficult to identify representative samples.

287 Ensemble models generally out-performed individual models in modeling current distributions,
288 suggesting ensemble models are more reliable and applicable for fisheries management purposes
289 than individual models. AUC values from ensemble models indicated high discrimination ability
290 of presence/absence data in the validation stage. Recent work has suggested that high
291 performance of ensemble models may not necessarily translate into high predictive ability
292 outside of the model's spatial and temporal extent (Crimmins *et al.*, 2013). An important
293 question currently receiving attention is how to assess species distribution models' predictive
294 performance for environmental change scenarios. Numerous studies have emphasized the use of
295 temporally and spatially independent data to evaluate transferability of a model across time
296 (Rapacciuolo *et al.*, 2012; Smith *et al.*, 2013) and space (Wenger & Olden, 2012). Transferability
297 of models over long time periods (>20 years) is difficult to assess for fish species, as
298 standardized sampling programs are typically less than thirty years old. Similarly, environmental
299 data, such as land use, are often unavailable at large scales for historic periods. While assessment
300 of transferability will provide some understanding of a model's validity for these purposes,
301 questions will remain regarding a model's response to climate variables with no current
302 analogue.

303 Although our results suggest lower rates of prevalence result in higher model performance,
304 ecological trait associations with model performance have been evaluated for other taxa groups -
305 including plants, insects, birds, and reptiles - with mixed results (Mitchell *et al.*, 2001; Pearce *et*
306 *al.*, 2001; Huntley *et al.*, 2004). Decreased performance in the form of AUC values with

307 increasing species prevalence has been documented previously (Segurado & Araujo, 2004;
308 Allouche *et al.*, 2006). This performance metric is independent of prevalence and thus our
309 results support the hypothesis that distributions of rare species are more predictable than those of
310 common species. Similarly, range size has been found to be negatively correlated with AUC
311 values of plants and birds (McPherson & Jetz, 2007; Syphard & Franklin, 2010). Still other
312 studies showed little to no influence of prevalence or range size on AUC values (McPherson *et*
313 *al.*, 2004; Franklin *et al.*, 2009).

314 Overall, climatic, land use, and geophysical categories of variables all appeared to be relevant to
315 modeling these species. Distribution models for fish species regularly include variables of
316 temperature, precipitation, drainage area, elevation, stream slope, discharge, land use, width,
317 depth and geology (Leathwick *et al.*, 2005; Lassalle *et al.*, 2008; Buisson & Grenouillet, 2009;
318 Lyons *et al.*, 2010; Steen *et al.*, 2010; Bond *et al.*, 2011; Labay *et al.*, 2011; Wenger *et al.*, 2011).
319 It is possible particular variables were removed for having little influence on model performance
320 that actually have strong impacts on species distributions at different scales. For example,
321 developed land use is well known to have negative consequences on aquatic communities (Wang
322 *et al.* 1997; Allan 2004); however, it likely failed to strongly influence model results for some
323 species because it represents a small percentage of the overall study area.

324 Response curves of environmental variables for each species were examined and compared to
325 available published relationships. For example, 10 of the 14 species have published estimated
326 upper thermal limits for water temperatures (Table III). Although we used mean annual
327 maximum air temperature in our models averaged over a ten year period, air and water
328 temperatures are generally related (Stefan & Preud'homme, 1993). Many of the modeled species'
329 response plots do not provide an estimate of an upper thermal limit, likely because their limit is

330 beyond our maximum value of mean annual maximum air temperature (34 °C). Response plots
331 for smallmouth bass, bigmouth shiner, slender madtom, and orangespotted sunfish provide an
332 estimate of upper thermal limits of air temperature similar to published estimates of upper
333 thermal water temperatures. If using these models to project distributions under climate change
334 scenarios, projections will likely be more realistic for species in which an upper thermal limit
335 was captured in response curves. For the models that are unable to estimate an upper thermal
336 limit, model projections under warming temperatures may overestimate future distributions.

337 Of the fourteen species we modeled, smallmouth bass is the most well-studied. Our strongest
338 predictor for smallmouth bass was 90th percentile of discharge, which was strongly correlated to
339 drainage area. A relative abundance model for riverine smallmouth bass in Missouri was strongly
340 influenced by stream size, most likely due to the association between flow variability and stream
341 size (Brewer *et al.*, 2007). Headwater streams typically have greater hydrologic variability as
342 compared to larger streams (Jackson *et al.*, 2001), and smallmouth bass are well known to be
343 associated with larger streams with stable hydrology (Poff & Allan, 1995; Zorn *et al.*, 2002).
344 Discharge, particularly during the spawning and rearing time period, has been found repeatedly
345 to be an important factor on riverine smallmouth bass populations (Cleary, 1956; Simonson &
346 Swenson, 1990; Bovee *et al.*, 1994; Smith *et al.*, 1995; Peterson & Kwak, 1999) as variation in
347 stream discharge negatively influences young of the year smallmouth bass through displacement
348 of eggs and fry, food availability and ability to forage.

349 The majority of species modeled have few known habitat associations other than local habitat
350 preferences in limited stream studies, making it difficult to validate response curves observed
351 from our models. Small-scale variables important to structuring fish communities are often
352 unavailable at the scale of this study. For example, coarse substrates have been associated with

353 the occurrence of freckled madtom (Bouska & Whitley, 2014) and in-stream vegetation is
354 often associated with presence of plains topminnow (Fischer & Paukert, 2008b). In general, there
355 appears to be a need for additional study of environmental requirements of warmwater, non-
356 game species in order to validate response curves. Basic ecology will be particularly important in
357 assessing climate impacts as warmwater fishes are largely expected to expand their ranges under
358 climate change scenarios (Comte *et al.*, 2013).

359 Impoundments have drastically changed the physical and biological template of many river
360 systems. Although the number of upstream dams was included as predictor variables, it did not
361 improve the explanatory power of any model, which may be a consequence of the spatial scale of
362 study. Impoundments often support non-native species that compete with native species. For
363 example, plains topminnow distribution has been suggested to be negatively influenced by the
364 presence of non-native species such as largemouth bass (*Micropterus salmoides*) and
365 mosquitofish (*Gambusia affinis*) (Fischer & Paukert, 2008a; Pasbrig *et al.*, 2012). Impoundments
366 have also been associated with reduced habitat and populations of suckermouth minnow,
367 slenderhead darter, and bigmouth shiner (Quist *et al.*, 2005; Heimann *et al.*, 2007). Further,
368 fluvial dependent species, such as plains minnow have been shown to require 115 km of free-
369 flowing river for population persistence (Perkin & Gido, 2011).

370 Current range projections were assessed by comparing them with known distribution maps. Both
371 NatureServe and the U.S. Geological Survey's Biodiversity Information Serving Our Nation
372 (BISON) product provide coarse-scale ranges and occurrence data for many species found
373 throughout the United States (NatureServe, 2010; U.S. Geological Survey, 2013). Both
374 resources allow for qualitative assessments of current range projections. Unfortunately, due to
375 incomplete coverage of sampling data, we cannot quantify the degree of over- or under-

376 projection. Based on comparisons with these layers, our current range projections for all species
377 appear to correlate well with known occurrences. Similarly, we can compare current projections
378 to other modeling efforts. For example, occurrence models of smallmouth bass in Missouri
379 closely match our projections (Brewer *et al.*, 2007).

380

381 *Limitations and applications*

382 Forecasting how species distributions will be influenced under scenarios of future environmental
383 change is key to developing potential conservation strategies. Reliability of distribution models
384 should be carefully evaluated before predictions are used to inform management decisions
385 (Fielding & Bell, 1997). Caution should be used when using model projections in decision-
386 making, with thorough understanding of the assumptions and uncertainties that limit modeling
387 procedures.

388 One of the primary assumptions of species distribution models is that the system being modeled
389 is in a static or equilibrium state (Guisan & Zimmerman, 2000; Franklin, 2009). In systems such
390 as streams and rivers, however, natural disturbances such as floods and drought are relatively
391 frequent and are important community structuring events (Leopold, 1994; Dodds *et al.*, 2004).
392 When dynamic systems are modeled with static models, there are limitations in their ability to
393 accurately predict species distributions using environmental predictors (Austin, 2002). More
394 mechanistic models are increasingly being developed, yet often require extensive datasets and
395 therefore cannot be used at large spatial scales (Kearney & Porter, 2009). Our species
396 distribution models, as with most, are also unable to account for biological interactions,
397 dispersal, phenotypic plasticity or evolutionary changes (Lavergne, 2010). Therefore, models

398 often assume these factors have non-significant impacts on species distribution when in reality
399 these are likely important factors to species adaptability. As models have progressed, numerous
400 studies have demonstrated how the inclusion of these factors can improve model fit, increase
401 accuracy and provide insight into mechanisms of community assembly (Araujo & Luoto, 2007;
402 Leathwick *et al.*, 2008; Morin & Thuiller, 2009; Van der Putten *et al.*, 2010; Boulangeat *et al.*,
403 2012).

404 The types of predictor variables used are also an important factor when assessing model
405 projections. Predictor variables used in species distribution models generally fall into one of
406 three categories: resource, direct and indirect (Austin, 1985). Resource variables represent matter
407 and energy consumed by organisms whereas direct variables have influence on an organisms'
408 physiology, but are not consumed. Indirect variables do not influence organisms physiologically,
409 but are often correlated with species distribution patterns, easy to measure, more accessible and
410 used as proxies for resource and direct variables (Guisan & Zimmerman, 2000; Austin, 2007).
411 Due to correlations between indirect environmental gradients and species occurrence, the use of
412 indirect predictor variables may result in error-prone predictions (Austin, 2002; Elith &
413 Leathwick, 2009). For example, occurrence of a fish species may be correlated with channel
414 slope due to its relationship with a stream discharge gradient. Slope may be a relevant substitute
415 for discharge for predicting a species' current distributions, but may no longer be relevant under
416 climate change scenarios, which may change discharge patterns, but not slope. Therefore, the
417 use of resource and direct variables is recommended for large spatial scales as these variables
418 lead to a more general, mechanistic model with greater interpretability (Austin, 2002). When
419 resource and direct variables are not available, however, indirect variables that serve as spatial
420 surrogates may be the only option.

421 **Conclusions**

422 Species distribution models are important tools for conservation and management of freshwater
423 fish species. We were able to estimate probability of occurrence of fourteen fish species to
424 almost 700,000 stream segments within our study area. Validation of our ensemble models
425 indicated high performance and qualitative comparisons with known species' ranges suggest the
426 models produced reliable species distribution projections. Although environmental associations
427 with most of our modeled species are unknown, numerous response curves for environmental
428 variables were supported by the literature. Given the high sensitivity of models to climate and
429 land use variables, we expect distributions to be strongly influenced by these variables.

430

431 **Acknowledgements**

432 The authors thank the state natural resource agencies that shared fish data with us to make this
433 project feasible: Arkansas Department of Environmental Quality, Illinois Department of Natural
434 Resources, Kansas Department of Wildlife and Parks, Minnesota Department of Natural
435 Resources, Minnesota Pollution Control Agency, Missouri Department of Conservation,
436 Missouri Department of Natural Resources, Nebraska Game and Fish Commission, North
437 Dakota Department of Health, North Dakota Game, Fish and Parks, and Wisconsin Department
438 of Natural Resources.

439

440 **References**

441

- 442 Aguirre-Gutierrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. & Biesmeijer, J.C.
443 (2013) Fit-for-purpose: Species distribution model performance depends on evaluation criteria -
444 Dutch Hoverflies as a case study. *PLoS One*, **8**, e63708.
- 445 Allan, J.D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual*
446 *Review of Ecology, Evolution, and Systematics* 35:257-284.
- 447 Allouche, O., A. Tsoar & Kadmon, R. (2006) Assessing the accuracy of species distribution models:
448 prevalence, kappa and the true skill statistic. *Journal of Applied Ecology*, **43**, 1223-1232.
- 449 Araujo, M.B. & New, M. (2006) Ensemble forecasting of species distributions. *Trends in Ecology and*
450 *Evolution*, **22**, 42-47.
- 451 Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of*
452 *Biogeography*, **33**, 1677-1688.
- 453 Araujo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions
454 under climate change. *Global Ecology and Biogeography*, **16**, 743-753.
- 455 Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some
456 possible new approaches. *Ecological Modelling*, **200**, 1-19.
- 457 Austin, M.P. (1985) Continuum concept, ordination methods and niche theory. *Annual Review of*
458 *Ecology, Evolution and Systematics*, **16**, 39-61.
- 459 Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and
460 statistical modelling. *Ecological Modelling*, **157**, 101-118.
- 461 Babet-Massin, M., Thuiller, W. & Jiguet, F. (2010) How much do we overestimate future local extinction
462 rates when restricting the range of occurrence data in climate suitability models? *Ecography*, **33**,
463 878-886.
- 464 Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change
465 on the future of biodiversity. *Ecology Letters*, **15**, 365-377.
- 466 Bond, N., Thomson, J., Reich, P. & Stein, J. (2011) Using species distribution models to infer potential
467 climate change-induced range shifts of freshwater fish in south-eastern Australia. *Marine and*
468 *Freshwater Research*, **62**, 1043-1061.
- 469 Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to
470 disentangle the drivers of species distributions and their abundances. *Ecology Letters*, **15**, 584-
471 593.
- 472 Bouska, K. & Whitley, G. (2014) Habitat associations of fish assemblages in the Cache River, Illinois.
473 *Environmental Biology of Fishes*, **97**, 27-42.
- 474 Bovee, K.D., Newcomb, T.J. & Coon, T.G. (1994) Relations between habitat variability and population
475 dynamics of bass in the Huron River, Michigan. Biological Report 21. In, p. 63. U.S. Geological
476 Survey, Washington, DC.
- 477 Brewer, S.K., Rabeni, C.F., Sowa, S.P. & Annis, G. (2007) Natural landscape and stream segment
478 attributes influencing the distribution and relative abundance of riverine smallmouth bass in
479 Missouri. *North American Journal of Fisheries Management*, **27**, 326-341.
- 480 Brosse, S. & Lek, S. (2000) Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear
481 techniques. *Freshwater Biology*, **44**, 441-452.
- 482 Buisson, L. & Grenouillet, G. (2009) Contrasted impacts of climate change on stream fish assemblages
483 along an environmental gradient. *Diversity and Distributions*, **15**, 613-626.
- 484 Buisson, L., Thuiller, W., Lek, S., Lim, P. & Grenouillet, G. (2008) Climate change hastens the turnover of
485 stream fish assemblages. *Global Change Biology*, **14**, 2232-2248.

486 Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting
487 of species distribution. *Global Change Biology*, **16**, 1145-1157.

488 Cianfrani, C., Le Lay, G., Maiorano, L., Satizabal, H.F. & Loy, A. (2011) Adapting global conservation
489 strategies to climate change at the European scale: the otter as a flagship species. *Biological*
490 *Conservation*, **144**, 2068-2080.

491 Cleary, R.E. (1956) Observations on factors affecting smallmouth bass production in Iowa. *The Journal of*
492 *Wildlife Management*, **20**, 353-359.

493 Comte, L., Buisson, L., Daufresne, M. & Grenouillet, G. (2013) Climate-induced changes in the
494 distribution of freshwater fish: observed and predicted trends. *Freshwater Biology*, **58**, 625-639.

495 Coutant, C.C. (1977) Compilation of temperature preference data. *Journal of the Fisheries Research*
496 *Board of Canada*, **34**, 739-745.

497 Crimmins, S.M., Dobrowski, S.Z. & Mynsberge, A.R. (2013) Evaluating ensemble forecasts of plant
498 species distributions under climate change. *Ecological Modelling*, **266**, 126-130.

499 Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond predictions:
500 biodiversity conservation in a changing climate. *Science*, **332**, 53-58.

501 Diniz-Filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogues-Bravo, D. & Araujo, M.B. (2009)
502 Partitioning and mapping uncertainties in ensembles of forecasts of species turnover of climate
503 change. *Ecography*, **32**, 897-906.

504 Dodds, W.K., Gido, K., Whiles, M.R., Fritz, K.M. & Matthews, W.J. (2004) Life on the edge: the ecology of
505 great plains prairie streams. *BioScience*, **54**, 205-216.

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

509 Eaton, J.G. & Scheller, R.M. (1996) Effects of climate warming on fish thermal habitat in streams of the
510 United States. *Limnology and Oceanography*, **41**, 1109-1115.

511 Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction
512 across space and time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677-697.

513 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F.,
514 Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C.,
515 Nakamura, M., Nakazawa, Y., Overton, J., Townsend Peterson, A., Phillips, S.J., Richardson, K.,
516 Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmerman, N.E.
517 (2006) Novel methods improve prediction of species' distributions from occurrence data.
518 *Ecography*, **29**, 129-151.

519 Esselman, P.C., Infante, D.M., Wang, L., Cooper, A. & Taylor, W.W. (2011) An initial assessment of
520 integrated landscape disturbance on river fish habitats in the conterminous United States.
521 *Restoration Ecology*, **23**, 133-151.

522 Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in
523 conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.

524 Fischer, J.R. & Paukert, C.P. (2008a) Historical and current environmental influences on an endemic
525 Great Plains fish. *American Midland Naturalist*, **159**, 364-377.

526 Fischer, J.R. & Paukert, C.P. (2008b) Habitat relationships with fish assemblages in minimally disturbed
527 Great Plains regions. *Ecology of Freshwater Fish*, **17**, 597-609.

528 Franklin, J. (2009) *Mapping species distribution: spatial inference and prediction*. Cambridge University
529 Press, Cambridge, UK.

530 Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J. & Fisher, R.N. (2009) Effect of species rarity on
531 the accuracy of species distribution models for reptiles and amphibians in southern California.
532 *Diversity and Distributions*, **15**, 167-177.

533 Grenouillet, G., Buisson, L., Casajus, N. & Lek, S. (2011) Ensemble modelling of species distribution: the
534 effects of geographical and environmental ranges. *Ecography*, **34**, 9-17.

535 Guisan, A. & Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological*
536 *Modelling*, **135**, 147-186.

537 Guisan, A., Graham, C.H., Elith, J., Huettmann, F. & Group, N.S.D.M. (2007) Sensitivity of predictive
538 species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332-340.

539 Guisan, A., and W. Thuiller (2005) Predicting species distribution: offering more than simple habitat
540 models. *Ecology Letters*, **8**, 993-1009.

541 Heimann, D.C., Licher, S.S. & Schalk, G.K. (2007) Effects of impoundments and land-cover changes on
542 streamflows and selected fish habitat in the Upper Osage River Basin, Missouri and Kansas.
543 Scientific Investigations Report 2007-5175. In, p. 96. U.S. Geological Survey

544 Heino, J., Virkkala, R. & Toivonen, H. (2009) Climate change and freshwater biodiversity: detected
545 patterns, future trends and adaptations in northern regions. *Biological Reviews*, **84**, 39-54.

546 Hernandez, P.A., C.H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species
547 characteristics on performance of different species distribution models. *Ecography*, **29**, 773-785.

548 Hokanson, K.E.F. 1977. Temperature requirements of some percids and adaptations to the seasonal
549 temperature cycle. *Journal of the Fisheries Board of Canada* 34:1524-1550.

550 Huntley, B., Green, R.E., Collingham, T.C., Hill, J.K., Willis, S.G., Bartlein, P.J., Cramer, W. & Hagemerijer,
551 W.J.M. (2004) The performance of models relating species geographical distributions to climate
552 is independent of trophic level *Ecological Letters*, **7**, 417-426.

553 Jackson, D.A., Peres-Neto, P.R. & Olden, J.D. (2001) What controls who is where in freshwater fish
554 communities - the role of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and*
555 *Aquatic Science*, **58**, 157-170.

556 Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of
557 climatic envelope models. *Ecological Applications*, **13**, 853-867.

558 Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data
559 to predict species' ranges. *Ecology Letters*, **12**, 334-350.

560 Kostyack, J., Lawler, J.J., Olden, J.D. & Scott, J.M. (2011) Beyond reserves and corridors: policy solutions
561 to facilitate the movement of plants and animals in a changing climate. *BioScience*, **61**, 713-719.

562 Lassalle, G., Beguer, M., Beaulaton, L. & Rochard, E. (2008) Diadromous fish conservation plans need to
563 consider global warming issues: An approach using biogeographical models. *Biological*
564 *Conservation*, **141**, 1105-1118.

565 Labay, B., Cohen, A.E., Sissel, B., Hendrickson, D.A., Martin, F.D. & Sarkar, S. (2011) Assessing Historical
566 Fish Community Composition Using Surveys, Historical Collection Data, and Species Distribution
567 Models. *Plos One*, **6**, e25145.

568 Lavergne, S., N. Mouquet, W. Thuiller and O. Ronce (2010) Biodiversity and climate change: integrating
569 evolutionary and ecological responses of species and communities. *Annual Review of Ecology,*
570 *Evolution and Systematics*, **41**, 321-350.

571 Lawler, J.J. & Olden, J.D. (2011) Reframing the debate over managed relocation. *Frontiers in Ecology and*
572 *the Environment*, **9**, 569-574.

573 Leathwick, J.R., Elith, J. & Hastie, T. (2006) Comparative performance of generalized additive models and
574 multivariate adaptive regression splines for statistical modelling of species distributions.
575 *Ecological Modelling*, **199**, 188-196.

576 Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005) Using multivariate adaptive
577 regression splines to predict the distributions of New Zealand's freshwater diadromous fish.
578 *Freshwater Biology*, **50**, 2034-2052.

579 Leathwick, J.R., Elith, J., Chadderton, W.L., Rowe, D. & Hastie, T. (2008) Dispersal, disturbance and the
580 contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species.
581 *Journal of Biogeography*, **35**, 1481-1497.

582 Leopold, L. (1994) *The view of the river*. Harvard University Press, Cambridge, MA.

583 Lyons, J., Stewart, J.S. & Mitro, M. (2010) Predicted effects of climate warming on the distribution of 50
584 stream fishes in Wisconsin, U.S.A. *Journal of Fish Biology*, **77**, 1867-1898.

585 Maloney, K.O., Weller, D.E., Michaelson, D.E. & Ciccotto, P.J. (2013) Species distribution models of
586 freshwater stream fishes in Maryland and their implications for management. *Environmental*
587 *Modeling & Assessment*, **18**, 1-12.

588 Manel, S., Williams, H.C. & Omerod, S.J. (2001) Evaluating presence-absence models in ecology: the
589 need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

590 Marmion, M., Parvianinen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus
591 methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

592 Mawdsley, J.R., O'Malley, R. & Ojima, D.S. (2009) A review of climate-change adaptation strategies for
593 wildlife management and biodiversity conservation. *Conservation Biology*, **23**, 1080-1089.

594 McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models.
595 *Ecography*, **30**, 135-151.

596 McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of
597 distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*,
598 **41**, 811-823.

599 Mitchell, M.S., Lancia, R.A. & Gerwin, J.A. (2001) Using landscape-level data to predict the distribution of
600 birds on a managed forest: effects of scale. *Ecological Applications*, **11**, 1692-1708.

601 Morin, A. & Thuiller, W. (2009) Comparing niche- and process-based models to reduce prediction
602 uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301-1313.

603 NatureServe (2010) Digital distribution maps of the freshwater fishes in the conterminous United States.
604 Version 3.0. In: (ed. NatureServe), Arlington, VA.

605 Olden, J.D., Kennard, M.K., Lawler, J.J. & Poff, N.L. (2011) Challenges and opportunities in implementing
606 managed relocation for conservation of freshwater species. *Conservation Biology*, **25**, 40-47.

607 Ostrand, K.G. & Wilde, G.R. (2001) Temperature, dissolved oxygen, and salinity tolerances of five prairie
608 stream fishes and their role in explaining fish assemblage patterns. *Transactions of the American*
609 *Fisheries Society*, **130**, 742-749.

610 Pasbrig, C.A., Koupal, K.D., Schainost, S. & Hoback, W.W. (2012) Changes in range-wide distribution of
611 plains topminnow *Fundulus sciadicus*. *Endangered Species Research*, **16**, 235-247.

612 Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using
613 logistic regression. *Ecological Modelling*, **133**, 225-245.

614 Pearce, J., Ferrier, S. & Scotts, D. (2001) An evaluation of the predictive performance of distributional
615 models for flora and fauna in north-east New South Wales. *Journal of Environmental*
616 *Management*, **62**, 171-184.

617 Pearson, R.G., Thuiller, W., Araujo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L.,
618 Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range
619 prediction. *Journal of Biogeography*, **33**, 1704-1711.

620 Perkin, J.S. & Gido, K.B. (2011) Stream fragmentation thresholds for a reproductive guild of Great Plains
621 fishes. *Fisheries*, **26**, 371-383.

622 Peterson, J.T. & Kwak, T.J. (1999) Modeling the effects of land use and climate change on riverine
623 smallmouth bass. *Ecological Applications*, **9**, 1391-1404.

624 Poff, N.L. & Allan, J.D. (1995) Functional organization of stream fish assemblages in relation to
625 hydrological variability. *Ecology*, **76**, 606-627.

626 Poulos, H.M., Chernoff, B., Fuller, P.L. & Butman, D. (2012) Ensemble forecasting of potential habitat for
627 three invasive fishes. *Aquatic Invasions*, **7**, 59-72.

628 Quist, M.C., Hubert, W.A. & Rahel, F.J. (2005) Fish assemblage structure following impoundment of a
629 Great Plains river. *Western North American Naturalist*, **65**, 53-63.

630 Rapacciuolo, G., Roy, D.B., Gillings, S., Fox, R., Walker, K. & Purvis, A. (2012) Climatic associations of
631 British species distributions show good transferability in time but low predictive accuracy for
632 range change. *Plos One*, **7**, e40212.

633 Segurado, P. & Araujo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal*
634 *of Biogeography*, **31**, 1555-1568.

635 Sharma, S., Vander Zanden, M.J., Magnuson, J.J. & Lyons, J. (2011) Comparing climate change and
636 species invasions as drivers of coldwater fish population extirpations. *Plos One*, **6**, e22906.

637 Simonson, T.D. & Swenson, W.A. (1990) Critical stream velocities for young-of-year smallmouth bass in
638 relation to habitat use. *Transactions of the American Fisheries Society*, **119**, 902-909.

639 Smale, M.A. & Rabeni, C.F. (1995) Hypoxia and hyperthermia tolerances of headwater stream fishes.
640 *Transactions of the American Fisheries Society*, **124**, 698-710.

641 Smith, A.B., Santos, M.J., Koo, M.S., Rowe, K.M.C., Rowe, K.C., Patton, J.L., Perrine, J.D., Beissinger, S.R.
642 & Moritz, C. (2013) Evaluation of species distribution models by resampling of sites surveyed a
643 century ago by Joseph Grinnell. *Ecography*, **36**, 1017-1031.

644 Smith, S.M., Odenkirk, J.S. & Reeser, S.J. (1995) Smallmouth bass recruitment variability and its relation
645 to stream discharge in three Virginia rivers. *North American Journal of Fisheries Management*,
646 **25**, 1112-1121.

647 Steen, P.J., Wiley, M.J. & Schaeffer, J.S. (2010) Predicting Future Changes in Muskegon River Watershed
648 Game Fish Distributions under Future Land Cover Alteration and Climate Change Scenarios.
649 *Transactions of the American Fisheries Society*, **139**, 396-412.

650 Stefan, H.G. & Preud'homme, E.B. (1993) Stream temperature estimation from air temperature. *Journal*
651 *of the American Water Resources Association*, **29**, 27-45.

652 Syphard, A.D. & Franklin, J. (2010) Species traits affect the performance of species distribution models
653 for plants in southern California. *Journal of Vegetation Science*, **21**, 177-189.

654 Thuiller, W. (2007) Climate change and the ecologist. *Nature*, **448**, 550-552.

655 Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M.B. (2009) BIOMOD - a platform for ensemble
656 forecasting of species distributions. *Ecography*, **32**, 369-373.

657 Tobin, B.D. & Weary, D.J. (2004) *Digital engineering aspects of karst map: a GIS version of Davies, W.E.,*
658 *Simpson, J.H., Ohlmacher, G.C., Kirk, W.S., and Newton, E.G., 1984, Engineering aspects of karstL*
659 *U.S. Geological Survey, National Atlas of the United States of America, Scale 1:7,500,000.*
660 Available at: <http://pubs.usgs.gov/of/2004/1352/> (accessed 2012).

661 Townsend Peterson, A. (2006) Uses and requirements of ecological niche models and relation
662 distributional models. *Biodiversity Informatics*, **3**, 59-72.

663 U.S. Geological Survey (2013) *Biodiversity Information Serving Our Nation (BISON)*. Available at:
664 (accessed 2014).

665 Van der Putten, W.H., Macel, M. & Visser, M.E. (2010) Predicting species distribution and abundance
666 responses to climate change: why it is essential to include biotic interactions across trophic
667 levels. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 2025-2034.

668 Vanni, M. (2002) Nutrient cycling by animals in freshwater ecosystems. *Annual Review of Ecology,*
669 *Evolution and Systematics*, **33**, 341-370.

670 Wang, L., Lyons, J., Kanehl, P., & Gatti, R. 1997. Influences of watershed land use on habitat quality and
671 biotic integrity in Wisconsin streams. *Fisheries* 22:6-12.

672 Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: an underappreciated
673 aspect of statistical validation. *Methods in Ecology and Evolution*, **3**, 260-267.

- 674 Wenger, S.J., Isaak, D.J., Luce, C.H., Neville, H.M., Fausch, K.D., Dunham, J.B., Dauwalter, D.C., Young,
675 M.K., Elsner, M.M., Rieman, B.E., Hamlet, A.F. & Williams, J.E. (2011) Flow regime, temperature,
676 and biotic interactions drive differential declines of trout species under climate change.
677 *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14175-
678 14180.
- 679 Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & Group, N.P.S.D.W. (2008)
680 Effects of sample size on the performance of species distribution models. *Diversity and*
681 *Distributions*, **14**, 763-773.
- 682 Zorn, T.G., Seelbach, P.W. & Wiley, M.J. (2002) Distributions of stream fishes and their relationship to
683 stream size and hydrology in Michigan's Lower Peninsula. *Transactions of the American Fisheries*
684 *Society*, **131**, 70-85.

685 **Tables**

686 **Table I.** The final list of species selected for model development and the number of occurrence
687 records out of 8,675 sampling events.

Common Name	Scientific Name (Nomenclature Authority)	Occurrence records
Bigeye shiner	<i>Notropis boops</i> (Gilbert, 1984)	182
Bigmouth shiner	<i>Hybopsis dorsalis</i> (Agassiz, 1854)	1639
Freckled madtom	<i>Noturus nocturnus</i> (Gilbert, 1886)	154
Largescale stoneroller	<i>Campostoma oligolepis</i> (Hubbs and Greene, 1935)	260
Orangespotted sunfish	<i>Lepomis humilis</i> (Girard, 1958)	934
Orangethroat darter	<i>Etheostoma spectabile</i> (Agassiz, 1854)	1021
Plains minnow	<i>Hybognathus placitus</i> (Girard, 1856)	69
Plains topminnow	<i>Fundulus sciadicus</i> (Cope, 1865)	212
River carpsucker	<i>Carpionodes carpio</i> (Rafinesque, 1820)	916
Shortnose gar	<i>Lepisosteus platostomus</i> (Rafinesque, 1820)	375
Slenderhead darter	<i>Percina phoxocephala</i> (Nelson, 1876)	549
Slender madtom	<i>Noturus exilis</i> (Nelson, 1876)	342
Smallmouth bass	<i>Micropterus dolomieu</i> (Lacepède, 1802)	1824
Suckermouth minnow	<i>Phenacobius mirabilis</i> (Girard, 1958)	551

688

689

690 **Table II.** Average and standard deviation of the coefficient of variation of the probabilities of
 691 occurrence over the five types of individual models for all stream segments where each species is
 692 projected to be present and absent. Higher coefficients of variation suggest greater uncertainty in
 693 model results.

Species	Presence		Absence	
	Mean CV	SD	Mean CV	SD
Bigeye shiner	0.41	0.08	2.12	0.56
Bigmouth shiner	0.45	0.10	1.36	0.54
Freckled madtom	0.49	0.08	0.89	0.14
Largescale stoneroller	0.64	0.14	1.21	0.20
Orangespotted sunfish	0.55	0.17	0.90	0.07
Orangethroat darter	0.59	0.15	1.68	0.62
Plains minnow	1.16	0.13	1.73	0.25
Plains topminnow	0.96	0.14	2.32	0.55
River carpsucker	0.59	0.16	1.15	0.30
Slenderhead darter	0.62	0.15	1.27	0.35
Slender madtom	0.79	0.25	2.29	0.69
Smallmouth bass	0.57	0.07	1.61	0.51
Shortnose gar	0.32	0.17	1.13	0.30
Suckermouth minnow	0.57	0.09	1.06	0.45

694

695 **Table III.** Published estimates of upper thermal limits for fish species modeled in this study.

Species	Estimated upper thermal limit (°C)	Response curve upper limit (°C)	Source
Bigmouth shiner	36.6	~32 °C	(Smale & Rabeni, 1995)
Orangespotted sunfish	32.5 - 36.4	~34 °C	(Smale & Rabeni, 1995; Eaton & Scheller, 1996)
Orangethroat darter	36.5	Not identified	(Smale & Rabeni, 1995)
Plains minnow	39.7	Not identified	(Ostrand & Wilde, 2001)
Plains topminnow	37.0	Not identified	(Smale & Rabeni, 1995)
River carpsucker	32.1 - 34.5	Not identified	(Coutant, 1977; Eaton & Scheller, 1996)
Shortnose gar	34.5	Not identified	(Coutant, 1977)
Slender madtom	36.5	~33 °C	(Smale & Rabeni, 1995)
Smallmouth bass	29.5 - 36.9	~31 °C	(Smale & Rabeni, 1995; Eaton & Scheller, 1996)
Suckermouth minnow	32.1	Not identified	(Eaton & Scheller, 1996)

696

Figure captions

Fig. I Study area (shaded in gray) and location of sub-sampled fish sampling events (denoted by black circles) used to develop and evaluate species distribution data. This map was produced using Albers equal-area conic projection.

Fig. II Study design for the development and evaluation of ensemble species distribution models.

Fig. III Model performance by model type and model ensemble performance for 14 species distribution models.

Fig. IV Variable importance averaged across all model runs for each species.

Fig. V Response curves of the probability of occurrence of smallmouth bass, river carpsucker, and slender madtom to mean annual maximum air temperature. Each panel illustrates the agreement across ten model runs for each of the five individual model types.

Fig. VI On the left of each panel is the projected probability of occurrence of bigeye shiner, bigmouth shiner, freckled madtom, and largescale stoneroller throughout the study area. Threshold probability, or the probability value above which a species is predicted to be present, are given in parentheses for each species. On the right of each panel is the coefficient of variation of the probabilities of occurrence over the five individual models, providing a measure of agreement across model types.

Fig. VII On the left of each panel is the projected probability of occurrence of orangespotted sunfish, orangethroat darter, plains minnow, and plains topminnow throughout the study area. Threshold probability, or the probability value above which a species is predicted to be present,

are given in parentheses for each species. On the right of each panel is the coefficient of variation of the probabilities of occurrence over the five individual models, providing a measure of agreement across model types.

Fig. VIII On the left of each panel is the projected probability of occurrence of river carpsucker, shortnose gar, slenderhead darter, and slender madtom throughout the study area. Threshold probability, or the probability value above which a species is predicted to be present, are given in parentheses for each species. On the right of each panel is the coefficient of variation of the probabilities of occurrence over the five individual models, providing a measure of agreement across model types.

Fig. IX On the left of each panel is the projected probability of occurrence of smallmouth bass and suckermouth minnow throughout the study area. Threshold probability, or the probability value above which a species is predicted to be present, are given in parentheses for each species. On the right of each panel is the coefficient of variation of the probabilities of occurrence over the five individual models, providing a measure of agreement across model types.

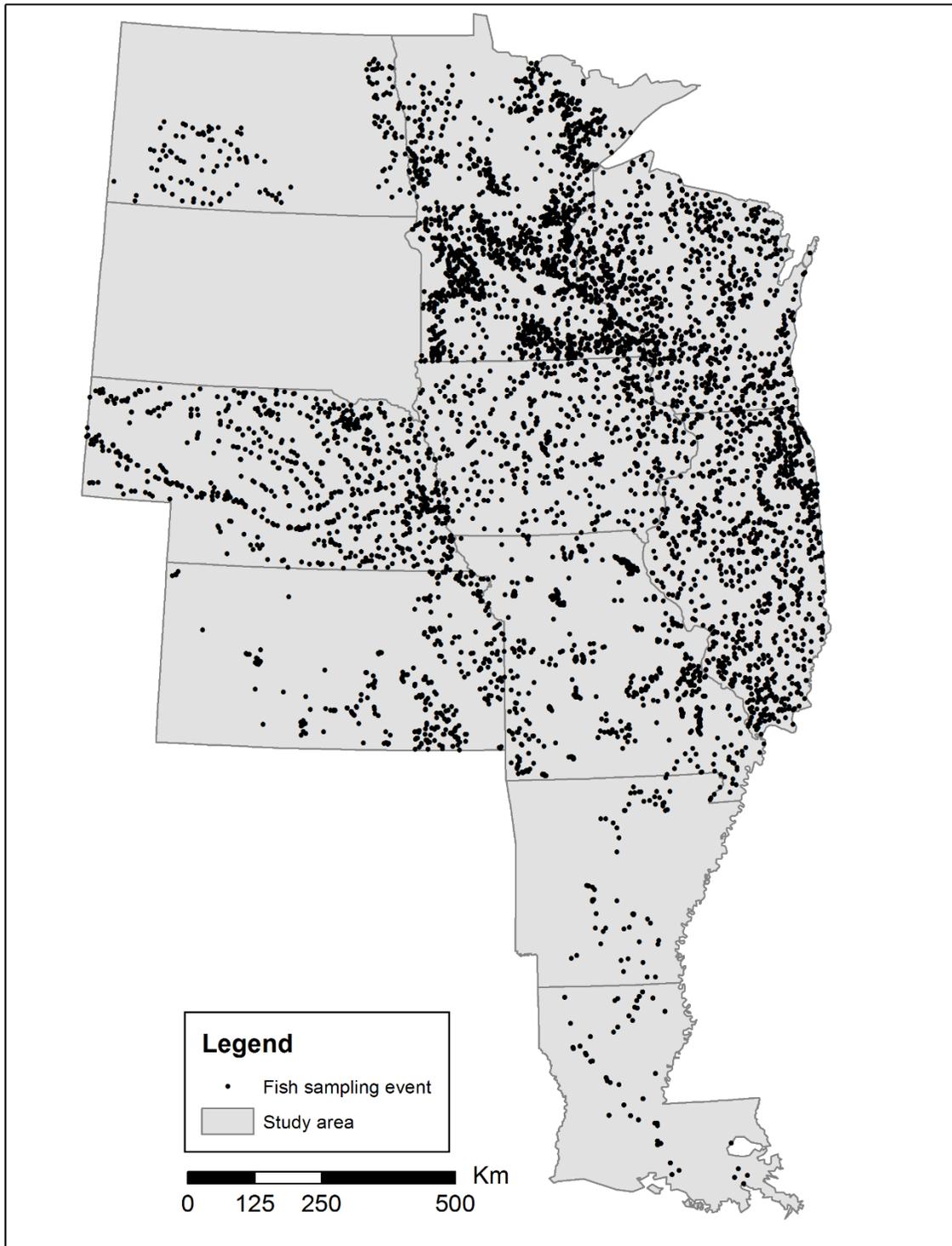


Figure 1.

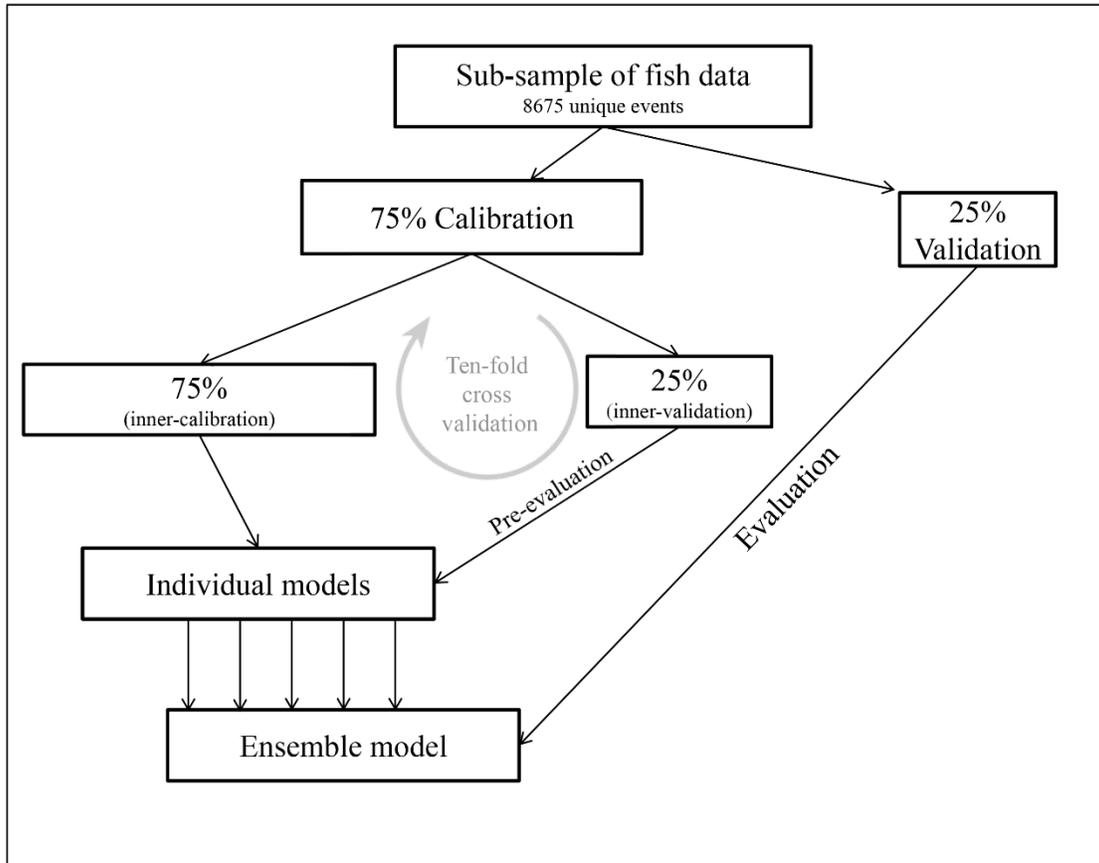


Figure 2.

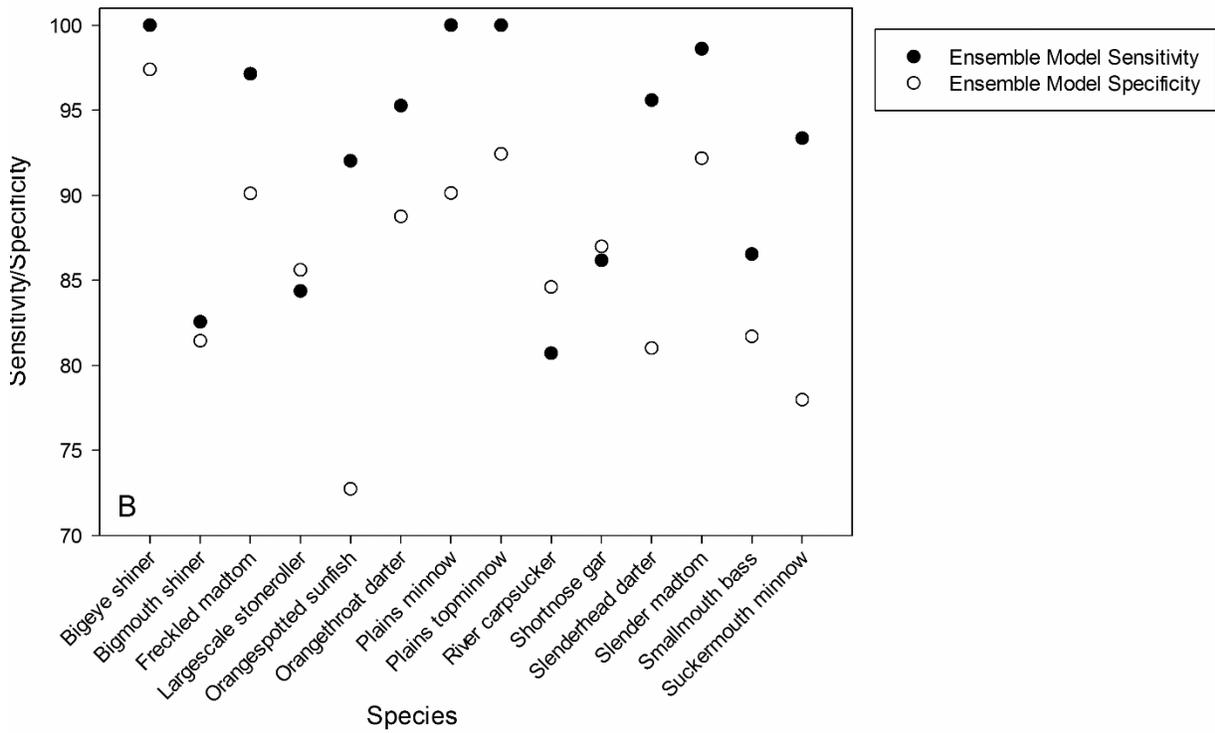
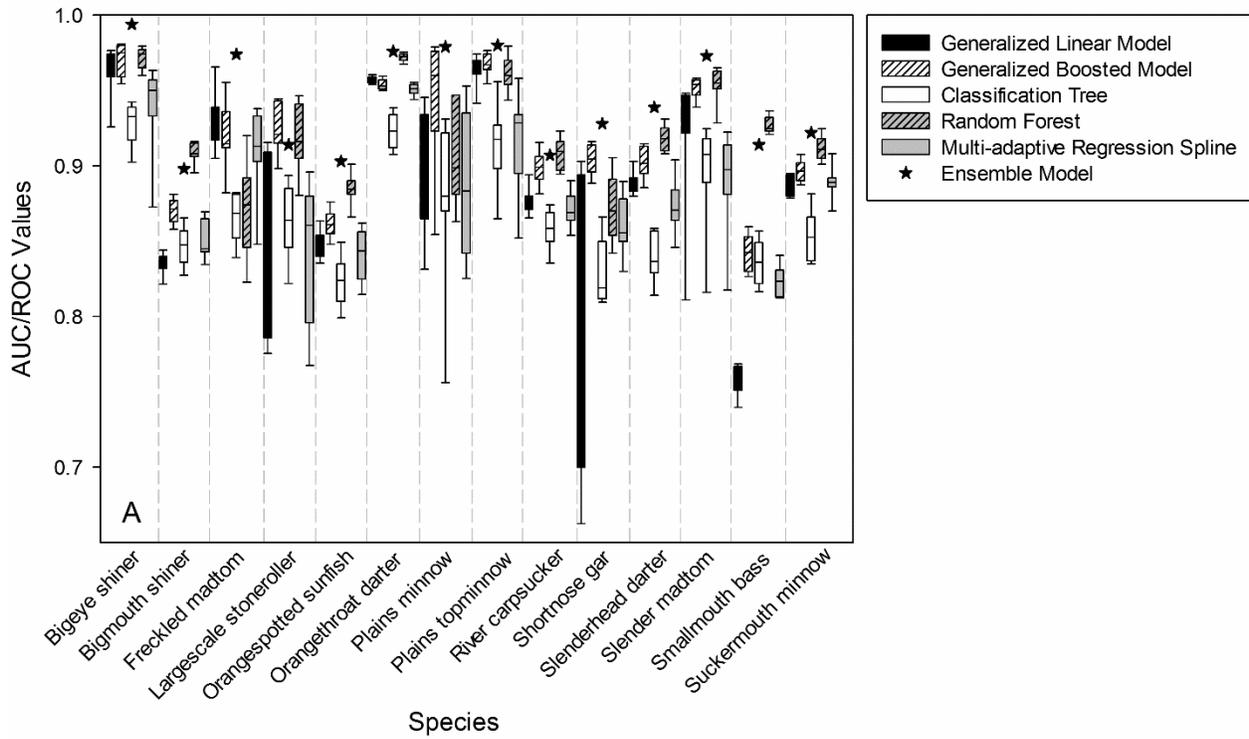


Figure 3.

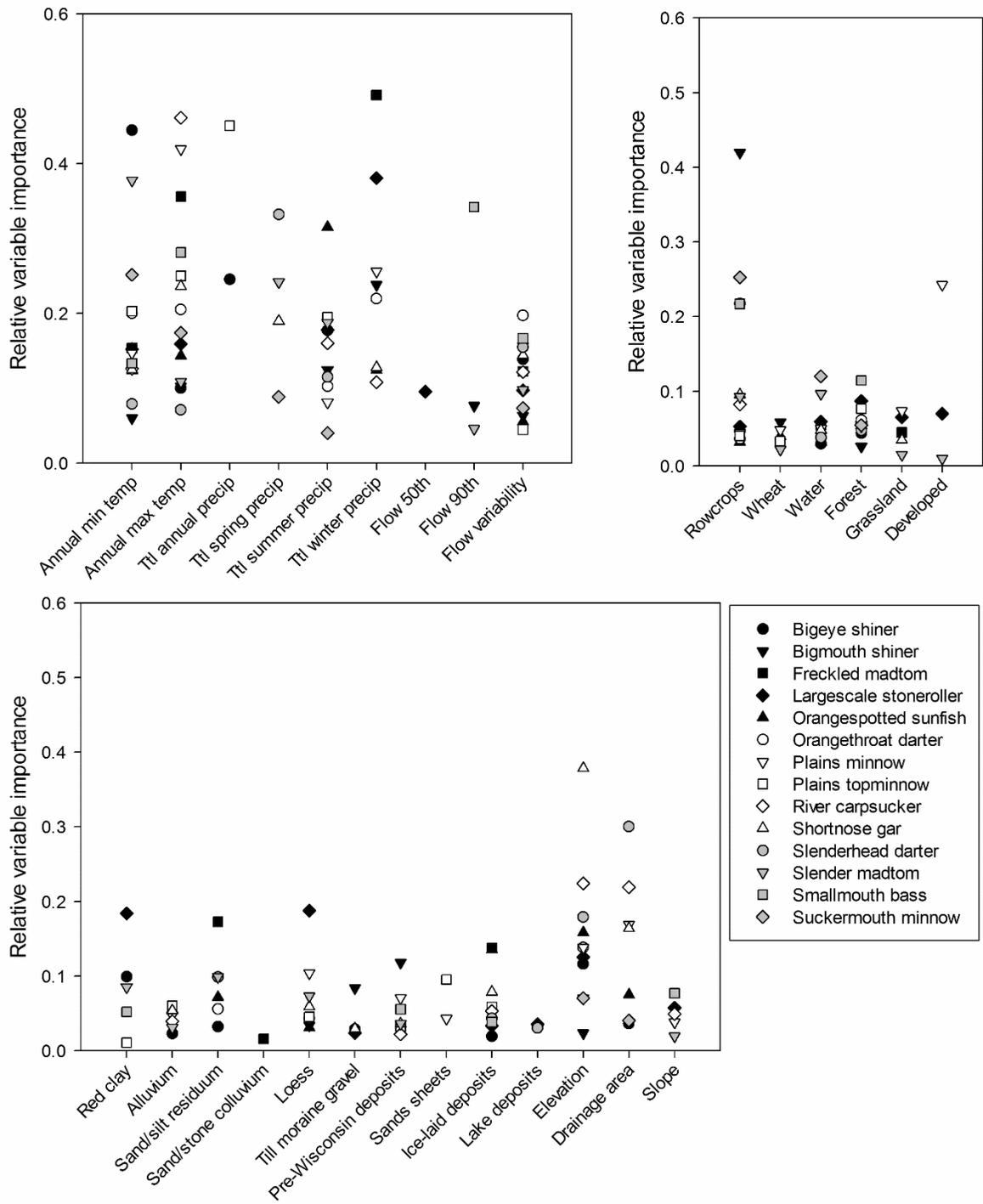


Figure 4.

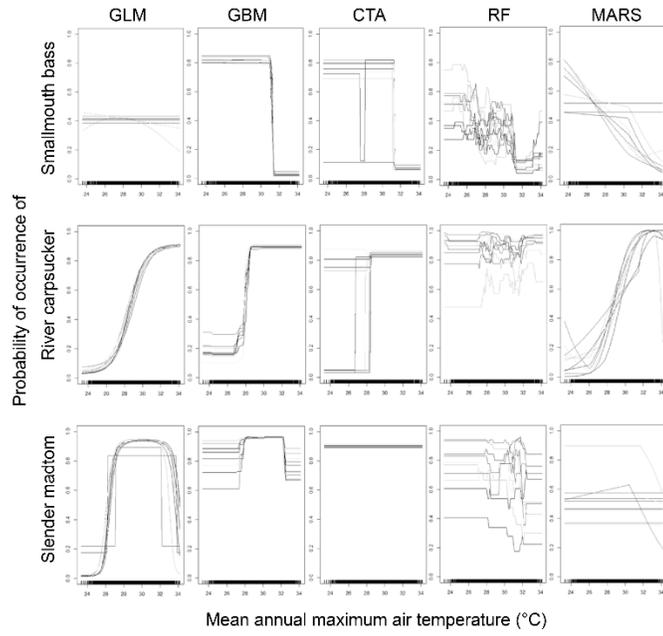


Figure 5.

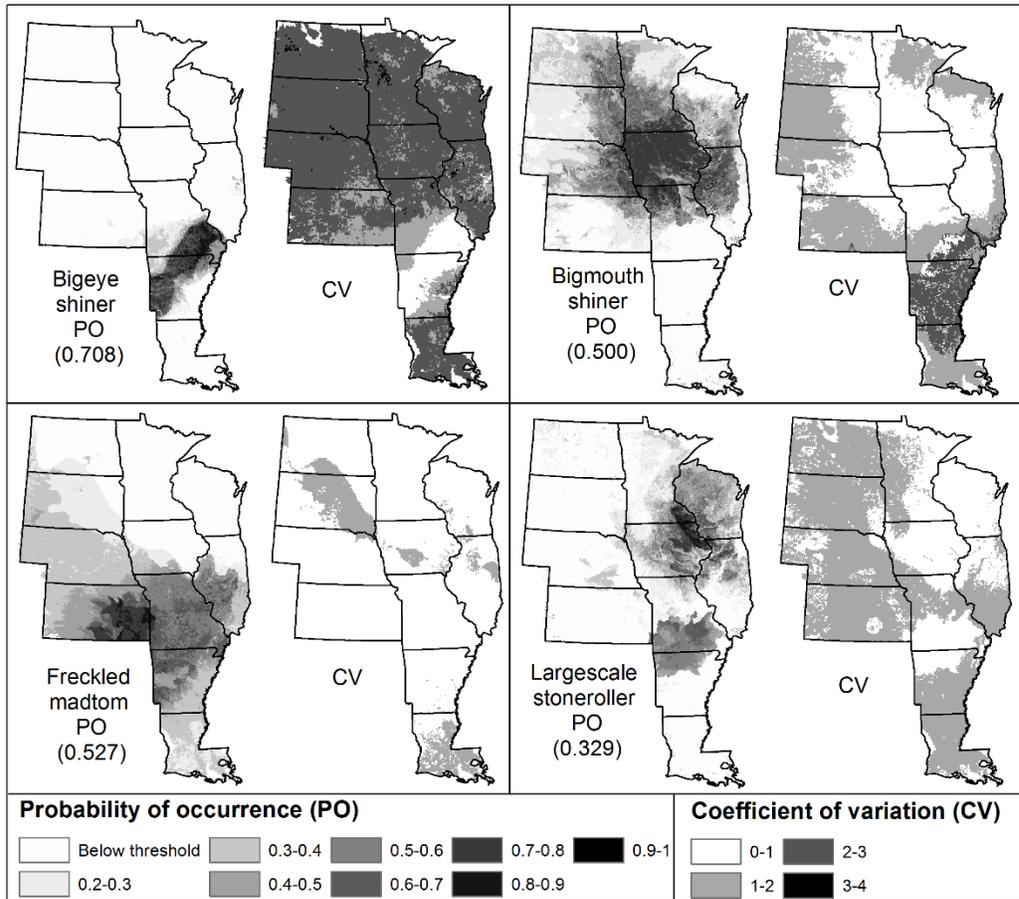


Figure VI.

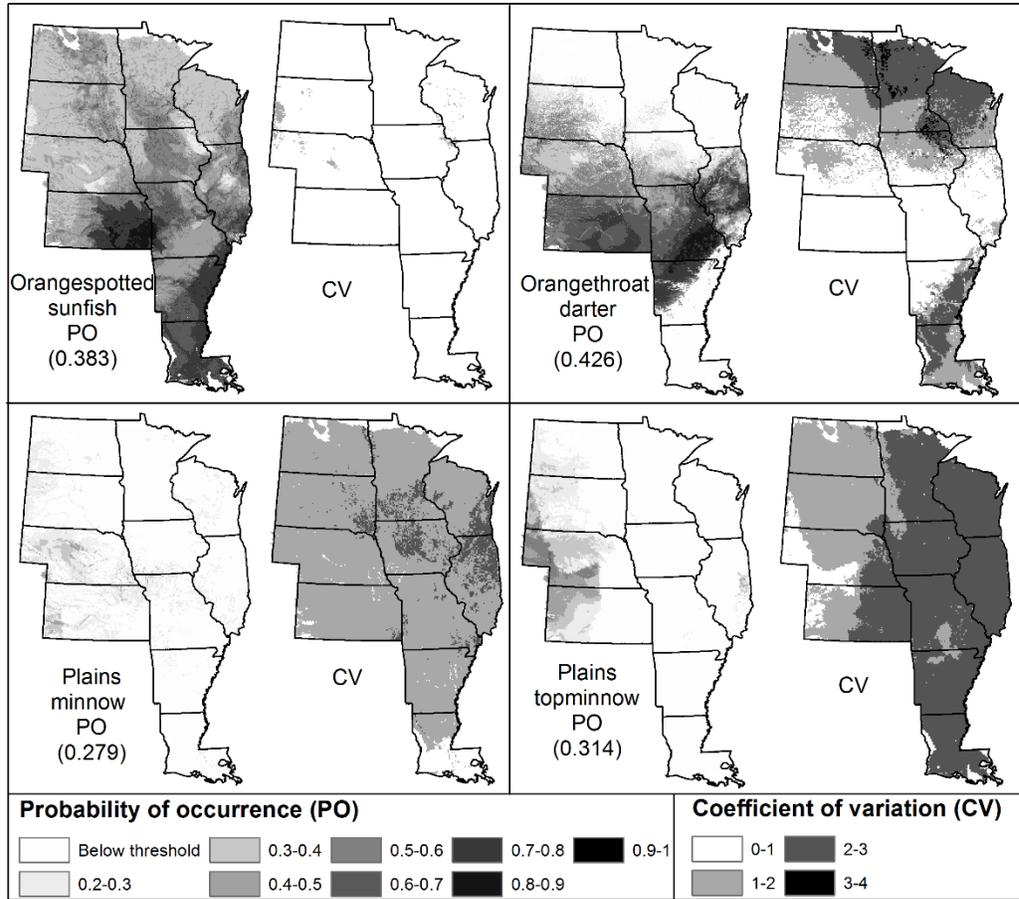


Figure VII.

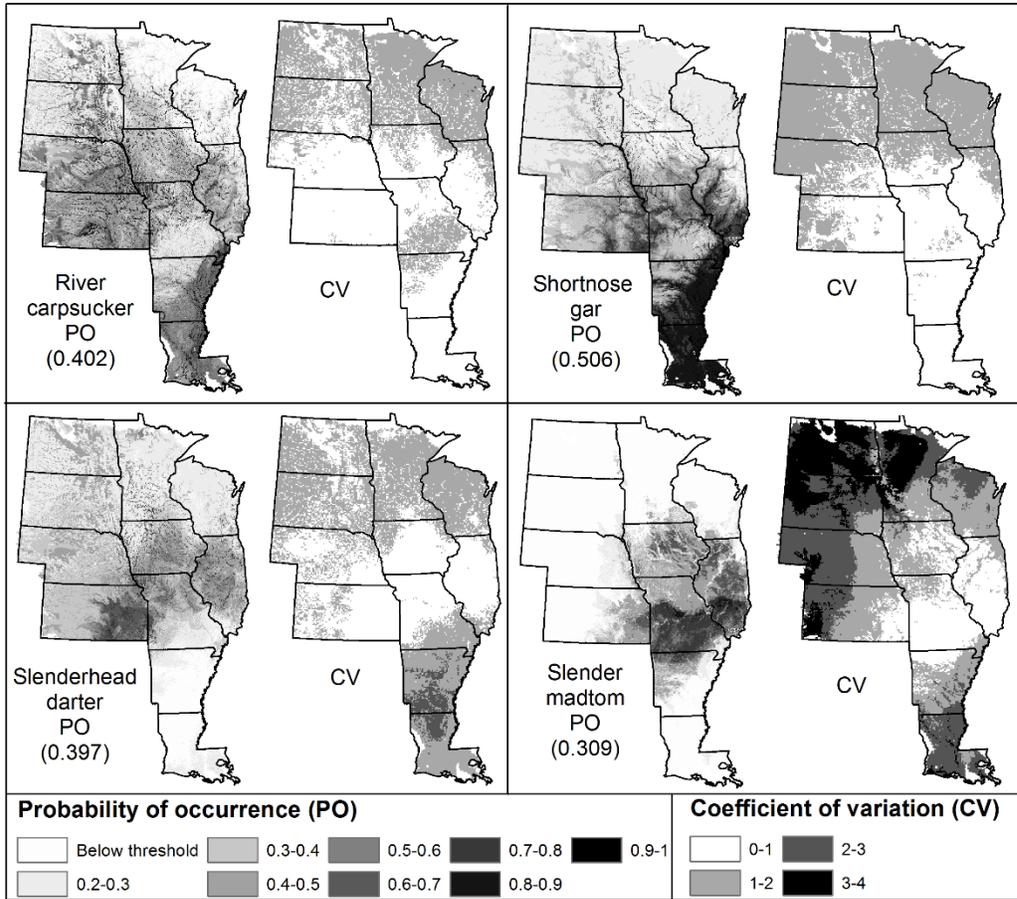


Figure VIII.

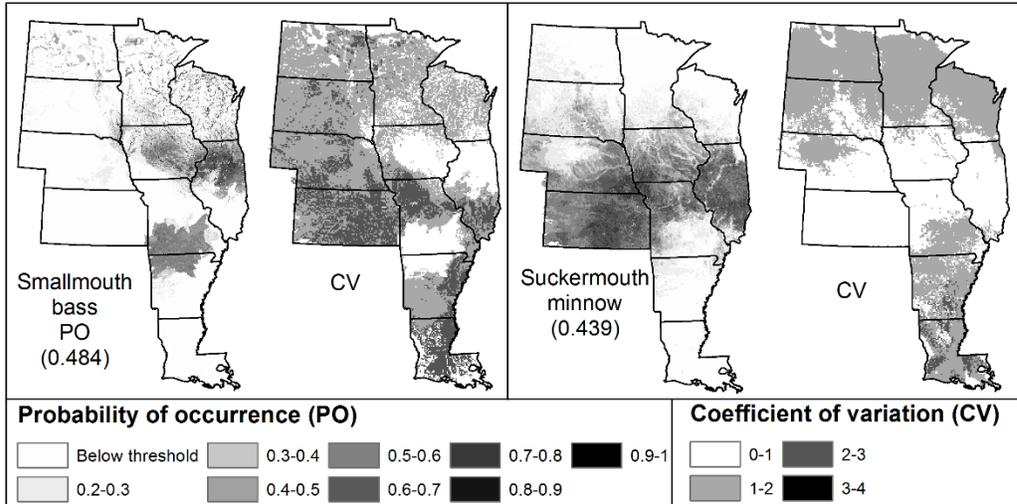


Figure IX.