Biodiversity Informatics, 18, 2024, pp. 43-55

SPECIES DISTRIBUTION MODEL ACCURACY IS STRONGLY INFLUENCED BY THE CHOICE OF CALIBRATION AREA

Sergio Luna[1], Alexander Peña-Peniche[2,3] & Roberto Mendoza-Alfaro[1*]

[1]*Universidad Autónoma de Nuevo León, Facultad de Ciencias Biológicas, Laboratorio de Ecofisiología, Nuevo León, México*
[2]*Universidad Autónoma de Nuevo León, Facultad de Ciencias Biológicas, Laboratorio de Biología de la Conservación y Desarrollo Sustentable, Nuevo León, México*
[3]*Centro de Investigación Científica de Yucatán, A. C., Unidad de Recursos Naturales, Mérida, Yucatán, México*

**Corresponding author: Roberto Mendoza-Alfaro,*
Email: roberto.mendoza@yahoo.com

Abstract. Species distribution models (SDM) are widely used tools in ecology and conservation aimed at predicting the potential distribution of a species based on its environmental requirements and occurrence data. SDM face many challenges and uncertainties that influence their accuracy. Selecting the ideal calibration area is one of these difficulties. This study analyzes the influence of the extent of the calibration area on the accuracy of SDM through simulations with virtual species. Using bioclimatic variables, 100 virtual species were gener-ated. Occurrence probabilities were determined based on environmental suitability, spatial sampling bias, and accessible areas. SDM were built using MaxEnt, varying size of calibration area, spatial filtering of occurrence records, predictor collinearity treatment, and regularization parameter. Model performance was assessed in terms of functional accuracy (true model accuracy) and discrimination accuracy (model ability to separate oc-currence from random sites). Results show that the extent of the calibration area was the most influential factor (explaining 50% of the variance in functional accuracy), while regularization multiplier, predictor collinearity, and spatial thinning had minimal impact (about 4% of explained variance combined). Overall, larger calibration areas generally led to higher functional accuracy, although it varies across species. The correlation between functional and discrimination accuracy was relatively low, indicating that models performing well in one metric may not excel in the other. In conclusion, this research advances the discussion on calibration area selection, providing insights on its substantial effects on model accuracy. Our findings demonstrate that the size of the calibration area is one of the most critical factors affecting the accuracy of models, surpassing the influence of other factors. These insights highlight the importance of select appropriate calibration areas to improve model predictions and ensure more reliable applications of the models.

Key words: species distribution models, ecological niche, calibration area, model accuracy, virtual species, simulations, Maxent

Introduction

Species distribution models (SDM) are widely used tools in ecology and conservation aimed at predicting the potential distribution of a species based on its environmental requirements and occurrence data (Peterson and Soberón 2012). SDM provide valuable information for assessing biodiversity patterns (Weaver et al. 2006), identifying conservation priorities (Srinivasulu et al. 2021), evaluating climate change impacts (Brodie et al. 2022), and forecasting species invasions (Duque-Lazo et al. 2016). However, SDM also face many challenges and uncertainties that may impair their reliability (see Sillero et al. (2021) and Sillero and Barbosa (2021) for details). One significant challenge is the occurrence thinning process, which aims to reduce artificial clustering in species occurrence records. While thinning is necessary to mitigate biases introduced by uneven sampling efforts, it can also reduce the number of data points available for modeling, potentially affecting the model's performance and accuracy (Veloz, 2009; Boria et al. 2014; Varela et al. 2014). Another critical challenge that significantly impacts model performance is managing multicollinearity among predictor variables. High levels of correlation between variables can lead to overfitting, where the model becomes excessively tailored to the training data, reducing its ability to generalize to new data. Additionally, multicollinearity can result in unreliable response curves, where the effect of one variable is confounded by its correlation with others, making it difficult to discern the true influence of each predictor (Dormann et al. 2013; De Marco and Nóbrega, 2018; Feng et al. 2019).

Furthermore, the challenge of selecting the ideal calibration area adds another layer of complexity, as it directly influences the geographic region where the model is calibrated using observed data (Machado-Stredel et al. 2021). This is particularly important for some presence-background algorithms, such as Ecological Niche Factor Analysis (ENFA; Hirzel et al. 2002), Genetic Algorithm for Rule-set Production (GARP; Stockwell and Noble, 1992), and Maximum Entropy (MaxEnt; Phillips et al. 2006; Merow et al. 2013), which are influenced by the selection of the calibration area. It is important to emphasize that while the selection of the calibration area is a critical challenge for some algorithms, not all methods face this problem (Sillero et al. 2021). For instance, Mahalanobis distance (Clark et al. 1993) and BIOCLIM (Booth et al. 2014) do not require a calibration area due to their focus on occurrence data.

Presence-background methods evaluate the environmental conditions across the study area (background) and compare them with the conditions where the species is found, as indicated by its occurrence records (Phillips et al. 2006; Merow et al. 2013). The choice of calibration area can significantly influence model accuracy, as different areas may have different environmental conditions and sampling biases (Acevedo et al. 2012; Owens et al. 2013). The ideal calibration area is where the species is in equilibrium with its environment (Guisan and Zimmermann, 2000). This means that the species has adapted to the current environmental conditions and as a result its distribution reflects these conditions (Araújo and Pearson 2005; Barve et al. 2011). Under the Biotic-Abiotic-Mobility framework (BAM), the M fulfills this assumption (Soberon and Peterson 2005; Barve et al. 2011); however, for most species, its estimation is complicated (Acevedo et al. 2012).

Despite its significance, there has yet to be a consensus on how to select an appropriate calibration area, and different criteria and methods have been used in the literature (Machado-Stredel et al. 2021; Rotllan-Puig and Traveset 2021). Some approaches include selecting political regions, polygons, or rectangles around occurrence records (Shcheglovitova and Anderson 2013), selecting bioclimatic or hydrogeographical regions occupied by the species (Espindola et al. 2019; Sillero et al. 2021), or generating buffers around occurrence records (Zhu et al. 2014). In many studies, the calibration area used is not explicitly stated (e.g., Masin et al. 2014). These different methods can result in regions of varying sizes, which can significantly impact subsequent model characteristics.

According to several studies, either too constrained or overly expansive calibration areas may compromise the accuracy of model predictions (VanDerWal et al. 2009, Acevedo et al. 2012). Other studies have found that smaller calibration areas may yield superior model accuracy, as they mitigate the risk of overfitting to conditions near occupied localities or exclude regions with suitable conditions that remain unoccupied due to dispersal limitations and biotic interactions (Anderson and Raza 2010). Furthermore, selecting a background from larger areas leads to changes in variable importance, resulting in models becoming increasingly simplified and dominated primarily by just a few variables (VanDerWal et al. 2009).

Considering this context, there is a strong need for a more systematic evaluation of the effect of calibration area on SDM performance. Throughout this study, we refer to "model performance" in terms of functional accuracy (the model's ability to infer the true occurrence probability) and discrimination accuracy (the model's ability to correctly differentiate between presences and background in geographic space) (Warren et al. 2020). In this research, we address this through a simulation approach. Virtual species simulation allows us to create artificial species with known occurrence probabilities and environmental responses, allowing the accurate isolation of the effects of targeted factors (Meynard et al. 2019). Our objectives were 1) to compare models calibrated over areas of various sizes in terms of functional and discrimination accuracy (Warren et al. 2020); 2) to test whether knowing the accessible area to a species leads to improved models; and 3) to compare the relative contribution of calibration area size to model performance against other factors such as occurrence thinning and collinearity. By using virtual species with known environmental responses and probabilities of occurrence, this study provides insights into how several factors influence SDM accuracy and highlights the importance of careful calibration area selection.

Methods

Virtual species simulations

We simulated 100 virtual species using the *virtualspecies* R package (Leroy et al. 2016) and the 19 bioclimatic variables from WorldClim 2.0 (Fick and Hijmans 2017). These variables encompass critical climatic factors influencing species distributions and are widely utilized in SDM studies. We used a resolution of 2.5 arc-minutes to simulate the species' niches, as this resolution provides a balance between spatial detail and computational efficiency. The region considered was restricted to the American continent due to its broad range of habitats and climatic conditions. We chose to include variables that combine precipitation and temperature, despite a recent trend of excluding them from SDM studies (Booth 2022). These variables have been demonstrated to significantly influence species distributions. For instance, precipitation during the wettest and driest months are key factors shaping global plant distributions (Huang et al. 2021), while the mean temperature of the driest quarter has been identified as the most influential variable in explaining continental fish distributions (Castillo-Torres et al. 2017).

The ecological niche concept is based on the principle that each species has an optimal set of environmental conditions where it can survive, grow, and reproduce (Peterson and Soberon, 2012). To define each species' environmental preferences, two to five of the 19 bioclimatic variables were randomly selected and then subjected to a principal component analysis (PCA) using singular value decomposition, and the two principal components derived were used to define the species niche. Different species are influenced by different environmental factors (Huang et al. 2021), and the primary drivers of their distributions can differ significantly. Randomly selecting 2 to 5 variables helps simulate this variability. For the first two principal components, a Gaussian response curve was defined, with randomly selected values for the mean (representing the most suitable values for the species) and the standard deviation (indicating the niche breadth, or physiological tolerance of the species).

The initial suitability function, which describes the species–environment relationship and can have any range of values, was converted to an occurrence probability with a logistic function. After this transformation, the probability of occurrence is bounded between 0 and 1 (i.e., the likelihood that the species is present at a site given the specific set of environmental conditions on that site). Specifically, we used a logistic transformation with β (inflexion point) fixed at 0.5 and α (steepness of the slope) randomly selected. This method serves as an equivalent to applying thresholds but offers a more ecologically realistic representation by smoothing the transition between suitable and unsuitable conditions and simulating stochastic processes acting on species occurrences (Leroy et al. 2016; Meynard et al. 2019).

The magnitude of the spatial sampling bias was simulated using the Human Influence Index dataset (WCS and CIESIN, 2005), representing the likelihood of selection per cell. These values were divided by the maximum value, resulting in a sampling bias range between 0 and 1.

An essential aspect, according to niche theory, is the distinction between the fundamental, potential, and realized niches. The fundamental niche represents the full range of environmental conditions under which a species can survive and reproduce, while the potential niche corresponds to the subset of those conditions that actually exists at a given time (Jackson and Overpeck, 2000). The potential niche includes areas that are environmentally suitable but may not be occupied by the species due to additional factors such as dispersal constraints and/or the influence of biological interactions (Barve et al. 2011; Soberón and Nakamura, 2009). The realized niche is a subset of the potential niche after considering these factors (Jiménez and Soberón 2022; Soberón and Arroyo-Peña 2017). Dispersal capability, in particular, can limit a species' ability to explore favorable regions. To account for this, we defined the accessible area of each species by programming a cellular automaton (CA) to find potential areas where the species would be at environmental equilibrium. To achieve this, each species was assigned a random dispersal capability within the range of 50 to 250 km. This dispersal capability is fixed for each species and represents the maximum distance it can disperse in each iteration. Using this information, the accessible area was simulated by selecting cells through a Bernoulli trial based on their occurrence probability. The point with the highest occurrence probability across the landscape served as the initial selection point, and a buffer was generated around it with a radius equivalent to the assigned dispersal capability. Subsequent cells were selected within this buffer through additional Bernoulli trials, generating a new buffer. This process was repeated until no additional cells could be added (environmental equilibrium). The resulting buffer was considered the accessible area for the species, representing the sites it has potentially explored. The CA function is based on the presence of cells with sufficiently high probabilities of occurrence. In regions with high occurrence probabilities, the buffer will continue to expand, allowing the species to explore more adjacent areas. Conversely, in regions where occurrence probabilities are lower, fewer cells are added, and the buffer's growth tends to diminish. This process is analogous to real ecological dynamics, where species are more likely to colonize and establish in regions that provide optimal conditions for survival and reproduction.

For each species, 100 occurrence records were obtained by randomly selecting cells (with replacement) with a Bernoulli trial with a probability of success equal to:

$$p(x) = s(x)b(x)r(x)$$

where $p(x)$ is the sampling probability of the cell, $s(x)$ is the occurrence probability of the species in that cell, $b(x)$ is the relative strength of spatial sampling bias, and $r(x)$ is a binary variable with values 1 inside the accessible area of the species and 0 outside (Warren et al. 2020). This process was repeated five times for each species to account for stochasticity in the sampling process.

SDM

The simulation of each virtual species aimed to represent realistic aspects affecting real species occurrences. When testing the models, our goal was to reflect real-world modeling scenarios, acknowledging that the exact parameters influencing species distributions are often unknown (Meynard et al. 2019). For example, including spatial bias in the simulation of each virtual species occurrence was intentional to mimic real-world uneven sampling conditions due to factors like accessibility and observer effort. The thinning process applied during modeling (see below) is one of the existing methods intended to reduce this bias (Aiello-Lammens et al. 2015). Similarly, variable selection is a critical step in species distribution modeling (SDM), yet the true explanatory variables driving species distributions are rarely known (Inman et al. 2021). By excluding variables based on collinearity rather than relying on using the true influencing factors, we aimed to reflect the inherent uncertainty that exists in real-world variable selection scenarios. While simulating all aspects that influence species distributions is conceptually and computationally challenging, we aimed to include some essential aspects to provide a more realistic evaluation framework.

SDM were built for each of the occurrence datasets using MaxEnt v3.4.4 (Phillips et al. 2006) in the *dismo* R package (Hijmans et al. 2017). The different levels of the four factors set out below involved the evaluation of 2000 models per species (400 models x 5 occurrence replicates). The remaining configurations were set to their default settings.

Calibration area.—Given that Maxent works with a maximum entropy principle, it is necessary to provide random data from the environment for its characterization (Merow et al. 2013). As the extent of the calibration area directly influences the range of environmental conditions available for model training, it is of great importance to carefully consider the geographic space in model calibration. Given that the accessible area for a species is usually unknown, the effect of the calibration area extent was evaluated by randomly sampling 10,000 background cells (default value) within a radius of 25, 50, 100, 200, 300, 500, 700, and 1000 km around the occurrence records. All available cells were used for calibration areas with less than 10,000 background points.

Spatial filtering of occurrence records.—Spatial filtering of occurrence records is one of the main methods used to reduce the sampling bias in datasets derived from biological collections (Taylor et al. 2020). Occurrence records were filtered using the *spThin* package (Aiello-Lammens et al. 2015) with the following filtering distances: 0 (no filtering), 5, 10, 15, and 20 km.

Predictor collinearity.—There is still a lack of consensus regarding how predictor collinearity should be treated in SDM, given that using highly correlated variables may influence model performance (Feng et al. 2019). We compared the effect of dealing with predictor collinearity by calibrating models using all available variables or selecting variables with Pearson correlation coefficients below 0.7 using the *vifcor* function of the *usdm* package (Naimi et al. 2014).

Regularization parameter (RM).—Maxent uses Lasso regularization to constrain the modeled distributions to lie within a specific interval around the empirical mean instead of matching it exactly. This overfitting can be reduced by specifying a RM value that penalizes the use of additional parameters (Phillips et al. 2006; Warren and Seifert 2011). The effect of this factor was evaluated by using RM values of 0.5, 1 (default), 2, 3, and 5.

In addition to assessing the relative contributions of the previously mentioned factors, we examined the accuracy of a model for each species constructed under a scenario where the ecological characteristics of the species are well understood (hereafter termed unbiased model). The considerations for building this model included utilizing 1000 occurrence data points (instead of 100) sampled without spatial sampling bias, using the accessible area of the species as the calibration area, employing only the environmental variables that precisely define a species' niche as predictors, and evaluating the regularization parameter with the same 5 previously defined values.

Model performance

Model performance was evaluated using both functional and discrimination accuracy (Warren et al. 2020). Functional accuracy (true model accuracy) was calculated as the Spearman rank correlation between the true occurrence probability and the occurrence probability inferred from MaxEnt (with the complementary log-log (cloglog) transformation) across the accessible area. For large areas, 25,000 cells were selected at random due to computational constraints. Preliminary trials showed that using 25,000 randomly selected cells produced results very close to those obtained using the entire area, with a Pearson correlation of 0.997 between values from the whole area and those from the 25,000 randomly selected cells.

Discrimination accuracy was calculated using cross-validation, where the data were divided into four groups according to two criteria: randomly and by geographic blocks, using the *ENMeval* R package (Kass et al. 2021). The Boyce index was used as the evaluation metric (Hirzel et al. 2006), calculated for each of the four groups. The average Boyce index across these groups was used as the overall measure of discrimination accuracy.

The Boyce index was selected as an additional metric to assess whether it is possible to identify the best models based on discrimination accuracy, particularly when considering variations in the datasets that might not be perfectly aligned with the "known" truth. While we have a 'true' model, the use of the Boyce index allows us to evaluate how models perform in a comparative context, offering an additional perspective on model performance in scenarios that simulate real-world conditions.

Data analysis

We chose not to rely on p-values to evaluate the significance of our findings due to well-documented criticisms of their use (Hurlbert et al. 2019) and their inappropriateness in the context of simulation studies (White et al. 2014). Instead, we adopted an approach that focuses on the relative contributions and relationships among the examined factors through linear models, utilizing the LMG method implemented in the *relaimpo* R package. This method is based on sequential R^2 and addresses the dependency of regressor orderings by averaging over these orderings using simple unweighted averages (Grömping 2006). The response variable was the functional accuracy of SDM models, and the explanatory variables were the extension of the calibration area, the spatial filtering of occurrence records, the predictor collinearity, and the regularization parameter β treated as categorical variables. This approach allowed us to analyze the variability and contributions of different factors influencing model performance more robustly, rather than relying on traditional hypothesis testing with p-values.

Additionally, we evaluated the correlation between functional and discrimination accuracy using Pearson correlation to assess the capacity of discrimination metrics to select the best-performing models using withheld data. These analyses were performed individually for each species, given that we do not expect all species to be affected by the evaluated factors in the same way; some may experience more pronounced sampling bias, others may have smaller accessible areas, etc.

Results

Overall, the unbiased models exhibited high accuracy within the species accessible area, with 80% of these models showing functional accuracy values

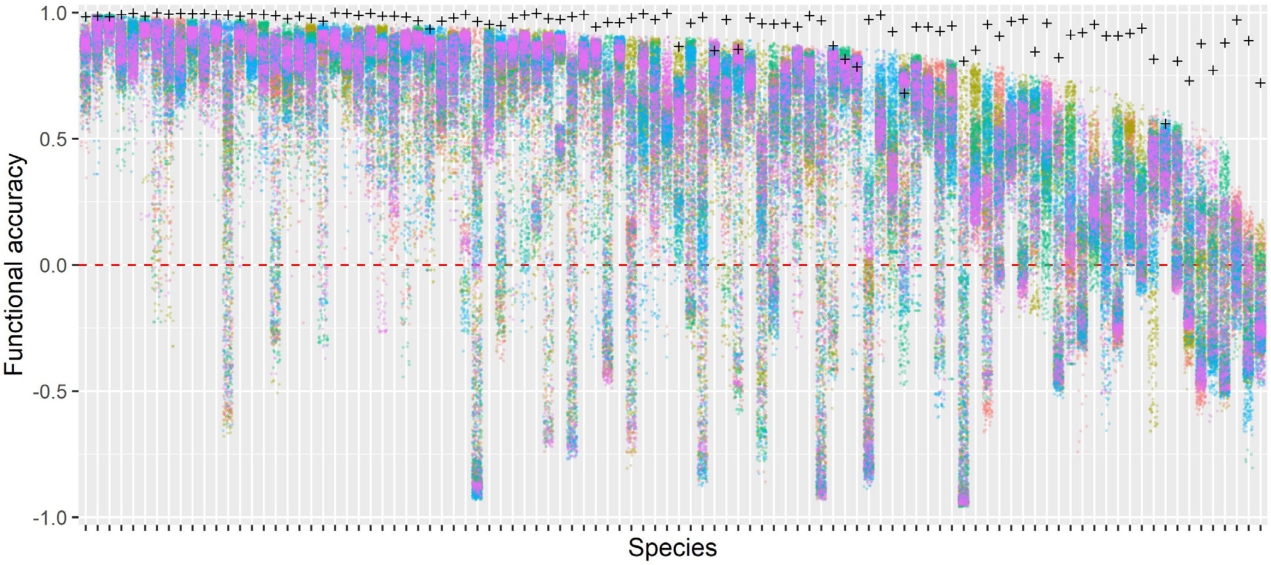

Figure 1. Functional accuracy of the models per species (400 models × 5 replicates). The Y-axis represents functional accuracy (measured as the Spearman rank correlation between true and predicted suitability) and the X-axis represents different species (without specific names as they are virtual). Data points are represented with five different colors (one for each replicate). Species are ordered from highest to lowest based on their best-performing model. The accuracy of the unbiased model is denoted for each species with a "+" symbol. The red dashed line represents the expected value by chance.

exceeding 0.9. The median functional accuracy of these models was high, with a Spearman correlation of 0.968, and the range of accuracy scores varied from 0.560 to 0.999 (Fig. 1). In contrast, the rest of the models demonstrated more variable performance within and across species. The maximum functional accuracy across species ranged from 0.178 to 0.996, with a median of 0.902. The minimum functional accuracy across species ranged from -0.959 to 0.699, with a median of -0.256. This considerable variability in accuracy was also reflected in the range of functional accuracy values within species (i.e. the difference between the maximum and minimum value for each species), showing values from 0.291 to 1.87, with a median of 1.045. Overall, these findings highlight the considerable range in model accuracy between species.

Among the analyzed species, 30 consistently exhibited models with positive functional accuracy, characterized by Spearman correlation coefficients greater than 0. The other 70 species showed at least one model with negative functional accuracy. Furthermore, 13 species were particularly notable, as most of their models yielded negative functional accuracy values. This disparity in model accuracy underscores the diverse responses of species to the calibration area and other factors, leading to variations in the accuracy of the generated models.

Explained variance

The extent of the calibration area turned out to be the most important factor in terms of true model accuracy, with a substantial median of 50.46% explained variance (range: 2.49% to 92.99%). Following this, the Regularization parameter (RM) played a less prominent role, with a median explained variance of 3.65% (range: 0.01% to 48.05%). Predictor collinearity and spatial thinning exhibited a negligible impact on true model accuracy, each contributing with a median explained variance of 0.41% and 0.04%, respectively (Fig. 2).

Functional and discrimination accuracy

We explored the correlation between functional and discrimination accuracy for the 100 species under two different data partitioning scenarios (Fig. 3). When considering random data partitioning, a wide range of correlations was observed. The minimum and maximum values were -0.58 and 0.82. The median correlation between these two metrics was 0.46. Remarkably, 19 species exhibited negative correlations. When data partitioning was based on

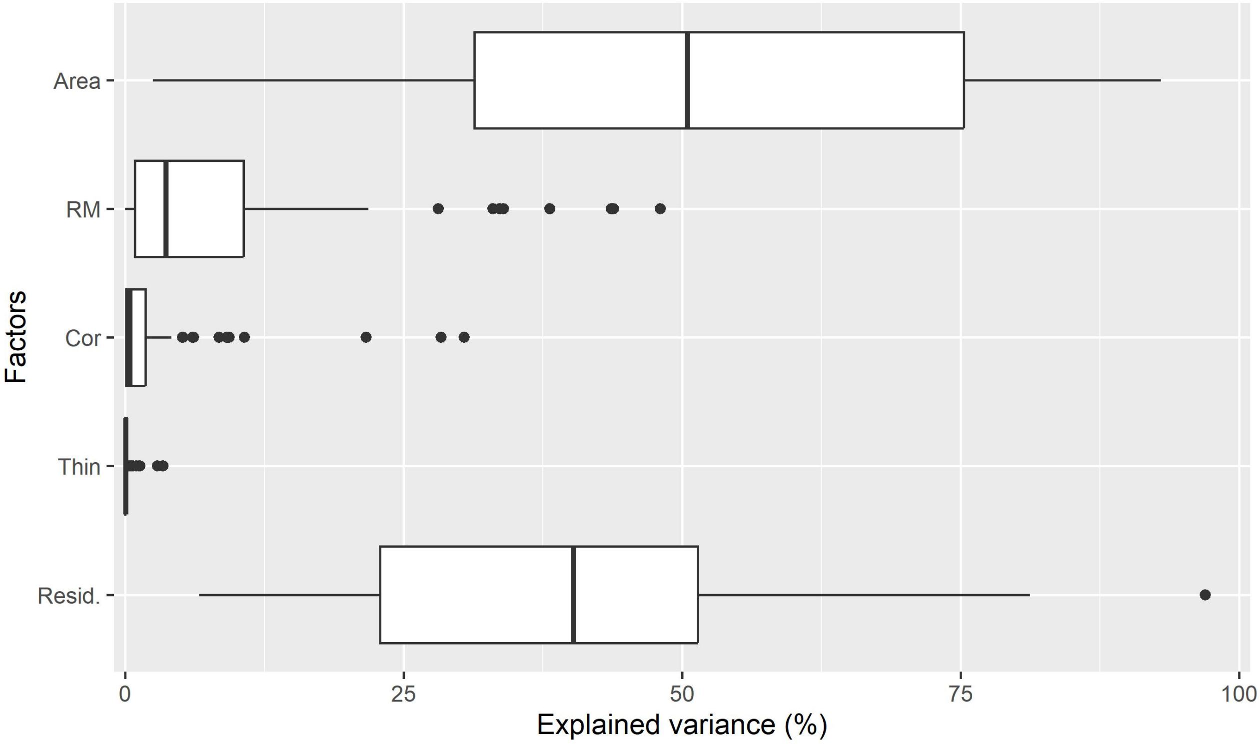

Figure 2. Percentage of variance explained by the four factors analyzed and the residual (unexplained) variance. The X-axis represents the percentage of variance explained, and the Y-axis shows the factors analyzed. The boxplots summarize the distribution of these percentages for all species. Area: Calibration area; RM: regularization multiplier; Cor: Predictor collinearity; Thin: Occurrence thinning.

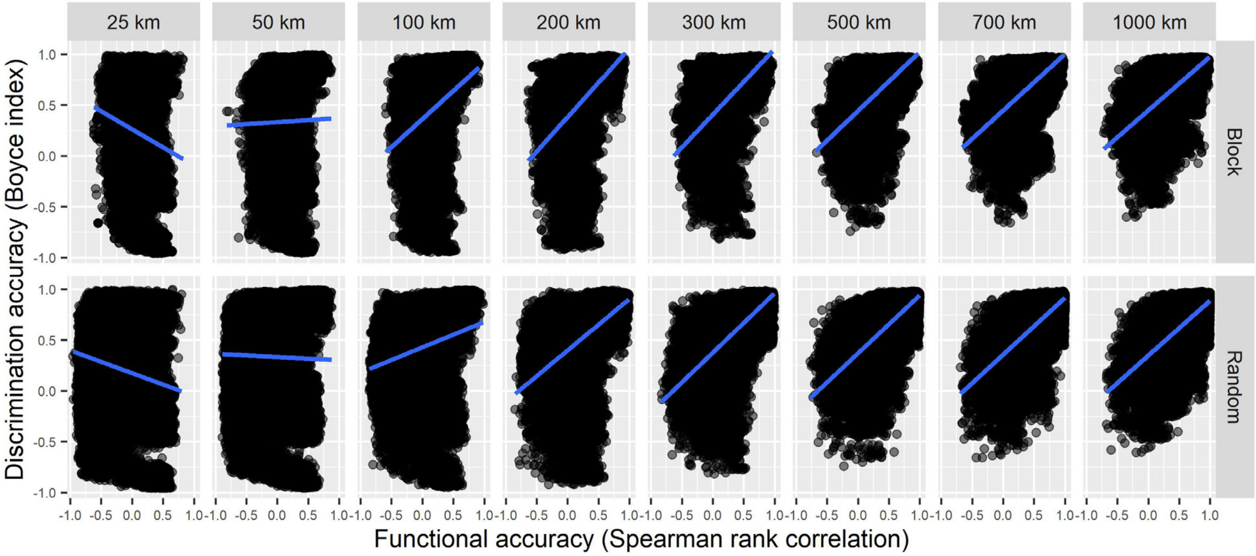

Figure 3. Relationship between model functional and discrimination accuracy for the 100 species under varying calibration area extents and partitioning methods. The x-axis represents the functional accuracy, divided by the extent of the calibration area, the y-axis represents discrimination accuracy, divided by the partitioning method (random and geographical blocks). The blue lines, obtained through Ordinary Least Squares (OLS) regression, provide a visual representation of the correlation between the two accuracy metrics for each subset of points.

geographical blocks, the correlations also displayed variability, ranging from -0.73 to 0.69, with a median correlation of 0.32. Here, 25 species showed negative correlations between functional and discrimination accuracy.

However, a noteworthy finding emerged despite the relatively low correlation between functional and discrimination accuracy. For 76 species, the best models based on the random data partitioning evaluation showed functional accuracies exceeding 0.5. In parallel, for 71 species, the best models selected with the geographical block data partitioning strategy achieved functional accuracies exceeding 0.5.

Discussion

Our research advances a more systematic evaluation of how varying extents of calibration area affect SDM accuracy. Indeed, the observed variability in model performance within and across species underscores the need for tailored approaches, considering species-specific characteristics. In this context, the most critical factor influencing the accuracy of species distribution models turned out to be the size of the calibration area. Overall, models calibrated with larger areas tend to show higher functional accuracy than models calibrated with smaller areas.

Although it is not straightforward to compare different algorithms, due to their reliance on different types of data (e.g., presence-only vs. presence-absence), statistical methodologies (e.g., classification vs. regression), or evaluation strategies (e.g., ROC-AUC vs Boyce Index), results are often compared in the literature (Bucklin et al. 2015; Valavi et al. 2022) and our major findings are in line with other previous research. VanDerWal et al. (2009), for instance, used buffers with increasing distances ranging from 10 to 500 km around the species' occurrences to study the impact of various calibration areas working with rainforest vertebrate from the Australian Wet Tropics (AWT). They found a rapid increase in accuracy as the background size expanded from 10 to 100 km (ROC-AUC > 0.93), with subsequent expansions beyond this threshold showing only marginal improvements (ROC-AUC > 0.99). However, an important drawback, acknowledged by the authors, was the potential overestimation of model accuracy when assessed over a large geographical extent (Lobo et al. 2008). VanDerWal et al. (2009) recognized this phenomenon and used a fixed evaluation area to calculate AUC values across all species. Their findings showed that the "fixed-area" accuracy was maximum at a background size around 200 km, and it gradually decreased as points were generated from larger regions. It is important to note that the use of simulated data with known occurrence probabilities allows us to directly assess the accuracy of the models without relying solely on the discrimination metrics, thereby ensuring that our findings are not artifacts of these metrics.

In a similar way, Acevedo et al. (2012) conducted a study using data from four ungulate species in Spain to evaluate the predictive accuracy of SDM calibrated over varying extents of calibration areas. Their results showed that while calibration accuracy (Miller's statistic) declined with the expansion of the calibration region, discrimination accuracy (ROC-AUC) increased. This approach allowed for the generation of purely environmental models that, when projected onto a new scenario, depicted the potential distribution of the species.

More recently Feng (2023), working with 87 hummingbird species, evaluated the effect of a series of buffers created around occurrences (from 5 to 5000 km) as calibration areas. The models calibrated with spatial buffers were compared with models calibrated with regions considered areas accessible to species. As a result, discrimination accuracy increased when the size of the calibration area was larger, but it reached a species-specific saturation threshold. Although the evaluation method affected this criterion, it was typically estimated to be less than 200 km. Surprisingly, model accuracy based on areas accessible to species was comparable to the saturation accuracy of models when spatial buffers were used.

In the present study, the comparison between unbiased models within each species' accessible areas and the rest of the models highlights a significant challenge: determining the accessible area of a species. Unbiased models consistently demonstrated excellent performance, achieving high functional accuracy across all species. In contrast, the rest of the models exhibited more variable accuracy, with some species displaying even negative functional accuracies. This discrepancy underscores the importance of considering species-specific characteristics and selecting appropriate calibration areas to ensure accurate model predictions. The findings suggest that modeling within species' accessible areas can mitigate biases and improve model reliability, highlighting the potential benefits of adopting unbiased approaches in SDM studies.

In contrast to our findings, Lobo and Tognelli (2011) reported different results. They investigated the impacts of spatial sampling bias, and the number and location of pseudo-absences on model accuracy using virtual species. Their results indicated that the number of pseudo-absences and the presence of spatial bias in sampling localities, along with their interaction, exerted a substantial influence on model accuracy (interpreted here with ROC-AUC, but they also evaluated sensitivity and specificity). As expected, higher number of pseudo-absences coupled with an absence of spatial bias yielded superior models. The location of pseudo-absences, whether distributed across the entire study area or restricted to regions outside the environmental envelope of the species, had a relatively smaller effect on model accuracy. They acknowledged that this might be attributed to the low relative occurrence area (only 3.5% of the total study area inhabited by the species). When contrasting our work with theirs, some differences stand out. They did not account explicitly for the dispersal capabilities of the species, potentially resulting in an overestimation of the realized distribution (Araújo and Pearson 2005). In addition, they employed thresholded maps instead of considering the more accurate occurrence probability (Leroy et al. 2016; Meynard et al. 2019). Also, the SDM was calibrated using the same bioclimatic variables that were used to create the virtual species niche. They used a threshold once more for model evaluation, restricting the use of data pertaining to the actual probability of occurrence. Finally, they simulated a single virtual species, while our study encompassed the results of 100 virtual species (in our work the lower relative contribution for the calibration area was 2.5% and the highest accounted for 93%). These distinctions highlight the complexities involved in modeling species distributions and the importance of considering multiple factors to enhance the robustness and ecological relevance of such models.

Increasing the extent of the calibration area involves incorporating data that are environmentally more distant (on average) from the occurrences. Consequently, the discrimination accuracy of the model may increase due to the ease to parameterize models with good discrimination capacity but that are low in useful information (Barve et al. 2011; Acevedo et al. 2012). This could be the result of larger calibration areas covering places with appropriate environmental conditions that are unoccupied because of biotic interactions and/or dispersal constrains, which could induce overfitting to conditions close to the occupied localities (Anderson and Raza 2010). On the other hand, the importance of coarse-scale factors such as climate may be underestimated at small calibration areas (Barve et al. 2011; Acevedo et al. 2012).

Our results confirm previous research highlighting the impact of the calibration area on model accuracy and show the complexity involved in its selection (VanDerWal et al. 2009). The environmental equilibrium assumption, wherein the species is adapted to its current environmental conditions, emphasizes the importance of choosing a calibration area that accurately reflects these conditions (Araújo and Pearson 2005). Nonetheless, our findings, along with earlier research, indicate that there is no agreement on the ideal calibration area (Rotllan-Puig and Traveset 2021; Machado-Stredel et al. 2021).

As some correlative methods for estimating ecological niches rely on contrasting the environmental characteristics of known occurrence sites with those from the available conditions across the study area, it becomes imperative to delineate and comprehend the potential range the species might have explored. This is crucial because the absence of a species outside its accessible area is not necessarily due to abiotic or biotic factors. Instead, a species may be absent from suitable regions simply due to its inability to disperse and reach those areas (Anderson and Raza 2010; Barve et al. 2011).

The way in which SDM handle collinearity between the predictor variables (Feng et al. 2019), sample bias (Ranc et al. 2017; Inman et al. 2021), and model complexity are additional aspects that could impact the model accuracy (Merow et al. 2014). However, when viewed in a multifactorial way, our results demonstrate that these factors have a significantly smaller impact than the selection of the calibration area. Concerning this, Barbet-Massin et al. (2012) observed that the impact of different methodological decisions in model quality varied depending on the specific SDM employed. For machine learning techniques (boosted regression trees and random forest), the number of pseudoabsences explained a greater amount of deviance (between 20 and 85%) than the weighting scheme and the method for selecting pseudo-absences (less than 15%). Thus, exploring a range of modeling techniques beyond Maxent is needed to further understand their differential responses and implications for SDM techniques.

Regarding the correlation between functional and discrimination accuracy, our results show that this correlation is relatively low. Discrimination accuracy based on both data partition schemes were a misleading measure of functional accuracy. However, contrary to our expectations, models selected with random partitioning demonstrated a better correlation between functional and discrimination accuracy. This is surprising because we expected lower correlation in this scenario due to the lower degree of independence between data used for evaluation and calibration. These findings suggest that random partitioning might be more effective in selecting models that accurately predict the true suitability values across the landscape, despite the theoretical advantages of block partitioning in creating geographically independent evaluation sets. One possible reason is that certain ranges of environmental values may be geographically clustered and so are not utilized during calibration in block partitioning, unlike random partitioning, which avoids this stratification (Kass et al. 2021). This contrasts with the findings of Warren et al. (2020), who observed a better correlation using geographical block partitioning.

Despite the relatively low correlation between functional and discrimination accuracy. The fact that the best models selected based on random and block data partitioning exhibited functional accuracies exceeding 0.5 highlights that, in certain cases, the choice of data partitioning strategy can lead to the selection of models with reasonably high functional accuracy, even when their overall discrimination accuracy showed limited alignment with the functional accuracy metric.

The inappropriate selection of the calibration area has significant implications for modeling applications (Barve et al. 2011). The consequences extend to critical aspects such as the inaccurate estimation of the extent of occurrence and area of occupancy, which may lead to misguided conservation priorities (VanDerWal et al. 2009), failure to generate appropriate mechanistic hypotheses about the parameters governing species distributions (VanDerWal et al. 2009), and distorted response curves (Thuiller et al. 2004). Therefore, it is essential to give careful thought and choose the calibration area to guarantee the validity and robustness of SDM.

Constructing spatial buffers around known occurrences, reflecting the potential spatial range a species could explore, offers a straightforward method for delineating a calibration area (Feng 2023). This approach aligns more closely with the theoretical considerations of species' mobility (Holloway and Miller 2017), providing a more realistic foundation for SDM exercises. In line with other authors (VanDerWal et al. 2009; Barve et al. 2011; Feng 2023), we recommend that species distribution modeling exercises should initiate with exploratory analyses of the calibration area, assessing the extent that can yield both the most accurate results and a biologically meaningful fit between species occurrence and predictor variables.

To sum up, this research advances the discussion of calibration areas, shedding light on their nuanced impacts on SDM accuracy. Acknowledging the complexity of these considerations, our findings contribute to the ongoing refinement of SDM practices, emphasizing the need for tailored approaches in different ecological contexts. The present study has demonstrated that the area of calibration is one of the most important factors affecting the functional accuracy of species distribution models using Maxent. Other factors, such as the value of the regularization parameter and the presence of collinearity between the predictor variables, have a much smaller impact.

ACKNOWLEDGMENTS

SL thanks the National Council of Research and Technology (CONAHCyT) for the doctoral scholarship provided. We also thank Alejandra Arreola-Triana and two anonymous reviewers for their helpful suggestions to earlier versions of this manuscript.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

REFERENCES

Acevedo, P., A. Jiménez-Valverde, J. M. Lobo, and R. Real. 2012. Delimiting the geographical background in species distribution modelling. J. Biogeogr. 39:1383–1390.

Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38:541–545.

Anderson, R. P., and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela: effect of study region on models of distributions. J. Biogeogr. 37:1378–1393.

Araújo, M. B., and R. G. Pearson. 2005. Equilibrium of species' distributions with climate. Ecography 28:693–695.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3:327–338.

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol. Model. 222:1810–1819.

Booth, T. H. 2022. Checking bioclimatic variables that combine temperature and precipitation data before their use in species distribution models. Austral Ecol. 47:1506–1514.

Booth, T. H., H. A. Nix, J. R. Busby, and M. F. Hutchinson. 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Divers. Distrib. 20:1–9.

Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol. Model. 275:73–77.

Brodie, S., J. A. Smith, B. A. Muhling, L. A. K. Barnett, G. Carroll, P. Fiedler, S. J. Bograd, E. L. Hazen, M. G. Jacox, K. S. Andrews, C. L. Barnes, L. G. Crozier, J. Fiechter, A. Fredston, M. A. Haltuch, C. J. Harvey, E. Holmes, M. A. Karp, O. R. Liu, M. J. Malick, M. Pozo Buil, K. Richerson, C. N. Rooper, J. Samhouri, R. Seary, R. L. Selden, A. R. Thompson, D. Tommasi, E. J. Ward, and I. C. Kaplan. 2022. Recommendations for quantifying and reducing uncertainty in climate projections of species distributions. Glob. Change Biol. 28:6586–6601.

Bucklin, D. N., M. Basille, A. M. Benscoter, L. A. Brandt, F. J. Mazzotti, S. S. Romañach, C. Speroterra, and J. I. Watling. 2015. Comparing species distribution models constructed with different subsets of environmental predictors. Divers. Distrib. 21(1):23–35.

Castillo-Torres, P. A., E. Martínez-Meyer, F. Córdova-Tapia, and L. Zambrano. 2017. Potential distribution of native freshwater fish in Tabasco, Mexico. Rev. Mex. Biodivers. 88:415–424.

Clark, J. D., J. E. Dunn, and K. G. Smith. 1993. A multivariate model of female black bear habitat use for a Geographic Information System. J. Wildl. Manage. 57(3):519–526.

De Marco, P. Jr, and C. C. Nóbrega. 2018. Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. PLoS ONE 13(9):e0202403.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. García Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46.

Duque-Lazo, J., H. van Gils, T. A. Groen, and R. M. Navarro-Cerrillo. 2016. Transferability of species distribution models: the case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. Ecol. Model. 320:62–70.

Espindola, S., J. L. Parra, and E. Vázquez-Domínguez. 2019. Fundamental niche unfilling and potential invasion risk of the slider turtle *Trachemys scripta*. PeerJ 7:e7923.

Feng, X. 2023. A test of species' mobility hypothesis in ecological niche modelling. J. Biogeogr. 50:1955–1966.

Feng, X., D. S. Park, Y. Liang, R. Pandey, and M. Papeş. 2019. Collinearity in ecological niche modeling: Confusions and challenges. Ecol. Evol. 9:10365–10376.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Grömping, U. 2006. Relative importance for linear regression in *R*: the package relaimpo. J. Stat. Softw. 17(1):1–27.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135:147–86.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. Package 'dismo'. Circles 9(1):1–68.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83(7):2027–2036.

Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Model. 199:142–152.

Holloway, P., and J. A. Miller. 2017. A quantitative synthesis of the movement concepts used within species distribution modelling. Ecol. Model. 356:91–103.

Huang, E., Y. Chen, M. Fang, Y. Zheng, and S. Yu. 2021. Environmental drivers of plant distributions at global and regional scales. Glob. Ecol. Biogeogr. 30:697–709.

Hurlbert, S. H., R. A. Levine, and J. Utts. 2019. Coup de grâce for a tough old bull: "statistically significant" expires. Am. Stat. 73:352–357.

Inman, R., J. Franklin, T. Esque, and K. Nussear. 2021. Comparing sample bias correction methods for Species Distribution Modeling using virtual species. Ecosphere 12(3): e03422.

Jiménez, L., and J. Soberón. 2022. Estimating the fundamental niche: accounting for the uneven availability of existing climates in the calibration area. Ecol. Model. 464:109823.

Jackson, S. T., and J. T. Overpeck. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology 26(Suppl): 194–220.

Kass, J. M., R. Muscarella, P. J. Galante, C. L. Bohl, G. E. Pinilla-Buitrago, R. A. Boria, M. Soley-Guardia, and R. P. Anderson. 2021. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods Ecol. Evol. 12:1602–1608.

Leroy, B., C. N. Meynard, C. Bellard, and F. Courchamp. 2016. virtualspecies, an R package to generate virtual species distributions. Ecography 39:599–607.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17:145–151.

Lobo, J. M., and M. F. Tognelli. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. J. Nat. Conserv. 19:1–7.

Machado-Stredel, F., M. E. Cobos, and A. T. Peterson. 2021. A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. Front. Biogeogr. 13(4):e48814.

Masin, S., A. Bonardi, E. Padoa-Schioppa, L. Bottoni, and G. F. Ficetola. 2014. Risk of invasion by frequently traded freshwater turtles. Biol. Invasions 16:217–231.

Merow, C., M. J. Smith, and J. A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36:1058–1069.

Merow, C., M. J. Smith, T. C. Edwards, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann, and J. Elith. 2014. What do we gain from simplicity versus complexity in Species Distribution Models? Ecography 37(12):1267–81.

Meynard, C. N., B. Leroy, and D. M. Kaplan. 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? Ecography 42:2021–2036.

Naimi, B., N. A. S. Hamm, T. A. Groen, A. K. Skidmore, and A. G. Toxopeus. 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37:191–203.

Owens, H. L., L. P. Campbell, L. L. Dornak, E. E. Saupe, N. Barve, J. Soberón, K. Ingenloff, A. Lira-Noriega, C. M. Hensz, C. E. Myers, and A. T. Peterson. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol. Model. 263:10–18.

Peterson, A. T., and J. Soberón. 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. Nat. Conserv. 10:102–107.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.

Ranc, N., L. Santini, C. Rondinini, L. Boitani, F. Poitevin, A. Angerbjörn, and L. Maiorano. 2017. Performance tradeoffs in target-group bias correction for species distribution models. Ecography 40(9):1076–87.

Rotllan-Puig, X., and A. Traveset. 2021. Determining the minimal background area for species distribution models: MinBAR package. Ecol. Model. 439:109353.

Sillero, N., and A. M. Barbosa. 2021. Common mistakes in ecological niche models. Int. J. Geogr. Inf. Sci. 35(2):213–226

Sillero, N., S. Arenas-Castro, U. Enriquez-Urzelai, C. Gomes Vale, D. Sousa-Guedes, F. Martínez-Freiría, R. Real, and A. Márcia Barbosa. 2021. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecol. Model. 456:109671.

Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers. Inform. 2:1–10.

Soberón, J., and B. Arroyo-Peña. 2017. Are fundamental niches larger than the realized? Testing a 50-year-old prediction by Hutchinson. PLoS ONE 12(4):e0175138.

Shcheglovitova, M., and R. P. Anderson. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecol. Model. 269:9–17.

Soberon, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers. Inform. 2:1–10.

Soberón, J., and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proc. Natl. Acad. Sci. USA 106(2):19644–19650.

Srinivasulu, A., B. Srinivasulu, and C. Srinivasulu. 2021. Ecological niche modelling for the conservation of endemic threatened squamates (lizards and snakes) in the Western Ghats. Glob. Ecol. Conserv. 28:e01700.

Stockwell, D. R. B., and I. R. Noble. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Math. Comput. Simul. 33:385–390.

Taylor, A. T., T. Hafen, C. T. Holley, A. González, and J. M. Long. 2020. Spatial sampling bias and model complexity in stream-based species distribution models: a case study of Paddlefish (*Polyodon spathula*) in the Arkansas River basin, USA. Ecol. Evol. 10:705–717.

Thuiller, W., L. Brotons, M. B. Araújo, and S. Lavorel. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27:165–172.

Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecol. Monogr. 92(1):e01486.

Varela, S., R. P. Anderson, R. García-Valdés, and F. Fernández-González. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37:1084–1091.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J. Biogeogr. 36:2290–2299.

VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Model. 220:589–594.

Warren, D. L., N. J. Matzke, and T. L. Iglesias. 2020. Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. J. Biogeogr. 47:167–180.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21:335–342.

Weaver, K. F., T. Anderson, and R. Guralnick. 2006. Combining phylogenetic and ecological niche modeling approaches to determine distribution and historical biogeography of Black Hills mountain snails (Oreohelicidae). Divers. Distrib. 12:756–766.

White, J. W., A. Rassweiler, J. F. Samhouri, A. C. Stier, and C. White. 2014. Ecologists should not use statistical significance tests to interpret simulation model results. Oikos 123:385–388.

Wildlife Conservation Society, and Center for International Earth Science Information Network. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC).

Zhu, G.-P., D. Rédei, P. Kment, and W.-J. Bu. 2014. Effect of geographic background and equilibrium state on niche model transferability: predicting areas of invasion of *Leptoglossus occidentalis*. Biol. Invasions 16:1069–1081.