

The influence of the number and distribution of background points in presence-background species distribution models

Anna M. Whitford^a, Benjamin R. Shipley^{b,*}, Jenny L. McGuire^{b,c,d}

^a H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

^b School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

^c School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

^d Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Original content: The characteristics of 6 simulated species used in this research, along with their generated IUCN ranges (Original data)

Keywords:

Species ranges
Model evaluation
North America
Transferability
Boyce index
MaxEnt
Simulation

ABSTRACT

Species distribution models (SDMs), which relate recorded observations (presences) and absences or background points to environmental characteristics, are powerful tools used to generate hypotheses about the biogeography, ecology, and conservation of species. Although many researchers have examined the effects of presence and background point distributions on model outputs, they have not systematically evaluated the effects of various methods of background point sampling on the performance of a single model algorithm across many species. Therefore, a consensus on the preferred methods of background point sampling is lacking. Here, we conducted presence-background SDMs for 20 vertebrate species in North America under a variety of background point conditions, varying the number of background points used, the size of the buffer used to constrain the background points around the occurrences, and the percentage of background points sampled within the buffer ("spatial weighting"). We evaluated the accuracy and transferability of the models using Boyce index, overlap with expert-generated range maps, and area overpredicted and underpredicted by the SDM (and AUC for comparability with other studies).

SDM performance is highly dependent on the species modelled but is affected by the number and spread of background points. Models with little spatial weighting had high accuracy (overlap values), but extreme extrapolation errors and overprediction. In contrast, SDMs with high transferability (high Boyce index values and low overprediction) had moderate-to-high spatial weighting. These results emphasize the importance of both background points and evaluation metric selection in SDMs. For other, more successful metrics, using many background points with spatial weighting may be preferred for models with large extents. These results can assist researchers in selecting the background point parameters most relevant for their research question, allowing them to fine-tune their hypotheses on the distribution of species through space and time.

1. Introduction

The geographic range of a species is a crucial aspect of its ecology, reflecting how the species interacts with landscapes, climates, and biological communities (Borges et al., 2019; Holt, 2003; Sonne et al., 2016). Understanding, modelling, and predicting species ranges are integral to our understanding of community ecology, biogeography, and ecosystem functioning (Elith et al., 2010). Species Distribution Models (or SDMs) are a key innovation used in the evaluation and prediction of species ranges. SDMs relate species location data (known presences and/or known absences) to information about the environmental characteristics at those locations (Elith and Leathwick, 2009). High performance

computing and easily accessible data have led to the rapid proliferation of SDM methods and techniques, in part due to the development of easy-to-use programs that conduct such modelling (for example, MaxEnt; Phillips et al., 2006, biomod; Thuiller et al., 2009, ecospat; Di Cola et al., 2017).

In particular, presence-background SDMs are widely used in ecology and conservation. Instead of requiring true absence data, which are challenging to acquire and often highly inaccurate (Jarnevich et al., 2015; Lobo et al., 2010), presence-background SDMs only require occurrence points and a set of generated background or pseudo-absence points (but see Sillero and Barbosa, 2021 on the distinction between the two). Often, these background points may sample the climates

* Corresponding author.

E-mail address: benranier@gmail.com (B.R. Shipley).

<https://doi.org/10.1016/j.ecolmodel.2023.110604>

Received 6 June 2023; Received in revised form 30 November 2023; Accepted 20 December 2023

Available online 27 December 2023

0304-3800/© 2023 Elsevier B.V. All rights reserved.

“available” to a species or the climate throughout the study area of interest (Grimmett et al., 2020). Presence-background SDMs are increasingly used to verify the current distribution of a species on a landscape (e.g., Stirling et al., 2016), to assess and highlight areas of particular conservation importance (e.g., Sarkar et al., 2009), to predict species’ future ranges under dynamic change (e.g., Schloss et al., 2012), and to understand the basic ecological drivers of species ranges through time (e.g., Fenker et al., 2020), amongst many other topics.

Because of the preponderance of research applying presence-background SDMs (Cobos et al., 2019), many researchers have examined the effects of varying model inputs on model accuracy (The ability of a model to distinguish between areas of high and low habitat suitability in the training domain; Elith et al., 2010; Phillips et al., 2009; Radosavljevic and Anderson, 2014) and transferability (the ability of a model trained in one place or time to predict distributions in a different place or time; Iturbide et al., 2018; Rousseau and Betts, 2022). Much effort has been dedicated to evaluating the effects of occurrence and background points on presence-background models, finding that the density and arrangement of both occurrence and background points affects model results and transferability (Barbet-Massin et al., 2012; Chefaoui and Lobo, 2008; Fourcade et al., 2014; Iturbide et al., 2018, 2015; Liu et al., 2019; Wisz et al., 2008; Grimmett et al., 2020). For example, spatial and environmental bias in occurrence points including those downloaded from online repositories like GBIF may compromise the ability of an SDM to fully capture either the fundamental or realized niche of a species (Boakes et al., 2010; Fourcade et al., 2013; Ruiz-Gutiérrez and Zipkin, 2011; Varela et al., 2014; Yackulic et al., 2013). Spatially and environmentally subsampling occurrence points before modelling has quickly become standard practice for mitigating these biases (Castellanos et al., 2019; Varela et al., 2014).

In contrast to occurrence points, which are usually provided to the SDM (either from online repositories or from field observations), background points are usually generated and assembled *de novo* in presence-background models. Because of this, the published literature employs many different, often contradictory, methods of background point generation (Barbet-Massin et al., 2012). For example, the number of background points varies widely across SDM research, despite their clear influences on model results (Fourcade et al., 2014). Some research uses “large number” strategies, selecting a large, arbitrary number of background points to generate (e.g., Bennett et al., 2019 uses 10,000). In contrast, other studies (especially those modelling multiple species) use a “multiplier” strategy, where the number of background points is tied to the number of occurrence points based on an (often small) multiplied coefficient (e.g., 10 times the number of occurrences, Chefaoui and Lobo, 2008). These varying methods exist despite research indicating that the number of background points used significantly affects the accuracy of presence-background SDMs (Barbet-Massin et al., 2012; Chefaoui and Lobo, 2008; Grimmett et al., 2020; Lobo et al., 2010; Phillips et al., 2009). The spatial extent of the training area (i.e., where background points are sampled from) similarly varies across SDM studies. Often, the training extent of an SDM is defined by political, geographic, or ecological regions, like the Cerrado in South America (Fenker et al., 2020) or Great Britain (Holloway et al., 2016). Other studies instead limit SDMs to match the expected range of the species of interest, using buffers of variable widths around the species occurrences or a bounding polygon (Stirling et al., 2016; VanDerWal et al., 2009).

Finally, species occurrences are often spatially biased towards areas that are easily accessible by humans (Kramer-Schadt et al., 2013; Phillips et al., 2009). Therefore, randomly sampling background points across the entire study region may provide inaccurate estimates of habitat suitability, especially in fragmented landscapes (Ruiz-Gutiérrez and Zipkin, 2011). Environmentally-subsampling the occurrences and background points reduces the spatial bias inherent in the occurrences (Castellanos et al., 2019; Varela et al., 2014), as does sampling background points closer to the occurrences, e.g., from within a buffer around them (Barve et al., 2011; Fourcade et al., 2014; M. Iturbide et al.,

2018; Lobo et al., 2010). However, spatially-constrained background points are susceptible to overfitting and extrapolation, especially when transferring an SDM trained on a small spatial extent to a larger one (Radosavljevic and Anderson, 2014). In fact, other researchers have employed the opposite strategy, sampling background points exclusively from *outside* the buffer around the occurrences (e.g., the SRE method in the “biomod2” R package Thuiller et al., 2009). To avoid these extrapolation issues, Shipley et al. (2022) developed a new strategy (hereafter called “spatial weighting”), whereby a certain percentage of background points are sampled from within a buffer, and the rest are sampled from across the entire study region. Anecdotal evidence suggests that spatially weighting background points leads to better-performing models, but this hypothesis has not yet been formally tested (Shipley et al., 2022).

Because each researcher has their own preferred methods of generating background points, little research has compared different methods of background point sampling on real study data using the same modelling parameters and occurrence point sampling. As a result, a strong consensus has yet to be reached on the best practices of background point sampling, especially when conducting multi-taxa research across large geographical extents (Barbet-Massin et al., 2012; Fourcade et al., 2014; M. Iturbide et al., 2015; Lobo et al., 2010; Machado-Stredel et al., 2021; Senay et al., 2013).

In this study, we evaluate different methods of background point selection, examining how the number of background points, the size and shape of the geographic area available to a species, and the spatial weighting affect presence-background SDM results. We systematically vary these model inputs and examine their effects on the modelled ranges of 20 vertebrate species and 6 simulated species across North America (Table 1). These methods, introduced in the R package megaSDM (Shipley et al., 2022), allow for increased flexibility in the degree of spatial bias of the background points but have not yet been comprehensively examined in the existing literature. Then, we evaluate the accuracy of these models using multiple quantitative metrics and examine the background point conditions under which presence-background models perform best.

We expect that sampling background points randomly across the study region will over-predict species’ presences near the occurrence points (see Chefaoui and Lobo, 2008), but we expect that limiting the sample of the background point to buffers around the occurrence points will lead to over-prediction in areas farther from the occurrence points. Fourcade et al. (2014) suggests that the area where the background points are selected should be biologically meaningful. We hypothesize this will occur with small to moderate sized buffers, as most species are unable to access the entire study region of North America. Finally, following the results of Barbet-Massin et al. (2012), we hypothesize that the best performing models will have a large number of background points.

2. Methods

2.1. Data collection

All analyses were conducted using R v. 4.0.5 (R Core Team, 2021), using the package megaSDM (Shipley et al., 2022). megaSDM uses the MaxEnt algorithm (from the Java script software provided by Phillips et al., 2004) to model species distributions across a variety of dynamic conditions. This package is efficient at handling many species, time periods, and use cases simultaneously, and it natively incorporates environmental subsampling of both occurrences and background points as a way to partially mitigate spatial and environmental bias (Shipley et al., 2022). For a diagram of our modelling workflow, see Fig. 1. A short glossary of the data required for species distribution modelling with megaSDM is provided in Supplementary Appendix 1.

We selected 20 species to model from sets of species that had been grouped according to geographic range size, taxonomy, and ecology (Table 1). We obtained species observation data from GBIF ([2](http://</p>
</div>
<div data-bbox=)

Table 1

List of species used in the analysis, along with the number of occurrence points that remained after vetting, the taxonomic class of the species, and the size of the IUCN range polygon in km². Species are ordered according to the number of occurrence points after environmental subsampling.

Species	Common Name	Occurrence Points	Class	Range Size (km ²)
<i>Canis latrans</i>	Coyote	5327	Mammal	9,560,176
<i>Procyon lotor</i>	Raccoon	5038	Mammal	8,647,235
<i>Trachemys scripta</i>	Pond slider	2776	Reptile	2,853,727
<i>Anaxyrus fowleri</i>	Fowler's toad	2423	Amphibian	2,013,427
<i>Pantherophis alleghaniensis</i>	Eastern rat snake	1922	Reptile	769,274
<i>Ensatina eschscholtzii</i>	Ensatina	1646	Amphibian	322,190
<i>Kinostemon subrubrum</i>	Eastern mud turtle	1312	Reptile	1,699,848
<i>Cryptotis parva</i>	North American least shrew	840	Mammal	3,591,180
<i>Scaphiopus holbrookii</i>	Eastern spadefoot	707	Amphibian	1,276,964
<i>Malaclemys terrapin</i>	Diamondback terrapin	497	Reptile	302,835
<i>Desmognathus quadramaculatus</i>	Blackbelly salamander	496	Amphibian	84,675
<i>Necturus maculosus</i>	Common mudpuppy	419	Amphibian	1,880,503
<i>Deirochelys reticularia</i>	Chicken turtle	416	Reptile	999,813
<i>Zapus trinotatus</i>	Pacific jumping mouse	342	Mammal	162,699
<i>Microtus townsendii</i>	Townsend's vole	235	Mammal	192,001
<i>Thomomys monticola</i>	Mountain pocket gopher	173	Mammal	52,769
<i>Sceloporus woodi</i>	Florida scrub lizard	140	Reptile	77,269
<i>Callospermophilus saturatus</i>	Cascade golden-mantled ground squirrel	84	Mammal	73,997
<i>Arborimus pomo</i>	Sonoma tree vole	60	Mammal	21,852
<i>Plethodon kentucki</i>	Cumberland Plateau salamander	36	Amphibian	29,151

www.gbif.org) and manually vetted the occurrences for geographical and observational accuracy. For a detailed description of the vetting procedures, see Supplementary Appendix 7. We limited the training extent of our study to North America (−152° to −58° longitude, 7° to 68° latitude), and projected the models to the extent of the continental United States (−125° to −65° longitude, 25° to 50° latitude). We chose to standardize the extent of the study region in order to compare models across species and to determine the effects of background point methods on both model accuracy and transferability when decoupled from study extent (Sarkar et al., 2009). However, because the extent of the study area may influence model evaluation (Lobo et al., 2008; VanDerWal et al., 2009), we additionally modelled the species with varying background extents based on bounding boxes around each of the background buffers used in the models. We then compared the results of these varying-extent models with the fixed-extent models (see *Varying-Extent Models* in the Results and Supplementary Appendix 4).

We used four bioclimatic variables obtained from WorldClim (Fick and Hijmans, 2017) at 2.5-minute resolution, relating to the mean and variance of temperature and precipitation across North America: Mean Annual Temperature (Bio1), Mean Annual Precipitation (Bio12), Temperature Annual Range (Bio7), Coefficient of Precipitation Seasonality (Bio15). We limited our modelling to these four climatic variables because of their simple biological and climatic interpretations and to avoid the effect of predictor collinearity on model transferability (Feng et al., 2019). The temperature annual range and mean annual temperature were correlated (Kendall's $\tau = -0.62$), because of the cold, highly seasonal climates of central and northern Canada (which were only included for training the models). All other variables were only moderately correlated to each other (Kendall's $\tau < 0.41$). Other environmental variables are likely to be influential in certain species distributions (Williams et al., 2012; Petitpierre et al., 2017; but see Bucklin et al., 2015; Braunisch et al., 2013). However, we held these four climatic variables constant across all species so that we can compare the models made from different species without introducing a potential masking variable and to the reduce the chances of over-parameterization for any one model. We accounted for the differing effects of environment on each species after generating the SDMs (see *Species Normalization*).

Occurrence points are often spatially and environmentally biased because of study design and non-random sampling (Fourcade et al., 2014). These observational biases may in turn lead to biased or inaccurate models of species distributions (Yackulic et al., 2013). To limit

environmental and spatial biases in the occurrence points, we used the methods developed by Varela et al. (2014) and expanded by Castellanos et al. (2019), which environmentally filters occurrence data into a set number of bins, and removes points that are environmentally clustered. Models constructed using environmental subsampling often outperform those that do not (Castellanos et al., 2019; Fourcade et al., 2014; Varela et al., 2014). We used 25 bins per environmental variable for the environmental filtering.

In addition to the real observations of the 20 North American vertebrate species, we simulated 6 virtual species to evaluate the various background point methods. The virtual species were simulated using the “virtualspecies” package in R (Leroy et al., 2016), using the same four bioclimatic variables we include in the real-species models to create response curves and to identify areas of suitability. These virtual species occurrences are intrinsically absent the spatial, environmental, and sampling biases that come with real occurrences, and therefore can provide additional information on the performance of the different techniques of background point sampling we examine. Although these occurrence points are not environmentally biased, we environmentally subsampled the virtual species following the same methods as the real species for consistency. See the supplementary data on FigShare (<https://figshare.com/account/items/24680163>) for the characteristics of each virtual species.

2.2. Background points and modelling

We manipulated three variables governing the prevalence and distribution of background points in the model: the number of background points, the size of the buffer within which background points may be generated (hereafter, buffer size), and a new variable describing the proportion of background points sampled from within the buffer (hereafter, spatial weighting). Each variable had five treatment levels (see Table 2), and for each pairwise combination of the three variables ($n = 125$), we generated sets of background points for each species.

2.3. Number of background points

The first variable we manipulated was the number of background points used in the model. Despite the evident effects of background point sample size on SDM performance, species distribution modelers continue to use a wide variety of background point numbers and densities. We evaluated five different treatments for the number of

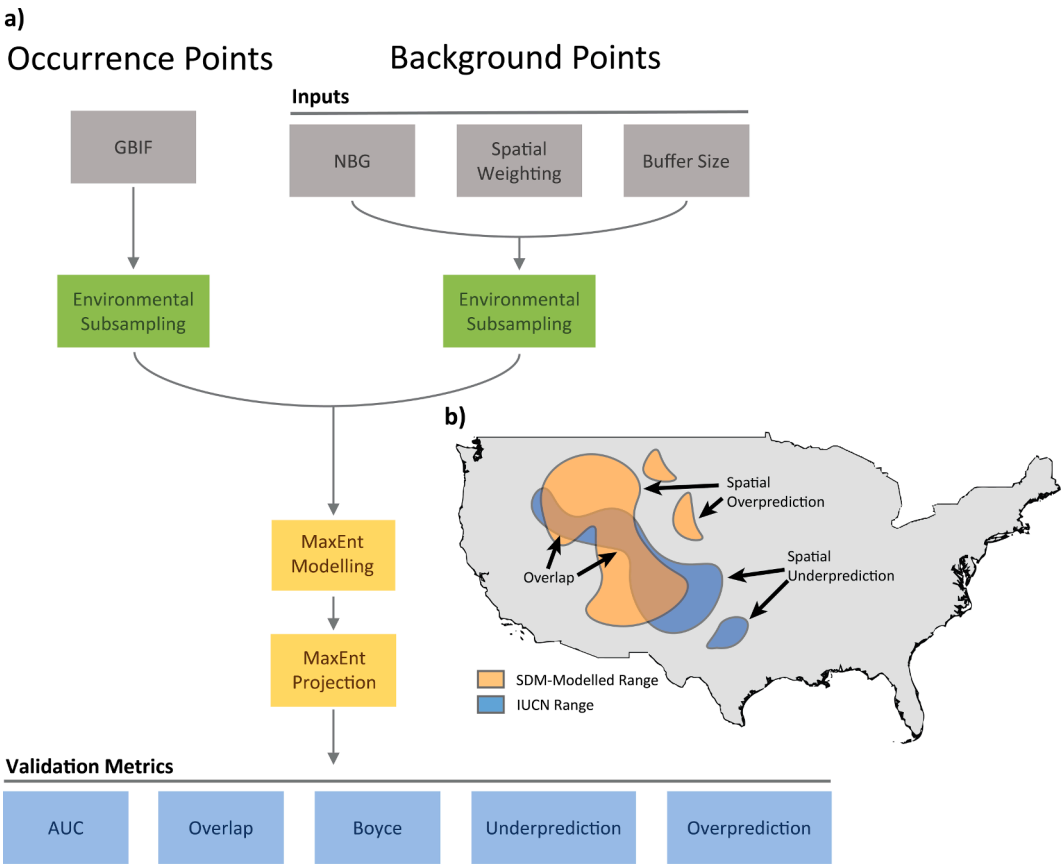


Fig. 1. a) A diagram of our workflow, including the inputs we varied to create different background point distributions (Table 2), the environmental subsampling to reduce spatial bias in the occurrence and background points (Castellanos et al., 2019; Varela et al., 2014), and the five evaluation metrics we used after generating the model (Table 3). b) Conceptual figure describing the differences between the three evaluation metrics we used that compare the modelled species distribution to a species range based by expert opinion (IUCN, 2020). Spatial overprediction is the area predicted by the SDM but not covered by the expert range. Spatial overprediction is the area predicted by the expert range but not covered by the SDM. Overlap is the area covered by both the SDM and expert range. Percent overlap is calculated relative to the expert range (Overlap area / IUCN range area) and is therefore mathematically related to spatial underprediction.

Table 2
Table of the background point parameters we varied to generate sets of background points to be used in the SDMs. A description of the parameter modified, the values tested, and hypothesized “preferred values” of the parameters for high model performance are provided.

Name	Description	Values Tested	Preferred Values
Number of Background Points	The number of pseudo-randomly generated background points generated by megaSDM and used in model evaluation	1000, 5000, 10,000, 1 * number of occurrences, 5 * number of occurrences	Varies (Barbet-Massin et al., 2012, etc.)
Buffer Size	Size of the buffer from which background points may be preferentially sampled. Buffer size must be ≥ 0 and may be any positive number.	100 km, 200 km, 95% quantile of nearest neighbour distance, IUCN range, WWF ecoregion boundary	Not recommended to use very small (close to 0) or very large buffers (Barve et al., 2011; Fourcade et al., 2014).
Spatial Weighting	Value between 0 and 1 determining the proportion of background points sampled within the buffer.	0, 0.25, 0.5, 0.75, 1	Unknown (Shipley et al., 2022)

background points: 5000, 10,000, $1 \times$ (the number of occurrences), $5 \times$ (the number of occurrences), and $10 \times$ (the number of occurrences). These five treatments were chosen to examine the impact of the number of background points on the resulting model. The two fixed-number treatments (5000, 10,000) examine how the absolute number of background points impacts model performance (Barbet-Massin et al., 2012). The three variable-number treatments (1x, 5x, 10x) examine the impact of using background points proportional to the number of occurrences (i. e., varying prevalence) at a low, medium, and high number of background points (Chefaoui and Lobo, 2008).

2.4. Buffer size

The size of the buffer from which the spatially-weighted background points are sampled influences the range of environments a species could

live in but does not. Often, an arbitrary, fixed-width buffer is selected for background point generation (e.g., 100 or 500 km). However, the habitat available to a given species depends on the ecology of the species itself, and therefore the ideal buffer width may vary according to the species modelled. In this study, we compared two fixed-width buffers (widths of 100 and 200 km) to three novel methods that vary the size and shape of the buffer based on the species itself. First, we used megaSDM’s functionality to tailor the buffer to the density of occurrences for each species. The width of the buffer was set proportionally to the 95% quantile of the nearest neighbour distance for each point (Shipley et al., 2022). Therefore, species with more widely distributed occurrences that were observed more sparsely across the landscape will have a wider background buffer than those with observations tightly clustered in space.

In addition to the fixed-width and variable buffer sizes, we used two

techniques that incorporate ecologically defined buffers for our models. In theory, the spatially constrained background points should define an “accessible area” for the species—where a species could theoretically disperse but does not (Barve et al., 2011; Machado-Stredel et al., 2021). These areas are predominantly defined by their ecology, and as such using biologically defined regions to constrain background points may lead to a more accurate assessment of this accessible area. Using World Wildlife Fund ecoregions (Olson et al., 2001) we developed a new method that generates background buffers from the combined extent of all ecoregions that contained at least one occurrence point of the species. For our final buffer treatment, we used the IUCN expert-generated range of the species (IUCN, 2020). IUCN ranges are known to overestimate the extent of occupancy of a species, especially when compared to survey data (Hurlbert and Jetz, 2007; Hurlbert and White, 2005), but they estimate the species’ extent of occupancy and the area potentially accessible to a species (Barve et al., 2011; Broekman et al., 2022). Therefore, they may provide a suitable extent from which to generate background points. For the six virtual species, we generated fake IUCN ranges by simplifying the presence-absence rasters to 50-km resolution and vectorizing them (see supplementary data on FigShare).

2.5. Spatial weighting

To avoid overfitting and over-extrapolation from spatially-weighted background points, some researchers might choose to sample background points from both within and outside of an occurrence point buffer, creating a version of an *a priori* bias grid (Phillips et al., 2009). Spatial weighting is a technique to preferentially bias background points inside of a given buffer, without exclusively sampling from that area (Shipley et al., 2022). In this study, spatial weighting ranges from 0 to 1, where 0 has the background points distributed completely randomly and 1 exclusively samples the background points from within a given buffer. Numbers between 0 and 1 describe the proportion of background points sampled exclusively within the buffer (as opposed to being randomly distributed across the whole study area). We evaluated five different spatial weighting proportions: 0, 0.25, 0.5, 0.75, and 1, to test how the varying levels of spatial bias in background points affects the model performance.

Once we generated the background point sets for each species and treatment combination, we environmentally subsampled each set of background points in the same way as the occurrence points. Using each of the background point sets for each species, we modelled the species distributions using four replicates in MaxEnt, randomly subsampling the training occurrence points each time to 80% of the original data and evaluating the model on the remaining 20%. We set the regularization to 1 for all replicates of all models; although optimizing the regularization parameter for each species leads to better performing individual models (Merow et al., 2013), we kept it constant to more effectively compare model performance across species.

2.6. Model evaluation

How to accurately evaluate the quality of a species distribution model is subject to ongoing discussion, with a variety of different evaluation methods used (e.g., Leroy et al., 2018). No one evaluation method perfectly describes a model’s fidelity to actual biogeographic patterns. Therefore, to capture a variety of model behaviors and to allow comparability with a variety of other studies, we used five distinct evaluation metrics.

AUC (area under the receiver operating curve) values were long considered the standard for characterizing SDM, but are nevertheless unreliable indicators of a model’s accuracy, especially when evaluating presence-background models (Lobo et al., 2008). Ranging from 0 to 1, the AUC evaluates the sensitivity (percent of correctly predicted presences) and specificity (percent of correctly predicted absences) of the model (Table 3). However, the receiver operating curve was designed for true negatives instead of background points and therefore provides undue weight to specificity when, in reality, a species cannot be definitively absent from a background point (Lobo et al., 2008). In addition, AUC values are influenced by the geographical extent of the model and the proportion of background points to presence points (Lobo et al., 2008; Yackulic et al., 2013). Therefore, despite their ubiquity in the presence-background SDM literature, AUC values do not adequately confer information about the transferability of the model, nor does it measure overfitting in areas outside of the training domain. In order to compare our results with the previous literature (most of which uses AUC), we calculated and averaged the test AUC values (from the 20% of data held back) for each model replicate, but we limit discussion of the results of the models when evaluated via AUC to Supplementary Appendix 3.

One common alternative for presence-background models is the **Boyce index**, which measures the ratio of predicted presences to expected presences, based on the continuous habitat suitability generated by the model (Boyce et al., 2002; Hirzel et al., 2006). In contrast to AUC, the Boyce index is not dependent on absences and may therefore be less strongly influenced by the number and distribution of background points than AUC values are (but see Jiménez and Soberón, 2020). Furthermore, the Boyce index measures the model’s habitat suitability after it has projected into geographic space, making it an appropriate method for evaluating the model’s transferability outside of its training extent (e.g., Petitpierre et al., 2017).

The Boyce index measures how well the model predicts presences and absences given the data provided to it. However, because we are interested in not only the discriminant ability of the model within its training extent, but also its transferability in geographic space, we considered three additional evaluation methods that directly compare the SDMs and the expert-generated range maps from IUCN. These range maps are coarse-grained and likely overestimate species occupancy (Hurlbert and Jetz, 2007), but are less influenced by spatial and climatic bias than species observations, especially when the species has not been extensively sampled (Fourcade et al., 2013). In fact, Broekman et al. (2022) compared IUCN ranges with the GPS-tracked movements of 49

Table 3
Table of the accuracy metrics used to evaluate the generated SDMs. A short description of the metric and which values indicate higher-performing models are also provided. Although the results for AUC values are not reported in the paper itself, we include AUC here for completeness.

Name	Description	Preferred Values
AUC Value	The area under the receiver operating curve. Ranges from 0 to 1, where 0.5 is a model that is no better than random and 1 is a perfectly discriminant model.	Larger values. AUC values ≥ 0.7 are generally accepted as “good” models (e.g., González-Ferreras et al., 2016)
Boyce Index	The correlation between observed and expected frequencies of occurrence points based on area, ranging from -1 to 1 where 0 indicates a random model (Hirzel et al., 2006).	Larger values
Spatial Overprediction	Total area (km ²) of the SDM-generated range that falls outside of the bounds of the IUCN range map (see Fig. 1)	Smaller values
Spatial Underprediction	Total area (km ²) inside the IUCN range that the SDM range does not occupy (see Fig. 1)	Smaller values
Percent Overlap	Proportion of the area within the SDM generated model that overlaps the IUCN range	Larger values (close to 1)

terrestrial mammal species, finding high correspondence between the two estimates of species occupancy. Therefore, expert ranges may be suitable for evaluating model predictions. Following similar methods as Fourcade et al. (2013), we constructed binary presence-absence maps from the continuous habitat suitability maps generated from megaSDM, setting the threshold for the suitability values as the “test sensitivity plus specificity” value. This threshold was chosen because of its insensitivity to occurrence prevalence and its high sensitivity in common species (Liu et al., 2015). We then aggregated the four replicates to create a “consensus” distribution map.

After constructing the binary distribution maps, we overlaid them with the IUCN expert generated range maps. From this overlay, we determined the percent of the IUCN range that overlaps with the modelled range (% overlap, used in Fourcade et al., 2013; Duan et al., 2022), in addition to two metrics evaluating the mismatch between the IUCN and SDM ranges (**spatial underprediction and overprediction**; Fig. 1, Table 3). Spatial overprediction is the area included in the projected SDM range that does not overlap the IUCN expert generated range map (i.e., a false positive, Fig. 1). In contrast, spatial underprediction highlights the area included in the IUCN range map not included in the modelled SDM range (i.e., a false negative, Fig. 1). To account for the relationship between species range size and underprediction (see Fig. 2), we additionally calculated the overprediction and underprediction as a proportion of the species IUCN range size (see Supplementary Appendix 6). The % overlap metric measures the accuracy of the model within the expected species range, whereas the spatial overprediction and underprediction metrics provide greater insight into the transferability of the model and its ability to correctly identify areas outside of the species range. To examine how the five evaluation metrics relate to each other, we conducted non-parametric correlations between each of them (Supplementary Appendices 2, 3, 4, 5).

2.7. Species normalization

Each species is influenced by a different suite of climate characteristics, leading to differences in model performance. To account for this, we normalized each evaluation metric by species, centring and scaling the evaluation metrics of each set of models to unit variance. Therefore, instead of an absolute measure of model performance, our evaluation metrics indicate model performance relative to the other models generated for that species. This normalization accounts for the effect of species identity on the evaluation results. It allowed us to find to the combination of background point variables that led to the most accurate models across all species, without being influenced by the specific ecology of the species.

2.8. Random forests and grid graphs

Once we evaluated and standardized the models, we assessed the importance of each of the three background point variables to the model accuracy. We then constructed random forests using the R packages “randomForest” (Liaw and Wiener, 2002) and “caret” (Kuhn, 2008), linking the evaluation metrics one at a time to the background point variables, taxon of the species, and species range size. Using the percent increase in mean squared error (%MSE), we evaluated the importance of each background point variable to the evaluation metric. To further remove the effects of taxon and species range size on the evaluation metrics, we incorporated the taxonomic class and the range size into each random forest model. Finally, we created graphs (henceforth, “grid graphs”; Fig. 4), to show visually how model accuracy was affected by treatment. The graphs show the average quantile of each normalized metric for each species, when one of the variables is held constant – a value of 1 indicates that the variable-treatment combination of interest leads to consistently high-performing models for that species, whereas a value of 0 means that the models consistently create inaccurate and non-transferrable species distributions. We created grid graphs for each

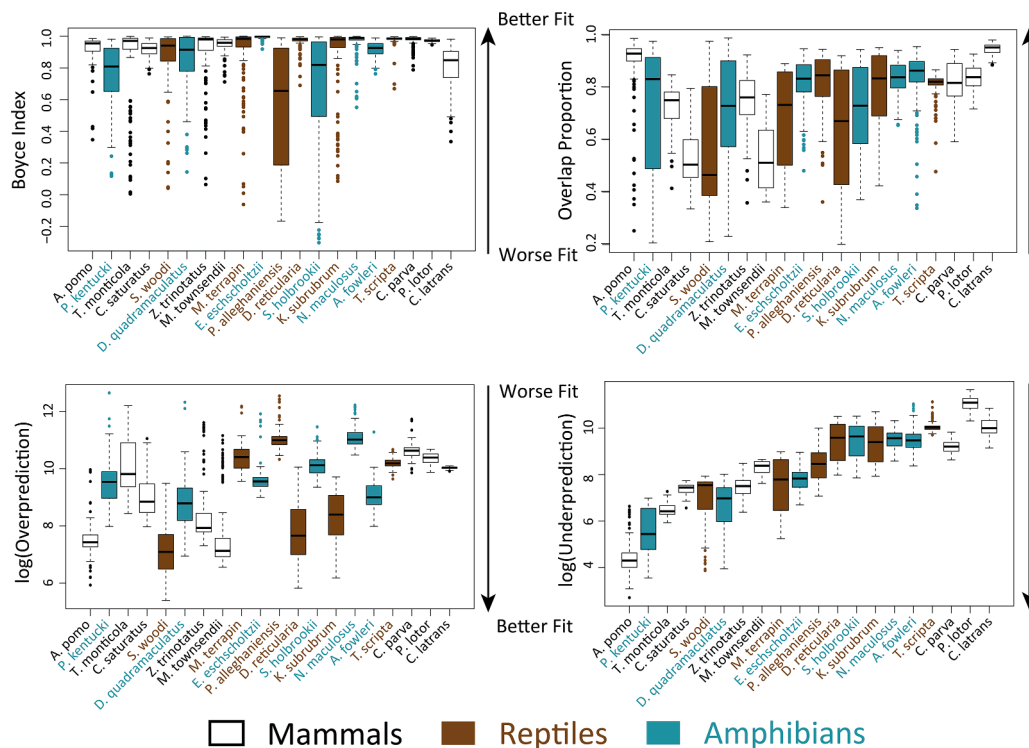


Fig. 2. Boxplots showing the distribution of values for each evaluation metric by species. In total, 125 models were run for each species (5 treatments of 3 variables). Species are arranged from smallest range size (*A. poma*) to largest range size (*C. latrans*). Arrows point towards better-performing models for each evaluation metric (i.e., higher Boyce index and % Overlap; lower spatial underprediction and overprediction).

treatment and variable combination, and for each evaluation metric.

3. Results

Overall, the accuracy of each model was highly influenced by the species modelled. All four evaluation metrics (Boyce index, % overlap, spatial underprediction, and spatial overprediction) varied significantly across species (Kruskal-Wallis tests, all $\chi^2 > 539.4$, $df = 19$, all $p < 0.001$, Fig. 2). Per species, the Boyce index ranged from 0.55 (*P. alleghaniensis*) to 0.99 (*E. eschscholtzii*), the % overlap ranged from 0.54 (*M. townsendii*) to 0.94 (*C. latrans*), the spatial underprediction ranged from 2572 (*A. pomo*) to 1.44×10^6 (*C. saturatus*) km², and the spatial overprediction ranged from 39,694 (*S. woodi*) to 1.62×10^6 (*N. maculosus*) km². Despite the significant influence of species on the average values of the evaluation metrics used, we found no consistent relationships between the absolute range size of a species (calculated from its IUCN range) and Boyce index, % overlap, or spatial overprediction (all Kendall's $\tau \leq \pm 0.3$), although spatial underprediction was positively correlated to absolute range size (Kendall's $\tau = 0.71$, see Fig. 2).

When aggregated across all species and treatments, all correlations between the four metrics used were less than ± 0.20 . However, when we conducted the same correlations for each species individually and averaged them, some relationships strengthened, most notably spatial overprediction and % overlap (from $\tau = -0.01$ to $\tau = 0.38$; Supplementary Appendix 2, also see Fig. 1). In addition, when accounting for species, spatial underprediction was almost perfectly rank-order correlated to % overlap ($\tau = -0.97$). Because we mitigate the potential random effect of species by normalizing each metric by species and because of the strong correlation between species range size and spatial underprediction, we removed spatial underprediction from subsequent analyses.

The random forest models we created to evaluate the relative

importance of each background point variable on model evaluation displayed several general trends. First, for all evaluation metrics, the degree of spatial weighting (how many background points were sampled exclusively from within the buffer) was strongly influential in predicting the overall value of the metrics (Fig. 3). In fact, the removal of spatial weighting led to the highest increase in MSE for all random forest models. The size of the buffer used to constrain background points significantly influenced the predictions of % overlap, and spatial overprediction, but was not useful for predicting the Boyce index values. In fact, for Boyce index, the removal of buffer size as a variable in the random forest model led to a decrease in MSE, suggesting that the number of background points did not provide any new information to the model (Fig. 3). Finally, the number of background points used influenced the values of the Boyce index (for which it was the 2nd most important variable), but had little effect on either overlap or spatial overprediction (Fig. 3).

The grid graphs show the specific background point treatments that lead to over- and under-performing models compared to the median for each species (Fig. 4). We found a clear negative relationship between the degree of spatial weighting and the % overlap values, with the highest overlap values occurring when randomly sampling background points across the entire continent (spatial weighting = 0; Fig. 4). In addition, % overlap was high when using ecoregions as the buffer (Fig. 4). Similar results were found in AUC values (Supplementary Appendix 3). However, when sampling either entirely within the buffer (spatial weighting = 1) or entirely randomly across the continent (spatial weighting = 0), the amount of spatial overprediction was also the greatest. Neither % overlap nor spatial overprediction were greatly affected by the number of background points used, although for larger-ranged species, fewer background points tended to increase overlap and led to spatial overprediction.

Unlike the metrics comparing IUCN ranges to the modelled ranges,

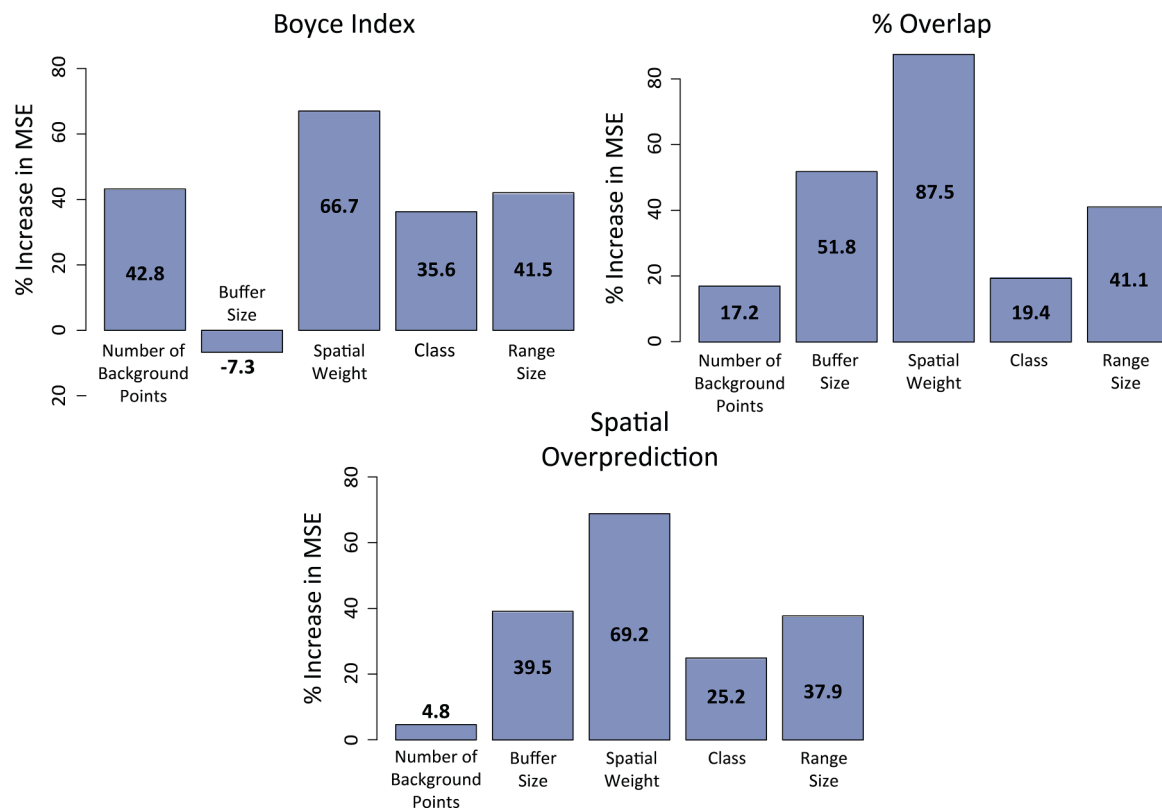


Fig. 3. Barplots describing the importance of each variable (calculated by % increase in mean squared error) in the creation of random forest models for each evaluation metric. In addition to the three background point variables examined (number of background points, buffer size, and spatial weight), the taxonomic class and range size of each species was added to the random forest models as potential confounding factors.

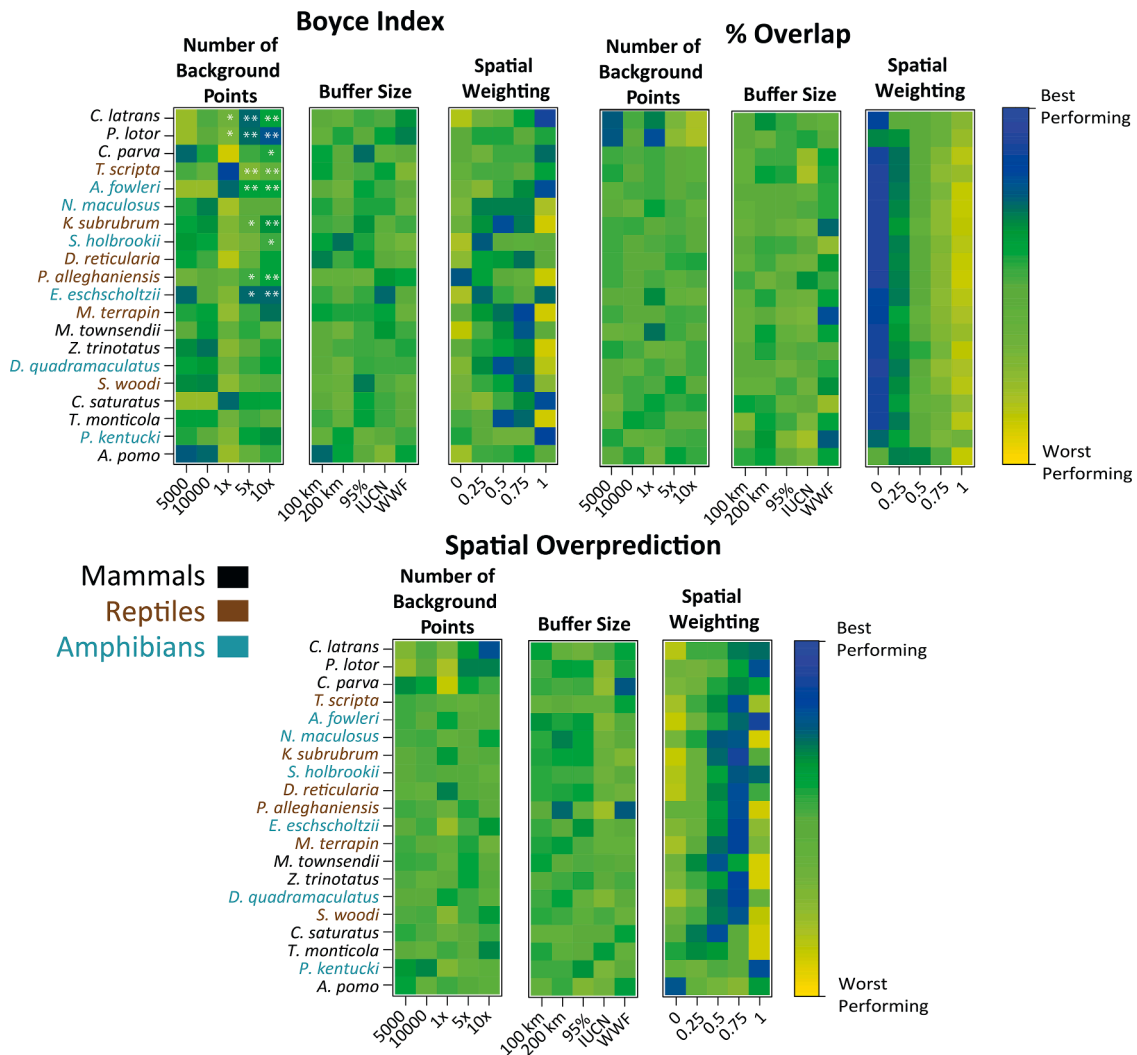


Fig. 4. The average quantiles of each evaluation metric, given each background point treatment. A quantile of 1 means that the models created using the given treatment have on average higher values than the models created using any other treatment. Likewise, a quantile of 0 suggests that the models created using the given treatment have lower values than models using any other treatment. Spatial overprediction values are multiplied by -1 so that darker values indicate better-performing models. Species are arranged from smallest (bottom) to largest (top) range sizes. The white asterisks in the top left plot show at which point the number of background points using the small multiplier technique is greater than 5000 (one asterisk) or 10,000 (two asterisks).

the Boyce index values were smallest with models created with fewer background points and were little affected by the size of the buffers. However, the SDMs in general performed worse (as evaluated by the Boyce index) when background points were sampled exclusively within the buffer (spatial weighting = 1) or randomly throughout the study region (spatial weighting = 0), with the highest Boyce index values in 14 out of 20 cases (70%) occurring when the models were created with a spatial weighting between 0.25 and 0.75 (Fig. 4).

3.1. Varying-Extent models

The results from the fixed-extent models, in which we kept the study extent constant across all species and model treatments, differed in several notable ways from the varying-extent models, in which we used the bounding box around the background buffers as the study extent for each model. First, although the correlations amongst the five evaluation metrics used remained relatively consistent with the fixed-extent correlations (Fig. S2), the relative importance of each treatment variable differed between the two methods. First, spatial weighting had much less influence on measures of accuracy and transferability when study extents were variable than when they were fixed. These changes were

particularly striking for measures of transferability (Boyce index and spatial overprediction), in which spatial weighting dropped from the most to the least important variable (Fig. S6). Instead, the number of background points and the size of the buffer used tended to have greater importance on model transferability (Fig. S6). For example, buffers made with IUCN boundaries and using the 95% distance between occurrences reduced spatial overprediction, but decreased IUCN overlap in varying-extent models (Fig. S7), in contrast to the fixed-extent models (which demonstrate only minor influence of buffer size on either metric). Finally, the tradeoff between spatial weighting, transferability, and accuracy disappeared when using varying extents –% overlap was similarly influenced by spatial weighting in both the fixed and varying-extent models (with lower spatial weighting leading to higher performance), but Boyce index and spatial overprediction demonstrated no relationship with spatial weighting (Fig. S7).

3.2. Virtual species

When using virtual species instead of the real species, we again observed different results, despite similar correlations between the metrics (Fig. S8). The pattern of within-range accuracy (% overlap) was

somewhat consistent to the real-species pattern, with high variable importance for the number of background points and spatial weighting in the random forests (Fig. S10). In addition, as with the true species, the size of the buffer used to sample the background points was influential in modulating the overlap and spatial overprediction of the model (Figs. S9). However, the metrics of transferability showed markedly different relationships with the evaluation metrics. The number of background points were not influential for the Boyce index and spatial overprediction at all, with negative variable importance in the random forests and no observable relationship in the grid graphs (Fig. S9). In contrast, spatial weighting and buffer size had a greater influence on Boyce index and spatial overprediction than the models made from real species (Fig. S9). See Supplementary Appendix 5 for a more detailed discussion of these results.

4. Discussion

Overall, we found significant variation in the three metrics used to evaluate model performance (Boyce index, % overlap, spatial overprediction). The accuracy and transferability of the models varied according to all three inputs we manipulated, including the number of background points, the size of the buffer used to constrain the background points, and the degree of spatial weighting (percent of points sampled from within the buffer). However, the metrics all differed in their responses to changes in the background points (Fig. 4) and the study extent of the model (Supplementary Appendix 4). These results emphasize the importance of background point selection on presence-background SDMs and the measures used to evaluate them, particularly when strong sampling bias exists in the species occurrence records and when species' fundamental niche is not well-estimated.

Across the four metrics used in this study, spatial weighting was on average the most important variable in driving model performance in the fixed-extent models (Fig. 3). However, spatial weighting affected model accuracy and model transferability in different ways. As expected, the measure of within-domain accuracy (% overlap) was higher in models without spatial weighting (i.e., background points sampled from the entire study area) than those where the background points were exclusively sampled from within the buffer (Fig. 4). However, these models generally overpredicted areas of suitable habitat, as evidenced by the higher spatial overprediction and lower Boyce index for

models with no spatial weighting (Fig. 4). In contrast, an intermediate degree of spatial weighting (i.e., between 0.25 and 0.75) led to increased transferability, with low spatial overprediction. This tradeoff between model accuracy and transferability supports previous literature. Chefaoui and Lobo (2008) found that models with more dispersed background points had higher AUC values than those that had spatially restricted background points, but that the models overpredicted the realized niche of the species. Similarly, Thuiller et al. (2004) observed overestimation in models without absences beyond the known species distributions, even while the accuracy of the model (measured using AUC values) remained consistently high. AUC values are dependent only on the accuracy of classifying presences and absences; therefore, if background points are more climatically distant from presence points, the AUC values will be artificially inflated, even as the modelled distribution diverges from reality (Chefaoui and Lobo, 2008; Lobo et al., 2010).

The limitations of relying exclusively on measures of within-domain model accuracy, especially for presence-background models, were further highlighted in the dramatic differences between % overlap and AUC values (high to moderate) and Boyce index (low) in fixed-extent models with high spatial weighting. Exclusively sampling within a buffer may lead to extrapolation in areas with climates not represented by the background. For example, Fig. 5 shows an SDM of *Z. trinotatus* (Pacific jumping mouse) created by exclusively sampling background points within a tight buffer around the occurrences (black oval). The overestimation of habitat suitability in the southeast of the study region is not reflected in the % overlap value, which it is only based on the climatic extent of the presence and background points and not the entire study area. In contrast, the Boyce index is low, reflecting the inability of the model to be transferred across space or time. The discordance between AUC, % overlap, and Boyce index values found throughout this study indicate that the SDM's performance when measured using % overlap and AUC values has little bearing on its transferability across space (also see Petitpierre et al., 2017). As a result, when modelling a small species range within a large extent, our results suggest having a spatial weight between 0 and 1 leads to the most high-performing SDMs. However, if the study extent of the model is only marginally larger than the species range itself, the advantages of sampling outside of the buffer disappear (Supplementary Appendix 4). In those cases, sampling exclusively within a buffer may be preferable (Lobo et al., 2010).

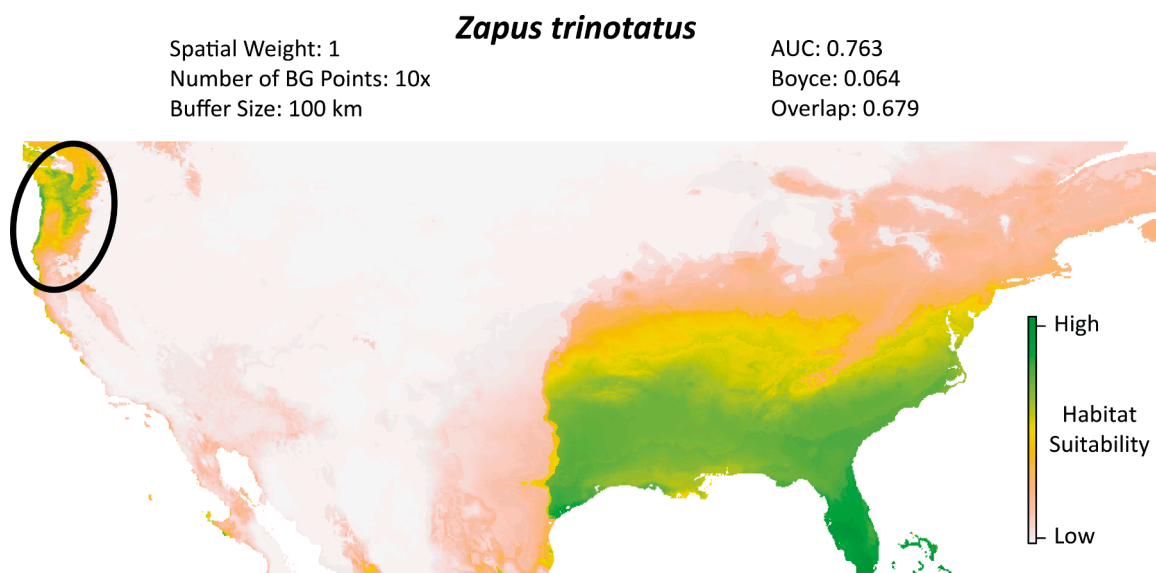


Fig. 5. Habitat suitability map for *Z. trinotatus* (Pacific jumping mouse), using 10 * the number of occurrence points for background points, a buffer of 100 km around each point, and sampling exclusively within the buffer. The true range of the species is contained within the black oval in the upper left, with a large area of overestimation in the bottom right. Despite this clear overestimation, the AUC and overlap values for this model suggests that it is at worst a fair model, in contrast to the Boyce index, which shows it to be little better than random.

In addition to the degree of spatial weighting, the buffer used to weight distance also affected fixed-extent model accuracy, as measured by the % overlap between the IUCN range and the model results. World Wildlife Fund ecoregions had the highest % overlap, in addition to the least amount of spatial overprediction. This suggests that using these ecoregion-based buffers may confer a slight increase in overall model accuracy, providing support for the “available habitat” interpretation of background points (Barve et al., 2011; Fourcade et al., 2014). However, more research into using biologically defined buffers for background point generation is needed to explore the implications in using ecoregion boundaries to model species distributions. In contrast, the number of background points did not substantially affect the discrepancy between the SDM-generated ranges and the IUCN ranges when evaluating % overlap, likely because the threshold we used to define presence-absence maps varies according to the conditions of the model (Fig. 3). However, when the model's extent varied with the climate available to the species, lower prevalence led to higher overlap (Fig. S7). As expected, AUC values were more highly-influenced by background points (with higher numbers of background points leading to larger AUC values), and showed the same preference for WWF ecoregions as the constraining buffer (Supplementary Appendix 3).

The transferability of the fixed-extent models (measured using the Boyce index and spatial overprediction) reacted differently to background inputs than the model accuracy. Aside from spatial weighting, the other background point variables we modified (number of background points, size of the buffer) had comparably little effect on the overall transferability of the models. Therefore, the transferability of these SDMs appears to be predominantly influenced by spatial weighting and relatively robust to other background point parameters.

The models created with study extent tailored to the buffers around each set of species occurrences (varying-extent models) demonstrated the confounding effect of study extent on model performance. Buffer size (which directly corresponded to the study extent of the model) significantly affected model performances, with higher % overlap but more spatial overprediction in models with larger buffers (i.e., 200 km and WWF ecoregion buffers). The results are expected, because using absences that are more geographically (and therefore more environmentally) distant from occurrence points often leads to artificially lower commission errors (Lobo et al., 2008). In addition, despite its clear importance for fixed-extent models, the effect of spatial weighting was nearly absent from the performance of varying-extent models. Therefore, a moderate amount of spatial weighting appears to be useful when extrapolation errors might occur (e.g., modelling a small species range on a large extent), but is not useful for models with a tightly-bounding study extent. Although the results of the fixed-extent models echo those of (Thuiller et al., 2004) in supporting the use of absences well outside the species climate envelope when extrapolating species responses to climate (but see VanDerWal et al., 2009), the performance of the varying-extent models suggests that these absences are not useful when the observed species range fills up much of the study extent.

The clear differences between our results when using real species and those when using virtual species suggests fundamental differences in the assumptions of presence-background modelling when the species completely fills its fundamental niche. First, the virtual species were created with known and definite relationships to the environmental variables modelled, whereas the species used in our analyses have no such *a priori* relationships and are influenced by other factors such as anthropogenic effects, species interactions, and geologic history. The virtual species therefore live everywhere within the study region that is suitable, even if there is a barrier to dispersal or a large swath of uninhabitable area between suitable regions. These patterns lead to models that are better able to estimate the species' “true” distribution in areas without occurrences (i.e., have better transferability), leading to changes in the Boyce index and overprediction patterns. In contrast, we observed less discrepancy between virtual and real species in measures of accuracy (e.g., % overlap), because these measures are broadly

unaffected by the amount of niche filling a species undergoes. In addition, the modelling methods we have employed in this study assume biased environmental sampling. Environmental filtering attempts to mitigate environmental bias, but, in the absence of such a bias, it may in fact introduce bias, by weighting areas of marginal suitability more heavily than areas within the centre of a species' niche. Likewise, the various buffer sizes and spatial weighting used for the background points attempt to account for biased spatial sampling, but for a virtual species, there is no spatial bias. The discrepancy between the results of the real species and virtual species suggest that these methods of background sampling may not be necessary or appropriate to apply in the absence of imperfectly sampled data. Therefore, in the exceedingly rare case of a species with little environmental bias and a well-known fundamental niche, we would recommend eschewing spatial weighting and constraining background points with a buffer altogether. Instead, simply manipulating the number of background points and the spatial extent of the analysis should be the priority for species distribution modelling in those cases.

Different SDM algorithms are likely to respond differently to changes in background point prevalence and distribution (Grimmett et al., 2020). No one algorithm consistently performs the best for every modelling case, and they are each influenced by varying inputs in different ways (Barbet-Massin et al., 2012; Valavi et al., 2022). We chose to use only MaxEnt in this study for a few reasons. First, MaxEnt is shown to be a “generalist” and a “jack of all trades” model, performing consistently above average across many different modelling scenarios (Elith et al., 2010; Giovanelli et al., 2010; Grimmett et al., 2020; Valavi et al., 2022). In addition, the discriminant ability of MaxEnt is somewhat less influenced by number and prevalence of occurrences (Grimmett et al., 2020) and the calibration extent (Giovanelli et al., 2010) than other algorithms like random forests and support vector machines. MaxEnt's relative resilience to occurrence number and extent allows us to more effectively examine how the distribution of background points affects SDMs across multiple species. However, a study of how these same background point changes influence other presence-background algorithms is needed to fully benchmark how background might influence species distribution models as a whole. Given the discrepancy between our results using real species and virtual species, we might expect models that require pseudo-absences (i.e., regression methods like generalized linear or additive models) to react differently to pseudo-absence distribution. For example, Barbet-Massin et al. (2012) evaluated the true skill statistic (TSS) of several SDM algorithms generated with various pseudo-absence point distributions, finding support for a large number of pseudo-absence points and (in the case of classification and machine learning methods) weighting pseudo-absence points farther from the occurrences. We suspect these differences arise from the conceptual differences between background points (which attempt to sample the climate available to the species) and pseudo-absences (which attempt to stand in for real absences), but more research is necessary to confirm this suspicion.

4.1. Conclusions and recommendations

All models are inherently defined by the quality of the data they are provided, the model inputs and parameters used when running them, and the scale, extent, and resolution of the analysis. Presence-background SDMs relying on machine learning techniques (such as Maxent) are highly influenced by the inputs provided (Radosavljevic and Anderson, 2014). This research constitutes a “deep-dive” into the effect of methodological changes in one such input – background point generation – on the accuracy and transferability of a single, presence-background model algorithm. Our results clearly show that not all modelling inputs are created equal, nor are methods for evaluating the accuracy of an SDM. We identified a discordance between the best-discriminating models within their training domains (as measured with the overlap metrics) and the most-transferrable models across

space (as measured with the Boyce index and overprediction metrics), supporting previous literature (M. Iturbide et al., 2018; Petitpierre et al., 2017). Therefore, the specific research question of the study may influence the selection of model inputs more than with the hypothetical accuracy or transferability of the model (regardless of the metric used to evaluate it). In fact, the presence of overprediction and underprediction in a model may provide valuable insight into the ecological processes structuring species ranges (Márcia Barbosa et al., 2013). For example, in an ecological study evaluating the fundamental niche of a species (the overall habitat a species could potentially inhabit), spatial overprediction may in fact be preferred and selected for by sampling background points randomly throughout the study area (Chefaoui and Lobo, 2008; Lobo et al., 2010; Márcia Barbosa et al., 2013). Another study examining occupancy of a species in a fragmented conservation hotspot may instead opt for a model that minimizes spatial overprediction. Furthermore, the discrepancies between the fixed-extent models and the varying-extent models serve to highlight the well-known importance of considering interactions between model extent, background points, and SDM performance. Studies that look to project small-ranged species into a big extent may choose different background point parameters than those that tailor the extent of the study to the species range (i.e., using moderate levels of spatial weighting instead of sampling exclusively within the buffer). Finally, the contrasting results of the virtual species underscores that these methods are designed to reduce bias in the occurrence data and if little bias occurs in the data, they may not be appropriate to apply.

Finding proper background point distributions is dependent on the specific study system and research question and is often more of an art than a science. The results of SDMs are, by their definition, hypotheses used to test ecological theory and provide guidance for conservation actions. These hypotheses may well be falsified, and can only truly be supported and evaluated by additional observation (Jarnevich et al., 2015). Here, we provide evidence that researchers must test multiple combinations of model inputs to “fine-tune” their models to answer their specific ecological questions. However, we found several general trends that may be useful for developing accurate SDMs. First, supporting the research conducted over the last decade, we find that AUC values are a deeply flawed metric with respect to presence-background models (see Fig. 5), being strongly affected by the number of background points used in the model, but unaffected by extrapolation errors that occur outside of the climate sampled by the background (Supplementary Appendix 3). Using the Boyce index or metrics comparing the modelled distribution to an expert hypothesis may help to more accurately evaluate both the accuracy and transferability of these models. When using these evaluation metrics, we find support for using spatially weighted background points (especially for small-ranged species on a large extent), in addition to using an ecologically based buffer like ecoregions to mitigate bias in the occurrence records. SDMs have the unparalleled ability to model our changing biological communities on a large scale. These models are highly versatile tools and should be treated as such. Understanding how modelling inputs affects the results of these SDMs is the first step towards refining these tools, creating better models to be used for the complex questions biogeographers answer today.

CRedit authorship contribution statement

Anna M. Whitford: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Benjamin R. Shipley:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – review & editing, Visualization. **Jenny L. McGuire:** Conceptualization, Investigation, Resources, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The characteristics of the simulated species and their ranges are on Figshare: [The influence of the number and distribution of background points in presence-background species distribution models](#) (Original data). Other data will be available upon request.

Acknowledgments

We would like to thank Julia Schap for helping to devise the procedure for occurrence point vetting, Amna Amir, Audrey Dods, and Megan Wright for vetting the occurrence points used in this study, and the Spatial Ecology and Paleontology Lab for their insights and assistance throughout the project. Funding for this work was provided by the National Science Foundation (DGE-2039655, 1945013).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2023.110604](https://doi.org/10.1016/j.ecolmodel.2023.110604).

References

- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* 3, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>.
- Bennett, M., Marquet, P.A., Sillero-Zubiri, C., Marino, J., 2019. Shifts in habitat suitability and the conservation status of the Endangered Andean cat *Leopardus jacobita* under climate change scenarios. *Oryx* 53, 356–367. <https://doi.org/10.1017/S0030605317000503>.
- Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol.* 8, e1000385. <https://doi.org/10.1371/journal.pbio.1000385>.
- Borges, C.M., Terribile, L.C., Oliveira, G.de, Lima-Ribeiro, M.de S., Dobrovolski, R., 2019. Historical range contractions can predict extinction risk in extant mammals. *PLoS One* 14, e0221439. <https://doi.org/10.1371/journal.pone.0221439>.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecol. Model.* 157, 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4).
- Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., Bollmann, K., 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* 36, 971–983. <https://doi.org/10.1111/j.1600-0587.2013.00138.x>.
- Broekman, M.J.E., Hilbers, J.P., Huijbregts, M.A.J., Mueller, T., Ali, A.H., Andrén, H., Altmann, J., Aronsson, M., Attias, N., Bartlam-Brooks, H.L.A., van Beest, F.M., Belant, J.L., Beyer, D.E., Bidner, L., Blaum, N., Boone, R.B., Boyce, M.S., Brown, M. B., Cagnacci, F., Cerne, R., Chamaillé-Jammes, S., Dejid, N., Dekker, J., Desbiez, A., L. J., Díaz-Muñoz, S.L., Fennessy, J., Fichtel, C., Fischer, C., Fisher, J.T., Fischhoff, I., Ford, A.T., Fryxell, J.M., Gehr, B., Goheen, J.R., Hauptfleisch, M., Hewison, A.J.M., Hering, R., Heurich, M., Isbell, L.A., Janssen, R., Jeltsch, F., Kaczensky, P., Kappeler, P.M., Krofel, M., LaPoint, S., Latham, A.D.M., Linnell, J.D.C., Markham, A. C., Mattisson, J., Medici, E.P., de Miranda Mourão, G., Van Moorter, B., Morato, R. G., Morellet, N., Mysterud, A., Mwiu, S., Odden, J., Olson, K.A., Ornicans, A., Pagon, N., Panzacchi, M., Persson, J., Petroelje, T., Rolandsen, C.M., Roshier, D., Rubenstein, D.I., Saïd, S., Salemgareyev, A.R., Sawyer, H., Schmidt, N.M., Selva, N., Sergiel, A., Stabach, J., Stacy-Dawes, J., Stewart, F.E.C., Stiegler, J., Strand, O., Sundaresan, S., Svoboda, N.J., Ullmann, W., Voigt, U., Wall, J., Wikelski, M., Wilmers, C.C., Zięba, F., Zwijacz-Kozica, T., Schipper, A.M., Tucker, M.A., 2022. Evaluating expert-based habitat suitability information of terrestrial mammals with GPS-tracking data. *Glob. Ecol. Biogeogr.* 31, 1526–1541. <https://doi.org/10.1111/geb.13523>.
- Bucklin, D.N., Basille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Románach, S.S., Speroterra, C., Watling, J.L., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Divers. Distrib.* 21, 23–35. <https://doi.org/10.1111/ddi.12247>.
- Castellanos, A.A., Huntley, J.W., Voelker, G., Lawing, A.M., 2019. Environmental filtering improves ecological niche models across multiple scales. *Methods Ecol. Evol.* 10, 481–492. <https://doi.org/10.1111/2041-210X.13142>.

- Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Model.* 210, 478–486. <https://doi.org/10.1016/j.ecolmodel.2007.08.010>.
- Cobos, M.E., Peterson, A.T., Barve, N., Osorio-Olvera, L., 2019. kuenm: an R package for detailed development of ecological niche models using Maxent. *PeerJ* 7, e6281. <https://doi.org/10.7717/peerj.6281>.
- 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., Guisan, A., 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 40, 774–787. <https://doi.org/10.1111/ecog.02671>.
- Duan, H., Yu, X., Xia, S., Liu, Y., 2022. Comparison of IUCN and species distribution modeling-estimated ranges of shorebirds in Coastal Mainland China. *Glob. Ecol. Conserv.* 38, e02236. <https://doi.org/10.1016/j.gecco.2022.e02236>.
- Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. *Methods Ecol. Evol.* 1, 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Feng, X., Park, D.S., Liang, Y., Pandey, R., Papes, M., 2019. Collinearity in ecological niche modeling: confusions and challenges. *Ecol. Evol.* 9, 10365–10376. <https://doi.org/10.1002/ece3.5555>.
- Fenker, J., Domingos, F.M.C.B., Tedeschi, L.G., Rosauer, D.F., Werneck, F.P., Colli, G.R., Ledo, R.M.D., Fonseca, E.M., Garda, A.A., Tucker, D., Sites, J.W., Breitman, M.F., Soares, F., Giugliano, L.G., Moritz, C., 2020. Evolutionary history of Neotropical savannas geographically concentrates species, phylogenetic and functional diversity of lizards. *J. Biogeogr.* 47, 1130–1142. <https://doi.org/10.1111/jbi.13800>.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>.
- Fourcade, Y., Engler, J.O., Besnard, A.G., Rödder, D., Secondi, J., 2013. Confronting expert-based and modelled distributions for species with uncertain conservation status: a case study from the corncrake (*Crex crex*). *Biol. Conserv.* 167, 161–171. <https://doi.org/10.1016/j.biocon.2013.08.009>.
- Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One* 9, e97122. <https://doi.org/10.1371/journal.pone.0097122>.
- Giovannelli, J.G.R., de Siqueira, M.F., Haddad, C.F.B., Alexandrino, J., 2010. Modeling a spatially restricted distribution in the Neotropics: how the size of calibration area affects the performance of five presence-only methods. *Ecol. Model.* 221, 215–224. <https://doi.org/10.1016/j.ecolmodel.2009.10.009>.
- González-Ferreras, A.M., Barquín, J., Peñas, F.J., 2016. Integration of habitat models to predict fish distributions in several watersheds of Northern Spain. *J. Appl. Ichthyol.* 32, 204–216. <https://doi.org/10.1111/jai.13024>.
- Grimmett, L., Whitted, R., Horta, A., 2020. Presence-only species distribution models are sensitive to sample prevalence: evaluating models using spatial prediction stability and accuracy metrics. *Ecol. Model.* 431, 109194. <https://doi.org/10.1016/j.ecolmodel.2020.109194>.
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>.
- Holloway, P., Miller, J.A., Gillings, S., 2016. Incorporating movement in species distribution models: how do simulations of dispersal affect the accuracy and uncertainty of projections? *Int. J. Geogr. Inf. Sci.* 30, 2050–2074. <https://doi.org/10.1080/13658816.2016.1158823>.
- Holt, R.D., 2003. On the evolutionary ecology of species' ranges. *Evol. Ecol. Res.* 5, 159–178.
- Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl. Acad. Sci.* 104, 13384–13389. <https://doi.org/10.1073/pnas.0704469104>.
- Hurlbert, A.H., White, E.P., 2005. Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecol. Lett.* 8, 319–327. <https://doi.org/10.1111/j.1461-0248.2005.00726.x>.
- Iturbide, M., Bedia, J., Gutiérrez, J.M., 2018. Background sampling and transferability of species distribution model ensembles under climate change. *Glob. Planet. Change* 166, 19–29. <https://doi.org/10.1016/j.gloplacha.2018.03.008>.
- Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., Gutiérrez, J.M., 2015. A framework for species distribution modelling with improved pseudo-absence generation. *Ecol. Model.* 312, 166–174. <https://doi.org/10.1016/j.ecolmodel.2015.05.018>.
- IUCN, 2020. In: The IUCN Red List of Threatened Species, 2019th-3. ed. <https://www.iucnredlist.org/resources/spatial-data-download>.
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. *Ecol. Inform.* 29, 6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>.
- Jiménez, L., Soberón, J., 2020. Leaving the area under the receiving operating characteristic curve behind: an evaluation method for species distribution modelling applications based on presence-only data. *Methods Ecol. Evol.* 11, 1571–1586. <https://doi.org/10.1111/2041-210X.13479>.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366–1379. <https://doi.org/10.1111/ddi.12096>.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Leroy, B., Delsol, R., Hugué, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M., Bellard, C., 2018. Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *J. Biogeogr.* 45, 1994–2002. <https://doi.org/10.1111/jbi.13402>.
- Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. Virtual species, an R package to generate virtual species distributions. *Ecography* 39, 599–607. <https://doi.org/10.1111/ecog.01388>.
- Liaw, A., Wiener, M., 2002. Classification and regression by random forest. *R News* 2, 18–22.
- Liu, C., Newell, G., 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. <https://doi.org/10.1111/ecog.03188>.
- Liu, C., Newell, G., White, M., 2015. On the selection of thresholds for predicting species occurrence with presence-only data. *Ecol. Evol.* 6, 337–348. <https://doi.org/10.1002/ece3.1878>.
- Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33, 103–114. <https://doi.org/10.1111/j.1600-0587.2009.06039.x>.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>.
- Machado-Stredel, F., Cobos, M.E., Peterson, A.T., 2021. A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. *Front. Biogeogr.* 13, e48814. <https://doi.org/10.21425/F5FBG48814>.
- Márcia Barbosa, A., Real, R., Muñoz, A.R., Brown, J.A., 2013. New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Divers. Distrib.* 19, 1333–1338. <https://doi.org/10.1111/ddi.12100>.
- Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* 51, 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2).
- Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., Guisan, A., 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Glob. Ecol. Biogeogr.* 26, 275–287. <https://doi.org/10.1111/geb.12530>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. <https://doi.org/10.1890/07-2153.1>.
- Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: *Twenty-First International Conference on Machine Learning - ICML '04*. Banff, Alberta, Canada. ACM Press, p. 83. <https://doi.org/10.1145/1015330.1015412>.
- Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* 41, 629–643. <https://doi.org/10.1111/jbi.12227>.
- Rousseau, J.S., Betts, M.G., 2022. Factors influencing transferability in species distribution models. *Ecography* e06060. <https://doi.org/10.1111/ecog.06060>, 2022.
- Ruiz-Gutiérrez, V., Zipkin, E.F., 2011. Detection biases yield misleading patterns of species persistence and colonization in fragmented landscapes. *Ecosphere* 2, art61. <https://doi.org/10.1890/ES10-00207.1>.
- Sarkar, S., Sanchez-Cordero, V., Cecilia Londono, M., Fuller, T., 2009. Systematic conservation assessment for the Mesoamerica, Choco, and Tropical Andes biodiversity hotspots: a preliminary analysis. *Biodivers. Conserv.* 18, 1793–1828. <https://doi.org/10.1007/s10531-008-9559-1>.
- Schloss, C.A., Nuñez, T.A., Lawler, J.J., 2012. Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proc. Natl. Acad. Sci.* 109, 8606–8611. <https://doi.org/10.1073/pnas.1116791109>.
- Senay, S.D., Worner, S.P., Ikeda, T., 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PLoS One* 8, e71218. <https://doi.org/10.1371/journal.pone.0071218>.
- Shipley, B.R., Bach, R., Do, Y., Strathearn, H., McGuire, J.L., Dilkina, B., 2022. megaSDM: integrating dispersal and time-step analyses into species distribution models. *Ecography* e05450. <https://doi.org/10.1111/ecog.05450>.
- Sillero, N., Barbosa, A.M., 2021. Common mistakes in ecological niche models. *Int. J. Geogr. Inf. Sci.* 35, 213–226. <https://doi.org/10.1080/13658816.2020.1798968>.
- Sonne, J., Martín González, A.M., Maruyama, P.K., Sandel, B., Vizin-Bugoni, J., Schleuning, M., Abrahamczyk, S., Alarcón, R., Araújo, A.C., Araújo, F.P., Mendes de Azevedo, S., Baquero, A.C., Cotton, P.A., Ingversen, T.T., Kohler, G., Lara, C., Guedes Las-Casas, F.M., Machado, A.O., Machado, C.G., Maglianesi, M.A., Moura, A.C., Nogueira-Bravo, D., Oliveira, G.M., Oliveira, P.E., Ornelas, J.F., Rodrigues, L., da, C., Rosero-Lasprilla, L., Rui, A.M., Sazima, M., Timmermann, A., Varassin, I.G., Wang, Z., Watts, S., Fjeldsø, J., Svenning, J.C., Rahbek, C., Dalsgaard, B., 2016. High

- proportion of smaller ranged hummingbird species coincides with ecological specialization across the Americas. *Proc. R. Soc. B Biol. Sci.* 283, 20152512 <https://doi.org/10.1098/rspb.2015.2512>.
- Stirling, D.A., Boulcott, P., Scott, B.E., Wright, P.J., 2016. Using verified species distribution models to inform the conservation of a rare marine species. *Divers. Distrib.* 22, 808–822. <https://doi.org/10.1111/ddi.12447>.
- Thuiller, W., Brotons, L., Araújo, M.B., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27, 165–172. <https://doi.org/10.1111/j.0906-7590.2004.03673.x>.
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>.
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* 92, e01486. <https://doi.org/10.1002/ecm.1486>.
- VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol. Model.* 220, 589–594. <https://doi.org/10.1016/j.ecolmodel.2008.11.010>.
- Varela, S., Anderson, R.P., García-Valdés, R., Fernández-González, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37, 1084–1091. <https://doi.org/10.1111/j.1600-0587.2013.00441.x>.
- Williams, K.J., Belbin, L., Austin, M.P., Stein, J.L., Ferrier, S., 2012. Which environmental variables should I use in my biodiversity model? *Int. J. Geogr. Inf. Sci.* 26, 2009–2047. <https://doi.org/10.1080/13658816.2012.698015>.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>.
- Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243. <https://doi.org/10.1111/2041-210x.12004>.