Articles

Assessing Effects of Variation in Global Climate Data Sets on Spatial Predictions from Climate Envelope Models

James I. Watling,* Robert J. Fletcher, Jr., Carolina Speroterra, David N. Bucklin, Laura A. Brandt, Stephanie S. Romañach, Leonard G. Pearlstine, Yesenia Escribano, Frank J. Mazzotti

J.I. Watling, C. Speroterra, D.N. Bucklin, Y. Escribano, F.J. Mazzotti
University of Florida, Ft. Lauderdale Research and Education Center, 3205 College Avenue, Ft. Lauderdale, Florida 33314

R.J. Fletcher, Jr.
University of Florida, Department of Wildlife Ecology and Conservation, 318 Newins-Ziegler Hall, Gainesville, Florida 32611

L.A. Brandt
U.S. Fish and Wildlife Service, 3205 College Avenue, Ft. Lauderdale, Florida 33314

S.S. Romañach
U.S. Geological Survey, 3205 College Avenue, Ft. Lauderdale, Florida 33314

L.G. Pearlstine
National Park Service, Everglades National Park, South Florida Natural Resources Center, 950 North Krome Avenue, Homestead, Florida 33030

Abstract

Climate change poses new challenges for natural resource managers. Predictive modeling of species–environment relationships using climate envelope models can enhance our understanding of climate change effects on biodiversity, assist in assessment of invasion risk by exotic organisms, and inform life-history understanding of individual species. While increasing interest has focused on the role of uncertainty in future conditions on model predictions, models also may be sensitive to the initial conditions on which they are trained. Although climate envelope models are usually trained using data on contemporary climate, we lack systematic comparisons of model performance and predictions across alternative climate data sets available for model training. Here, we seek to fill that gap by comparing variability in predictions between two contemporary climate data sets to variability in spatial predictions among three alternative projections of future climate. Overall, correlations between monthly temperature and precipitation variables were very high for both contemporary and future data. Model performance varied across algorithms, but not between two alternative contemporary climate data sets. Spatial predictions varied more among alternative general-circulation models describing future climate conditions than between contemporary climate data sets. However, we did find that climate envelope models with low Cohen's kappa scores made more discrepant spatial predictions between climate data sets for the contemporary period than did models with high Cohen's kappa scores. We suggest conservation planners evaluate multiple performance metrics and be aware of the importance of differences in initial conditions for spatial predictions from climate envelope models.

* Corresponding author: watlingj@ufl.edu

Introduction

Climate change creates new challenges for natural resource managers by changing the dynamics between species and their environment (Thomas et al. 2004). For example, species ranges may shift and current protected area networks may no longer encompass suitable climate space for the species for which they were designed (Hannah et al. 2007). In recognition of these potential effects, many efforts have been initiated to assess vulnerability of species to climate change and develop strategies for responding to climate change in ways that continue to protect species and habitats (Povilitis and Suckling 2010; Rowland et al. 2011; Reece and Noss 2013). The use of modeling for forecasting potential impacts of climate change and aiding in development of strategies is becoming increasingly common (Araújo and Peterson 2012).

Climate envelope models (CEMs) are one type of model widely used to forecast potential climate change effects on biodiversity (Thomas et al. 2004; Lawler et al. 2009). A general approach to climate envelope modeling is to describe the climate conditions currently experienced by a species (the climate envelope) and forecast the future spatial distribution of the climate envelope according to projections of future climate change, assuming the contemporary species–climate relationship will hold true (Franklin 2009). There are numerous sources of uncertainty in predictions from CEMs (Dormann et al. 2008), including errors in the georeferenced species occurrences used in model training (Guisan et al. 2007), differences in the algorithm used to model species–environment relationships (Elith et al. 2006), potential sample biases in data used for model building (Kadmon et al. 2004; McCarthy et al. 2012), and differences among the future climate scenarios being modeled (Real et al. 2010).

Although the calibration of initial conditions can have substantive impacts on model performance and predictions (Araújo and New 2007), few studies have evaluated the consequences of using alternative climate data sets to define initial conditions in CEMs (but see Morin and Chuine 2005 and Parra and Monahan 2008). In this paper we address that gap by quantifying performance and spatial predictions of CEMs using two different contemporary data sets and then directly contrasting variability in spatial predictions from CEMs developed with different contemporary climate data sets to the variability in spatial predictions across three alternative projections of future climate.

Climate change is expected to be an important driver of ecological change in the 21st century and beyond (Rosenzweig et al. 2008), and there are many data sets available to describe historical, contemporary, and future climate conditions (http://www.ipcc-data.org). Given uncertainty in the course of future climate change (often described in the context of alternative carbon dioxide [CO_2] emissions scenarios; Nakicenović and Swart [2000]), and in the projections of alternative Atmospheric–Oceanic General Circulation Models (hereafter, GCMs; Diniz-Filho et al. 2009), an increasingly common approach in predictive modeling uses ensemble forecasts that draw on multiple projections from different models to describe some of the uncertainty in projections of future conditions (Araújo and New 2007). Much less attention has been directed at understanding how variation in contemporary climate data sets influences model predictions. However, a previous study found considerable variation in model performance and spatial predictions between two contemporary climate data sets for predictions of mammals in California using a single modeling algorithm (Parra and Monahan 2008).

Models are an important tool for natural resource decision-making, but uncertainties and limitations of models should be described as explicitly as possible when used to guide conservation policy (Real et al. 2010). Because protection of threatened and endangered species is a global environmental policy priority, our study provides a real-world example for quantifying variability in the climate data used to construct CEMs, and understanding effects of that variability on performance and spatial predictions of models describing species responses to climate change. We created CEMs for 12 species of federally listed threatened and endangered terrestrial vertebrates occurring in peninsular Florida (U.S. Endangered Species Act, ESA 1973, as amended; Table 1). The species differ in range size and degree of ecological specialization, allowing us to test performance and predictions of CEMs across a range of ecological contexts. We selected these species because their precarious status in the United States makes them likely candidates for vulnerability assessments (USFWS 2010), for which climate envelope models may provide key information. Development of robust, data-driven CEMs can play an integral role in determining future management strategies for these and other species in the face of climate change.

Because little is known about how differences in contemporary climate data sets affect CEMs, our overarching objective was to examine variation in spatial predictions from models calibrated using two alternative climate data sets. We compared variation in spatial predictions attributable to differences in the contemporary climate data set on which models were calibrated to the variation in predictions made using different GCMs describing future climate. To make our comparisons as generalizable as possible, we used three different

Table 1. Species and climate predictors used in construction of climate envelope models compiled in 2011.

Common name	Scientific name	Predictor variables
Mammals		
Florida panther	*Puma concolor coryi*	Temperature: July, October
		Precipitation: February, April, May, August, September, October
Birds		
Florida grasshopper sparrow	*Ammodramus savannarum floridanus*	Temperature: September
		Precipitation: February, May, June, October
Florida scrub jay	*Aphelocoma coerulescens*	Temperature: July, November
		Precipitation: January, March, May, June, August, October
Piping plover	*Charadrius melodus*	Temperature: July, October
		Precipitation: February, May, September, October
Wood stork	*Mycteria americana*	Temperature: July, October
		Precipitation: March, May, July, September, October
Audubon crested caracara	*Polyborus plancus audubonii*	Temperature: August, November
		Precipitation: April, May, September, December
Everglades snail kite	*Rostrhamus sociabilis plumbeus*	Temperature: July, September, November
		Precipitation: February, April, May, September, October
Whooping crane	*Grus americana*	Temperature: July, October
		Precipitation: March, May, August, October
Red-cockaded woodpecker	*Picoides borealis*	Temperature: August
		Precipitation: March, May, June, July, September
Reptiles and amphibians		
American crocodile	*Crocodylus acutus*	Temperature: May, December
		Precipitation: January, April, May, July, September, October, November
Sand skink	*Neoseps reynoldsi*	Temperature: September, December
		Precipitation: February, April, May, June, August, September, October
Eastern indigo snake	*Drymarchon corais couperi*	Temperature: July, September
		Precipitation: February, April, May, August, September, October

modeling algorithms and 12 different species in our study. We asked four specific questions, the first about the climate data themselves, followed by three questions about the effects of climate data on the outputs of the species' CEMs. First, is the discrepancy between contemporary climate data sets greater or less than the discrepancy among data sets describing future climate? Second, does CEM performance differ between models constructed using two alternative data sets of contemporary climate? Third, is the discrepancy in spatial predictions from CEMs using two alternative contemporary climate data sets less than the discrepancy in spatial predictions of future conditions made using data from three alternative GCMs? Fourth, is the discrepancy in spatial predictions using two alternative climate data sets associated with model performance?

Methods

We used two alternative climate data sets—the WorldClim data set (Hijmans et al. 2005) and the Climate Research Unit (CRU) data set (New et al. 2002)—to describe the contemporary climate conditions on which models were calibrated. The WorldClim data set is global in extent, covers a time span of approximately 50 years (1950–2000) and is available at spatial resolutions ranging from approximately 30 arc-seconds to 10 arc-minutes. The CRU data set is also global in extent, covers a time span of 29 years (1961–1990) and is available at a resolution of approximately 10 arc-min (all models described herein were created at the common resolution shared by the two data sets, 10 arc-min, or 1/6th of a degree). Both data sets draw on observations from ground-based climate stations and are widely used by researchers to forecast climate change effects on species. A preliminary comparison between the two data sets was made by Hijmans et al. (2005), who suggested that differences between the WorldClim and CRU data may be attributable to differences in the elevation data used as covariates in the interpolation of climate surfaces, differences in the number of weather stations included in the two studies, and differences in the way the climate surfaces were described for each 10 arc-min grid cell (an average value was used for the WorldClim data, whereas estimated climate in the middle of the grid cell was used for the CRU data). We used untransformed monthly climate variables rather than derived bioclimate variables because previous work showed that model performance and predictions were similar whether monthly or bioclimate data were used (Watling et al. 2012), and we assumed that many users would be more likely to use untransformed data.

Data on future climate conditions were extracted from downscaled projections of 21st century climate change (Tabor and Williams 2010). We used average projections for the years 2041–2060 from three GCMs (the Geophysical Fluid Dynamics Laboratory Coupled Model version 2.0, the National Center for Atmospheric Research Community Climate System Model version 3.0, and the Hadley Center for Climate Prediction, United Kingdom Meteorological Office coupled model version 3.0) under the A1B CO_2 emission scenario (a high-emissions scenario that assumes a balance between fossil-intensive and nonfossil future energy sources; Nakicenović and Swart 2000).

We compiled georeferenced species occurrences from a variety of online databases and the primary literature (Table S1, *Supplemental Material*). Occurrences were preprocessed to remove erroneous data, dubious records, or occurrences falling far outside the native geographic range of species. Occurrences from coastal areas that fell just outside the domain of the climate data were 'snapped' to the nearest terrestrial grid cell, and all duplicate observations per grid cell were removed. We used a modification of the target group approach (Phillips et al. 2009) to define an ecologically relevant domain for model development. Briefly, the target group approach specifies that model domain be defined by the composite geographic range of ecologically similar species sampled using similar methods as the focal species. Details on the delineation of the target group domain for each species are included in Text S1 (*Supplemental Material*). Georeferenced observations of all target group species were obtained from a single online database (the Global Biodiversity Information Facility; www.gbif.org), data were preprocessed as described for the species being modeled, and the 100% minimum convex polygon defining each target group was used as a mask to extract climate data from the two climate databases.

Predictor variables for each model were drawn from a pool of 24 candidate variables (12 monthly observations each of mean temperature and precipitation) using ecological niche factor analysis (Hirzel et al. 2002). To reduce the effects of colinearity among predictor variables, we removed highly correlated ($r > 0.85$) variables from the analysis based on inspection of the cluster diagram of variable correlation in the Biomapper program (Hirzel et al. 2002). When multiple intercorrelated variables were included in a cluster, we retained the variable that was most associated with species presence (based on its marginality, the extent to which a species occurs in areas where climate differs most significantly from average conditions in the study area; Hirzel et al. 2002) but less correlated ($r < 0.85$) with other selected variables. We arbitrarily chose to use the CRU data set to identify predictor variables for modeling, after which the corresponding subset of predictor variables was extracted from the WorldClim data set.

Climate envelope models were constructed for each species using three different algorithms: Maximum entropy (Phillips et al. 2006; Phillips and Dudík 2008), Generalized Linear Models (GLMs; McCullugh and Nelder 1989) and Random Forests (RF; Cutler et al. 2007). All three algorithms are capable of generating high-performance models (Elith et al. 2006; Guisan et al. 2007; Elith and Graham 2009). Maximum entropy models were coded in R (R Development Core Team 2005) but run in the program Maxent (Phillips et al. 2006), and the remaining algorithms were run in the program R. Because we lack true absence data for all species, we randomly generated 10,000 pseudo-absences (Chefaoui and Lobo 2008) for model development. Although the use of 10,000 pseudo-absences is somewhat arbitrary, many studies use this number of points for model training (e.g., Guisan et al. 2007; Phillips et al. 2009; VanDerWal et al. 2009). For all models, 75% of the presence data were used for model training and 25% used for model testing. For Maxent, we used default settings except to define randomization conditions, which were the same as for models run in R (see below). For GLMs, we transformed the binary (presence or absence) response using the logit transformation (calculated as the natural log of the odds of presence or absence), and considered only additive effects of predictor variables (e.g., we did not include interactions among predictors). For RFs, we 'grew' 500 trees per species and the number of predictors used to construct each individual tree was p/3 where p = number of predictors in the full model (which varied for each species, Table 1; Liaw and Wiener 2002). In all cases, we included linear predictor combinations only, ran 100 replicate models using random subsets of the presence data for the training–testing split, and calculated performance metrics for each replicate model run. Prediction maps were created using all species observations.

The total number of observations for a species differed slightly because the climate grids had slightly different spatial configurations and therefore a different number of grid cells was occupied by each species (e.g., three nearby coordinates may occupy three different grid cells in one climate grid, but only two cells in another grid because the grids themselves are not perfectly aligned). In addition, the CRU data set contained slightly fewer cells than the WorldClim data set (see Figure 1 for a visual comparison of the two data sets). We allowed the number of observations to differ between data sets because we used the same process (snapping of occurrences to nearest grid cell and removal of duplicate observations from all grid cells) for both climate data sets, and therefore our results represent those that would be obtained using a single data set.

We used two performance metrics for model evaluation. The area under the receiver–operator curve (AUC) measures the tendency for a random occupied grid cell to have a higher suitability than a random pseudo-absence cell (Fielding and Bell 1997). High AUC values indicate models that are best able to discriminate between sites occupied by a species and random points. Cohen's kappa (hereafter, kappa) is a measure of agreement between predicted and observed presence or absence that corrects for agreement resulting from random chance (Fielding and Bell 1997). Kappa requires that the user define a threshold beyond which a

Figure 1. Differences in spatial coverage between two grid-based contemporary climate data sets, Climate Research Unit (CRU) and WorldClim, compiled in 2014. A close up of the northern Caribbean (Cuba and the Bahamas) and adjacent southern Florida shows slight differences in resolution between the two data sets.

probability is interpreted as presence. There are many approaches to defining this threshold (Freeman and Moisen 2008) and we chose one of the most robust methods by identifying the threshold that resulted in maximum kappa. To identify this threshold, we made five replicate model runs using random subsets of the species occurrence data as described above for each 0.01-unit change in threshold between 0.01 and 0.99 and calculated kappa for each randomization. We calculated the average kappa for each incremental change in the threshold to identify the threshold at which kappa was maximized. This threshold was used in the 'full' series of 100 randomizations described above. We summarized model performance by averaging results across the 100 randomizations for each species by algorithm and climate data set partition. Example code for running species models is included as Text S2 (*Supplemental Material*), and occurrence data sets used for analyses of CRU and WorldClim data are included as Tables S2 and S3 (*Supplemental Material*), respectively.

Statistical analyses

To answer our first research question (Is the discrepancy between contemporary climate data sets greater or less than the discrepancy between projections of future climate?), we calculated spatial correlations (wherein the observation of a given cell in one map was paired with the corresponding cell from a second map, and Pearson's correlation calculated across all cells in the two maps; Syphard and Franklin 2009) between CRU and WorldClim maps as contemporary estimates of monthly precipitation and monthly mean temperature and among pairwise combinations of GCM projection maps of future climate. We then used *t*-tests to assess spatial correlation between time steps by comparing the average spatial correlation between contemporary climate data sets with the average spatial correlation across GCM projections for temperature and

precipitation separately ($N = 12$ monthly observations of temperature and precipitation). We also used *t*-tests to compare the average spatial correlation among temperature variables to the average spatial correlation among precipitation variables for each time step (contemporary and future) separately.

To evaluate the potential for spatial autocorrelation in the climate data layers to influence our results, we used the Clifford et al. correction (Clifford et al. 1989) to test whether correlations between CRU and WorldClim temperature and precipitation maps were significantly different from zero. The Clifford et al. correction reduces the effect of spatial autocorrelation by adjusting the estimate for the degrees of freedom when conducting *t*-tests to determine significance of a spatial correlation. Code for implementing the Clifford et al. correction was taken from Plant (2012), and we used 20 distance classes to generate correlograms (plots showing the degree of autocorrelation in each distance class) for each data set.

To answer question two (Does performance differ between models constructed using two alternative data sets of contemporary climate?), we used a linear mixed-effects model. Mixed-effects models include both fixed terms (e.g., the explanatory variables that one is interested in modeling) as well as random effects (variables treated as random samples from a larger population). In general, fixed effects have informative levels that one is interested in comparing, whereas comparisons of different levels of a random effect are not of interest (Crawley 2007). In our study we tested for significant fixed effects of algorithm (Maxent, GLM, or RF), climate data set (CRU or WorldClim), or their interaction on AUC and kappa, while specifying species as a random effect. Separate models were run for the two model performance metrics. The effects of algorithm were coded as dummy variables, and the significance of fixed effects were tested as the likelihood ratio between the full model and models with the effect being tested removed (Fox 2002).

To answer question three (Is the discrepancy in spatial predictions from models using two alternative contemporary climate data sets less than the discrepancy in spatial predictions of future conditions made using data from three alternative GCMs?), we calculated the spatial correlation between contemporary prediction maps constructed from WorldClim and CRU input data for each species and algorithm separately. We also calculated the average spatial correlation among the future projection maps across the three GCMs based on models built with data from the CRU data set (results were qualitatively similar when projections were compared with CEMs constructed from the WorldClim data). We used a linear mixed-effects model to test for significant fixed effects of algorithm, time step (contemporary or future) or their interaction on the spatial correlation between prediction maps, again specifying species as a random effect. The use of time in our model allows us to directly compare the variation in spatial prediction maps between contemporary climate data sets (CRU and WorldClim) with the variation in spatial predictions across the GCMs used to describe future climate. Model

Table 2. Correlation coefficients (*r*) by month, compiled in 2011, between two sources of contemporary global climate data, the Climate Research Unit and WorldClim data sets, and three general circulation models describing future climate (the Geophysical Fluid Dynamics Laboratory Coupled Model version 2.0, the National Center for Atmospheric Research Community Climate System Model version 3.0, and the Hadley Center for Climate Prediction, United Kingdom Meteorological Office coupled model version 3.0). All projections are for the A1B CO_2 emission scenario from the Intergovernmental Panel on Climate Change.

Month	Contemporary climate (CRU-WorldClim)		Future climate (three general circulation models)	
	Precipitation	Temperature	Precipitation	Temperature
January	0.981	0.999	0.972	0.998
February	0.980	0.998	0.977	0.997
March	0.977	0.998	0.968	0.998
April	0.976	0.998	0.975	0.998
May	0.974	0.998	0.970	0.997
June	0.967	0.996	0.970	0.996
July	0.966	0.995	0.960	0.992
August	0.966	0.994	0.956	0.992
September	0.965	0.995	0.959	0.996
October	0.963	0.996	0.964	0.997
November	0.966	0.998	0.957	0.997
December	0.979	0.999	0.969	0.997
Average *r* (± 1 SD)	0.972 ± 0.007	0.997 ± 0.002	0.966 ± 0.007	0.996 ± 0.002

significance was tested as described for question two above. We used the Clifford et al. correction as described for question one to evaluate the potential for spatial autocorrelation in CRU and WorldClim prediction maps to influence our results.

To answer question four (Is the discrepancy in spatial predictions using two alternative climate data sets associated with model performance?), we used linear regression to associate the spatial correlation between CRU and WorldClim prediction maps with AUC and kappa scores for each species model. Performance metrics were averaged between CRU and WorldClim models. We also test for an association between sample size (the number of presences used in each model) and the spatial correlation between contemporary prediction maps. Unless indicated otherwise, all statistical analyses were conducted in R (R Development Core Team 2005).

Results

Is the discrepancy between contemporary climate data sets greater or less than the discrepancy between projections of future climate?

Spatial correlations between contemporary climate data sets averaged $r = 0.972 \pm 0.007$ and $r = 0.997 \pm 0.002$ for monthly precipitation and monthly temperature respectively, and spatial correlations across future climate projections averaged $r = 0.966 \pm 0.007$ and $r = 0.996 \pm 0.002$ for precipitation and temperature, respectively (Table 2). Use of the Clifford et al. correction to test for significance of spatial correlations indicated that all correlations were significant after adjusting for spatial autocorrelation (all $P < 0.001$). There was no difference in the average spatial correlation between contemporary and future climate data sets for either temperature ($t = -0.948$, $df = 22$, $P = 0.350$) or precipitation ($t = -1.854$, $df = 22$, $P = 0.080$). Average

spatial correlations were greater among temperature variables than among precipitation variables for both the contemporary ($t = -12.611$, $df = 12.40$, $P < 0.001$) and future time steps ($t = -14.003$, $df = 12.91$, $P < 0.001$).

Does performance differ between models constructed using two alternative data sets of contemporary climate?

Between 5 and 9 uncorrelated predictor variables were selected to create CEMs for each species (Table 1). Model performance based on the AUC metric was universally high, exceeding 0.90 in all but one case (Table 3). Kappa scores were more variable, ranging from 0.060 to 0.920. The linear mixed-effects model indicated a significant effect of algorithm on performance of species models using both AUC and kappa as response metrics (all $P < 0.01$), whereas the effect of climate data set was not significant in either case ($\chi^2 = 0.13$, $df = 1$, $P = 0.71$ and $\chi^2 = 0.70$, $df = 1$, $P = 0.40$ for models of AUC and kappa, respectively). Random forests were the best-performing models, followed by maximum entropy, with generalized linear models showing the lowest performance. In contrast, spatial correlations between prediction maps created describing climate suitability for the 12 species did not vary significantly among algorithms ($\chi^2 = 0.21$, $df = 1$, $P = 0.65$ and $\chi^2 = 3.09$, $df = 1$, $P = 0.08$ for tests of RF vs. GLM and Maxent vs. GLM, respectively) and ranged from 0.246 to 0.980 (Table 4).

Is the discrepancy in spatial predictions from models using two alternative contemporary climate data sets less than the discrepancy in spatial predictions of future conditions made using data from three alternative GCMs?

The effect of time (contemporary or future) in the mixed model was significant ($\chi^2 = 5.11$, $df = 1$, $P =$

Table 3. Summary model performance metrics compiled in 2011 (AUC and kappa, average values across 100 random partitions of occurrence data into testing-training subsets) for species, algorithms (generalized linear models [GLM]; Maximum entropy [Max]; and random forests [RF]), and contemporary climate data sets (Climate Research Unit [CRU] and WorldClim).

	CRU						WorldClim					
	AUC			Kappa			AUC			Kappa		
Common name	GLM	Max	RF	GLM	Max	RF	GLM	Max	RF	GLM	Max	RF
Mammals												
Florida panther *Puma concolor coryi*	0.981	0.995	0.990	0.262	0.668	0.890	0.984	0.994	0.993	0.326	0.610	0.897
Birds												
Florida grasshopper sparrow *Ammodramus savannarum floridanus*	0.987	0.998	0.983	0.065	0.310	0.773	0.987	0.998	0.974	0.119	0.365	0.816
Florida scrub jay *Aphelocoma coerulescens*	0.998	0.990	0.999	0.718	0.526	0.902	0.997	0.992	0.999	0.601	0.502	0.890
Piping plover *Charadrius melodus*	0.865	0.927	0.989	0.168	0.378	0.836	0.857	0.927	0.990	0.172	0.409	0.784
Wood stork *Mycteria americana*	0.910	0.947	0.980	0.422	0.507	0.863	0.912	0.902	0.980	0.430	0.488	0.870
Audubon crested caracara *Polyborus plancus audubonii*	0.957	0.993	0.999	0.114	0.712	0.897	0.954	0.993	1.000	0.091	0.731	0.900
Everglades snail kite *Rostrhamus sociabilis plumbeus*	0.994	0.995	0.999	0.322	0.540	0.663	0.991	0.994	0.999	0.316	0.575	0.903
Whooping crane *Grus americana*	0.918	0.986	0.989	0.102	0.351	0.673	0.920	0.980	0.989	0.092	0.348	0.708
Red-cockaded woodpecker *Picoides borealis*	0.964	0.958	0.981	0.273	0.341	0.707	0.964	0.957	0.988	0.273	0.334	0.812
Reptiles and amphibians												
American crocodile *Crocodylus acutus*	0.931	0.942	0.979	0.078	0.126	0.522	0.931	0.947	0.986	0.060	0.117	0.664
Sand skink *Neoseps reynoldsi*	0.998	0.996	0.997	0.234	0.111	0.734	0.996	0.997	0.993	0.416	0.144	0.756
Eastern indigo snake *Drymarchon corais couperi*	0.979	0.996	0.999	0.354	0.552	0.895	0.979	0.987	0.999	0.317	0.592	0.920

0.02), indicating that spatial correlations of predictions with contemporary climate data sets were greater (mean $r = 0.845 \pm 0.099$ for the best-performing random forest algorithm) than correlations of predictions among future climate projections (mean r for random forest models = 0.766 ± 0.094; Table 4). Use of the Clifford et al. correction indicated that all correlations between prediction maps were significant after adjusting for spatial autocorrelation (all $P < 0.001$). Contemporary prediction maps for the 12 species using CRU and WorldClim data are included in Figure 2.

Is the discrepancy in spatial predictions using two alternative climate data sets associated with model performance?

There was no association between AUC scores (mean of AUC from CRU and WorldClim models) and spatial correlations of either contemporary or future prediction maps (both $P > 0.343$). Spatial correlations between contemporary prediction maps were positively associated with kappa scores however ($F_{1,10} = 9.01$, $P = 0.013$; Figure 3), although there was no relationship between kappa and spatial correlations among future prediction maps ($F_{1,10} = 0.08$, $P = 0.778$). Spatial

correlations between CRU and WorldClim prediction maps were independent of sample size ($F_{1,10} = 0.21$, $P = 0.660$).

Discussion

Neither performance nor spatial predictions of CEMs for 12 species of threatened and endangered vertebrates in the southeastern United States varied significantly between two different contemporary climate data sets. In contrast, CEM performance was significantly affected by the choice of modeling algorithm, and spatial predictions were significantly more discrepant for the future than for the contemporary period. We therefore conclude that the choice of contemporary climate data set from which CEMs are constructed contributes relatively little to uncertainty in model predictions compared with other parameters such as algorithm and GCM selection. However, species with relatively low kappa scores (Everglades snail kite *Rostrhamus sociabilis plumbeus*, whooping crane *Grus americana*, red-cockaded woodpecker *Picoides borealis*, and American crocodile *Crocodylus acutus*) had more discrepant prediction maps between climate data sets than species with high kappa

Table 4. Spatial correlations (*r*) between prediction maps from two contemporary climate data sets compiled in 2011 (the Climate Research Unit [CRU] and WorldClim) and three general circulation models describing future climate (the Geophysical Fluid Dynamics Laboratory Coupled Model version 2.0, the National Center for Atmospheric Research Community Climate System Model version 3.0, and the Hadley Center for Climate Prediction, United Kingdom Meteorological Office coupled model version 3.0) using three modeling algorithms (generalized linear models [GLM]; Maximum entropy [Max]; and random forests [RF]) and 12 species. All future climate data are for the A1B CO_2 emission scenario from the Intergovernmental Panel on Climate Change.

| Common name	Contemporary climate (CRU –WorldClim)			Future climate (three general circulation models)		
	GLM	Max	RF	GLM	Max	RF
Mammals						
Florida panther *Puma concolor coryi*	0.776	0.934	0.890	0.576	0.829	0.680
Birds						
Florida grasshopper sparrow *Ammodramus savannarum floridanus*	0.897	0.840	0.844	0.819	0.807	0.618
Florida scrub jay *Aphelocoma coerulescens*	0.866	0.953	0.917	0.739	0.861	0.899
Piping plover *Charadrius melodus*	0.956	0.932	0.787	0.856	0.754	0.714
Wood stork *Mycteria americana*	0.971	0.980	0.884	0.941	0.917	0.901
Audubon crested caracara *Polyborus plancus audubonii*	0.922	0.936	0.943	0.945	0.758	0.720
Everglades snail kite *Rostrhamus sociabilis plumbeus*	0.565	0.948	0.916	0.350	0.858	0.798
Whooping crane *Grus americana*	0.939	0.776	0.845	0.869	0.703	0.654
Red-cockaded woodpecker *Picoides borealis*	0.945	0.978	0.864	0.821	0.902	0.878
Reptiles and amphibians						
American crocodile *Crocodylus acutus*	0.844	0.903	0.701	0.835	0.790	0.800
Sand skink *Neoseps reynoldsi*	0.246	0.501	0.617	0.842	0.861	0.770
Eastern indigo snake *Drymarchon corais couperi*	0.872	0.951	0.933	0.563	0.787	0.764
Average ± 1 SD	0.817 ± 0.211	0.886 ± 0.135	0.845 ± 0.099	0.763 ± 0.178	0.819 ± 0.064	0.766 ± 0.094

scores. Herein we discuss the implications of our results for applied conservation.

Although spatial correlations between contemporary prediction maps constructed from WorldClim and CRU data were generally high, and greater than spatial correlations among future prediction maps constructed from three alternative GCMs, there was interspecific variation in correlations for contemporary prediction maps (Table 4). We found that the most discrepant maps resulted from the models with the lowest average kappa scores. Kappa describes a model's ability to correctly classify contemporary presences and absences (Fielding and Bell 1997). Because different applications of CEMs may differ in their tolerance for misclassified presences or absences (omission and commission errors; Fielding and Bell 1997), low kappa scores per se do not necessarily mean that a model is not useful for a particular application. For example, a model may receive a low kappa score because it misclassifies many absences (i.e., it has low

specificity), but if a user prioritizes correctly classified presences when using CEMs to define suitable areas for restoration or translocation, for example, the model's inability to correctly classify absences may be of little concern. However, users of CEMs for applied conservation should also be aware that a low kappa score may also indicate the sensitivity of prediction maps to variation in climate data inputs. We noted no obvious ecological correlates uniting species for which kappa scores were low, although identifying such correlates would be useful for prioritizing species for assessment using CEMs.

One important message resulting from our work is the importance of considering multiple performance metrics when evaluating models. Here we found that kappa, but not AUC, was associated with spatial discrepancy in prediction maps resulting from models with uniformly high AUC scores. Although the utility of AUC as a performance metric has been questioned (Lobo et al. 2008) it remains an important means by which CEMs are
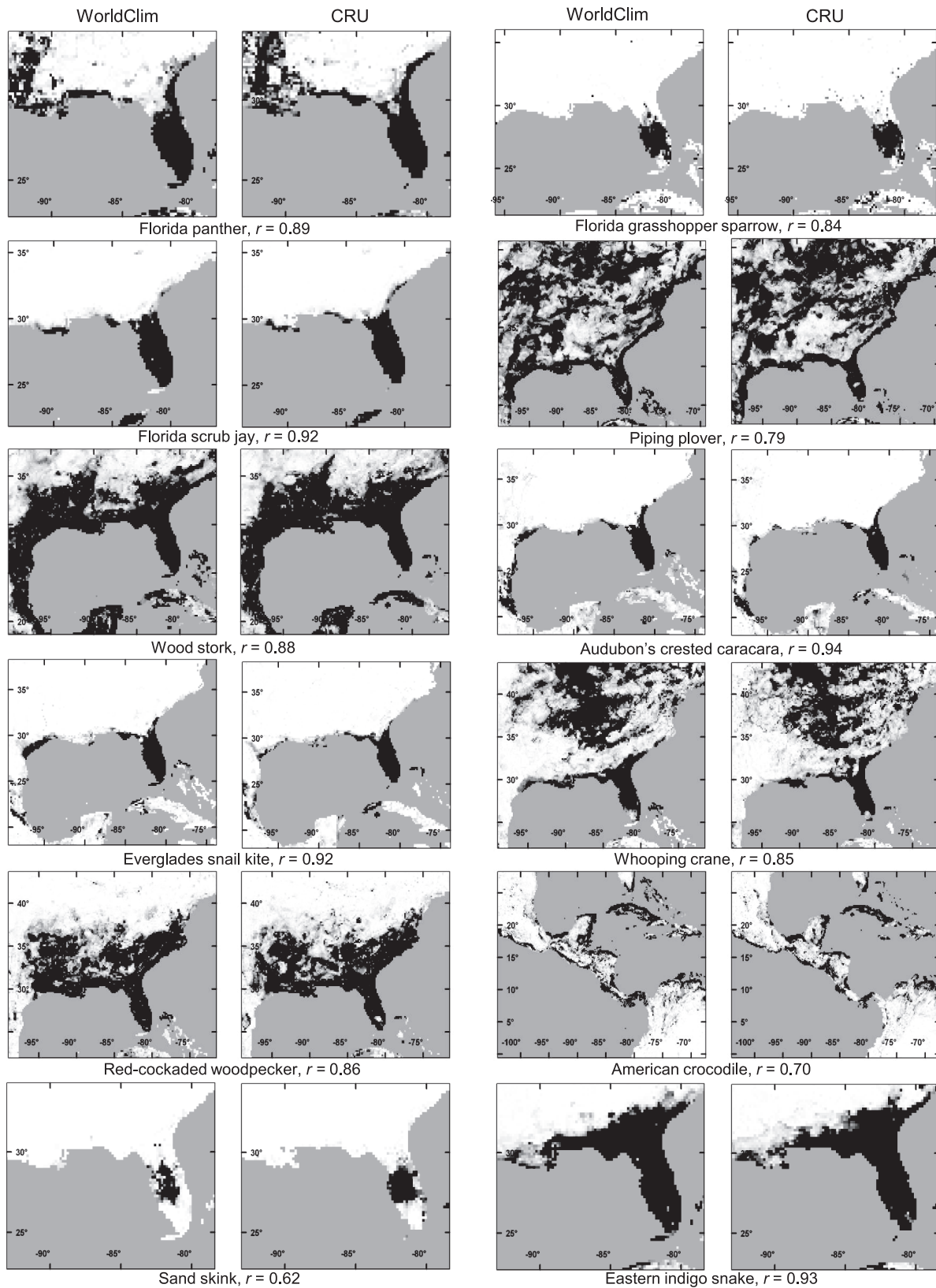
Figure 2. Prediction maps of climate suitability for 12 species compiled in 2014, based on contemporary climate data drawn from two climate data sets (Climate Research Unit, CRU and WorldClim) using the random forest algorithm. The spatial correlation between each species-specific pair of maps is included following the species name. In all maps, darker colors indicate greater climate suitability.
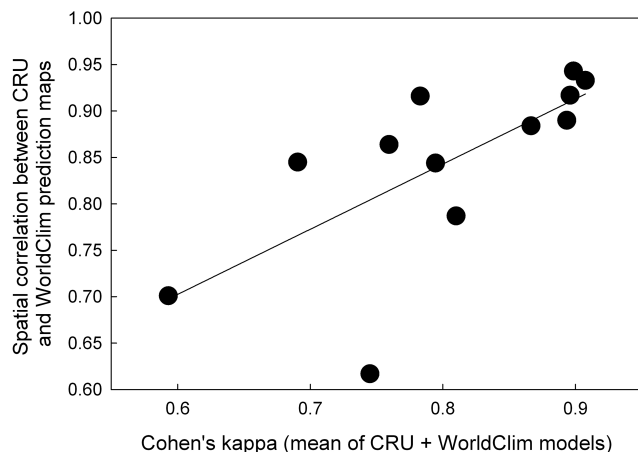
Figure 3. Scatterplot illustrating the relationship between the mean kappa score for each species model using the random forest algorithm and the correlation between prediction maps made using Climate Research Unit (CRU) and WorldClim contemporary climate data compiled in 2013.

evaluated and compared with one another. Kappa is less frequently reported in the literature than AUC, but we found that it was a more sensitive indicator of spatial discrepancy in alternative CEM prediction maps. We recommend using multiple metrics to assess CEM performance because alternative metrics may provide different insight into model behavior.

More generally, our research illustrates that similarly high-performing models (using the widely applied AUC metric) can sometimes make inconsistent spatial predictions. This observation may have important implications if CEMs are to be used to make management decisions in support of species adaptation to climate change. Our results underscore the importance of thoughtful approaches to model training, algorithm selection, and the selection of climate data projections. Models trained using different sources of contemporary data may show considerable differences in spatial predictions (Figure 2), translating into different interpretations regarding the spatial implementation of conservation actions, such as designing natural corridors through the landscape that may allow species to track shifts in climate suitability (Williams et al. 2005) or protecting areas in anticipation of them being suitable climate space in the future. Our work suggests that kappa may be a potential indicator of the extent to which spatial predictions from CEMs are sensitive to the data on which they are constructed. Because of the implications of alternative spatial predictions for applied conservation, users may want to evaluate alternative models when kappa scores are low to assess potential differences in alternative prediction maps (Jones-Farrand et al. 2011).

Random forests models are only recently gaining widespread use in predictive modeling of climate change effects on species, but they showed high performance as measured by our two evaluation metrics (Table 3). Although another methods comparison found somewhat lower performance of RF compared with other methods using simulated species distribution data (Elith

and Graham 2009), our comparison of real occurrence data indicated consistently high performance of RF relative to other algorithms, and other studies confirm the high performance of random forest models (Prasad et al. 2006; Watling et al. 2012). Furthermore, although performance is often lower for widespread species compared with range-limited species (Hernandez et al. 2006; Guisan et al. 2007), model performance was generally high for even the most geographically widespread species we modeled.

We acknowledge some key limitations of our study. First, the use of independent survey data (e.g., McCarthy et al. 2012) to validate model performance may result in lower performance for all algorithms than reported here. We also acknowledge that our results may differ if statistical models that attempt to account for spatial autocorrelation were used (Dormann et al. 2007). However, we saw little change in the spatial correlations between CRU and WorldClim prediction maps when grid cells were removed from prediction maps to reduce spatial autocorrelation. By focusing on relative comparisons between prediction maps, we assumed that any bias resulting from spatial autocorrelation is present in all maps and would have less impact on our results than if we were comparing absolute species–climate relationships. Finally, we have focused on climate predictors in this study because climate is believed to play an important role in limiting the coarse-scale distribution of many species (Jiménez-Valverde et al. 2011). However, finer scale distributions of species may be related to many other factors besides climate (e.g., topography, habitat availability, etc.), and we have not included such additional explanatory variables here because our primary purpose was to make methodological comparisons.

In conclusion, across all species there were no significant differences in model performance or spatial predictions between CEMs constructed with WorldClim or CRU data. Thus, the importance of the contemporary climate data set used to create models contributes relatively little to model uncertainty compared with other factors. However, there were differences in prediction maps between CRU and WorldClim models for some species, and the most discrepant maps came from models with low kappa scores. We advise users to be aware that a low kappa score may indicate a model that is particularly sensitive to the conditions on which it is trained. When using CEMs for conservation planning, it is wise to evaluate alternative models that make competing spatial predictions so that a range of alternatives may be evaluated.

Supplemental Material

Please note: The *Journal of Fish and Wildlife Management* is not responsible for the content or functionality of any supplemental material. Queries should be directed to the corresponding author for the article.

Table S1. The number and source of species observations used for climate envelope model construction.
Found at DOI: 10.3996/072012-JFWM-056.S1 (15 KB DOCX)

Table S2. Species occurrences (presences and pseudo-absences) used in construction of climate envelope models from Climate Research Unit contemporary climate data.

Found at DOI: 10.3996/072012-JFWM-056.S2 (9.8 MB ZIP)

Table S3. Species occurrences (presences and pseudo-absences) used in construction of climate envelope models from WorldClim contemporary climate data.

Found at DOI: 10.3996/072012-JFWM-056.S3 (10 MB ZIP)

Text S1. Procedure used to define target areas from which pseudo-absence data were drawn for construction of climate envelope models.

Found at DOI: 10.3996/072012-JFWM-056.S4 (13 KB DOCX)

Text S2. Example R code for running species models described in the manuscript.

Found at DOI: 10.3996/072012-JFWM-056.S5 (26 KB DOCX)

References

Araújo MB, New M. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22:42–47.

Araújo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93:1527–1539.

Chefaoui RM, Lobo JM. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modeling 210:478–486.

Clifford P, Richardson S, Harmon D. 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45:123–134.

Crawley MJ. 2007. The R book. Chichester, England, United Kingdom: John Wiley and Sons.

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Diniz-Filho JA, Bini LM, Rangel TF, Loyola RD, Hof C, Nogués-Bravo D, Araújo MB. 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography 32:897–906.

Dormann CF, McPherson, JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distribution data: a review. Ecography 30:609–628.

Dormann CF, Purschke O, García Márquez JR, Lautenbach S, Schröder B. 2008. Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. Ecology 89:3371–3386.

Elith J, Graham CH. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66–77.

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

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Fox J. 2002. Appendix to an R and S-PLUS companion to applied regression. Available: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf (January 2014).

Franklin J. 2009. Mapping species distributions: spatial inference and prediction. New York: Cambridge University Press.

Freeman EA, Moisen GG. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modeling 217:48–58.

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT. 2007. What matters most for predicting the occurrence of trees: techniques, data, or species' characteristics? Ecological Monographs 77:615–630.

Hannah L, Midgley G, Andelman S, Araújo M, Highes G, Martinez-Meyer E, Pearson R, Williams P. 2007. Protected area needs in a changing climate. Frontiers in Ecology and the Environment 5:131–138.

Hernandez PA, Graham CH, Master LL, Albert DL. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

Hijmans RS, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Hirzel AH, Hausser J, Chessel JD, Perrin N. 2002. Ecological niche-factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83:2027–2036.

Jiménez-Valverde A, Barve N, Lira-Noriega A, Maher SP, Nakazawa Y, Papeş M, Soberón J, Sukumaran J, Townsend Peterson A. 2011. Dominant climate influences on North American bird distributions. Global Ecology and Biogeography 20:114–118.

Jones-Farrand DT, Fearer TM, Thogmartin WE, Thompson III FR, Nelson MD, Tirpak JM. 2011. Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble prediction. Ecological Applications 21:2269–2282.

Kadmon R, Farber O, Danin A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14:401–413.

Lawler JJ, Shafer SL, White D, Kareiva P, Maurer EP, Blaustein AR, Bartlein PJ. 2009. Projected climate-induced faunal change in the Western Hemisphere. Ecology 90:588–597.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News 2/3:18–22.

Lobo JM, Jiménez-Valverde A, Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145–151.

McCarthy KM, Fletcher RJ Jr, Rota CT, Hutto RL. 2012. Predicting species distributions from samples collected along roadsides. Conservation Biology 26:68–77.

McCullugh P, Nelder JA. 1989. Generalized linear models. Second edition. New York: Chapman and Hall.

Morin X, Chuine I. 2005. Sensitivity analysis of the tree distributions model PHENOFIT to climate input characteristics: implications for climate impact assessment. Global Change Biology 11:1493–1503.

Nakicenović N, Swart R, editors. 2000. Intergovernmental Panel on Climate Change special report on emissions scenarios. England, United Kingdom: Cambridge University Press.

New M, Lister D, Hulme M, Makin I. 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21:1–25.

Parra JL, Monahan WB. 2008. Variability in 20th century climate change reconstructions and its consequences for predicting geographic responses of California mammals. Global Change Biology 14:2215–2231.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190:231–259.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Plant RE. 2012. Spatial data analysis in ecology and agriculture using R. Boca Raton, Florida: CRC Press.

Povilitis A, Suckling K. 2010. Addressing climate change threats to endangered species in U.S. recovery plans. Conservation Biology 24:372–376.

Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199.

R Development Core Team. 2005. R: a language and environment for statistical computing, reference index version 2.2.1. Vienna: R Foundation for Statistical Computing. Available: www.R-project.org (January 2014).

Real R, Márquez AL, Olivero J, Estrada A. 2010. Species distribution models in climate change scenarios are still not useful for informing policy planning: an uncertainty analysis using fuzzy logic. Ecography 33:304–314.

Reece JS, Noss RF. 2014. Prioritizing species by conservation value and vulnerability: a new index applied to species threatened by sea-level rise and other risks in Florida. Natural Areas Journal 34:31–45.

Rosenzweig C, Karoly D, Vicarelli M, Neofotis P, Wu Q, Casassa G, Menzel A, Root TL, Estrella N, Seguin B, Tryjanowski P, Liu C, Rawlins S, Imeson A. 2008. Attributing physical and biological impacts to anthropogenic climate change. Nature 453:353–357.

Rowland EL, Davison JE, Graumlich LJ. 2011. Approaches to evaluating climate change impacts on species: a guide to initiating the adaptation planning process. Environmental Management 47:322–337.

Syphard AD, Franklin J. 2009. Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. Ecography 32:907–918.

Tabor K, Williams JW. 2010. Globally downscaled climate projections for assessing the conservation impacts of climate change. Ecological Applications 20:554–565.

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, Ferreira de Siqueira M, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsvels AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE. 2004. Extinction risk from climate change. Nature 427:145–148.

[ESA] U.S. Endangered Species Act of 1973, as amended, Pub. L. No. 93-205, 87 Stat. 884 (Dec. 28, 1973). Available at: http://www.fws.gov/endangered/esa-library/pdf/ESAall.pdf

U.S. Fish and Wildlife Service. 2010. Rising to the Urgent Challenge: Strategic Plan for Responding to Accelerating Climate Change. Available at: http://www.fws.gov/home/climatechange/pdf/CCStrategicPlan.pdf (accessed January 2014).

VanDerWal, J, Shoo LP, Graham C, Williams SE. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecological Modelling 220:589–594.

Watling JI, Romañach SS, Bucklin DN, Speroterra C, Brandt LA, Pearlstine LG, Mazzotti FJ. 2012. Do bioclimate variables improve performance of climate envelope models? Ecological Modelling 246:79–85.

Williams P, Hannah L, Andelman S, Midgley G, Araújo MB, Hughes G, Manne L, Martinez-Meyer E, Pearson R. 2005. Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. Conservation Biology 19:1063–1074.