Ecological Modelling 330 (2016) 50–59

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

On the dangers of model complexity without ecological justification in species distribution modeling

David M. Bell[a,*], Daniel R. Schlaepfer[b,c]

[a] Pacific Northwest Research Station, U.S. Forest Service, Corvallis, OR, USA
[b] Department of Botany, University of Wyoming, Laramie, WY, USA
[c] Department of Environmental Sciences, Section of Conservation Biology, University of Basel, Switzerland

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form 18 March 2016
Accepted 20 March 2016

Keywords:
Prediction
Extrapolation
Model fitting
Species distribution modeling
Transferability

ABSTRACT

Although biogeographic patterns are the product of complex ecological processes, the increasing complexity of correlative species distribution models (SDMs) is not always motivated by ecological theory, but by model fit. The validity of model projections, such as shifts in a species' climatic niche, becomes questionable particularly during extrapolations, such as for future no-analog climate conditions. To examine the effects of model complexity on SDM predictive performance, we fit statistical models of varying complexity to simulated species occurrence data arising from data-generating processes that assume differing degrees of distributional symmetry in environmental space, interaction effects, and coverage in climate space. Mismatches between data-generating processes and statistical models (i.e., different functional forms) led to poor predictive performance when extrapolating to new climate-space and greater variation in extrapolated predictions for overly complex models. In contrast, performance issues were not apparent when using independent evaluation data from the training region. These results draw into question the use of highly flexible models for prediction without ecological justification.

Published by Elsevier B.V.

1. Introduction

Monumental increases in the availability of ecological data and computing resources allows increasingly complex ecological models to be leveraged for predicting changes in biogeography. Increasing complexity in ecological models developed to represent species distributions in both geographic and environmental space is supported by the fact that those same distributions depend on a suite of processes associated with physiology (Buckley et al., 2011), demography (Pagel and Schurr, 2011), dispersal (Elith and Leathwick, 2009; Iverson et al., 2004), and biotic interactions (Parmesan and Yohe, 2003; Vanderwel et al., 2013; Walther et al., 2002). However, model complexity is sometimes motivated by the maximization of predictive performance, not ecological theory, as has been noted for correlative species distribution modeling (Austin, 2002, 2007). Correlative species distribution models (SDMs) are commonly used to assess habitat suitability as it relates to key environmental gradients (Elith and Leathwick, 2009) and generate global change predictions at regional to continental scales, painting a portrait of extreme biogeographic change under most, if not all, future scenarios of climate change (Parmesan and Yohe, 2003; Walther et al., 2002). While the SDMs do not generally model ecological processes constraining species occurrences as mechanistic approaches might (Ibáñez et al., 2006), the advent of virtual species simulation as a method of testing key assumptions of these models offers opportunities to assess model robustness and appropriateness (Meynard et al., 2013; Zurell et al., 2010). Model assessments exploring the impacts of failing to meet assumptions on predictive performance are not only needed to guide ecologists in choosing SDMs, testing their reliability, and interpreting their results (Aguirre-Gutierrez et al., 2013; Austin, 2007; Elith and Graham, 2009; Jimenez-Valverde et al., 2008), but for any ecological models used for prediction.

In part, the diversity of SDMs available for modeling reflects differences in assumptions about how species respond to environmental gradients. The prevalence of unimodal patterns of species occurrences along environmental gradients is well-supported (Austin, 2005; Gauch and Whittaker, 1972) and is assumed to represent suitability declines as conditions depart from the optimum (Austin and Smith, 1989; Heikkinen and Mäkipää, 2010). Asymmetric and symmetric unimodal patterns are both common (Austin and Gaywood, 1994; Austin and Van Niel, 2011; Boucher-Lalonde et al., 2012; Ellenberg, 1953), suggesting that

* Corresponding author at: 3200 SW Jefferson Way, Corvallis, OR 97331, USA. Tel.: +1 5417507298.
E-mail address: dmbell@fs.fed.us (D.M. Bell).

http://dx.doi.org/10.1016/j.ecolmodel.2016.03.012
0304-3800/Published by Elsevier B.V.

both can be reasonable representations of reality. The degree of symmetry is often interpreted as evidence of certain mechanisms controlling species distributions in environmental space, such as physiological constraints producing asymmetric distributions (Austin and Gaywood, 1994; Austin and Smith, 1989).

An apparent asymmetry in a species distribution may arise from truncation in the climate-space (Normand et al., 2009). Given that no-analog climates are likely to be common in the future (Williams and Jackson, 2007), ecological models that perform well under contemporary conditions may be unable to predict future changes. For example, contemporary patterns of conifer budbreak dates in western North America are negatively correlated with temperatures (i.e., earlier budbreak in warmer regions), but budbreak under future conditions may be delayed as chilling requirements are no longer met (Harrington and Gould, 2015). Because there may be no contemporary analogs to some future climates, models sensitive to truncation in the climate space will struggle in extrapolating to future conditions.

Although not generally discussed in relation to species distribution modeling, interactions might project asymmetries from one environmental gradient to another, as noted for 23% of European tree species (Boucher-Lalonde et al., 2012). As a result, an asymmetric species distribution might arise because an asymmetry or truncation along one environmental gradient influences suitability along the other gradient. Therefore, the source of observed asymmetries in species distributions is not trivial and is not necessarily easily accounted for in SDMs.

The source of complexity in species distributions, both environmentally and geographically, is at the heart of the debate concerning the complexity of SDMs. In recent years, SDMs increasingly utilize highly flexible correlative statistical models capable of accommodating a diverse suite of species distributional responses to environmental gradients (Elith and Leathwick, 2009). Increasing flexibility seems to improve model performance based on traditional cross-validation techniques (Santika and Hutchinson, 2009), but compared to simple models such as the generalized linear model (GLM), more flexible models such as the generalized additive model (GAM), random forest (RF) models, maximum entropy (MaxEnt) models, or boosted regression trees (BRT) may not perform well in terms of predicting into other regions or the future (Araujo et al., 2005; Randin et al., 2006; Schibalski et al., 2014; Merow et al., 2014). This dichotomy indicates that complex models may be fitting spurious patterns that are difficult to identify if evaluating model performance using cross-validation within the same region used to train the models; for instance, spatial autocorrelation is a concern for cross-validation when predicting within the same region (Le Rest et al., 2014) and also when predicting into novel climate space (Crase et al., 2014). While species responses to environment may be highly conditional on local landscapes and communities, increased model complexity may be difficult to interpret or may explain random variation not related to any ecological processes (i.e., overfitting). Thus, our confidence in model predictions to novel climate space is based on the ability of a model to reproduce the underlying processes contributing to species distributions (Evans et al., 2013).

In this study, we examine the influence of process and model complexity on ecological inference and prediction. Our objectives were to (1) determine how model complexity impacted performance when predicting species occurrence within a training region as well as extrapolating to other regions and (2) to explore the factors contributing the variation in performance, such as the underlying process generating the data (including random and spatial error), the model employed, and the sampling of data. We used a virtual species approach to simulate presence and absence data (e.g.; Meynard and Quinn, 2007) based on four different underlying climatic suitability processes, we fit SDMs of varying complexity, and we evaluated model performance in terms of observed species occurrence and underlying suitability processes within a single region and across regions (i.e., independent validation and transferability, respectively).

2. Materials and methods

2.1. Study area and climate data

In this study, we focus on the dry domain of the United States because it encompasses a large, climatically complex region in which simulation of species distributions could produce complex patterns (Fig. 1). The dry domain the of U.S. encompasses ecosystems ranging from the eastern slope of the Sierra Nevada and Cascade Mountains to the western Great Plains, from lowland deserts to montane forests to alpine meadows (Bailey, 1995). We divided the region into sub-regions based on state boundaries so that we could fit models to data from a single region (Northern Rocky Mountains [NR]) and test transferability of these models to other regions representing different climate-spaces (Southern Rocky Mountains [SR], Southwest [SW], and the Great Plains [GP]). All regions overlap climatically, but none share similar climatic extents (Fig. 1b–e), ensuring that species distribution models developed in NR would need to extrapolate to predict species geographic and environmental distributions in SR, SW, and GP. Therefore, these regions provide an appropriate case for testing model transferability.

Climate data were extracted for a grid of sample locations, with points located uniformly at 1/32° intervals, resulting in 12,583 total sample points. To ensure climatic realism, we extracted 30-year climate normal (1981–2010) from the 30 arc-second (approximately 800 m) PRISM data set (PRISM Climate Group, 2012) and calculated log winter (November to March) precipitation (dm) and minimum annual temperature (°C). We chose these variables because (1) snowpack and extreme winter temperatures are often incorporated into plant species distribution models in the region (e.g., Rehfeldt et al., 2006), (2) the correlation between these variables was intermediate (Pearson correlations $r = 0.22$, -0.22, -0.58, and 0.63 for NR, SR, SW, and GP, respectively), and (3) these variables resulted in somewhat divergent climate-space among the four sub-regions (Fig. 1b–e). The coverage of the climate space for different regions was assessed by examining the similarity of univariate and multivariate climate within the training region (NR) to the projection regions (SR, SW, and GP) as measured with the NT1 and NT2 indices defined by Mesgaran et al. (2014). Both in terms of univariate and multivariate climate space, large portions of the projection regions were climatically similar to the training region, but that greatest dissimilarity between training and projection regions occurred in the southern portions of the study region (Fig. 2).

2.2. Simulation experiment design

To examine the influences of the species occurrence data generating process on model prediction, we employed a virtual ecologist approach (Meynard et al., 2013; Meynard and Quinn, 2007; Zurell et al., 2010) wherein we designed a simulation experiment that allowed us to vary the data generating process, the models used to describe species environmental distributions, and the climate-space provided to the models for statistical inference. While there are many SDM approaches reported on in the literature, in this research, we focus on GLM, GAM, RF, MaxEnt, and BRT models with and without interactions to represent a gradient in model complexity because (1) many ecologists and species distribution modelers are familiar with them, (2) computation with model fitting is relatively simple and fast, and (3) the main objective was to test the effect of model complexity and not to evaluate specific models. As a result, the current study uses a relatively small series of models

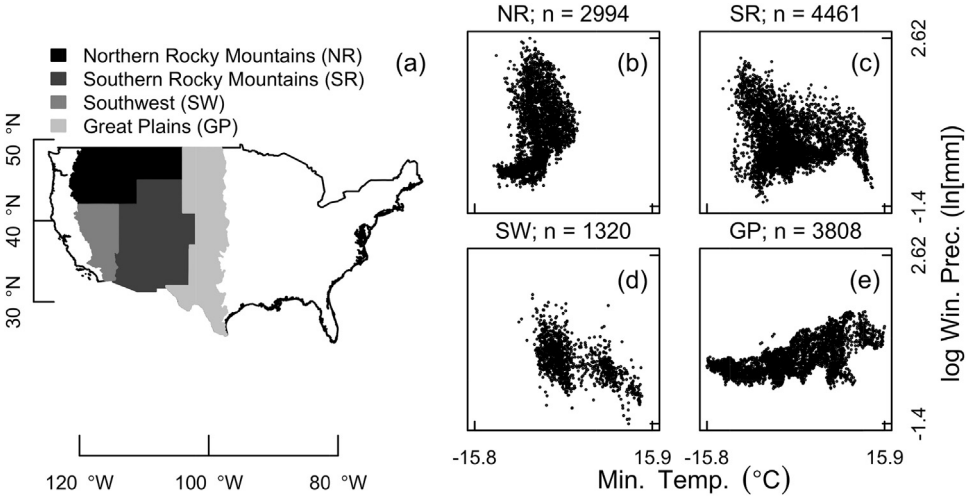

Fig. 1. Geographic and climatic space of the study region are presented. (a) Map of the study area and the four study regions: Northern Rocky Mountains (NR), Southern Rocky Mountains (SR), Southwest (SW), and Great Plains (GP). (b–e) Minimum annual temperature and log winter precipitation for each region show the degree of overlap between regions.

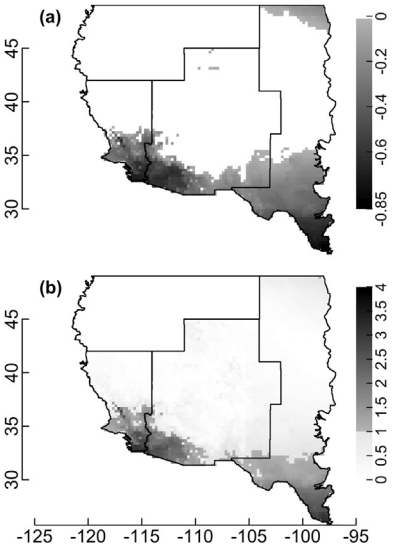

Fig. 2. Geographic distribution of (a) univariate (NT1) and (b) multivariate dissimilarity (NT2; Mesgeran et al., 2014) in climate space to climate within the NR region. NT1 takes 0 under climate similarity and negative values under dissimilarity; NT2 is 0-1 under similarity and >1 under dissimilarity. Dissimilarity was greatest in areas of dark gray and black whereas white and light gray indicated high climatic similarity, respectively.

to examine the consequences of model complexity on prediction for virtual species arising from differing underlying processes (for a total of $n = 320{,}000$ combinations). In the remainder of this section, we describe the details of our simulation study, in which we (1) defined four virtual species based on four climatic suitability processes that represent commonly assumed characteristics of biogeographic patterns, (2) simulated species occurrences across the study area, (3) fit 10 different statistical models to data arising from each of the four processes for region NR, and (4) evaluated model performance within region NR and across the other regions (SR, SW, and GP).

2.3. Virtual species: generating climatic suitabilities

The climatic suitability process may be symmetric or asymmetric, motivating different modeling approaches and highlighting different mechanisms (Austin and Smith, 1989; Elith and Leathwick, 2009). Species may be capable of occurring in portions of the climate space absent from a specific data set (i.e., sampling bias, Thuiller et al., 2004) or may not exist anywhere in the contemporary climate space (Williams and Jackson, 2007; Williams et al., 2007). We assumed that the geographic distribution of the virtual species was entirely in NR; we therefore covered the complete climate space of the species during model building (Barbet-Massin et al., 2010; Thuiller et al., 2004). Species distributional responses to one variable may be highly conditional on some other variable. Therefore, climatic suitability processes were defined in terms of three attributes: shape (symmetric S vs. asymmetric A), the importance of conditional responses to environments (conditional C vs. independent I), and the extent of climate-space (full F vs. truncated T) (Table 1). Specifically, we focused our attention on four scenarios combining different combinations of the shape, conditionality, and climate extent attributes (SIF, SIT, SCT, and AIF) because (1) they represented realistic scenarios for species distributions in climate space, (2) all scenarios except SIF could result in an observed skewed pattern in observational data (Fig. 3), and (3) they provided a range of complexity in the underlying process upon which the data would be simulated and different SDMs would be fit. Scenarios that were both asymmetric and truncated (ACT and AIT) were excluded from the current analysis because they would be particularly hard to fit statistically given the distributional climatic optimum would be outside of the climatic range of the training data. Because initial exploration of data-generating processes indicated that interactions in conjunction only produce asymmetric behavior when paired with climate truncation (SCT; Fig. 3), we also excluded models with interactions, but no truncation of the climate-space (ACF and SCF).

Symmetric, independent distributional response to climate variables when all suitable climate-space is represented in a sample (SIF) represents an idealized species niche where the two climate variables have independent effects. The fact that there is strong support for the SIF scenario in some taxa (Boucher-Lalonde et al., 2012) indicates that this simple model is a reasonable starting point for SDMs. We simulated climatic suitability for the SIF scenario $\theta_{SIF,i}$ for location i (Fig. 3) with an inverse logit function (details in the Appendix A).

Skewed species distributions are by no means rare, but a variety of processes could create a skewed pattern. For example, asymmetric, independent distributional response to climate variables when all suitable climate-space is represented in a sample (AIF) could

Table 1
Description of sources of variation.

Source of variation	Description	Levels
Data generating process (P)	The functional form used to simulate the data, including the symmetry (asymmetric A or symmetric S), the independence of one response to climate from another (independent I and conditional C), the climate space (full F or truncated T).	AIF, AIT, SCT, SIF
Error in process (E)	Additional variation added to the impact of climate on suitability in logit space (i.e., f(climate)) representing the presence or absence of random noise and spatial autocorrelation	None Random Spatial Random + spatial
Model (M)	Model functional forms with differing levels of complexity (see Fig. 3)	GLM GAM RF MaxEnt BRT
Use of interactions (I)	Whether or not interaction terms are to be included in the model fitting process	No Yes
Data realization (R)	40 independent, random draws from the binomial distribution based on P and E	–
Data splitting (D)	40 differing partitions of the data producing the training data set and the cross-validation data set	–

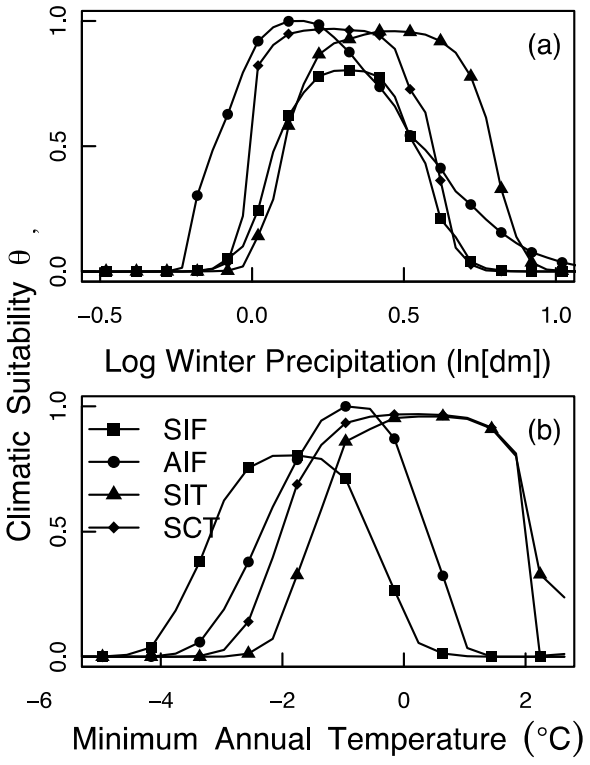

Fig. 3. Climatic suitability response curves for region NR, as represented by the 99th percentile values of climatic suitability $\theta_{P,i}$ based on process P at location i for a given value of (a) log winter precipitation and (b) minimum annual temperature for each of the climatic suitability processes.

cause skewed patterns in data, especially if physiology constrains species performance (Austin and Smith, 1989). We simulated climatic suitability for the AIF scenario $\theta_{AIF,i}$ for location i (Fig. 3) with a normalized version of the beta response function.

If the data do not include the entire climate space to which the species is adapted (i.e., truncated climate-space), occurrence data could appear skewed even if the underlying distributional response is symmetric. If climate variables have independent effects and the underlying response is symmetric, truncation with respect to one variable (SIT) should not influence the shape of the response to the other variable. Given that this process differs from SIF only in that the distribution was shifted toward the upper limit of t_i for the NR region, climatic suitability for the SIT scenario $\theta_{SIT,i}$ for location i (Fig. 3) was simulated using the same form as for SIF but with different parameter values.

If climate variables have conditional effects and the underlying response is symmetric (SCT), truncation with respect to one variable will influence the shape of the response to the other variable. We based climatic suitability for SCT for location i (Fig. 3) on SIF, but with the addition of an interaction term.

2.4. Virtual species: simulating occurrence data

To account for the stochastic nature of data simulation, we developed realizations of the occurrence data for each data generating process based on random draws from the Bernoulli distribution (Meynard et al., 2013). This probabilistic approach allows to sample from differently shaped response curves and reduces bias in performance measures. We simulated species occurrence data drawing $R = 40$ random samples to represent ($j = 1,\ldots, R$) different realizations of the Bernoulli process (i.e., binomial distribution with number of trials equal to one)

$$y_{P,ij} \sim Bernoulli(\theta_{P,i}) \quad (1)$$

where $y_{P,ij} = 1$ if present at location i from realization j based on process P, $y_{P,ij} = 0$ if absent, and $\theta_{P,i}$ is the climatic suitability for location i and process P. To explore the impacts of spatial autocorrelation and random variation, we simulated data with and without an autocovariate term and a random error term for our virtual species. As a result, $\text{logit}(\theta_{P,i}) = f(\text{climate}) + a_{P,i} + \varepsilon_{P,i}$, where f(climate) is the predicted climatic suitability conditioned only on climate in logit space, $a_{P,i}$ is an autocovariate transform, and $\varepsilon_{P,i}$ is a random error term. The autocovariate transform represents the weighted impact of observations within one degree latitude and longitude, such that $a_{P,i} = 2w_{P,i} - 0.5$, where w_{ij} is the average climatic suitability for process P within 1 degree latitude and longitude of observation i weighted by the distance in degrees (*sensu* Dormann et al., 2007). The error term represents random noise, such that $\varepsilon_{P,i} = N(0, 0.64)$. Both the autocovariate transform and random error term were selected to produce equivalent variability in the data: standard deviation approximately equal to 0.1 when $\text{logit}^{-1}[f(\text{climate})] = 0.5$. One could compare this data simulation to hiring 40 independent survey crews sent to survey the entire study region with slightly different plot locations. While they all sample the same, or at last similar, climate-space, they survey different portions of the landscape and, therefore, produce different records of species occurrence.

2.5. Model fitting and complexity

For each realization j of each process P (i.e., each independent sample), we fit each of the five models to $M = 50$ separate random

samples of 70% of the observations, retaining the remaining 30% as an independent validation data set (i.e., within-region validation comparable to traditional cross-validation, Fielding and Bell, 1997). Strategies to minimize dependence between test and training data are often necessary for species distribution modeling (Araujo et al., 2005; Brenning, 2005; Segurado et al., 2006), such as partitioning the dataset into spatial aggregates and randomly assigning aggregates as testing or training data (Bahn and McGill, 2013; Madon et al., 2013). Assessment within and outside the training region allows our virtual species approach to address this spatial dependence in accuracy assessment. By comparing the 50 fits, and their ability to predict within different regions (see Section 2.6), we attempted to assess how partitioning of the data to model training vs. model evaluation impacted inference.

We fit the data using glm, gam (mgcv package version 1.8-12, Wood, 2006), randomForest (randomForest package version 4.6-12, Liaw and Wiener, 2002), and gbm (gbm package version 2.1.1, Ridgeway, 2015) functions in the R programming environment (version 3.2.3, R Development Core Team, 2010), and with the java application 'maxent' version 3.3.3k (available from http://www.cs.princeton.edu/~schapire/maxent/; Phillips et al., 2004). Binomial/Bernoulli models with logit link functions were assumed for GAM, GLM, and BRT models. We included the main effects of the climate variables (all models) and quadratic effects to represent hump-shaped distributions (GLM and BRT) respectively with the quadratic features (MaxEnt). Interactions were implemented as an interaction effect between centered main effects (GLM), a tensor interaction product with cubic regression splines (GAM), the number of randomly selected variables as branching candidates (RF), the product features (MaxEnt), and the 'interaction depth' parameter (BRT). Interaction and quadratic effects used centered covariates to limit the effects of collinearity (GLM, GAM, BRT). A subset of the GLM and GAM functions (GAM without interactions, GLM without interactions, GLM without interactions, and GLM with interactions, respectively) correspond to the data generating processes (AIF, SIF, SIT, SCT, respectively) when no error was involved. For further details, see Appendix A.

A computer science approach to quantify complexity, which remains a challenge for very different models (Merow et al., 2014), is algorithmic complexity that was successfully introduced for SDM comparison by García-Callejas and Araújo (2016). We used computation time needed to fit a model as a numerical approximation of algorithmic complexity under the assumption that all models are comparably efficiently implemented (García-Callejas and Araújo, 2016). Differences in computation time arise not only from differences in model implementation within the R environment, but also from variation in complexity and flexibility of models (García-Callejas and Araújo, 2016). We implemented our time measurements with the R function 'system.time', i.e., we forced a garbage collection before the start of each model run. Our time measurements indicated that GLM was simplest, followed by GAM and RF, then MaxEnt, and finally BRT (Fig. 4). The presence of interaction terms impacted computation time, but the largest differences were observed between differing model types (except for GAM and RF). Therefore, differences in computation times highlight dramatic differences in model complexity among modeling methodologies consistent with general understanding of the models (Elith and Leathwick, 2009).

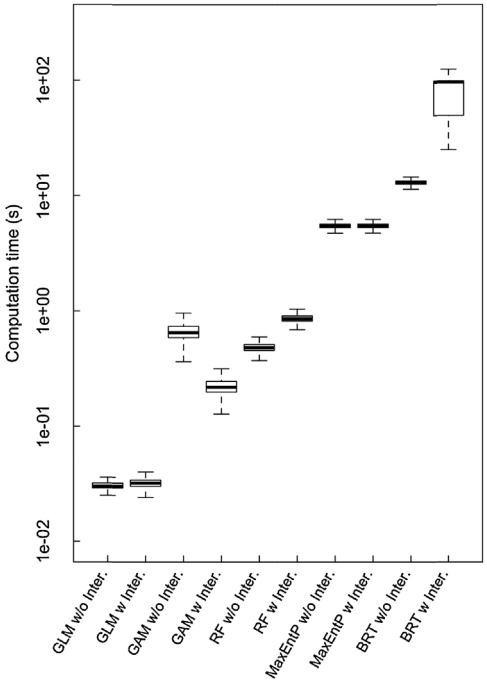

Fig. 4. Complexity of models represented by the computational time needed to be fit to the virtual species data.

2.6. Evaluating models

To evaluate models, we tested model fit in the region from which data for model fitting arose (NR) using the independent validation data (30% of the sample) and in each region to which the models might be used to extrapolate (SR, SW, and GP) using all data points. Model evaluation was based on (1) the ability of the model to fit the data (i.e., simulated absence/presence observations) and (2) the ability of the model to reproduce the underlying process (i.e., climatic suitability of the virtual species). Measures of data fit were True Skill Statistic (TSS), κ, and Area Under the ROC Curve (AUC). TSS measures the discrimination between hits and false alarms (Allouche et al., 2006); κ measures the degree to which the model performs better than random (Kraemer, 2006); AUC measures the overall ability to discriminate between presences and absences independent of thresholds identifying presence and absence (Fielding and Bell, 1997). TSS and κ scores reported in this paper are associated with presences and absences generated based on the probability threshold that maximizes the metric of interest. AUC scores are threshold-independent. While none of these metrics are perfect (e.g., κ is known to be sensitive to prevalence, Allouche et al., 2006), we chose them as metrics of model fit to provide results relevant to practicing modelers. Because we simulated the underlying processes, we were additionally able to assess the ability of the models to reproduce the underlying process by comparing predicted to observed climatic suitabilities graphically and statistically. We used root mean squared error (RMSE) and mean absolute error (MAE) as statistical measures of model performance with respect to the underlying process.

To understand how the underlying processes and modeling techniques influenced model performance, we performed a mixed-effects analysis of variance (ANOVA) on each of the five performance metrics (Table 1). We included the process, the model type, the inclusion of interaction terms in the models, inclusion of random and spatially variation (i.e., error), and interactions among these factors as the fixed effects, with the realization j of the underlying process as a random effect, to estimate the percentage of the variation in model performance that was explained by each of the components of our experimental design. ANOVAs were carried out using the aov function in R and variance partitioning was accomplished by calculating the sum of squares for each factor, the random effects (i.e., realizations of the data generation and error processes), and the residuals (i.e., data splitting for cross-validation) relative to the total sum of squares (Venables and Ripley,

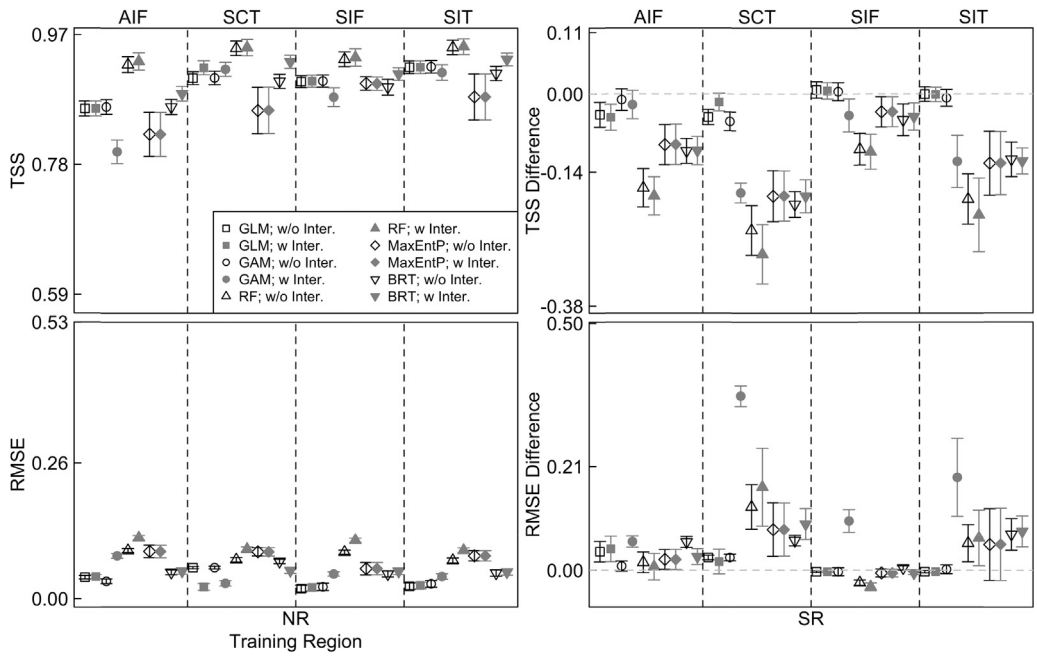

Fig. 5. Estimated means (± standard error) for the training region (NR) and estimated differences (± standard error) between the training region and one region (SR) to which models were transferred, for each model and process type of true skill statistics (TSS) and the root mean square error (RMSE). Other metrics (κ, AUC, and MAE) and transferability regions (SW and GP) in Fig. B1.

1997). Given the combination of fixed and random effects, the residual variation was interpreted to represent the influence of data-splitting for validation on performance metrics. We do not report *F* statistics and *p*-values here as the validity of statistical tests of significance on simulation data may be considered suspect (White et al., 2014).

3. Results

Within the training region NR, all models performed well, but there was substantial variation in measures of classification accuracy and representation of the underlying process (Fig. 5). For classification accuracy, RF performed best (e.g., high TSS) and MaxEnt performed worst (e.g., low TSS). The incorporation of interactions reduced performance for GAM models when the full climate space was available for model fitting (AIF and SIF), whereas interactions improved BRT performance regardless of the underlying process and GLM and GAM models when interactions were present in the underlying process (SCT). For representation of the underlying processes, patterns were similar, except that RF performed poorly (e.g., high RMSE). Models fitting the AIF process performed worse than models fit to SCT, SIT, and SIF. Results were qualitatively similar for other metrics (κ, AUC, and MAE) (Fig. B1).

For evaluations based on classification accuracy (TSS, κ, and AUC) in the training region (NR), the model type and use of interactions ($M + I + M \times I$) explained 42–76%, and data-splitting and realizations ($D + R$) explained 7–29%, and the data generating process ($P + E + P \times E$) explained 4–30% of the variation in evaluation metrics (Tables 2 and B1). For evaluations based on the underlying suitability process (RMSE and MAE), the model type and use of interactions ($M + I + M \times I$) explained 58–73%, and data-splitting and realizations ($D + R$) explained 4–5%, and the data generating process ($P + E + P \times E$) explained 7–19% of the variation in evaluation metrics (Tables 2 and B1). Error and spatial autocorrelation introduced into the data generation (E) contributed little to overall pattern for any evaluation metric.

Evaluations based on transferability indicated that models generally performed worse outside of the training region NR, especially more complex models. For most comparisons between the training regions and regions to which models were transferred, differences ([SR or SW or GP] – NR evaluation metrics) for evaluations based on observed presences and absences (TSS) were negative and differences for evaluations based on the underlying suitability process (RMSE) were positive (Fig. 5). The magnitudes of differences were largest for RF models and GAM models with interactions, especially for symmetric processes where the climate-space was truncated (SCT and SIT). Results were qualitatively similar for other metrics (κ, AUC, and MAE) and transferability regions (SW and GP) (Fig. B1).

Table 2
Percentages of variation in evaluation metrics for each process and region combination explained by the data generating process (P), error and spatial autocorrelation in the process (E), the model (M), the use of interactions (I), data realization (R) and data-splitting process (D) based on sums of squares. Factorial combinations of these factors are presented in the appendix (Table B1).

	Evaluation metric	Variation source					
		P	M	I	E	R	D
NR	TSS	29.0	38.8	0.0	0.3	7.1	9.1
	κ	13.8	73.6	1.0	0.2	1.1	5.9
	AUC	3.9	17.7	5.4	0.0	10.3	18.7
	RMSE	6.1	67.9	0.9	0.5	2.4	3.1
	MAE	18.8	51.2	1.8	0.4	2.0	2.5
SR	TSS	28.1	36.7	0.9	0.2	6.3	10.0
	κ	3.0	17.6	8.1	1.6	12.4	16.6
	AUC	13.1	26.5	7.6	0.0	8.0	11.5
	RMSE	25.0	17.6	6.2	0.1	7.8	9.0
	MAE	25.1	22.0	3.4	0.1	9.0	10.3
SW	TSS	12.2	24.8	4.9	0.2	9.9	16.5
	κ	9.1	19.2	9.1	0.7	8.7	12.4
	AUC	4.3	24.1	10.9	0.1	8.4	11.6
	RMSE	14.5	23.0	8.9	0.2	6.7	7.9
	MAE	16.8	24.4	5.9	0.2	8.3	9.7
GP	TSS	10.2	41.0	1.2	0.7	11.0	15.5
	κ	9.7	18.6	9.7	2.4	11.1	13.4
	AUC	5.0	23.1	10.8	0.1	13.7	16.8
	RMSE	19.7	18.6	8.3	0.1	9.3	9.7
	MAE	21.7	19.6	4.8	0.1	13.8	12.5

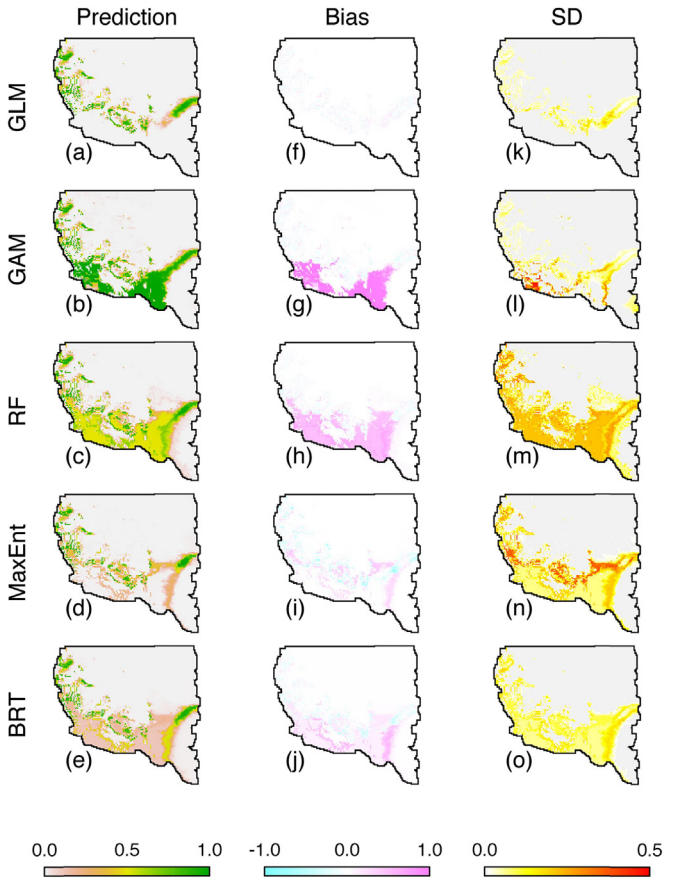

Fig. 6. Comparing the (a–e) mean predictions, (f–j) prediction bias, and (k–o) standard deviations of the predictions from the five models with interactions under the SCT scenario across the study region show geographic distribution of prediction uncertainties.

Differences in TSS for GP for AIF, SIF, and SIT processes tended to be positive for GAM and GLM models, indicating better performance in GP compared to NR. In the transferability regions SR, SW, and GP compared to the training region NR, the percent of the variation explained increased by 1–9% for use of interactions (I), 3–25% for the interacting effects of model type and interactions (M × I), and 0–22% for data-splitting and realizations (R + D + R × D) (excluding ROC; Tables 2 and B1). These results indicate that differences in model complexity and the specific data set used to fit models drive differences in predictive performance outside the training region NR. Furthermore, the percentage of variation explained by a certain factor varied among evaluation metrics by as much as an order of magnitude.

In some cases, complex models produced biased estimates of climatic suitability that tended to be highly uncertain, depending upon the realization of the data process and the data-splitting. In particular, the SCT process proved problematic for all models with interactions except GLM and GAM without interactions. Firstly, probabilities of occurrence were often over-predicted in the southern portion of the study area, especially GAM and RF (Fig. 6a–j). Secondly, the standard deviations in predictions from more complex models (RF, MaxEnt, and BRT) were quite high in much of SR, SW, and GP regions whereas the standard deviations in predictions for simpler models (GLM and GAM) were only high in geographic regions where intermediate suitabilities were predicted (Fig. 6k–o).

In addition, examining climatic response curves gives an idea on the univariate restrictions on species occurrence probability. While the climatic suitability curves with respect to temperature for the training region NR were largely consistent with each other, many models (GAM, RF, MaxEnt, and BRT) over-predicted climatic suitability at high values of minimum annual temperature (Fig. 7). Over-prediction of suitability by GAM at high temperature was evident in all projection regions, but declined as precipitation declined. The lack of an underlying parametric closed-functional form in RF, MaxEnt, and BRT models resulted in flat response curves that never approached zero. Only the GLM with interactions (i.e., the model that correctly represented the process) performed well at low log winter precipitation values in region GP. Similar patterns were observed at low log winter precipitation, except that (1) GAM response curves were more similar to GLM response curves and (2) flat portions of RF, MaxEnt, and BRT response curves were closer to zero (Fig. B2).

4. Discussion

This simulation study highlights the risks inherent in modeling species distributions using increasingly flexible correlative models when added complexity is not ecologically justified. Highly flexible statistical methods are gaining increasing popularity in species distribution modeling (Elith and Leathwick, 2009). However, the analytical advances have not been matched by theoretical justifications for that complexity (Austin, 2007). In particular, the added complexity that may improve model performance may also introduce issues of overfitting, and thus biases in model predictions. In the present study, increasing flexibility was particularly problematic when the entire climate-space in which a species might survive was not or could not be observed (e.g., SIT and SCT; Fig. 5). The truncation resulted in poor representation of the underlying process for climatic conditions beyond the truncation in the original dataset (Fig. 7), indicating that initial model fits are, to some extent, fitting spurious correlations rather than the underlying ecological processes (i.e., overfitting). In addition, the most flexible models (RF, MaxEnt, and BRT) all lack an underlying parametric form driving predictions, making extrapolation difficult when environmental constraints on species distributions are not well characterized by the data (Elith et al., 2010; Stohlgren et al., 2011). SDMs are utilized frequently in situations that promote apparent truncation such as modeling involving no-analog climates (Williams et al., 2007; Williams and Jackson, 2007), alien species (Fitzpatrick et al., 2007), and topographically rich landscapes. Predictions from highly flexible models therefore require greater scrutiny during model evaluation.

This is not to say that adding ecologically founded complexity, such as physiology, demography, or species interactions, should be avoided. While relatively simple models can represent physiological constraints on a species' niche, such as thresholds or declining function as one moves away from the species optimum (e.g., Boucher-Lalonde et al., 2012), ecological processes are clearly complex and predictions may depend upon this complexity (Ibáñez et al., 2006). However, adding complexity without ecological justification seems inappropriate and impacts our ability to understand fundamental ecological responses and predictions. The geographic distribution of poor predictions roughly matched geographic areas climatically dissimilar to the training region (Figs. 2 and 6) and likely arose from response curves that tended to over-predict probability of occurrence when extrapolating (Fig. 7). Advances in predictive ecological models will undoubtedly depend on incorporating ecological and evolutionary dynamics and mechanisms into statistical frameworks (Evans et al., 2013; McMahon et al., 2011). Our concern lies in leveraging increasingly flexible correlative approaches without properly examining the consequences of that added flexibility. More research using the virtual species framework is needed to better characterize the tradeoff between ecological theory and empirical flexibility.

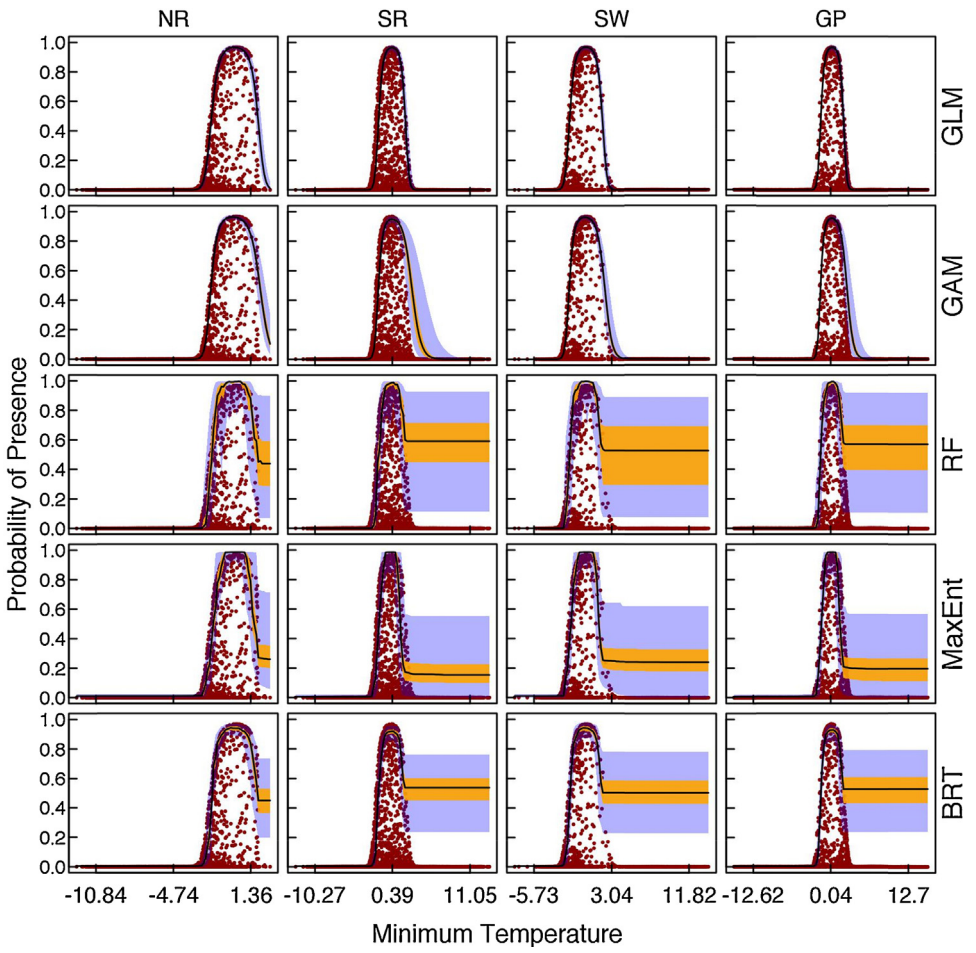

Fig. 7. Comparison of predicted climatic suitability curves (black, mean; 50% confidence intervals, orange; 90% confidence intervals, blue) with the true climatic suitabilities for scenario SCT (red dots) with respect to variation in minimum annual temperature under mean winter precipitation. As a result, the partial response curve represents the upper bound on predicted occurrence probabilities. Response curves are drawn only for the range of the covariate (*x*-axis) in a given region (see Fig. 1).

We were surprised to find that models did occasionally perform better outside of the training region. Model evaluations in region GP indicated improved performance compared to the training region. Region GP has a larger temperature range and a much smaller winter precipitation range than NR (Fig. 1). This narrow range of precipitation is comparable to the central portion of the geographic precipitation distribution in NR where most of the data for model fitting arose. Therefore, temperature may control most of the pattern and the relatively narrow influence of precipitation is well characterized in the training region. As a result, extrapolation errors do not appear to be much of a problem for region GP.

Our study benefited from simulated data in that we knew the underlying processes perfectly, but this will never be the case for SDMs fit to actual observations. Identifying overly complex models depends on observed presences and absences (Franklin, 2009). Measures of performance within the training region were largely explained by data generation (P and E), data sampling (R and D), and modeling (M and I). Outside the training region, the importance of interactions, data-splitting, and data realization (I, D, and R in Tables 2 and B1) was elevated. Given that species distribution modelers cannot control data generation and generally do not control data sampling, a solid knowledge when interaction terms are ecologically important and what functional form best represent the underlying process may prevent major errors in prediction. Our virtual species approach also indicated that uncertainties introduced by randomly sampling from the binomial distribution (R) were far greater than those introduced by simulated error and spatial autocorrelation (E) (Tables 2 and B1). This result contrasts with other studies, especially those highlighting the importance of spatial autocorrelation (Veloz, 2009; Crase et al., 2014; Le Rest et al., 2014). Future work should examine how much random or spatially explicit variation in the underlying suitability process is needed to overwhelm the importance of other factors examined in this research.

As indicated by our results, the danger of complex models does not become apparent until predictions are evaluated against novel climate-spaces, i.e., multivariate extrapolations. This is not a new idea (Araujo et al., 2005; Crase et al., 2014), but most scientists producing SDMs still validate models without assessing transferability. Our results show that cross-validation using data from the same region, and thus the same climate-space, as the training data is not sufficient for model selection or evaluation (Fig. 5). Given that climate truncation caused some of the greatest deviations in model performance, it appears that extrapolating models into unobserved climate-space contributes to the problem. Therefore, if validation data are selected from the same climate-space as the training data, evaluations cannot account for the extrapolation issue.

Extrapolation issues are particularly troubling when these models are used to predict future conditions because future landscapes will almost certainly include no-analog climates (Williams et al., 2007, 2012) and therefore suffer from the same extrapolation issues as we observed when applying SDMs across regions. Even when steps are taken to minimize model overfitting, such as penalized likelihood approaches to GAM models or using RF models (Breiman,

2001), resulting models can still be overly complex, resulting in biased predictions when extrapolating. Not only can these predictions be biased, but uncertainties may be so large (e.g., Figs. 6 and 7) that predictive maps become useless in guiding decision making. Presenting this uncertainty in a meaningful fashion could provide useful information to both ecologists and end-users (Wiens et al., 2009).

Acknowledgements

Funding for this work was provided by the NSF Postdoctoral Research Fellowship in Biology DBI-1202800 and the USDA Forest Service. We thank the University of Wyoming, Dr. W.K. Lauenroth, and Dr. J.B. Bradford for their support. This manuscript was improved by the comments of Dr. M.L. Mobley. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or the USDA Forest Service.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2016.03.012.

References

Aguirre-Gutierrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M., Biesmeijer, J.C., 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria – Dutch Hoverflies as a case study. PLoS ONE 8, e63708, http://dx.doi.org/10.1371/journal.pone.0063708.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232, http://dx.doi.org/10.1111/j.1365-2664.2006.01214.x.

Araujo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Global Change Biol. 11, 1504–1513, http://dx.doi.org/10.1111/j.1365-2486.2005.01000.x.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157, 101–118, http://dx.doi.org/10.1016/S0304-3800(02)00205-3.

Austin, M.P., 2005. Vegetation and environment: discontinuities and continuities. In: van der Maarel, E. (Ed.), Vegetation Ecology. Blackwell Publishing, Inc., Oxford, UK.

Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200, 1–19, http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005.

Austin, M.P., Gaywood, M.J., 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. J. Veg. Sci. 5, 473–482, http://dx.doi.org/10.2307/3235973.

Austin, M.P., Smith, A.E., 1989. A new model for the continuum concept. Vegetatio 83, 35–47.

Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. J. Biogeogr. 38, 1–8.

Bahn, V., McGill, B.J., 2013. Testing the predictive performance of distribution models. Oikos 122, 321–331, http://dx.doi.org/10.1111/j.1600-0706.2012.00299.x.

Bailey, R.G., United States Department of Agriculture 1995. Description of the Ecoregions of the United States, 2nd ed. rev. and enl, Miscellaneous Publication No. 1391.

Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? Ecography 33, 878–886, http://dx.doi.org/10.1111/j.1600-0587.2010.06181.x.

Boucher-Lalonde, V., Morin, A., Currie, D.J., 2012. How are tree species distributed in climatic space? A simple and general pattern. Global Ecol. Biogeogr. 21, 1157–1166, http://dx.doi.org/10.1111/j.1466-8238.2012.00764.x.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862.

Buckley, L.B., Waaser, S.A., MacLean, H.J., Fox, R., 2011. Does including physiology improve species distribution model predictions of responses to recent climate change? Ecology 92, 2214–2221, http://dx.doi.org/10.1890/11-0066.1.

Crase, B., Liedloff, A., Vesk, P.A., Fukuda, Y., Wintle, B.A., 2014. Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. Global Change Biol. 20, 2566–2579, http://dx.doi.org/10.1111/gcb.12598.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628, http://dx.doi.org/10.1111/j.2007.0906-7590.05171.x.

Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77, http://dx.doi.org/10.1111/j.1600-0587.2008.05505.x.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Syst. 40, 677–697, http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159.

Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1, 330–342.

Ellenberg, H., 1953. Physiologisches und ökologisches Verhalten derselben Pflanzenarten. Ber. Deutsch. Bot. Ges. 64, 350–361.

Evans, M.R., Bithell, M., Cornell, S.J., Dall, S.R.X., Diaz, S., Emmott, S., Ernande, B., Grimm, V., Hodgson, D.J., Lewis, S.L., Mace, G.M., Morecroft, M., Moustakas, A., Murphy, E., Newbold, T., Norris, K.J., Petchey, O., Smith, M., Travis, J.M.J., Benton, T.G., 2013. Predictive systems ecology. Proc. Biol. Sci. 280, 20131452, http://dx.doi.org/10.1371/journal.pone.0031824.

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

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J., Dunn, R.R., 2007. The biogeography of prediction error: Why does the introduced range of the fire ant over-predict its native range? Global Ecol. Biogeogr. 16, 24–33, http://dx.doi.org/10.1111/j.1466-8238.2006.00258.x.

Franklin, J., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, UK.

García-Callejas, D., Araújo, M.B., 2016. The effects of model and data complexity on predictions from species distributions models. Ecol. Model. 326, 4–12.

Gauch, H.G.J., Whittaker, R.H., 1972. Coenocline simulation. Ecology 53, 446–451.

Harrington, C.A., Gould, P.J., 2015. Tradeoffs between chilling and forcing in satisfying dormancy requirements for Pacific Northwest tree species. Front. Plant Sci. 6, 120, http://dx.doi.org/10.3389/fpls.2015.00120.

Heikkinen, J., Mäkipää, R., 2010. Testing hypotheses on shape and distribution of ecological response curves. Ecol. Model. 221, 388–399, http://dx.doi.org/10.1016/j.ecolmodel.2009.10.030.

Ibáñez, I., Clark, J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S., McBride, A., Welch, N.E., Wolosin, M.S., 2006. Predicting biodiversity change: outside the climate envelope, beyond the species-area curve. Ecology 87, 1896–1906.

Iverson, L.R., Schwartz, M.W., Prasad, A.M., 2004. How fast and far might tree species migrate in the eastern United States due to climate change? Global Ecol. Biogeogr. 13, 209–219, http://dx.doi.org/10.1111/j.1466-822X.2004.00093.x.

Jimenez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. Divers. Distrib. 14, 885–890, http://dx.doi.org/10.1111/j.1472-4642.2008.00496.x.

Kraemer, H.C., 2006. Correlation coefficients in medical research: from product moment correlation to the odds ratio. Stat. Methods Med. Res. 15, 525–545, http://dx.doi.org/10.1177/0962280206070650.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Global Ecol. Biogeogr. 23, 811–820, http://dx.doi.org/10.1111/geb.12161.

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

Madon, B., Warton, D.I., Araujo, M.B., 2013. Community-level vs species-specific approaches to model selection. Ecography 36, 1291–1298, http://dx.doi.org/10.1111/j.1600-0587.2013.00127.x.

McMahon, S.M., Harrison, S.P., Armbruster, W.S., Bartlein, P.J., Beale, C.M., Edwards, M.E., Kattge, J., Midgley, G., Morin, X., Prentice, I.C., 2011. Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. Trends Ecol. Evol. 26, 249–259, http://dx.doi.org/10.1016/j.tree.2011.02.012.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., Elith, J., 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography 37, 1267–1281.

Mesgaran, M.B., Cousens, R.D., Webber, B.L., 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. Divers. Distrib. 20, 1147–1159, http://dx.doi.org/10.1111/ddi.12209.

Meynard, C.N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. J. Biogeogr. 34, 1455–1469, http://dx.doi.org/10.1111/j.1365-2699.2007.01720.x.

Meynard, C.N., Kaplan, D.M., Silman, M., 2013. Using virtual species to study species distributions and model performance. J. Biogeogr. 40, 1–8, http://dx.doi.org/10.1111/jbi.12006.

Normand, S., Treier, U.A., Randin, C., Vittoz, P., Guisan, A., Svenning, J.-C., 2009. Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients. Global Ecol. Biogeogr. 18, 437–449, http://dx.doi.org/10.1111/j.1466-8238.2009.00451.x.

Pagel, J., Schurr, F.M., 2011. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. Global Ecol. Biogeogr. 21, 293–304, http://dx.doi.org/10.1111/j.1466-8238.2011.00663.x.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37–42, http://dx.doi.org/10.1038/nature01286.

Phillips, S.J., Dudik, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Canada, pp. 655–662.
PRISM Climate Group, 2012. http://prism.oregonstate.edu.
R Development Core Team, 2010. R: A Language and Environment for Statistical Computing.
Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. 33, 1689–1703.
Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., Evans, J.S., 2006. Empirical analyses of plant–climate relationships for the western United States. Int. J. Plant Sci. 167, 1123–1150, http://dx.doi.org/10.1086/507711.
Ridgeway, G., 2015. gbm: Generalized Boosted Regression Models. R Package Version 2.1.1. https://CRAN.R-project.org/package=gbm.
Santika, T., Hutchinson, M.F., 2009. The effect of species response form on species distribution model prediction and inference. Ecol. Model. 220, 2365–2379, http://dx.doi.org/10.1016/j.ecolmodel.2009.06.004.
Schibalski, A., Lehtonen, A., Schröder, B., 2014. Climate change shifts environmental space and limits transferability of treeline models. Ecography 37, 321–335, http://dx.doi.org/10.1111/j.1600-0587.2013.00368.x.
Segurado, P., Araújo, M.B., Kunin, W.E., 2006. Consequences of spatial autocorrelation for niche-based models. J. Appl. Ecol. 43, 433–444, http://dx.doi.org/10.1111/j.1365-2664.2006.01162.x.
Stohlgren, T.J., Jarnevich, C.S., Esaias, W.E., Morisette, J.T., 2011. Bounding species distribution models. Curr. Zool. 57, 642–647.
Thuiller, W., Brotons, L., Araujo, M.B., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27, 165–172, http://dx.doi.org/10.1111/j.0906-7590.2004.03673.x.
Vanderwel, M.C., Lyutsarev, V.S., Purves, D.W., 2013. Climate-related variation in mortality and recruitment determine regional forest-type distributions. Global Ecol. Biogeogr. 22, 1192–1203.
Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J. Biogeogr. 36, 2290–2299, http://dx.doi.org/10.1111/j.1365-2699.2009.02174.x.
Venables, W.N., Ripley, B.D., 1997. Modern Applied Statistics with S-PLUS. Springer-Verlag New York, Inc., New York, NY, USA, pp. 548.
Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. Nature 416, 389–395, http://dx.doi.org/10.1038/416389a.
White, J.W., Rassweiler, A., Samhouri, J.F., Stier, A.C., White, C., 2014. Ecologists should not use statistical significance tests to interpret simulation model results. Oikos 123, 385–388, http://dx.doi.org/10.1111/j.1600-0706.2013.01073.x.
Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. U. S. A. 106, 19729–19736.
Williams, J.W., Jackson, S.T., 2007. Novel climates, no-analog communities, and ecological surprises. Front. Ecol. Environ. 5, 475–482.
Williams, J.W., Jackson, S.T., Kutzbach, J.E., 2007. Projected distributions of novel and disappearing climates by 2100 AD. Proc. Natl. Acad. Sci. U. S. A. 104, 5738–5742, http://dx.doi.org/10.1073/pnas.0606292104.
Williams, A.P., Allen, C.D., Macalady, A.K., Griffin, D., Woodhouse, C.A., Meko, D.M., Swetnam, T.W., Rauscher, S.A., Seager, R., Grissino-Mayer, H.D., Dean, J.S., Cook, E.R., Gangodagamage, C., Cai, M., McDowell, N.G., 2012. Temperature as a potent driver of regional forest drought stress and tree mortality. Nat. Climate Change 3, 292–297, http://dx.doi.org/10.1038/nclimate1693.
Wood, S., 2006. Generalized Additive Models: An Introduction with R – Simon Wood – Google Books. Chapman & Hall/CRC Press, Boca Raton, FL.
Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., Grimm, V., 2010. The virtual ecologist approach: simulating data and observers. Oikos 119, 622–635, http://dx.doi.org/10.1111/j.1600-0706.2009.18284.x.