ECOGRAPHY



*Research article*

# How far can I extrapolate my species distribution model? Exploring shape, a novel method

**Santiago José Elías Velazco[1,2], Miranda Brooke Rose[3], Paulo De Marco Jr[4], Helen M. Regan[5] and Janet Franklin[3]**

[1]Instituto de Biología Subtropical, Universidad Nacional de Misiones-CONICET, Puerto Iguazú, Misiones, Argentina
[2]Programa de Pós-Graduação em Biodiversidade Neotropical, Universidade Federal da Integração Latino-Americana, Foz do Iguaçu, Paraná, Brazil
[3]Department of Geography, San Diego State University, San Diego, CA, USA
[4]Theory, Metacommunity and Landscape Ecology Lab, ICB V, Universidade Federal de Goiás, Goiânia, Goiás, Brazil
[5]Department of Evolution, Ecology, and Organismal Biology, University of California – Riverside, Riverside, CA, USA

**Correspondence: Santiago José Elías Velazco (sjevelazco@gmail.com)**

**Ecography**
**2023: e06992**
doi: 10.1111/ecog.06992

Subject Editor: Miguel Araújo
Editor-in-Chief: Miguel Araújo
Accepted 17 October 2023

Species distribution and ecological niche models (hereafter SDMs) are popular tools with broad applications in ecology, biodiversity conservation, and environmental science. Many SDM applications require projecting models in environmental conditions non-analog to those used for model training (extrapolation), giving predictions that may be statistically unsupported and biologically meaningless. We introduce a novel method, Shape, a model-agnostic approach that calculates the extrapolation degree for a given projection data point by its multivariate distance to the nearest training data point. Such distances are relativized by a factor that reflects the dispersion of the training data in environmental space. Distinct from other approaches, Shape incorporates an adjustable threshold to control the binary discrimination between acceptable and unacceptable extrapolation degrees. We compared Shape's performance to five extrapolation metrics based on their ability to detect analog environmental conditions in environmental space and improve SDMs suitability predictions. To do so, we used 760 virtual species to define different modeling conditions determined by species niche tolerance, distribution equilibrium condition, sample size, and algorithm. All algorithms had trouble predicting species niches. However, we found a substantial improvement in model predictions when model projections were truncated independently of extrapolation metrics. Shape's performance was dependent on extrapolation threshold used to truncate models. Because of this versatility, our approach showed similar or better performance than the previous approaches and could better deal with all modeling conditions and algorithms. Our extrapolation metric is simple to interpret, captures the complex shapes of the data in environmental space, and can use any extrapolation threshold to define whether model predictions are retained based on the extrapolation degrees. These properties make this approach more broadly applicable than existing methods for creating and applying SDMs. We hope this method and accompanying tools support modelers to explore, detect, and reduce extrapolation errors to achieve more reliable models.

NORDIC SOCIETY OIKOS

www.ecography.org

Keywords: environmental novelty, extrapolation, Mahalanobis distance, model prediction, non-analog environmental data, transferability

## Introduction

Species distribution and ecological niche models (hereafter SDMs), used in a wide range of disciplines (Franklin 2013), rely on ecological niche theories, wherein species distributions are determined by biotic, abiotic, and historical dispersal factors of the focal species (Austin 2002, Soberón and Peterson 2005, Peterson and Soberón 2012). The correspondence of the species niche in environmental and geographic space, known as Hutchinson's duality, is an essential assumption on which SDMs rely. According to Hutchinson's duality, each point in geographic space corresponds to a single point in environmental space; however, a single point in environmental space could correspond to one or more points in geographic space (Hutchinson 1957, Colwell and Rangel 2009). This property allows SDMs to estimate different distributions corresponding to occupied, potential, or invadable areas (Soberón and Peterson 2005). Therefore, these models can be used, for instance, to identify new populations of known species (Fois et al. 2015), detect areas at risk for invasive species (Montti et al. 2021), target habitats for species translocation or assisted dispersal (Regan et al. 2012), evaluate the effect of environmental change on species distributions (Calambás-Trochez et al. 2021), or explore the distribution of species in past periods (Bueno et al. 2016). These applications all involve applying SDMs to projection data (environmental data representing the predictor variables) that are temporally or geographically different from the training data (a.k.a. calibration data). When an SDM is used to make predictions for a different time period or region from the training data, this is referred to as projection or transferring in the SDM literature (Araújo et al. 2019), and often involves extrapolation to environmental conditions outside those used to train the models, e.g. non-analog conditions (Elith et al. 2010, Rousseau and Betts 2022). Consequently, model predictions projected to environmental conditions far from the training conditions may be statistically unsupported and ecologically meaningless, significantly limiting their utility (Araújo and Peterson 2012).

Extrapolation is a phenomenon that can affect predictive models of all kinds, and several methods for detecting, measuring, or avoiding extrapolation have been proposed for different research areas (Aniceto et al. 2016, Mahony et al. 2017, Meyer and Pebesma 2021). However, the SDM field has a unique combination of theoretical frameworks, methodological characteristics, and attributes: 1) SDM relies on niche theory, so it is expected that SDM can estimate the shape and environmental suitability pattern of different niche types (Soberón and Peterson 2005, Peterson and Soberón 2012); 2) SDM often uses categorical data types to characterize the response variable such as species presence, absence, pseudo-absence, and background points (Barbet-Massin et al. 2012, Liu et al. 2019), and/or other distributional information (Merow et al. 2017); 3) SDM uses different modeling methods (hereafter algorithms) that range from environmental envelope and distance-based models to machine learning and Bayesian methods (Norberg et al. 2019), and because of the degree of uncertainty generated by different modeling approaches, ensemble models are often implemented (Thuiller et al. 2019); 4) depending on the model purpose and amount of data, different partitioning methods are used for model validation (Valavi et al. 2019), and 5) frequently SDMs are constructed for rare or poorly sampled species (Breiner et al. 2015). Such features necessitate the development of methods to detect, quantify, and limit extrapolation that are adapted specifically for SDMs.

To illustrate this problem, we consider different SDMs constructed globally for the invasive species *Ligustrum lucidum* (Fig. 1). It is clear that suitability values are similar across algorithms within the ranges of environmental conditions used for model training (depicted by a white polygon in Fig. 1). However, these bivariate partial dependence plots often exhibit different trends under environmental conditions far from those used for model training (outside the white polygon in Fig. 1). For example, unrealistic increases in habitat suitability are predicted for extremely high precipitation and temperature (for three of the model types, GAM, RF and Maxent), and moderately high suitability is predicted for a range of temperature and precipitation combinations far from those conditions used in the model fitting (for GP and SVM; Fig. 1).

Several strategies and tools have been developed to control, explore, and measure extrapolation. Some algorithms control how estimations outside training conditions are made. For instance, for generalized additive models, penalty order for thin plate splines can be controlled, while for Maxent predictions outside training conditions can be 1) freely estimated, 2) kept constant (clamping), 3) reduced based on the difference between clamped and non-clamped predictions (fade by clamping), or 4) set to zero suitability value (no extrapolation) (Phillips et al. 2006). However, even using these approaches can lead to unrealistic predictions in non-analog conditions precisely because they control estimation but do not measure the degree of extrapolation.

Some extrapolation metrics used to estimate the degree of extrapolation and define non-analog conditions have been proposed in the SDM literature; for instance, the Multivariate environmental similarity surface (MESS; Elith et al. 2010) is perhaps the most extensively used approach because it was first implemented in the widely-used Maxent algorithm (Elith et al. 2011). MESS measures degree of extrapolation based on the environmental distance between projection data and the centroid of training data; univariate extrapolation limits are then defined using a rectilinear envelope (Elith et al. 2010). Alternatively, the Environmental overlap (EO) approach (Zurell et al. 2012) is a binary metric that

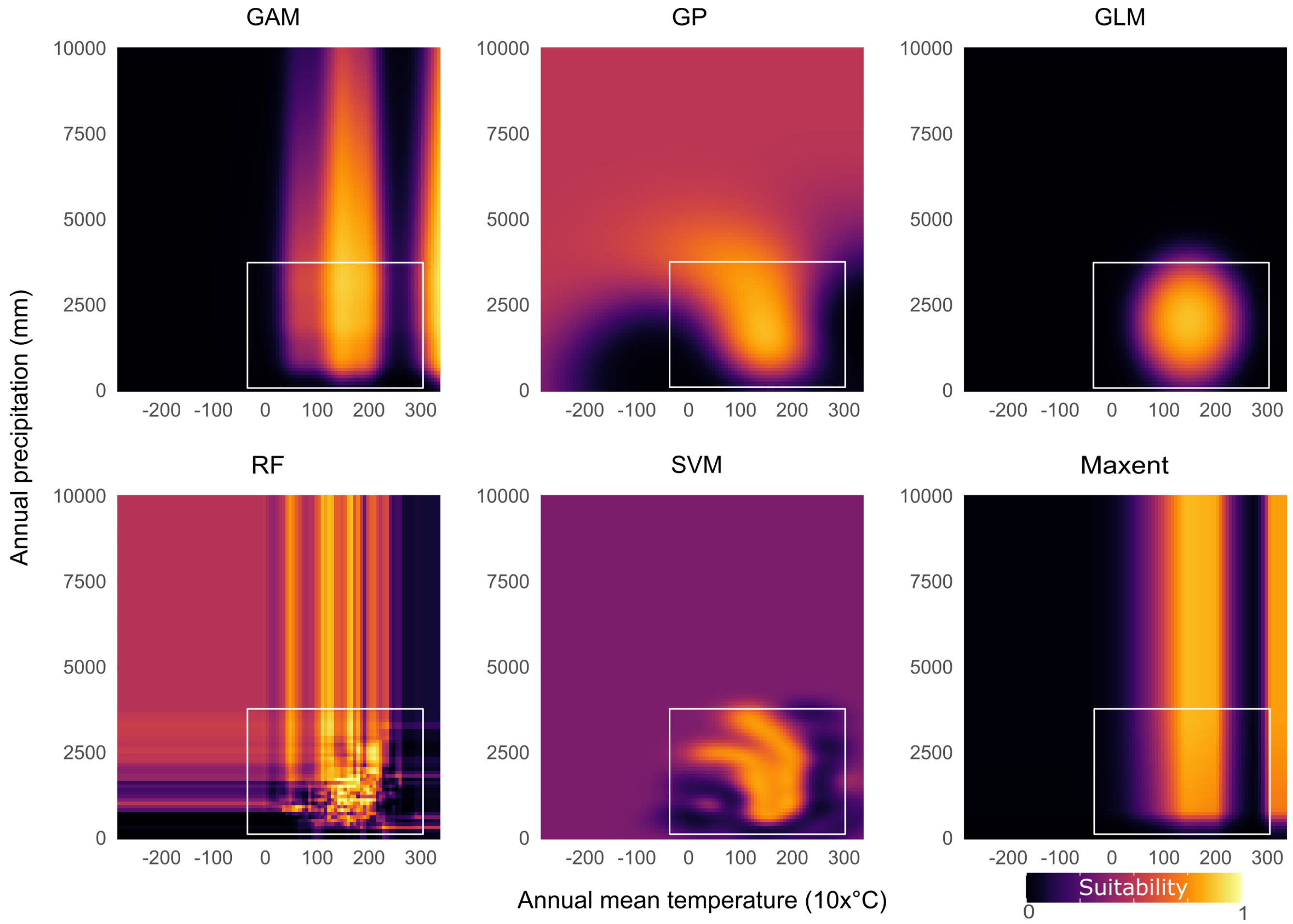


Figure 1. Bivariate partial dependence of *Ligustrum lucidum* predicted by six algorithms using annual mean temperature and annual precipitation, illustrating the differences across algorithms in environmental measurement space far from the training conditions (outside the white polygon). White polygons depict the limits of the training data (i.e. presences and pseudo-absences). Algorithms employed: generalized additive model (GAM), Gaussian process (GP), generalized linear model (GLM), random forest (RF), support vector machine (SVM), and maximum entropy (Maxent; predicted without clamping). SDMs were constructed with the same presence and pseudo-absence data from Montti et al. (2021).

splits environmental space into a specified number of bins based on the training data, grouping all projection data that fall outside the unique combination of environmental predictor values used in training data into a single category without reflecting how far the values of projection data lie beyond the range of training data. The Mobility-oriented parity approach (MOP; Owens et al. 2013) is similar to MESS in detecting strict extrapolation; however, MOP restricts extrapolation by averaging environmental distance to the user-specified nearest part of the training data. Extrapolation detection (EXDET; Mesgaran et al. 2014) uses two metrics to measure the degree of both univariate extrapolation (i.e. extrapolation outside the range of training conditions) and combinatorial extrapolation (i.e. extrapolation within the range of training conditions). Each of these methods has limitations. For example, none captures the complex relationships between training and projection data in environmental space; approaches such as MESS, MOP and EXDET discriminate between analog and non-analog projection data based on a rectilinear envelope, and MESS and EXDET measure extrapolation based on environmental condition of the training data centroid (further details about these metrics is provided under 'Properties of Shape in comparison with other approaches').

Given the underlying assumption of SDMs that species distributions are in equilibrium with environmental conditions that represent niche dimensions (Guisan and Zimmermann 2000) and the need to project these models into environmental conditions beyond the training data, a method to explore the degree of extrapolation that is model-agnostic (i.e. it does not depend on model approach), simple to interpret and implement, and capable of capturing the complex relationships between training and projections data in environmental space is called for. Furthermore, we are unaware of research that has compared the performance of existing methods under different modeling conditions, which is crucial for evaluating the rigor of SDMs used to make projections under novel environmental conditions. In this paper, we introduce a novel extrapolation metric, Shape, and explore the performance of this and other extrapolation metrics with an experiment based on virtual species. Finally, tools for measuring and exploring extrapolation implemented in 'flexsdm' R package (www.r-project.org) are provided.

## Material and methods

### The Shape method

As with other extrapolation metrics, Shape measures the degree of extrapolation (i.e. environmental novelty) in environmental space based on the relationship between the training and projection data. In the context of SDM, training data represents the environmental conditions derived from data used for model training or fitting (i.e. presence, presence–absence, presence–pseudo–absence, or presence–background points). In contrast, projection data represents the environmental conditions used for model prediction (e.g. data from a given geographical projection area, time period, or data resolution). Shape is a model-agnostic approach because its calculation is based solely on the environmental distance between training and projection data, i.e. the degree of extrapolation is independent of model parameters and predictions.

The Shape method measures environmental distance using Mahalanobis distance ($d$) based on predictor variables used in modeling (Eq. 1):

$$d(p,t) = \sqrt{(p-t)^T M^{-1} (p-t)} \quad (1)$$

Where $M^{-1}$ is an inverse covariance matrix based on the training data, $p$ is a vector of a projection point, $t$ is a vector of a training data point, and $T$ indicates the transpose of the vector. In a conventional Mahalanobis metric (from which the Shape method deviates), $t$ would typically represent the centroid of environmental conditions (i.e. the mean); however, $t$ represents each training data point in Shape.

Shape calculates the extrapolation metric ($S_{pi}$) for a given projection data point $p_i = (i = 1, \ldots, r)$ by its multivariate distance to the nearest training data point, where training data points are denoted by $t_j$ ($j = 1,\ldots, m$) (Eq. 2, Fig. 2A):

$$S_{pi} = \frac{\min_{j=1}^{m}\left\{d\left(p_i, t_j\right)\right\}}{A} \quad (2)$$

where $m$ is the total number of training data points, and $A$ is a dispersion factor that relativizes this distance (Williams et al. 2007, Fitzpatrick et al. 2018, Meyer and Pebesma 2021, Fig 2B). The higher the Shape value, the greater the environmental novelty of the projection point and, consequently, the lower the degree of reliability of a model prediction.

The factor $A$ is calculated as the averaged Mahalanobis distance between training data points $t_j$ and the centroid of the training data $c$ (Eq. 3, Fig. 2B):

$$A = \frac{\sum_{j=1}^{m} d\left(t_j, c\right)}{m} \quad (3)$$

where $m$ is the total number of training data points. The quantity $A$ reflects the dispersion of the training data in environmental space.

Extrapolation is a continuous phenomenon that depends on how different the projection data are from the training

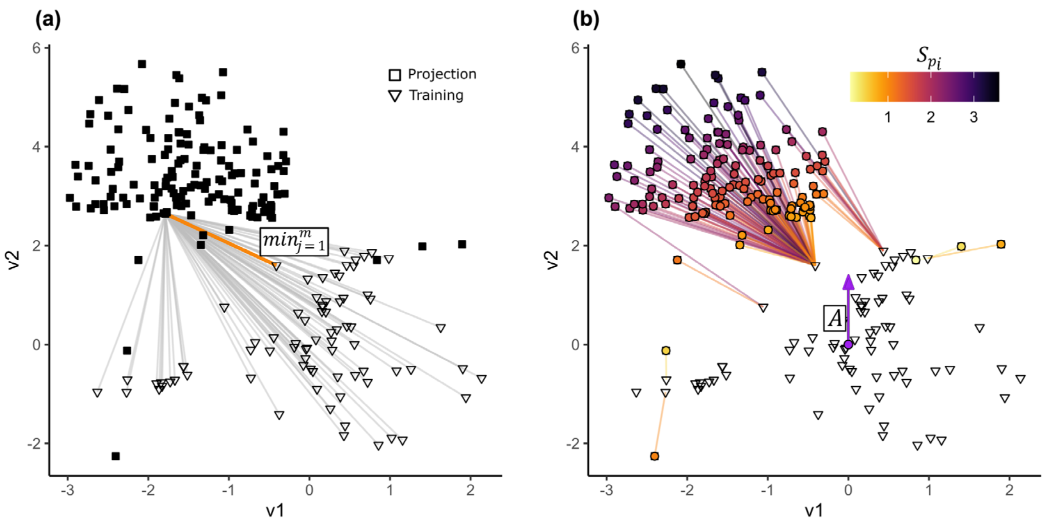


Figure 2. Illustration of Shape procedure to calculate degree of extrapolation in environmental space defined by two hypothetical variables, $v_1$ and $v_2$. In this example, $v_1$ and $v_1$ are the predictor variables. (A) For each projection data point, $p_i$, the Mahalanobis distances to training data, $t_j$, are calculated (gray lines), and the minimum of these distances is selected (orange line). (B) After calculating all minimum distances from projection to training data (lines colored from yellow to black), the Shape metric ($S_{pi}$) for each of the $r$ projection data points is derived as the ratio of the minimum distance to the training data for that point and a factor $A$ (depicted as the length of the purple arrow). The higher the extrapolation degree (Shape metric), the darker the color of the projection points.

data. Distinct from other extrapolation metrics like MOP, this approach keeps Mahalanobis distances unbounded but nonetheless relativized. Relativizing the extrapolation measure by the dispersion factor $A$ is extremely important because it standardizes how far a projection data point is from the training data. For instance, if a given projection datum has a corresponding Shape value of 100, it is 100 times further away from the nearest training datum than the magnitude represented by the dispersion factor. $A$.

### Properties of Shape in comparison with other approaches

We compared Shape to other metrics used to measure the degree of extrapolation commonly used in SDMs, i.e. Multivariate environmental similarity surface (MESS; Elith et al. 2010), Environmental overlap (EO; Zurell et al. 2012); Mobility-oriented parity (MOP; Owens et al. 2013), Extrapolation detection (EXDET; Mesgaran et al. 2014), and Area of applicability (AOA; Meyer and Pebesma 2021). This last approach, AOA, has not been as widely used in SDM; however, its similarity to the Shape metric and warrants comparison.

Exploring and comparing different extrapolation metrics in a simplified environmental space defined by two hypothetical variables shows that the 'Shape' metric measures the degree of extrapolation by following the shape of the training data in environmental space because it considers each training observation as a reference for its calculation (Eq. 2), hence its name (Fig. 3A). This way of calculating the degree of extrapolation addresses some of the limitations of previous extrapolation detection approaches. MESS limits extrapolation similarly to Bioclim, where the degree of extrapolation is measured using rectilinear envelopes (Fig. 3B), and conditions outside the rectilinear envelope encompassing the range of training conditions are considered non-analog (Fig. 4B, Elith et al. 2010). MESS also takes the centroid of the training data as a reference to calculate the degree of extrapolation, rather than each training data point as in Shape; consequently, lower extrapolation values (higher MESS values) will be near this centroid (Fig. 3B, Elith et al. 2010), which can be a poor reflection of the real environmental distance between the training and all projection data (Fig. 3B). Conversely, MOP measures the degree of combinatorial extrapolation by taking a portion of the training data as a reference, partially solving the problem of using centroids as a reference. However, MOP keeps the maximum extrapolation value of 0 (strict extrapolation) to limit SDM projections outside training condition ranges, making it impossible to evaluate extrapolation degree beyond the training conditions (Fig. 3D, Owens et al. 2013). EO divides environmental space into a certain number of bins based on the training data; then, projection data within (EO = 0) and outside (EO = 1) the unique combination of environmental predictor values are delimited as analogous and non-analog, respectively (Zurell et al. 2012). EO has the advantage over MESS and MOP in that it discriminates non-analog conditions based not only on the environmental range of predictors, but also on their combinations (Fig. 3C, 4C, Zurell et al. 2012). However, EO does not continuously measure the degree of extrapolation (Fig. 3C). EXDET measures combinatorial and univariate degree of extrapolation; for the former (i.e. projection data inside a rectilinear envelope), EXDET uses Mahalanobis distance between the projection data and the centroid of the training data (Mesgaran et al. 2014); consequently, it accounts for multicollinearity but still relies on the centroid and therefore suffers from the same limitation as MESS. For measuring univariate extrapolation (i.e. outside the training condition), EXDET uses Euclidean distance; however, it relies on the minimum and maximum limits of the training data, i.e. a rectilinear envelope, to measure extrapolation (Fig. 3E, Mesgaran et al. 2014, Bouchet et al. 2020). An approach called Area of applicability (AOA; Meyer and Pebesma 2021) was published recently (during the development of our Shape method). Despite some similarities between Shape and AOA (Fig. 2F), they differ in how the degree of extrapolation is calculated and the criteria for selecting an extrapolation threshold beyond which model projections are not made or considered to be unreliable (Supporting information). For instance, AOA weights the variables based on their importance to a given model, which could distort the distances between training and projection data and render them dependent on model type and parametrization, making this method impractical when using many algorithms or an ensemble approach.

The threshold extrapolation value used to truncate model predictions is another difference between Shape and other approaches (Fig. 4). Although MESS, EO, MOP and EXDET are calculated with different metrics, all of them truncate model predictions in any area with non-analog conditions (Fig. 4B–D). Of those three, EXDET alone measures the degree of combinatorial extrapolation; however, it will define as non-analog any projection data with extrapolation values > 1 by following the shape of the ellipse defined by the Mahalanobis distance to the training data centroid (Fig. 4D). AOA captures the shape of the training data space in a similar way to Shape (Fig. 4E); however, in AOA, the definition of non-analog data is affected by the environmental distance between training and testing data, and consequently, it is sensitive to the partition approach used for model validation (e.g. bootstrap, k-fold, or geographically structured partition method; Meyer and Pebesma 2021).

Extrapolation measures should be independent of the range of environmental conditions of the projection data (e.g. different environmental variable limits are expected for a geographic region restricted to one country than for others encompassing the whole world). This important property produces the same extrapolation metric values regardless of the projection conditions. Most approaches are not sensitive to the changing ranges of projection data, i.e. the metrics values do not change with changes in range of projection data (Supporting information). However, MOP, which standardizes the Euclidean distance to be bounded between 0 and 1, renders projection data more or less non-analog based on changes in the range of environmental variables (Supporting information).

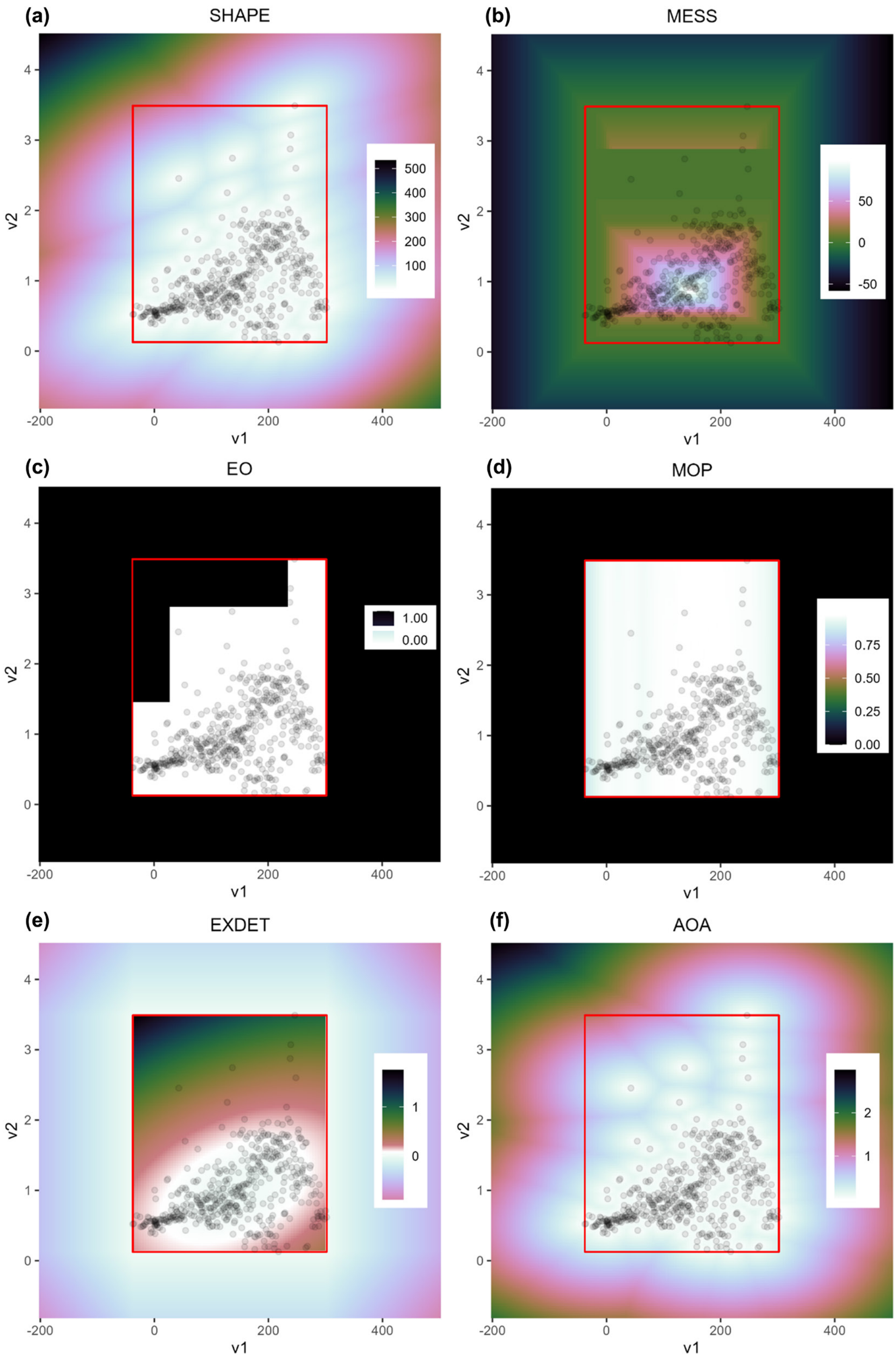


Figure 3. Patterns of different extrapolation metrics in environmental space defined by two hypothetical variables, $v_1$ and $v_2$. Black points and red boxes depict hypothetical training data and limits of training conditions, respectively. The higher the Shape, EXDET, EO and AOA values, the higher the degree of extrapolation. The lower the MOP and MESS, the higher the degree of extrapolation. AOA was calculated using the importance values measured in a random forest model.

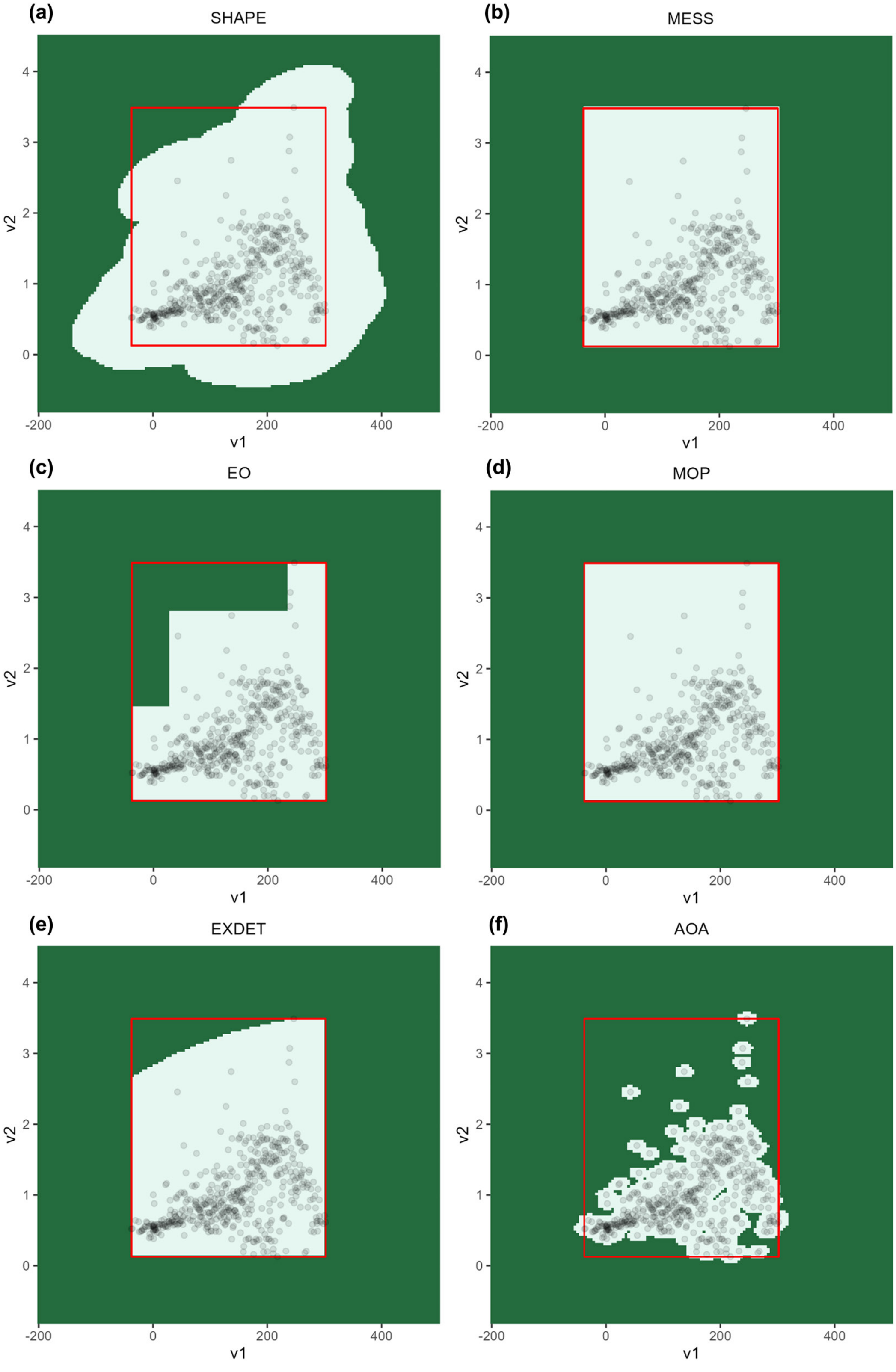


Figure 4. Binary non-analog (green) and analog (white) environmental condition patterns defined by thresholds applied to different extrapolation metrics in the environmental space of two hypothetical variables, $v_1$ and $v_2$. Black points and red boxes depict the training data and limits of training conditions, respectively. The extrapolation threshold used for Shape is 100. Spatial block cross-validation defined the AOA extrapolation threshold. The other methods assign areas outside the training range as non-analog and truncate SDM predictions regardless of metric value.

## Testing the performance of extrapolation metrics

To compare Shape's performance based on Mahalanobis and Euclidean distances and to existing extrapolation detection methods, we used a virtual species approach (Zurell et al. 2010) (see the Supporting information for details about how the Shape metric based on Euclidean distance is calculated). We tested extrapolation metrics under different modeling conditions determined by species niche tolerance, distribution equilibrium condition, sample size (number of presences), and algorithm; all of these affect the ability of an SDM to estimate species niche and, consequently, its susceptibility to extrapolation uncertainty (Table 1).

The niche tolerance (or niche breadth) represents the range of environmental conditions or resources that define the species niche and limit species range sizes. Therefore, species with broader niche tolerance often have larger distribution sizes than species with narrow tolerance (Slatyer et al. 2013). Models are expected to predict narrow niches better than broader ones (Connor et al. 2018, Andrade et al. 2019). The distributional equilibrium condition describes the extent to which a species inhabits (or not) all suitable areas (Peterson et al. 2011). Non-equilibrium distributions will pervasively affect the models' ability to reconstruct species niches because those conditions tend to underestimate niches (Peterson et al. 2018). In the same way, sample size directly affects the predictive performance of SDM; large samples tend to fit better models than small ones (Wisz et al. 2008, Gaul et al. 2020). The algorithm used is one of the primary sources of uncertainty in SDM (Dormann et al. 2008, Thuiller et al. 2019). Different algorithms predict distinct suitability patterns (Fig. 1), directly affecting the suitability pattern near or further from the environmental condition used for model training. In this experiment, we used six different SDM algorithms to estimate suitability: generalized additive model (GAM), generalized linear model (GLM), Gaussian process (GP), maximum entropy (Maxent), random forest (RF) and support vector machine (SVM). Niches were modeled by presence-only SDM, i.e. we used presences sampled from virtual species and pseudo–absences. We decided to use pseudo-absences instead of real absences because this data type is frequently used in SDMs (Guillera-Arroita et al. 2015; further model information in the Supporting information)

Table 1. Combination of factor levels used in the experiment to test the performance of different extrapolation metrics. GAM: generalized additive model; GLM: generalized linear model; GP: Gaussian process; Maxent: maximum entropy; RF: random forest; SVM: support vector machine.

| Niche tolerance | Distribution condition | Sample size | Final number of species | Algorithm |
|---|---|---|---|---|
| Broad | Equilibrium | 100 | 100 | GAM, GLM, GP, Maxent, RF and SVM |
| | | 20 | 100 | |
| | Non-equilibrium | 100 | 96 | |
| | | 20 | 100 | |
| Narrow | Equilibrium | 100 | 92 | |
| | | 20 | 97 | |
| | Non-equilibrium | 100 | 84 | |
| | | 20 | 91 | |

We used a factorial experimental design where all levels of evaluated factors were combined; therefore, each extrapolation metric was tested by 4560 SDMs (760 virtual species × 6 algorithms; Table 1). We used the same virtual species as Andrade et al. (2019) because their approach generates species distributions using realistic stochastic processes based on dispersal simulation and population dynamics at the cell level (see the Supporting information and Andrade et al. 2019 for details about the virtual species approach).

Extrapolation values in environmental and geographical spaces can be used to explore the relationship between degree of extrapolation and suitability values. However, it is common to use extrapolation metrics to truncate model predictions by assuming an extrapolation threshold beyond which projection data is assumed to be unacceptably extrapolative (Fig. 4), and suitability values below that threshold are set to zero (Thuiller et al. 2004, Stohlgren et al. 2011, Owens et al. 2013, Montti et al. 2021). We use truncation to measure how modeled performance can be improved using different extrapolation metrics. The improvement in SDM prediction was measured by root mean square error (RMSE) using worldwide cells ($n = 584\,521$) and comparing the values of the known niche of each virtual species with suitability predicted by SDMs before and after model truncation (i.e. projection data assumed to be unacceptably extrapolative were assigned a suitability value of 0). The lower the RMSE value, the better the method performance (Supporting information).

The extrapolation distance threshold used to truncate a model varies depending on the modeling approach: MESS $<= 0$, $EO = 1$, $MOP = 0$, and EXDET combinatorial extrapolation $>= 1$ and EXDET univariate extrapolation $< 0$ are assumed to be unacceptably extrapolative projection data (Fig. 3–4). The AOA metric weights variables by their importance in a given model to calculate degree of extrapolation; therefore, we extracted the variables' importance from each algorithm for weighting. The AOA extrapolation threshold is derived by identifying the maximum dissimilarity of the training data via cross-validation (Meyer and Pebesma 2021). Because this threshold selection method is sensitive to the validation partitioning approach, and following the recommendation of Meyer and Pebesma (2021), we used spatial cross-validation (band and block). The Shape approach does not assume a default threshold, so we tested values ranging from 20 to 300%. The higher the threshold values, the less restrictive a model is. Exploring a range of Shape thresholds combined with broad modeling conditions defined in our experiment is important to determining when lower or higher threshold values are needed and how our new approach performs in comparison with other methods.

## Data analysis

We used two approaches to analyze the RMSE arising from the experiment results. First, using generalized additive models for location, scale and shape (GAMLSS; Rigby et al.

2019), we performed a post hoc test comparing mean performance among extrapolation metrics under different modeling conditions for each algorithm separately (Supporting information). We also counted the number of cases in which a given extrapolation metric had the top performance (i.e. the lowest RMSE) for each combination of the niche tolerance, distribution condition, and sample size for each algorithm separately to detect which method performed best most frequently under different modeling conditions.

We used the R packages 'ecospat' (www.r-project.org, Di Cola et al. 2017), 'kuenm' (Cobos et al. 2019), 'dsmextra' (Bouchet et al. 2020), and 'CAST' (Meyer et al. 2022) for calculating MESS, MOP, EXDET, AOA, respectively. Codes available in Zurell et al. (2012) were used to calculate EO. We implemented codes for calculating the Shape metric in 'flexsdm' R package (Velazco et al. 2022). The 'flexsdm' package was used to construct SDMs, and the 'caret' (Kuhn 2008) package was used to extract variables' importance. GAMLSS models were fitted with 'gamlss' (Rigby and Stasinopoulos 2005) and 'emmeans' (Lenth 2022) and 'multcomp' (Hothorn et al. 2008) were used for the post hoc analysis.

## Results

In most cases, the Shape metric based on Mahalanobis and Euclidean distance performed similarly in terms of mean RMSE (Supporting information). The performance of Shape based on both distances varied regarding the algorithm, niche tolerance, and distribution conditions (Supporting information). However, because Mahalanobis distance has the advantage of incorporating correlation structure into the metrics (by the covariance matrix) and accounting for the scaling of variables, we found it more appropriate than Euclidean distance. Therefore, in the following results, we reported comparisons between Shape based on Mahalanobis distance and the other extrapolation metrics.

Although all the SDMs were constructed with the same environmental variables used to define species niches, and despite the simplicity of niche shapes, all models had trouble predicting species niches (mean RMSE for algorithms, niche tolerances, distribution conditions, and sample sizes was $0.248 \pm 0.182$, Supporting information). Modeling conditions characterized by broad niches, non-equilibrium distribution, and few occurrences showed the worst performance for most algorithms (Supporting information). However, we found a substantial improvement in model predictions (i.e. mean RMSE for algorithms, niche tolerances, distribution conditions, sample sizes, and extrapolation metrics was $0.119 \pm 0.086$) when model projections were truncated, independent of the extrapolation metric used (Fig. 5, Supporting information). MESS, EO, MOP and EXDET showed similar improvement for all algorithms, while AOA performed similarly to or worse than those three metrics (Fig. 5, Supporting information). Shape performance was dependent on the degree of extrapolation threshold used to truncate models. Because of this flexibility, our approach showed similar or better performance than the other approaches for some threshold values and can better deal with all modeling conditions and algorithms (Fig. 5, Supporting information).

The frequency of the highest performance of each extrapolation metric (i.e. the lowest RSME) shows that Shape generally outperforms other methods independently of algorithm (Fig. 6, Supporting information). However, Shape's performance depends on niche tolerance and threshold value; we found that lower threshold values are preferable when modeling species with narrower niches, and higher extrapolation thresholds are effective for broader niches (Fig. 6, Supporting information). Even higher extrapolation thresholds improved models for species with both broad niches and non-equilibrium distributions (Fig. 6, Supporting information). Whereas AOA methods resulted in the lowest RMSEs of all the extrapolation metrics when Maxent was applied to narrow niches (Supporting information).

## Discussion

Our novel extrapolation metric Shape is flexible enough to be used under different modeling situations because it considers the shape of training conditions when determining the degree of extrapolation to novel conditions. It is also possible to use any extrapolation threshold with Shape to define whether projection data are considered to be unacceptably extrapolative and model projections unreliable by flagging or truncating model predictions beyond the threshold. These two properties make our approach more broadly applicable than existing extrapolation metrics applied to SDM projections.

Frequently, non-analog environmental conditions, under which model projections are considered unreliable, are defined in a binary way using values of environmental predictors in the projection data that are beyond the range of the training data (Elith et al. 2010, Peterson et al. 2011, Briscoe et al. 2019). However, environmental novelty is a continuous phenomenon with a complex shape occurring any time a model is projected with data distinct from training data (Yates et al. 2018). As seen here, some metrics assume straight boundaries as limits of the training conditions (e.g. MESS, MOP, EXDET; Fig. 4), offering too simplistic a solution for controlling model extrapolation. For instance, the degree of extrapolation of projection data within rectilinear training conditions could be as great as projection data outside those limits (Supporting information). In the field of SDM, limits based on the range of the training data could be more problematic in cases where a species niche extends outside the training conditions because occurrences define only a portion of the fundamental niche (e.g. undersampled species or biased occurrence data), a portion of the fundamental niche is available for a species in the geographical space for a given time (i.e. existing niche), or as a consequence of niche truncation (Owens et al. 2013, Soberón and Arroyo-Peña 2017, Qiao et al. 2019, Sales et al. 2019). Therefore, it is important to measure the degree of extrapolation continuously and independently of the rectilinear limits of training

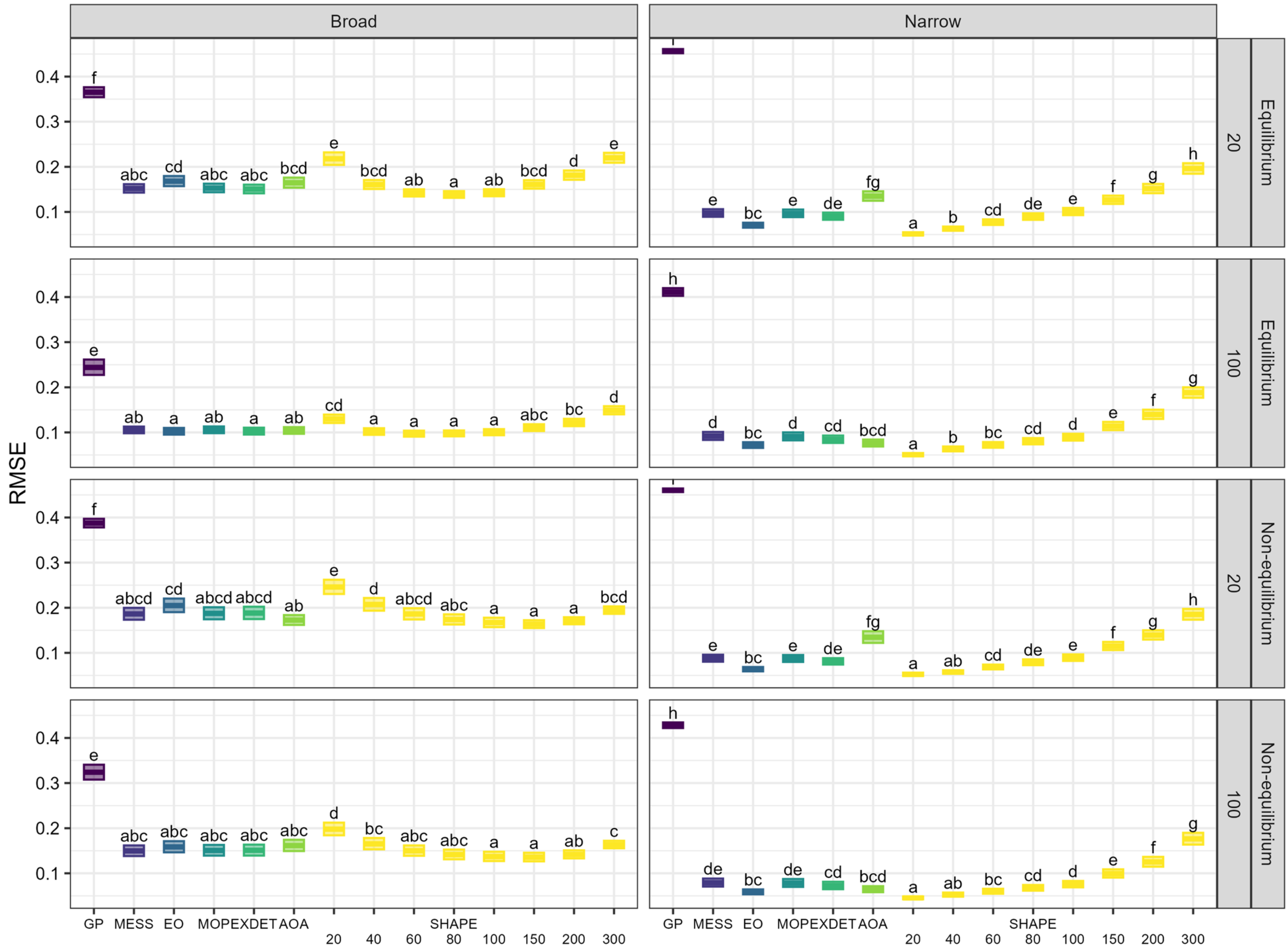


Figure 5. Average RMSE and confidence intervals for Gaussian process (GP) predictions and the effect of extrapolation metrics for different niche tolerances (Broad and Narrow), distribution conditions (Equilibrium and Non-equilibrium), and sample sizes (20 and 100), and for different percent thresholds for the Shape metric (from 20 to 300). Average RMSEs with same letters within each panel are not significantly different according to the HDS Tukey test ($p < 0.05$; see results for other SDM algorithms in the Supporting information).

data; this approach is consistent with Meyer and Pebesma (2021); therefore, we advise species distribution modelers to use extrapolation metrics like Shape or AOA.

Unpredictable model behavior in novel projection conditions affects SDM outcomes, highlighting that extrapolation is a severe problem that should be evaluated and addressed. We found that most algorithms examined, except GLM and GP, show a higher error when modeling species with broader niches, few occurrences, and at non-equilibrium (Supporting information). Such findings align with previous research showing that model transferability was worst for species with larger geographic distributions (Rousseau and Betts 2022), and that model accuracy increases with sample size (Gaul et al. 2020). All algorithms used in our experiment are frequently employed in SDMs and commonly have good model performance (e.g. Maxent, RF, SVM; Elith et al. 2006, Lorena et al. 2011, Qiao et al. 2015, Norberg et al. 2019, Valavi et al. 2022). However, we found that they all have large extrapolation errors, thus failing to predict species niches when projected globally. It has recently been found that most SDM algorithms have low transferability when evaluated in non-analog environmental conditions (Jiménez et al. 2019, Norberg et al. 2019, Qiao et al. 2019, Charney et al. 2021, Rousseau and Betts 2022). We highlight that our experiment imposed very favorable conditions for modeling, such as the simplicity of niche shape (i.e. multivariate Gaussian distribution) and using the same predictors for defining species niches and model fitting. Nevertheless, under real modeling conditions, species could have niche shapes that are different from a multivariate normal distribution (Austin and Gaywood 1994), and, in most cases, modelers simply infer which environmental variables affect species distributions. These typical modeling conditions could prevent the SDM from correctly describing the species niches and, consequently, predict unrealistically under novel projection conditions. We recommend new and expanded experiments using the virtual species approach to understand which niche features and model workflow


Figure 6. Frequency of the highest performance (i.e. the lowest RMSE across all metrics) for each extrapolation metric applied to random forests (RF) projections under different niche tolerances (Broad and Narrow), distribution conditions (Equilibrium and Non-equilibrium), and sample sizes (20 and 100), and for different percent thresholds for the Shape metric (from 20 to 300). Colors refer to extrapolation metric as indicated in the x-axis labels. See other algorithm results in the Supporting information.

decisions most hamper the SDMs predictive ability in non-analog conditions (e.g. training area size or number of pseudo-absences).

Extrapolation metrics can be used to explore the relationship between the degree of model extrapolation and predicted suitability in geographical or environmental space to reveal the degree of environmental novelty of projection data. This exploratory use is essential to limit the interpretation of a model to non-analog environmental conditions. However, extrapolation distance metrics can be binarized based on a threshold chosen to define where model projections are assumed to be unreliable and used for model truncation. Model truncation is extremely important when habitat suitability projected by SDMs is used in spatial decision support (e.g. for calculating proportion of invasive species ranges within and outside a protected area, or for performing spatial conservation prioritization with future conditions). Model truncation can help reduce the effect of extrapolation in problematic scenarios where extrapolation leads to an increase in model overprediction that would change an assessment result. All previous SDM extrapolation metrics use a default threshold to limit extrapolation (Elith et al. 2010, Engler and Rödder 2012, Zurell et al. 2012, Owens et al. 2013, Mesgaran et al. 2014). Our experiment showed that a flexible extrapolation distance threshold is necessary to deal with different modeling conditions. Based on our experiment, we can recommend that lower extrapolation thresholds are preferable for species with narrow niches and higher extrapolation thresholds for species with broad niches and in non-equilibrium (Fig. 5, Supporting information). Further research is necessary to develop new methods other than simple model truncation using an extrapolation distance threshold, such as down-weighting suitability values with higher extrapolation values.

While a flexible extrapolation distance threshold is an advantage of Shape, the choice of a threshold is subjective. Therefore, we recommend that modelers thoroughly explore the patterns of suitability and degree of extrapolation

Table 2. Tools for measuring and exploring extrapolation distance patterns for species distribution models available as a function in 'flexsdm' R package (https://sjevelazco.github.io/flexsdm/articles/v06_Extrapolation_example.html, Velazco et al. 2022).

| Function | Description |
|---|---|
| extra_eval() | measure extrapolation distance based on Shape method by comparing environmental data used for model training and projection conditions. It also calculates univariate and combinatorial extrapolation |
| extra_truncate() | truncate suitability predictions based on one or more extrapolation values |
| p_pdp() | partial dependence plot (a.k.a. response curves) |
| p_bpdp() | bBivariate partial dependence plot (i.e. bivariate response curves) |
| p_extra() | plot extrapolation distance in geographical and environmental space |

in geographical and environmental space to decide which threshold best fits a given model by 1) measuring extrapolation with a metric that captures the shape of training data realistically in environmental space (e.g. Shape and AOA), 2) constructing univariate or bivariate partial dependence plots to inspect algorithm behavior on average conditions (Fig. 1), 3) exploring degree of extrapolation and suitability pattern in environmental and geographical space, 4) exploring the relationship between extrapolation distance and suitability and 5) whenever possible, combining such explorations with knowledge of the organism's ecology (e.g. altitude range, environmental condition tolerances, detection of regions with environmental conditions unlikely for species existence). We provide tools in the 'flexsdm' R package (www.r-project.org) to facilitate such exploration (Table 2, Velazco et al. 2022). Figure 7 shows extrapolation distance patterns of projection data measured with Shape using training data in geographic space and in environmental space for pair-wise combination of variables used in model construction, illustrated for a virtual species dataset in a portion of California, USA.

In summary, we have introduced the novel method Shape to measure the degree of extrapolation and explored its properties relative to other previously published methods. Based on a virtual species experiment, we showed that Shape has similar or better performance than other extrapolation metrics and, due to its versatility, this new metric can be used in different modeling scenarios commonly encountered in SDM applications. We recommend using Shape when a model is projected for other time periods and geographical regions than the training data, or even in current environmental conditions and large training areas. We hope our method and tools support modelers to explore, detect, and reduce uncertainty in extrapolation to achieve more reliable results.

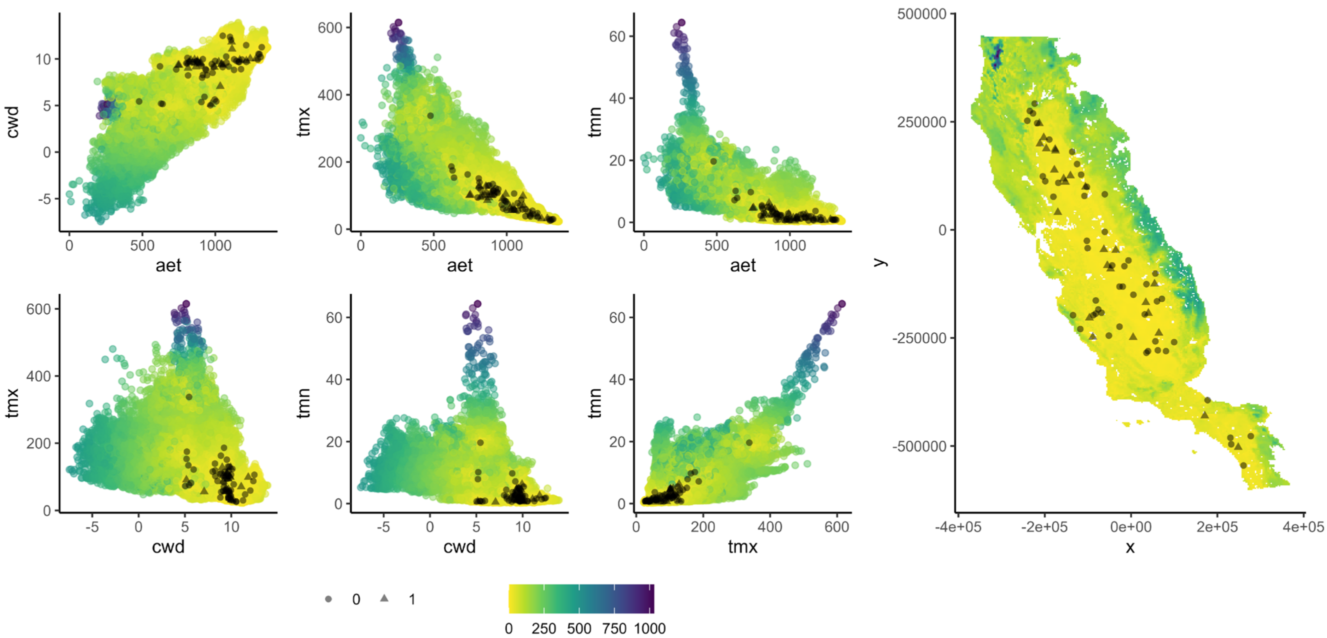


Figure 7. Example of extrapolation distance patterns of projection data measured with Shape (colors range between yellow and blue) using training data (circles and triangles depict absences and presences, respectively) in environmental (left six figures) and geographic (right) space where environmental space is shown for pair-wise combination of variables used in model construction, illustrated for a species dataset in a portion of California, USA. Geographical coordinates are NAD83 datum. Environmental variables shown are actual evapotranspiration (AET), climatic water deficit (CWD), maximum temperature of the warmest month (TMX), and minimum temperature of the coldest month (TMN). Plots were produced using the functions p_extra described in Table 2.

*Acknowledgements* – We acknowledge André F. A. de Andrade for providing the virtual species used in the experiment and Natalia A. Bedrij, Ignacio Mignoli, Giovani C. Andrella and Felipe M. Barbosa for reviewing and commenting on the draft of this manuscript.

*Funding* – SJEV, MBR, HMR and JF were supported by the National Science Foundation, USA (Award 1853697 to HMR and JF). PDM is supported by a productivity grant from CNPq, Brazil (310547/2020-2).

### Author contributions

**Santiago José Elías Velazco**: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **M. Brooke Rose**: Methodology (supporting); Software (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Paulo De Marco Jr.**: Conceptualization (supporting); Methodology (supporting); Writing – review and editing (supporting). **Helen M. Regan**: Funding acquisition (lead); Writing – original draft (supporting); Writing – review and editing (supporting). **Janet Franklin**: Conceptualization (supporting); Funding acquisition (lead); Investigation (supporting); Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting).

### Transparent peer review

The peer review history for this article is available at https://publons.com/publon/10.1111/ecog.06992.

### Data availability statement

Data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.r2280gbk5 (Velazco et al. 2023).

### Supporting information

The Supporting information associated with this article is available with the online version.

## References

Andrade, A. F. A., Velazco, S. J. E. and De Marco, P. 2019. Niche mismatches can impair our ability to predict potential invasions. – Biol. Invas. 21: 3135–3150.

Aniceto, N., Freitas, A. A., Bender, A. and Ghafourian, T. 2016. A novel applicability domain technique for mapping predictive reliability across the chemical space of a QSAR: reliability-density neighbourhood. – J. Cheminform. 8: 69.

Araújo, M. B. and Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modeling. – Ecology 93: 1527–1539.

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E. and Rahbek, C. 2019. Standards for distribution models in biodiversity assessments. – Sci. Adv. 5: eaat4858.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecol. Modell. 157: 101–118.

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

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

Bouchet, P. J., Miller, D. L., Roberts, J. J., Mannocci, L., Harris, C. M. and Thomas, L. 2020. dsmextra: extrapolation assessment tools for density surface models. – Methods Ecol. Evol. 11: 1464–1469.

Breiner, F. T., Guisan, A., Bergamini, A. and Nobis, M. P. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. – Methods Ecol. Evol. 6: 1210–1218.

Briscoe, N. J., Elith, J., Salguero-Gómez, R., Lahoz-Monfort, J. J., Camac, J. S., Giljohann, K. M., Holden, M. H., Hradsky, B. A., Kearney, M. R., McMahon, S. M., Phillips, B. L., Regan, T. J., Rhodes, J. R., Vesk, P. A., Wintle, B. A., Yen, J. D. L. and Guillera-Arroita, G. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. – Ecol. Lett. 22: 1940–1956.

Bueno, M. L., Pennington, R. T., Dexter, K. G., Kamino, L. H. Y., Pontara, V., Neves, D. M., Ratter, J. A. and de Oliveira-Filho, A. T. 2016. Effects of Quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. – Ecography 40: 403–414.

Calambás-Trochez, L. F., Velazco, S. J. E., Hoffmann, P. M., Gurski, E. M., Brum, F. T. and Carlucci, M. B. 2021. Climate and land-use changes coupled with low coverage of protected areas threaten palm species in south Brazilian grasslands. – Perspect. Ecol. Conserv. 9.

Charney, N. D., Record, S., Gerstner, B. E., Merow, C., Zarnetske, P. L. and Enquist, B. J. 2021. A test of species distribution model transferability across environmental and geographic space for 108 western North American tree species. – Front. Ecol. Evol. 9: 689295.

Cobos, M. E., Peterson, A. T., Barve, N. and Osorio-Olvera, L. 2019. kuenm: an R package for detailed development of ecological niche models using Maxent. – PeerJ 7: e6281.

Colwell, R. K. and Rangel, T. F. 2009. Hutchinson's duality: the once and future niche. – Proc. Natl Acad. Sci. USA 106: 19651–19658.

Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J., Wang, F. and Liu, J. 2018. Effects of grain size and niche breadth on species distribution modeling. – Ecography 41: 1270–1282.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N. and Guisan, A. 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. – Ecography 40: 774–787.

Dormann, C. F., Purschke, O., García Márquez, J. R., Lautenbach, S. and Schröder, B. 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. – Ecology 89: 3371–3386.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Elith, J., Kearney, M. and Phillips, S. 2010. The art of modelling range-shifting species: the art of modelling range-shifting species. – Methods Ecol. Evol. 1: 330–342.
Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. and Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists: statistical explanation of MaxEnt. – Divers. Distrib. 17: 43–57.
Engler, J. O. and Rödder, D. 2012. Disentangling interpolation and extrapolation uncertainties in ecological niche models: a novel visualization technique for the spatial variation of predictor variable collinearity. – Biodivers. Inform. 8: 30–40.
Fitzpatrick, M. C., Blois, J. L., Williams, J. W., Nieto-Lugilde, D., Maguire, K. C. and Lorenz, D. J. 2018. How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. – Global Change Biol. 24: 3575–3586.
Fois, M., Fenu, G., Cuena Lombraña, A., Cogoni, D. and Bacchetta, G. 2015. A practical method to speed up the discovery of unknown populations using species distribution models. – J. Nat. Conserv. 24: 42–48.
Franklin, J. 2013. Species distribution models in conservation biogeography: developments and challenges. – Divers. Distrib. 19: 1217–1223.
Gaul, W., Sadykova, D., White, H. J., Leon-Sanchez, L., Caplat, P., Emmerson, M. C. and Yearsley, J. M. 2020. Data quantity is more important than its spatial bias for predictive species distribution modelling. – PeerJ 8: e10411.
Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R. and Wintle, B. A. 2015. Is my species distribution model fit for purpose? Matching data and models to applications: matching distribution models to applications. – Global Ecol. Biogeogr. 24: 276–292.
Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.
Hothorn, T., Bretz, F. and Westfall, P. 2008. Simultaneous inference in general parametric models. – Biom. J. 50: 346–363.
Hutchinson, G. E. 1957. Concluding remarks. – Cold Spring Harb. Symp. Quant. Biol. 22: 415–427.
Jiménez, L., Soberón, J., Christen, J. A. and Soto, D. 2019. On the problem of modeling a fundamental niche from occurrence data. – Ecol. Modell. 397: 74–83.
Kuhn, M. 2008. Building predictive models in R using the caret Package. – J. Stat. Softw. 28: 1–26
Lenth, R. V. 2022. emmeans: estimated marginal. – https://CRAN.R-project.org/package=emmeans.
Liu, C., Newell, G. and White, M. 2019. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. – Ecography 42: 535–548.
Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., Giovanni, R. D., Lohmann, L. G., de Carvalho, A. C. P. L. F. and Yamamoto, M. 2011. Comparing machine learning classifiers in potential distribution modelling. – Expert Syst. Appl. 38: 5268–5275.
Mahony, C. R., Cannon, A. J., Wang, T. and Aitken, S. N. 2017. A closer look at novel climates: new methods and insights at continental to landscape scales. – Global Change Biol. 23: 3934–3955.
Merow, C., Wilson, A. M. and Jetz, W. 2017. Integrating occurrence data and expert maps for improved species range predictions: expert maps & point process models. – Global Ecol. Biogeogr. 26: 243–258.
Mesgaran, M. B., Cousens, R. D. and 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.
Meyer, H. and Pebesma, E. 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. – Methods Ecol. Evol. 12: 1620–1633.
Meyer, H.,Milà, C.,Ludwig, M. and Linnenbrink, J. 2022. CAST: "caret" applications for spatial-temporal models. – https://CRAN.R-project.org/package=CAST.
Montti, L., Velazco, S. J. E., Travis, J. M. J. and Grau, H. R. 2021. Predicting current and future global distribution of invasive *Ligustrum lucidum* W. T. Aiton: assessing emerging risks to biodiversity hotspots. – Divers. Distrib. 27: 1568–1583.
Norberg, A. et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. – Ecol. Monogr. 89: e01370.
Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E. and Peterson, A. T. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. – Ecol. Modell. 263: 10–18.
Peterson, A. T. and Soberón, J. 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. – Nat. Conserv. 10: 102–107.
Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. and Bastos Araújo, M. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.
Peterson, A. T., Cobos, M. E. and Jiménez-García, D. 2018. Major challenges for correlational ecological niche model projections to future climate conditions: climate change, ecological niche models, and uncertainty. – Ann. N. Y. Acad. Sci. 1429: 66–77.
Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Modell. 190: 231–259.
Qiao, H., Soberón, J. and Peterson, A. T. 2015. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. – Methods Ecol. Evol. 6: 1126–1136.
Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G. and Papeş, M. 2019. An evaluation of transferability of ecological niche models. – Ecography 42: 521–534.
Regan, H. M., Syphard, A. D., Franklin, J., Swab, R. M., Markovchick, L., Flint, A. L., Flint, L. E. and Zedler, P. H. 2012. Evaluation of assisted colonization strategies under global change for a rare, fire-dependent plant. – Global Change Biol. 18: 936–947.
Rigby, R. A. and Stasinopoulos, D. M. 2005. Generalized additive models for location, scale and shape. – Appl. Stat. 54: 507–554.
Rigby, R. A. et al. 2019. Distributions for modeling location, scale, and shape: using gamlss in R. – Chapman and Hall/CRC.
Rousseau, J. S. and Betts, M. G. 2022. Factors influencing transferability in species distribution models. – Ecography 2022: e06060.
Sales, L. P., Ribeiro, B. R., Pires, M. M., Chapman, C. A. and Loyola, R. 2019. Recalculating route: dispersal constraints will drive the redistribution of Amazon primates in the Anthropocene. – Ecography 42: 1789–1801.
Slatyer, R. A., Hirst, M. and Sexton, J. P. 2013. Niche breadth predicts geographical range size: a general ecological pattern. – Ecol. Lett. 16: 1104–1114.
Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. – Biodivers. Inform. 2: 1–10.

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

Stohlgren, T. J., Jarnevich, C. S., Esaias, W. E. and Morisette, J. T. 2011. Bounding species distribution models. – Curr. Zool. 57: 642–647.

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

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N. and Zimmermann, N. E. 2019. Uncertainty in ensembles of global biodiversity scenarios. – Nat. Commun. 10: 1446.

Valavi, R., Elith, J., Lahoz-Monfort, J. J. and Guillera-Arroita, G. 2019. blockCV : an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. – Methods Ecol. Evol. 10: 225–232.

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

Velazco, S. J. E., Rose, M. B., de Andrade, A. F. A., Minoli, I. and Franklin, J. 2022. flexsdm: an R package for supporting a comprehensive and flexible species distribution modelling workflow. – Methods Ecol. Evol. 13: 1661–1669.

Velazco, S. J. E., Brooke Rose, M., De Marco Jr., P., Regan, H. M. and Franklin, J.. 2023. Data from: How far can I extrapolate my species distribution model? Exploring shape, a novel method. – Dryad Digital Repository, https://doi.org/10.5061/dryad.r2280gbk5.

Williams, J. W., Jackson, S. T. and Kutzbach, J. E. 2007. Projected distributions of novel and disappearing climates by 2100 AD. – Proc. Natl Acad. Sci. USA 104: 5738–5742.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H. and Guisan, A. 2008. Effects of sample size on the performance of species distribution models. – Divers. Distrib. 14: 763–773.

Yates, K. L. et al. 2018. Outstanding challenges in the transferability of ecological models. – Trends Ecol. Evol. 33: 790–802.

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. and Grimm, V. 2010. The virtual ecologist approach: simulating data and observers. – Oikos 119: 622–635.

Zurell, D., Elith, J. and Schröder, B. 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions: predicting to new environments. – Divers. Distrib. 18: 628–634.