

Ensemble Species Distribution Model Identifies Survey Opportunities for At-Risk Bearded Beaksedge (*Rhynchospora crinipes*) in the Southeastern United States

Authors: Ramirez-Reyes, Carlos, Street, Garrett, Vilella, Francisco J., Jones-Farrand, D. Todd, Wiggers, M. Scott, et al.

Source: Natural Areas Journal, 41(1) : 55-63

Published By: Natural Areas Association

URL: <https://doi.org/10.3375/043.041.0108>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Ensemble Species Distribution Model Identifies Survey Opportunities for At-Risk Bearded Beaksedge (*Rhynchospora crinipes*) in the Southeastern United States

Carlos Ramirez-Reyes,^{1,5} Garrett Street,¹ Francisco J. Vilella,² D. Todd Jones-Farrand,³ M. Scott Wiggers,⁴ and Kristine O. Evans¹

¹Quantitative Ecology & Spatial Technologies Laboratory, Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Box 9690, Mississippi State, MS 39760

²U.S. Geological Survey, Mississippi Cooperative Fish and Wildlife Research Unit, Box 9691, Mississippi State, MS 39762

³U.S. Fish and Wildlife Service, 302 Natural Resources, University of Missouri, Columbia, MO 65211

⁴U.S. Fish and Wildlife Service, Mississippi Field Office, 6578 Dogwood View Parkway, Jackson, MS 39213

⁵Corresponding author: c.ramirez@msstate.edu; 608-520-1362

Associate Editor: John N. Williams

ABSTRACT

Locating additional occurrences of at-risk species can inform assessments of their status and conservation needs (including potential legal protections). The perennial bearded beaksedge (*Rhynchospora crinipes*) ranges from Mississippi to North Carolina, but known occurrences are limited. Because of the species' apparent rarity, a model to identify areas with suitable habitat conditions for the species will allow conservationists to effectively prioritize and allocate scarce surveying resources. We used known occurrence records, a suite of environmental datasets, and four species distribution modeling techniques (generalized additive, GAM; maximum entropy, MaxEnt; generalized boosted, GBM; and weighted ensemble) to generate maps to inform surveys for *R. crinipes*. The ensemble approach improved predictive performance (AUC-PR = 0.95) compared to other techniques (individual model AUC-PR ranged from 0.7 to 0.8). We also obtained quantitative insights on the species' habitat relationships, including the likelihood of *R. crinipes*'s presence near Atlantic white cedar (*Chamaecyparis thyoides*) habitat and floodplains, which is consistent with prior field observations. The ensemble model indicated that 3.6% of the study area could be suitable habitat, but only 0.38% had high suitability. Small stream riparian habitats and Atlantic swamp forests in Alabama, Florida, and Georgia had the highest proportion of suitable areas. Prioritizing surveys in areas with model-indicated high habitat suitability is expected to reveal additional *R. crinipes* occurrences. We suggest surveying efforts for other at-risk species may benefit from using an ensemble modeling approach to identify and prioritize survey areas and improve ecological knowledge of these species.

Index terms: at-risk species; ensemble modeling; habitat suitability; SDM

INTRODUCTION

At-risk species pose numerous challenges for conservation practitioners, particularly when determining whether or not such species may need active management or warrant legal protections. Many conservation challenges arise from limited available information for these species, including poorly described habitats and ecological needs, poorly defined ranges, and small number of populations that increase difficulties locating occurrences in the wild (Rutrough et al. 2019). As such, surveys can become resource intensive and overly burdensome (Pellet and Schmidt 2005). Tools that help identify and prioritize areas with a higher likelihood of finding an at-risk species can improve the efficiency and effectiveness of allocating scarce resources for surveys (Guisan et al. 2013; Tulloch et al. 2016).

Species distribution modeling (SDM) can be used to identify species' suitable habitat based on known occurrence locations and associated environmental data (Franklin 2013). This approach has advanced in recent years with various new modeling techniques being developed, which include regression-based and machine-learning methods (Elith and Graham 2009;

Jones-Farrand et al. 2011). The ample number and variety of available modeling methods, however, may overwhelm conservation practitioners interested in developing an SDM for a particular species, as each method has its own attendant requirements (e.g., use of presence-only data or presence and absence data) and limitations. Further, each method will also likely produce different results, making model interpretation a challenge for novice and experienced practitioners alike. Ensemble models can reduce the subjectivity associated with selection of any particular modeling approach by combining the output (e.g., suitability scores) of multiple models into a single final weighted model (Araújo and New 2007). These types of models are commonly used in climate and weather forecasting, where ensemble models are preferred over any one of the numerous available climate models as they better capture uncertainty associated with the various modeling approaches (Lutz et al. 2016). As such, ensemble species distribution models present an attractive option for practitioners as they simplify the process of choosing a particular model technique and reduce the uncertainty associated with individual model algorithms (Guisan et al. 2017).

Rhynchospora crinipes (Gale) is a perennial plant species that may benefit from improved understanding of its habitat distribution and more efficient, effective surveys (*sensu* Imm et al. 2001). *R. crinipes*, commonly known as bearded beaksedge, Alabama beaksedge, mosquito beaksedge, or hairy-peduncled beakrush, has a broad range within the southeastern United States with occurrences scattered across five states including southern portions of Mississippi, Alabama and Georgia, the Florida Panhandle, and central North Carolina (Schotz 2017). No occurrences are known from South Carolina (Kral 1996; McMillan 2006). This perennial plant species is typically found in the Atlantic and Gulf Coastal Plains (Kral 1996), often on banks and bars of blackwater streams (Kral 1996; Carter 2005), and in drained areas with high percentages of silt and clay (McMillan 2006; Chafin 2011). *Rhynchospora crinipes* has also been described as having a close association with Atlantic white cedar (*Chamaecyparis thyooides* (L.)) occurrences (Kral 1996).

As presently understood, *R. crinipes* is considered rare throughout its known range, although it can be locally abundant (Kral 1996; Schotz 2017). Continued discoveries of previously unknown populations since the 1980s, along with the apparent abundance of ostensible suitable habitat, have led some to contend that more populations may yet be found (Kral 1995; Sorrie et al. 1997; Schotz 2017). While focused surveys and fortuitous discoveries have continued to identify new locations of *R. crinipes* (e.g., Anderson 1988, 1991, 1995; Kral 1995; Sorrie et al. 1997), future surveying efforts to locate additional populations of this—and other at-risk species with similarly wide ranges—could become challenging and overly burdensome without a tool that informs such surveys. Accordingly, we sought to (1) produce and evaluate a suite of individual and ensemble models that can identify areas with greater potential suitability for *R. crinipes* and, consequently, allow conservation managers to prioritize future surveys; (2) identify habitat associations that can improve our ecological knowledge of the species; and (3) estimate the species' current representation on protected lands based on the best available model. We discuss how similar approaches to generating distribution models can be applied to other at-risk species to improve the efficiency and effectiveness of survey efforts.

METHODS

Study Area

We modeled *R. crinipes* habitat suitability across 87,148 km² of the coastal plain in the southeastern United States. The modeled area included the species' known range (portions of Alabama, Florida, Georgia, Mississippi, and North Carolina) and portions of South Carolina, as we wanted to detect potential suitable habitat in that state (Figure 1). To delimit the study area, we used six-digit hydrologic unit code (HUC6) basins in the coastal plain within these states and trimmed them to 320 km inland. We chose this distance since the farthest inland record for the species was approximately 320 km from the Gulf of Mexico in Georgia (the farthest record from the Atlantic coast was located in North Carolina at around 200 km inland) and since there is no previous indication that the species is likely to occur in Virginia. However, in Florida, we restricted the study

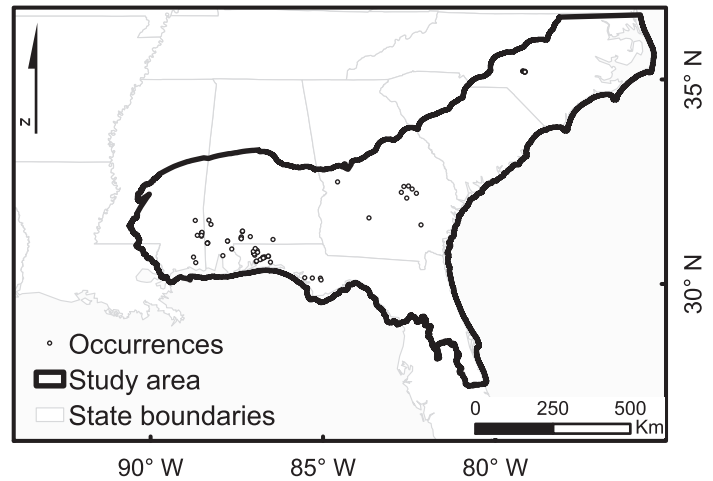


Figure 1.—Study area corresponding to the approximate *R. crinipes* range in the southeastern United States including portions of the states of Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina. The occurrence records used to build the models are depicted.

area to HUC6 basins across the Florida panhandle and northern portions of peninsular Florida, as no records have been reported at lower latitudes. A small portion of Louisiana was included as it overlapped with the basins and Kral (1996) suggested that the species could occur in the eastern portion of the state.

Data

We modeled habitat suitability using two types of data: (1) georeferenced occurrence records for the species (presences), and (2) environmental predictor variables presumed to be associated with conditions required for *R. crinipes* presence in the form of digital raster surfaces.

We obtained 165 presence records primarily from State Natural Heritage Programs, surveys (e.g., Kral 1995; Schotz 2017), and available herbarium records. We filtered these records to randomly select one record for every 500 m to prevent overfitting the models, resulting in a final set of 91 presence records for model training and evaluation (Figure 1). To compensate for the lack of absence data (both presence and absence data are necessary to calibrate SDM models), we generated a set of random points where the species is assumed to be absent or where the species may be present but was not observed (pseudoabsences; Phillips et al. 2009). Given the riparian nature of the species, we used a two-step sequence of operations to restrict the possible area for drawing pseudoabsence random points. First, we retained all areas within an arbitrary 50 km buffer of occurrence records to avoid areas distant from known locations. Second, within the retained area, we then created a 200 m (maximum distance to a stream within our occurrence dataset) buffer around all streams registered in the National Hydrography Dataset (NHD; USGS 2018a). We used this buffer as our pseudoabsence sampling area, where we drew 500 random points for model calibration and testing.

We compiled environmental variables that have been associated with *R. crinipes* (e.g., those described by Kral [1995] and Schotz [2017]). Therefore, we included variables related to soil characteristics including percent clay and soil organic

Table 1.—Environmental predictor variables. The first three columns correspond to the range (minimum, median, and maximum) of environmental values associated to the presence dataset. The last three columns show the relative importance of each environmental predictor variable for each modeling approach. Since each model algorithm (GAM, GBM, and Maxent) calculated its own model contribution index, we include the approximate importance of variables in terms of *P* values for GAM, where smaller values indicate variables that better inform the model. We also include the relative influence of the model variable (GBM), and the percent contribution to the model (MaxEnt), where larger values denote more informative variables in the respective model.

Variable (units)	Values for presences			Model contribution		
	Min	Median	Max	GAM	GBM	MaxEnt
Annual precipitation (mm/y)	1123	1606	1680	0.01	14.1	1.8
Soil percent clay (%)	0.5	10	42.8	0.07	5.6	1.7
Soil organic matter (%)	0.5	15	72	0.38	15.6	2.8
Distance to streams (m)	0	30	200	<0.001	8.2	7.6
Mean annual stream flow (cms)	0	0.02	0.056	0.03	6.4	1.6
Mean annual stream velocity (mps)	0	0.3	124.5	0.53	2.7	1.7
Distance to floodplain (m)	0	24.3	1714	0.04	7.5	35.6
Mean summer NDVI (units)	0.2	0.7	0.9	0.16	3.2	0.1
Tree canopy cover (%)	0	85	100	0.27	1.4	2
Atlantic white cedar suitability (HSI)	0.1	0.5	0.9	<0.001	34.7	44.9

matter. To incorporate the species' riparian habitat needs, we included distance to streams, and the flow and velocity of the nearest stream segment. We also incorporated vegetation variables that might influence light availability and plant associations, such as percent tree canopy and Normalized Difference Vegetation Index (NDVI) (complete set of variables in Table 1). We obtained environmental predictor variables from geospatial data repositories including WorldClim (Hijmans et al. 2005) for climatic variables, Landsat satellite imagery collections on Google Earth Engine (Gorelick et al. 2017) for vegetation and surface reflectance variables, SSURGO (USDA 2019a) for soils variables, and the National Hydrography Database (USGS 2018a) for stream variables. We resampled all environmental variables to a 30 m cell size. Since environmental variables could be associated with each other (e.g., elevation and temperature), we used Pearson's correlation analysis to calculate the degree of correlation among each combination of environmental variables and retained variables that were not correlated to each other ($|r| < 0.7$; Dormann et al. 2013).

Because of *R. crinipes*'s purported association with Atlantic white cedar (cf. Kral 1996), we generated an ensemble suitability map for Atlantic white cedar to use as a predictor variable for *R. crinipes*. The process to generate this ensemble map was similar to that described for *R. crinipes*, except for three key differences. First, we used 99 spatially filtered occurrence records for Atlantic white cedar obtained from the Global Biodiversity Information Facility (GBIF 2019). Second, we restricted sampling of pseudoabsence data to a 50 km buffer around presence data only (not to a buffer around streams). Third, the environmental predictor variables for this model were annual precipitation, average coldest winter temperature, distance to floodplains, soil organic matter, soil percent sand, tree canopy cover, NDVI, and distance to streams. The resulting habitat suitability ensemble map was not correlated to other predictor variables used for *R. crinipes* ($|r| < 0.7$).

Data Analysis

We used three individual modeling approaches that are commonly used in SDM, including a regression-based model (i.e., generalized additive model [GAM; Hastie and Tibshirani

1986]) and two machine-learning models (generalized boosted model, also known as boosted regression trees [GBM; Friedman et al. 2000], and maximum entropy [MaxEnt; Phillips et al. 2006]). These models are among the most frequently used and well-documented techniques encountered in SDM literature (e.g., Elith et al. 2006; Elith and Graham 2009; Heikkinen et al. 2012). All statistical analyses were completed using R 3.5.2 software (R Core Team 2017). We first tabulated the pixel values for all retained environmental layers at each presence and pseudoabsence point. With this table, we then calibrated each of the three models using the libraries *mgcv* for GAM (Wood 2018), *dismo* for MaxEnt (Hijmans et al. 2017), and *gbm* for GBM (Greenwell et al. 2019) with binary presence/pseudoabsence data as the response variable and environmental predictor data as independent variables. We then projected each calibrated model to the full study area to generate three separate prediction maps, each representing a habitat suitability index (HSI) for the species, with pixel values ranging from 0 (no suitability) to 1 (high suitability). Since we used presence-only records in our study, the habitat suitability index is understood as the similarity to species' used habitat (Latif et al. 2015).

We used a leave-one-out approach to evaluate the predictive power of each model. This approach did not require splitting data into separate training and validation datasets, and allowed all records that were included in model calibration ($n = 591$) to be also used for model validation by comparing them with a set of model-predicted presences and absences. We first calibrated a model using all records in the presence and pseudoabsence environmental response table minus one observation row (i.e., a training dataset of size $n - 1$, or 590 in our case). The omitted record's environmental data were then used to solve the calibrated model to obtain a predicted presence value (i.e., a test value ranging from 0 for predicted absence to 1 for predicted presence). We replaced the omitted row back into the environmental response table and removed the following row to repeat the process and obtain the next predicted test value. We iterated this sequence a total of 591 times to generate a list of predicted test values of size n to be contrasted with actual presence and pseudoabsence records. We used the AUC-PR (precision recall of the area under the receiver operating

Table 2.—Model performance metrics for each model algorithm, where higher values represent better association between model-predicted habitat suitability scores and actual presence and pseudoabsence records.

Metric	GAM	GBM	MaxEnt	Ensemble
AUC-PR	0.7	0.74	0.8	0.95
AUC-ROC	0.93	0.95	0.95	0.99
Pearson's correlation	0.69	0.8	0.76	0.79
Jaccard index	0.58	0.72	0.71	0.69

characteristic curve) as a metric to evaluate the agreement between actual and model-predicted presence and absence data. While AUC-ROC is among the most popular evaluation metrics in SDM literature, its AUC-PR variation can be more robust when working with rare species (Sofaer et al. 2019a). The range of possible AUC-PR values varies from 0.5 (poor model performance) to 1 (good model performance). As a robustness check, for each model we also calculated the AUC-ROC, the Jaccard index, and Pearson's correlation coefficients between model-predicted suitability scores and actual presence and pseudoabsence records.

We finally generated an ensemble model that reflected the agreement of individual models and also weighted the predictive performance of each individual modeling approach (Thuiller et al. 2009; Trolle et al. 2014). We calculated the weight of each model (GAM, GBM, and MaxEnt) by dividing its AUC-PR performance score by the sum of all three AUC-PR scores. As such, models with greater AUC-PR performance had higher weight in the final ensemble. We multiplied each map by its respective weight, and added the resulting products to form the ensemble. The final ensemble model for each species thus consisted of a map with pixel values of habitat suitability index (HSI) ranging from 0 (not suitable) to 1 (highly suitable). We calculated the performance metrics for the model ensemble in a similar fashion to individual models by contrasting the model predicted values with actual presence and pseudoabsence values.

RESULTS

The ensemble habitat suitability model for *R. crinipes* had a greater performance compared to its individual components in two of the four performance metrics (e.g., AUC-PR = 0.95 & AUC-ROC = 0.99). As such, the GAM model exhibited the lowest performance (e.g., AUC-PR = 0.7) compared to that of GBM (AUC-PR = 0.74) and MaxEnt (AUC-PR = 0.8) (Table 2; Figure 2). The environmental predictor variables that consistently exhibited greater importance to inform individual models were distance to floodplain, suitability for Atlantic white cedar, and annual precipitation, whereas NDVI and percent canopy cover ranked the lowest (Table 1). *R. crinipes* presences were frequently located close to streams and floodplains (medians of 30 and 24 m, respectively) and in the vicinity of low stream velocity (medians of 0.3 m/s) and low flow streams (median of 0.02 m³/s). Species presence locations also were associated with low percentages of soil clay and organic matter (medians of 10% and 15%, respectively), and high tree canopy cover and NDVI (medians of 85% and 0.7, respectively) (Table 1). In addition, *R. crinipes* presences were associated with habitat suitability for

Atlantic white cedar (median = 0.5 HSI; Figure 3) and we found a 35.5% overlap between the suitable areas of both species.

The ensemble model indicated that 284,863 ha (3.26% of the study area) could be deemed as suitable habitat for *R. crinipes* (HSI > 0.4, corresponding to the threshold of habitat suitability that maximized the sum of training sensitivity and specificity in our model as suggested by Liu et al. [2016]). We defined areas of low suitability as those with HSI = 0.4–0.6, medium with HSI = 0.6–0.8, and high with HSI > 0.8. As such, most of this suitable area belonged to the low suitability category (96.6%), 2% to medium, and only 0.38% to high suitability. Most of the suitable area was concentrated in Alabama, Florida, and Georgia, together accounting for nearly two-thirds (65.1%) of suitable area, while Mississippi, North Carolina, and South Carolina accounted for the remaining third (34.9%) (Table 3). When overlaying habitat suitability with the vegetation type map defined by LANDFIRE (USDA 2019b), three vegetation types accounted for nearly half (45.5%) of the suitable areas identified: eastern small stream riparian forests, Atlantic swamp forests, and eastern floodplain forests. Combined with seven other vegetation types (each accounting for <10% of identified suitable habitat), these accounted for 79.7% of all potentially suitable habitat identified (Table 4). Forty-nine other vegetation types accounted for the remaining 20.3% of suitable area (not included in Table 4).

Of the total suitable area, 19% (53,731 ha) overlapped with established protected areas registered in the U.S. protected areas network (PAD-US; USGS 2018b). However, this proportion varied by state, with Florida having the largest proportion of suitable area protected (46.5%) while the other states varied from 7.8% to 18.4% (Table 3). According to the PAD-US land ownership type classification, the largest proportions of suitable areas within protected areas corresponded to the state (35%), federal (29.7%), unknown (11.3%), or private categories (8.6%) (Table 5). With respect to vegetation types, only Atlantic coastal marshes and pine flatwoods had high proportions of protected area (46.9% and 56.5%, respectively; Table 4).

DISCUSSION

Modeling species distributions can be a valuable tool to improve monitoring efforts for at-risk species and we identified candidate areas that can be targeted for future surveys for *R. crinipes*. We found that the ensemble model had an overall greater predictive performance compared to individual model techniques, since it ranked highest in two of four performance metrics and was among the top for the other metrics. These results are similar to those reported by Marmion et al. (2009), who found that weighted ensemble models increased the accuracy of species distribution forecasts. Therefore, using the ensemble approach to produce models for rare species may reduce the bias of individual models and improve survey effectiveness (Araújo and New 2007; Lomba et al. 2010). For instance, GAM estimated a large proportion of habitat suitability (29.5% of the study area), while GBM, MaxEnt, and the ensemble estimated lower proportions (2.9%, 9.7%, and 3.2%, respectively; Figure 2). However, given that this model had the lowest predictive performance, much of this suitable area could

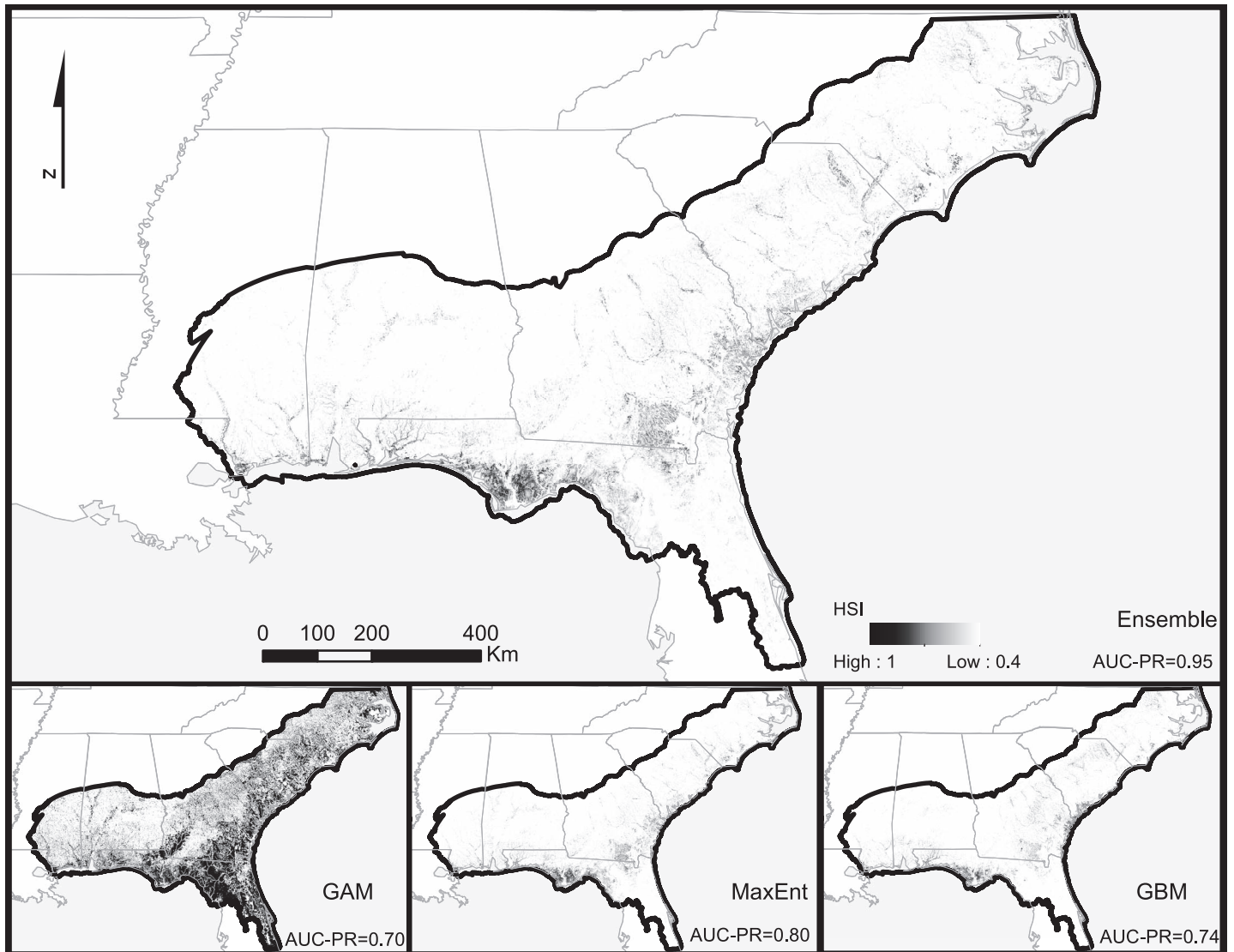


Figure 2.—Habitat suitability maps for *R. crinipes* using three modeling algorithms and the resulting ensemble. Areas with consistently higher habitat suitability scores in individual model approaches are carried out to the ensemble model, while areas deemed as suitable in only one approach are not represented in the ensemble.

be the result of overprediction, particularly in areas far off riparian zones. Accordingly, GAM's lower performance had a lower weight in the ensemble, resulting in most of those possibly overpredicted suitable areas not being represented in the final model.

In addition to identifying potentially suitable areas, we contributed to the understanding of the species' habitat, which, like many rare species, was mostly limited to descriptive habitat associations, and/or highly localized empirical observations. By relating environmental predictor variables with occurrence data, we gained greater insight into how a given species responds to environmental gradients. For example, in contrast to previous observations noting that *R. crinipes* was associated with high percentages of silt and clay (McMillan 2006; Chafin 2011), we found that occurrences were most commonly found in lower percentages of clay (mostly around 10%; Figure 3). Also, the species was reported in full sun or light shade (Chafin 2007),

while we found a median of 85% tree canopy cover, indicating a potential for higher occurrence frequency in shaded areas than expected. In addition, we corroborated that *R. crinipes* is more likely to occur near rivers and floodplains (Table 1; Figure 3), characteristics that have been previously noted (e.g., Kral 1996; Carter 2005).

We also observed a strong relationship between habitat suitability for Atlantic white cedar and that of *R. crinipes* (Figure 3), which aligns with previous assertions about the association of these two species (Kral 1996; Sheridan et al. 1997). However, only one-third of suitable habitat was common to the two species, which may indicate that while they are associated, *R. crinipes* is not dependent on Atlantic white cedar. Because of the lack of comprehensive data locating blackwater streams, we had to use various stream characteristics that may serve as proxies for them. Using these proxies, we detected that occurrences are more frequent near stream segments that have low velocities

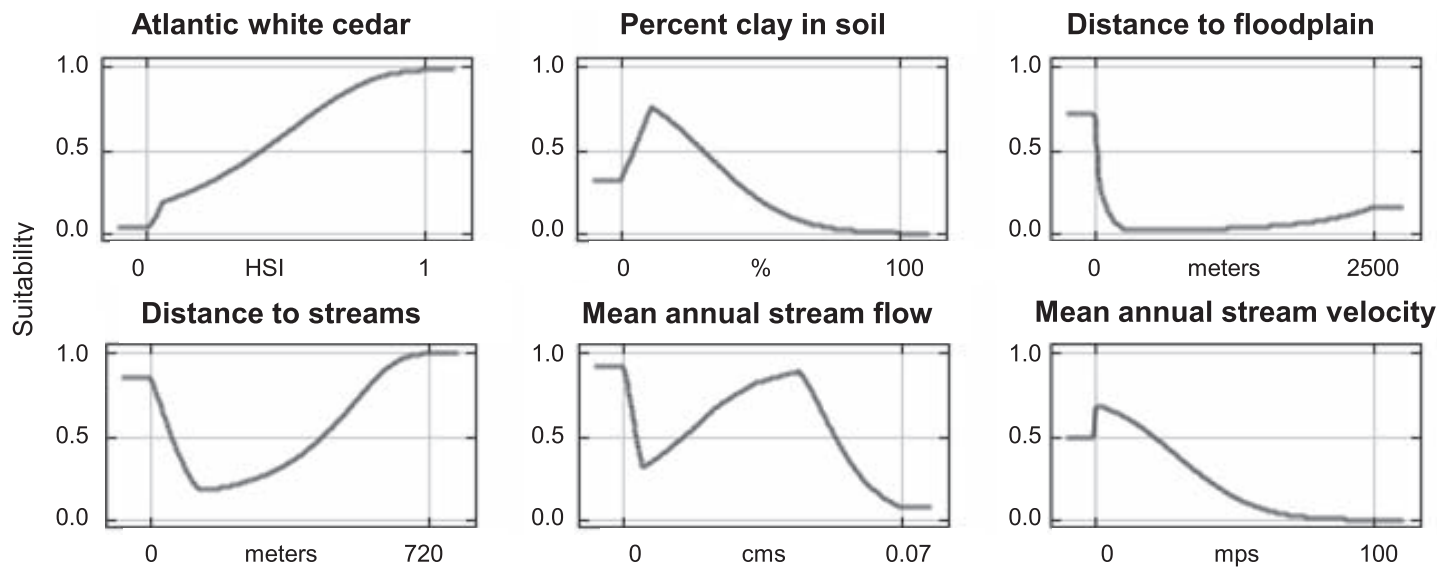


Figure 3.—Response curves for a selection of environmental predictor variables in the MaxEnt model. Each curve depicts how suitability for the species changes across each environmental variable gradient.

(medians of 0.3 m/s) and medium water flow volumes ($0.02 \text{ m}^3/\text{s}$) (Table 1, Figure 3). These stream characteristics, in combination with other environmental variables, allowed us to discriminate among the numerous riparian zones in the study area with only 3.2% of the riparian areas deemed suitable. Nevertheless, development of a spatially explicit characterization of blackwater streams (e.g., from remote sensing data) could represent a better predictor over proxy variables, as it could include other blackwater characteristics such as tannin concentrations.

Most of the rare plants in southeastern forests are herbs, which require specialized, connected habitats and are susceptible to disturbance (Imm et al. 2001). Within *R. crinipes*'s range, floodplain forests have historically been subjected to periodic logging and other anthropogenic disturbances (Chafin 2007; Schotz 2017). Of particular interest for *R. crinipes* conservation, we estimated that 19% of the species' predicted suitable habitat had some degree of protection, particularly on state- and federally-protected lands. Nevertheless, over one-third (35%) of suitable areas belonged to other protection categories, including private lands, local governments, and NGOs (Table 5), highlighting the potential conservation

Table 3.—Suitable areas for *R. crinipes* in the study area. We show the total suitable area per each state and the proportion of that area currently designated as protected areas (PA) and non-protected areas (non-PA).

State	Total suitable area in ha (% of study area)	Suitable area in PA in ha (% of state total)	Suitable area in non-PA in ha (% of state total)
Alabama	58,947 (20.7)	4582 (7.8)	54,365 (92.2)
Florida	63,132 (22.1)	29,371 (46.5)	33,761 (53.5)
Georgia	63,186 (22.2)	6145 (9.7)	57,041 (90.3)
Mississippi	25,154 (8.8)	4623 (18.4)	20,531 (81.6)
North Carolina	34,657 (12.2)	3836 (11.1)	30,820 (88.9)
South Carolina	39,788 (14)	5174 (13)	34,614 (87)
Total	284,863 (100)	53,731 (19)	231,132 (81)

impact that these protected areas could have for this (and other) at-risk species (Graves et al. 2019). In our study area, Alabama and Georgia had large shares of suitable habitat (20.7% and 22.2%, respectively), but also exhibited lower proportions of suitable habitat with land protection (7.8% and 9.7%, respectively), which present additional conservation challenges for the species in these states (Table 3). A similar pattern was observed in vegetation types, where the largest proportions of suitable areas were found in eastern small stream riparian forests (20.8%) and Atlantic swamp forests (14.8%); yet these were among the least protected vegetation types (11.6% and 24.3%, respectively). Strategic conservation of suitable habitat, particularly small stream riparian forests and Atlantic swamp forests, has the potential to improve long-term conservation of *R. crinipes* and other at-risk species.

Table 4.—Proportion of suitable habitat area occupied by different vegetation types as defined by LANDFIRE vegetation layer. We only show the 10 main classes (out of 59) that combined accounted for >75% of the total suitable area. We provide summaries for both protected (PA) and non-protected areas (non-PA).

Vegetation type	Total suitable area in ha (% of study area)	Suitable area in PA in ha (%)	Suitable area in non-PA in ha (%)
Eastern small stream riparian forests	61,626 (20.8)	7164 (11.6)	54,462 (88.4)
Atlantic swamp forests	43,839 (14.8)	10,636 (24.3)	33,203 (75.7)
Eastern floodplain forests	29,241 (9.9)	5812 (19.9)	23,429 (80.1)
Managed tree plantation	25,848 (8.7)	2410 (9.3)	23,438 (90.7)
Longleaf pine woodland	20,834 (7)	4589 (22)	16,244 (78)
Atlantic coastal marsh	15,950 (5.4)	7488 (46.9)	8462 (53.1)
Pine flatwoods	14,483 (4.9)	8187 (56.5)	6296 (43.5)
Coastal plain oak forest	8977 (3)	869 (9.7)	8107 (90.3)
Maritime forest	7340 (2.5)	1885 (25.7)	5454 (74.3)
Developed-upland herbaceous	5112 (1.7)	1411 (27.6)	3701 (72.4)

Table 5.—Ownership type for suitable habitat within protected areas as defined in the Protected Areas Database of the United States (PAD-US; USGS 2018b).

Ownership	Suitable area in ha (%)
State	19,148 (35.6)
Federal	15,967 (29.7)
Unknown	6056 (11.3)
Private	4628 (8.6)
Regional agency special district	3360 (6.3)
Designation	2332 (4.3)
Local government	1024 (1.9)
Non-governmental organization	1004 (1.9)
American Indian lands	211 (0.4)

Our modeling work met the standards recommended by Sofae et al. (2019b) and Araújo et al. (2019) for developing species distribution models to inform decision making. These included the use of quality presence data, performance of cross validation of the models, use of species-tailored predictors at an appropriate extent and spatial and temporal resolutions, use of multiple algorithms including a model ensemble, evaluation with multiple metrics, and model development with expert opinion. Nevertheless, as a cautionary note, our model does not represent the likelihood of species occurrence, but rather the degree of similarity of predicted areas to used habitat, which may be used to inform design and implementation of additional surveys for *R. crinipes* presence. Species distribution models, such as we have presented herein, are not inherently final and should be refined and recalibrated as new survey data become available. For instance, while *R. crinipes* has not been reported in South Carolina (McMillan 2006), the ensemble model showed opportunities for conducting surveys in the state (Table 3, Figure 2), which could then be used to supply additional calibration data (presence and absence data) to refine and improve current models. Another important avenue to improve these models will be the inclusion of true absence data (i.e., the location where the species has been surveyed for but not detected). However, logistical constraints often prevent sampling designs whereby absences may be recorded, resulting primarily in opportunistic observations of species presence. Structured sampling designs that allow recording of true absences will help further refine models by reducing the uncertainty caused by having to use pseudoabsences (Lobo et al. 2010).

Our study illustrates the potential utility of species distribution models (particularly ensemble models) for identifying gaps in representation of suitable habitats in current conservation land portfolios, which can subsequently inform conservation strategies for at-risk species. We produced a habitat suitability map that can inform monitoring and survey efforts for *R. crinipes* in the southeastern United States. A high-resolution digital map can be accessed at USGS ScienceBase-Catalog (<https://www.sciencebase.gov/catalog/item/5f57a9ef82cea1f95a61a6f3>). By combining the output of individual model approaches into an ensemble, we were able to improve the predictive model performance and improve our understanding of the species' habitat associations at the macroscale. Our ensemble model will be used to guide additional surveys for the species in the next survey season,

where new records will allow refining the models herein. However, further modeling work would be contingent on funding availability. Therefore, species distribution modeling can be an effective tool to incorporate our understanding of at-risk species into a transparent framework. Modeling potential distributions can serve a variety of applications beyond survey planning, because it can be the first step for further analyses, including informing protection and restoration, informing species translocations and reintroductions, and estimating impacts to species viability from climate and land use change (Franklin 2013; Urbina-Cardona et al. 2019). We encourage practitioners to take advantage of this approach to target survey efforts for other at-risk species.

ACKNOWLEDGMENTS

We thank our collaborators at State Natural Heritage Program offices for providing data used to develop these models. We also thank the valuable comments of the anonymous reviewers. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service or U.S. Geological Survey. Three of the authors are U.S. Government Employees. This work was prepared as part of their official duties. Title 17 U.S.C. §105 provides that “Copyright protection under this title is not available for any work of the United States Government.” Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Funding for this project was available thanks to cooperative agreement #F17AC00899 with the U.S. Fish and Wildlife Service. This material was also supported by McIntire-Stennis Funds #1019539 and the Forest and Wildlife Research Center at Mississippi State University.

Carlos Ramirez-Reyes is a conservation scientist with expertise in landscape ecology and land use science.

Garrett Street is a quantitative ecologist specializing in species distributions, animal movement and statistical model development and interpretation. His work emphasizes the mechanisms linking individuals to populations and linking ecological theory to emergent phenomena.

Francisco J. Vilella is Assistant Unit Leader of the Mississippi Cooperative Fish and Wildlife Research Unit and Professor of Conservation Biology in the Department of Wildlife, Fisheries and Aquaculture at Mississippi State University.

D. Todd Jones-Farrand is a science coordinator working in the South Atlantic Gulf and Mississippi Basin regions. His primary roles are supporting Species Status Assessments and landscape-scale conservation planning.

M. Scott Wiggers is a botanist with expertise in threatened and endangered species recovery and policy, conservation and fire ecology of longleaf pine savannas, and private lands conservation.

Kristine Evans is an Assistant Professor of Conservation Biology and co-director of the Quantitative Ecology and Spatial Technologies Lab in the Department of Wildlife, Fisheries and Aquaculture at Mississippi State University.

LITERATURE CITED

- Anderson, L.C. 1988. Status of endangered *Rhynchospora crinipes* (Cyperaceae). *Systematic Botany* 13:410.
- Anderson, L.C. 1991. Noteworthy plants from North Florida. *SIDA, Contributions to Botany* 14:467-474.
- Anderson, L.C. 1995. Noteworthy plants from North Florida. *SIDA, Contributions to Botany* 16:581-587.
- Araújo, M.B., R.P. Anderson, A. Márcia Barbosa, C.M. Beale, C.F. Dormann, R. Early, R.A. Garcia, A. Guisan, L. Maiorano, B. Naimi, et al. 2019. Standards for distribution models in biodiversity assessments. *Science Advances* 5:eaat4858.
- Araújo, M.B., and M. New. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22:42-47.
- Carter, R. 2005. An introduction to the sedges of Georgia. *Tipularia* 20:17-45.
- Chafin, L. 2011. Georgia's natural communities and associated rare plant and animal species: Thumbnail accounts. Unpublished. Available at <<https://georgiawildlife.com>>.
- Chafin, L.G. 2007. *Field Guide to the Rare Plants of Georgia*. State Botanical Garden of Georgia, Athens.
- Dormann, C.F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R.G. Marquez, B. Gruber, B. Lafourcade, P.J. Leitão, et al. 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27-46.
- Elith, J., and C.H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32:66-77.
- Elith, J., C. Graham, R. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J. Leathwick, A. Lehmann, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.
- Franklin, J. 2013. Species distribution models in conservation biogeography: Developments and challenges. *Diversity and Distributions* 19:1217-1223.
- Friedman, J., T. Hastie, and R. Tibshirani. 2000. Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics* 28:337-407.
- GBIF. 2019. Global Biodiversity Information Facility. <<https://www.gbif.org/>>
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202:18-27.
- Graves, R.A., M.A. Williamson, R.T. Belote, and J.S. Brandt. 2019. Quantifying the contribution of conservation easements to large-landscape conservation. *Biological Conservation* 232:83-96.
- Greenwell, B., B. Boehmke, and J. Cunningham. 2019. Package "gbm" - Generalized boosted regression models. Version 2.1.5. CRAN.
- Guisan, A., W. Thuiller, and N.E. Zimmermann. 2017. *Habitat Suitability and Distribution Models*. Cambridge University Press, Cambridge, UK.
- Guisan, A., R. Tingley, J.B. Baumgartner, I. Naujokaitis-Lewis, P.R. Sutcliffe, A.I.T. Tulloch, T.J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, et al. 2013. Predicting species distributions for conservation decisions. *Ecology Letters* 16:1424-1435.
- Hastie, T., and R. Tibshirani. 1986. Generalized additive models. *Statistical Science* 3:297-318.
- Heikkinen, R.K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276-288.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965-1978.
- Hijmans, R.J., S. Phillips, J. Leathwick, and J. Elith. 2017. Package "dismo" - Species distribution modeling. Version 1.1-4. CRAN.
- Imm, D.W., H.E. Shealy Jr., K.W. McLeod, and B. Collins. 2001. Rare plants of southeastern hardwood forests and the role of predictive modeling. *Natural Areas Journal* 21:36-49.
- Jones-Farrand, D.T., T.M. Fearer, W.E. Thogmartin, F.R. Thompson III, M.D. Nelson, and J.M. Tirpak. 2011. Comparison of statistical and theoretical habitat models for conservation planning: The benefit of ensemble prediction. *Ecological Applications* 21:2269-2282.
- Kral, R. 1995. Status report on *Rhynchospora crinipes* Gale. Unpublished report for the U.S. Fish and Wildlife Service.
- Kral, R. 1996. Supplemental notes on *Rhynchospora crinipes* and related species in section *Fuscae* (Cyperaceae). *SIDA, Contributions to Botany* 17:385-411.
- Latif, Q.S., V.A. Saab, K. Mellen-Mclean, and J.G. Dudley. 2015. Evaluating habitat suitability models for nesting white-headed woodpeckers in unburned forest. *Journal of Wildlife Management* 79:263-273.
- Liu, C., G. Newell, and M. White. 2016. On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution* 6:337-348.
- Lobo, J.M., A. Jiménez-Valverde, and J. Hortal. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103-114.
- Lomba, A., L. Pellissier, C. Randin, J. Vicente, F. Moreira, J. Honrado, and A. Guisan. 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation* 143:2647-2657.
- Lutz, A.F., H.W. ter Maat, H. Biemans, A.B. Shrestha, P. Wester, and W.W. Immerzeel. 2016. Selecting representative climate models for climate change impact studies: An advanced envelope-based selection approach. *International Journal of Climatology* 36:3988-4005.
- Marmion, M., M. Parviainen, M. Luoto, R.K. Heikkinen, and W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* 15:59-69.
- McMillan, P.D. 2006. A taxonomic and phytogeographic treatment of the *Rhynchospora* (Cyperaceae) of South Carolina and the eastern United States, including a revision of *Rhynchospora* section *Globulares*. Doctoral diss., Clemson University, Clemson, SC.
- Pellet, J., and B.R. Schmidt. 2005. Monitoring distributions using call surveys: Estimating site occupancy, detection probabilities and inferring absence. *Biological Conservation* 123:27-35.
- Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.
- Phillips, S.J., M. Dudík, J. Elith, C.H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* 19:181-197.
- R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria.
- Rutrough, A., I.V. Widick, and W.T. Bean. 2019. Reconstruction of the historical range alters niche estimates in an endangered rodent. *Ecography* 42:1742-1751.
- Schotz, A. 2017. Preliminary status assessment update of *Rhynchospora crinipes* (Cyperaceae), hairy-peduncled beakrush. Unpublished report for the U.S. Fish and Wildlife Service.

- Sheridan, P., S. Orzell, and E. Bridges. 1997. Some noteworthy vascular plant records from Atlantic white-cedar, *Chamaecyparis thyoides* (L.) B.S.P., habitats of western Georgia. P. 82 in T. Shear and K. O. Summerville, eds., Proceedings: Atlantic White Cedar: Ecology and Management. USDA Forest Service, Southern Research Station, Newport News, VA.
- Sofaer, H.R., J.A. Hoeting, and C.S. Jarnevich. 2019a. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution* 10:565-577.
- Sofaer, H.R., C.S. Jarnevich, I.S. Pearse, R.L. Smyth, S. Auer, G.L. Cook, T.C. Edwards Jr., G.F. Guala, T.G. Howard, J.T. Morissette, and H. Hamilton. 2019b. Development and delivery of species distribution models to inform decision-making. *BioScience* 69:544-557.
- Sorrie, B.A., B. van Eerden, and M.J. Russo. 1997. Noteworthy plants from Fort Bragg and Camp MacKall, North Carolina. *Castanea* 62:239-259.
- Thuiller, W., B. Lafourcade, R. Engler, and M. Araujo. 2009. BIOMOD, a platform for ensemble forecasting of species distributions. *Ecography* 32:369-373.
- Trolle, D., J.A. Elliott, W.M. Mooij, J.H. Janse, K. Bolding, D.P. Hamilton, and E. Jeppesen. 2014. Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling and Software* 61:371-379.
- Tulloch, A.I.T., P. Sutcliffe, I. Naujokaitis-Lewis, R. Tingley, L. Brotons, K.M.P.M.B. Ferraz, H. Possingham, A. Guisan, and J.R. Rhodes. 2016. Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. *Biological Conservation* 199:157-171.
- [USDA] U.S. Department of Agriculture. 2019a. SSURGO-Web Soil Survey 3.3.
- [USDA] U.S. Department of Agriculture. 2019b. LANDFIRE Project - Existing vegetation type layer.
- [USGS] U.S. Geological Survey. 2018a. National Hydrography Datasets - NHDPlusV2.
- [USGS] U.S. Geological Survey. 2018b. Protected areas database of the United States (PAD-US) Version 2.0.
- Urbina-Cardona, N., M.E. Blair, M.C. Londoño, R. Loyola, J. Velásquez-Tibatá, and H. Morales-Devia. 2019. Species distribution modeling in Latin America: A 25-year retrospective review. *Tropical Conservation Science* 12:194008291985405.
- Wood, S. 2018. Package “mgcv” - Mixed GAM computation vehicle with automatic smoothness estimation. Version 1.8-31. CRAN.