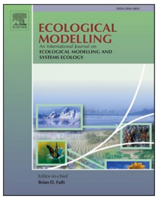
Ecological Modelling 507 (2025) 111177

Contents lists available at ScienceDirect

Ecological Modelling

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



Improving species distribution models by optimising background points: Impacts on current and future climate projections

Armand Rausell-Moreno [a,*], Núria Galiana [a], Babak Naimi [b], Miguel B. Araújo [a,c,d]

[a] *Department of Biogeography and Global Change, National Museum of Natural Sciences, Consejo Superior de Investigaciones Científicas (CSIC), Calle Jose Gutierrez Abascal, 2, 2008, Madrid, Spain*
[b] *Quantitative Biodiversity Dynamics (QBD), Department of Biology, University of Utrecht, Padualaan 8, Utrecht, 3584 CH, the Netherlands*
[c] *Biodiversity Chair, Mediterranean Institute for Agriculture, Environment and Development (MED), and Global Change and Sustainability Institute (CHANGE), Institute for Advanced Studies and Research (IIFA), Rua Dr. Joaquim Henrique da Fonseca, 7000-890 Évora, Portugal*
[d] *Theoretical Sciences Visiting Program, Okinawa Institute of Science and Technology Graduate University, Onna, 904-0495, Japan*

ARTICLE INFO

Keywords:
Species distribution models
Climate change
Virtual species
Background points
Performance
Presence-only

ABSTRACT

Species Distribution Models (SDM) are often fit using presence-background data due to the lack of reliable absence records. To calibrate these models, background records are required, yet the optimal number of records and if they should be proportional to study area or the number of occurrences remains uncertain. This study addresses three key questions: (i) how does varying background proportions affect predictive accuracy? (ii) How do background proportions influence future species distribution projections under climate change? and (iii) should the number of background records be determined based on study region size or presence record availability? To investigate these questions, we simulated 280 virtual species distributions worldwide under present and future climate conditions. Model outputs were evaluated against simulated "true" distributions under both present and future scenarios. Results indicate that sampling background records proportional to either presence points or study area yields comparable average performance. Optimal performance occurred with a 0.5–1 ratio of background records to presence points when sampled proportionally to presences, and with approximately 5 % of the study area sampled when proportional to region size. Species prevalence also modulated the optimal presence-background ratio. Increasing the number of background records across suitable and unsuitable areas had contrasting effects for both strategies tested, emphasizing the need to assess model performance separately for both. Notably, background proportions influenced baseline predictions but had minimal impact on future projections, where niche-related variables dominated model performance. These findings offer practical insights for SDM practitioners. Adjusting background sampling strategies enhances current prediction accuracy, while future projections remain robust across different sampling approaches, ensuring more reliable modelling outcomes.

1. Introduction

Species distribution models (SDMs) have emerged as a powerful tool for multiple biodiversity-related applications (Araújo and Peterson, 2012). They are instrumental in assessing the habitat of rare species (Pearson et al., 2007; Williams et al., 2009) or identifying the potential invasive ranges of introduced species (Mainali et al., 2015; Naimi et al., 2022). SDMs also assist conservation planning by providing predictions of species occurrences (Guisan et al., 2013) or evaluating species persistence under current conditions (Araújo et al., 2002) and future climate change scenarios (Araújo et al., 2011; Triviño et al., 2018). However, despite their wide applicability, significant uncertainties remain—particularly regarding the selection of background records when only presence data are available for model development.

Ideally, both presence and absence data are used to train SDMs. However, reliable absence data are rarely available in practice. As a result, background records are generated and used for model fitting, raising critical questions about the best way to generate these records, including sampling strategy and proportional balance. Conceptually, background points refer to locations sampled across the study area to

---

* Corresponding author at: Department of Biogeography and Global Change, National Museum of Natural Sciences, Consejo Superior de Investigaciones Científicas (CSIC), Calle Jose Gutierrez Abascal, 2, 2008, Madrid, Spain.

E-mail address: armand.rausell@mncn.csic.es (A. Rausell-Moreno).

https://doi.org/10.1016/j.ecolmodel.2025.111177
Received 3 February 2025; Received in revised form 8 May 2025; Accepted 8 May 2025

represent available environmental conditions, including presence locations and without assuming species absence.

Regarding the questions rising from background generation, Barbet-Massin et al. (2012) highlighted these challenges in the paper titled: *"Selecting pseudo-absences for SDMs: How, where and how many?"*. In the present, extensive literature is available addressing background from multiple perspectives. For instance, selecting points randomly within environmental space yields more stable and accurate models than sampling randomly in geographic space (Broussin et al., 2024; Hanberry et al., 2012; Steen et al., 2024; Wisz and Guisan, 2009). Others compared the performance of background data over presence-only and pseudo-absence data (Fernandez et al., 2022) and proved better performance for presence-background fitted SDMs.

Necessarily, many more studies have examined how the spatial distribution of pseudo-absences affected model performance. For example, Thuiller et al. (2004) found that limiting pseudo-absences to areas near known occurrences can result in overly conservative models. In contrast, Chefaoui and Lobo (2008) demonstrated that selecting background records from environmentally distant locations improved model accuracy, while VanDerWal et al. (2009) cautioned against drawing pseudo-absences from excessively small or overly large areas to avoid model instability or inflated statistical outputs.

Yet, despite the multiple approaches, a consensus on the number of background records and whether these records should be determined proportional to the size of the study area or relative to the number of presence points is still lacking. Only a handful of studies have explored the impact of varying the number of background records on model performance (Barbet-Massin et al., 2012; Hysen et al., 2022; Liu et al., 2019; Santini et al., 2021; Whitford et al., 2024), leaving a lack of clear guidelines on optimal sample sizes to improve SDM accuracy. Sampling 10,000 records has been established as common practice among SDM practitioners. However, works that employ that number often lack a proper justification on such selection (Baker et al., 2022; Glad and Mallard, 2022; Luna et al., 2024; Martínez-Fonseca et al., 2024; McCulloch-Jones et al., 2023; Sanczuk et al., 2022) or dismiss the particular conditions under which the seminal works from Barbet-Massin et al. (2012) and Phillips and Dudík (2008) arrived at those numbers. Recently, researchers have begun to use alternatives to the widely accepted 'magical number' of 10,000 background records, often regarded as suitable for any situation and set and used as default configuration in the popular Maxent (Phillips and Dudík, 2008). Instead, they have opted to create more balanced proportions between presence and background records (Egorova et al., 2024), as some works found better performing proportions with 1:1 ratio presence-background (Barker and MacIsaac, 2022; Tong et al., 2023).

Lastly, another underexplored question is how background record generation strategies and their proportions influence SDM performance under future climate projections. Santini et al. (2021) highlighted the need to investigate this issue, noting that models optimised for present conditions may not perform similarly under future climate scenarios (see also Araújo et al. 2005). Additionally, the challenges of validating future projections and the limitations of traditional metrics necessitate alternative approaches, such as the use of virtual species (e.g., Grimmett et al., 2021; Valladares et al., 2014).

To address these gaps, this study employs a global-scale virtual species experiment, simulating the distributions of 280 species under both current and future climate conditions. Specifically, we investigate how background sampling strategies (generating background records proportional to the area or proportional to the number of presence records) influence the predictions of SDMs in the current time and their projections in future, offering insights to enhance SDM performance and applicability in a range of applications, including conservation planning under climate change. By evaluating the effects of various background proportions using a robust performance assessment protocol, we provide practical recommendations for optimising SDMs.

2. Methods

2.1. Virtual species generation

To simulate realistic species distributions, we generated 280 virtual species with a global distribution using the R package '*virtualspecies*' (Leroy et al., 2016). The virtual species distributions were designed to emulate realistic biogeographical patterns by defining species range centroids under two primary conditions:

1. Real species richness as weights
   We used a real-world map as a template and bird species richness (Lumbierres et al., 2022) was used as probability weights for sampling 300 species-range centroids. Specifically, pixels with higher species richness were more likely to be selected for centroids, with disabled sample replacement. With this approach, we incorporate existing biodiversity information into the generation of virtual species distributions (Fig. 1).
2. Latitudinal variation in range size
   To mimic the observed trend of increasing distributional range size at higher latitudes as described by Rapoport's rule (Stevens, 1989), the virtual species range widths expanded with latitude. This was implemented using an equation (Eq. (1)) that defined the extent of range expansion based on the latitude of the centroid for each virtual species.

$$Extent\ increase = 8 \cdot \frac{log(|latitude|) - log(0.1)}{log(90) - log(0.1)} \tag{1}$$

Where 8 is the maximum extent increase allowed. Extents were also limited to a minimum of 4° (~440 km near the equator) and a maximum of 10° (~1110 km near the equator). Lastly, to add certain stochasticity to the range size selection, a random value between $4^\circ + Extent\ increase$ and $10^\circ + Extent\ increase$ was selected and finally applied to expand the range using as center point the species selected centroid. As a result, species at higher latitudes are more likely to have a wider range, though exceptions remain possible. For further details, see the virtual species section in the supplementary material.

After selecting the centroid of species ranges, we generated virtual species using the *generateSpFromBCA* function from the 'virtualspecies' R package (Leroy et al., 2016) as shown in Fig. 2. This function performs a Between Component Analysis (BCA) of the environmental conditions within the study area. A BCA initially performs a Principal Component Analysis (PCA) for both time periods to reduce dimensionality and summarize variance into principal axes. BCA is then applied to the PCA-transformed environmental data, treating the climate scenarios (current vs. future) as predefined groups. The goal of BCA is to find new axes that best separate these groups while retaining as much environmental variability as possible. From the BCA, which is a combination of both environmental sets, a random point is selected to serve as the mean of the species response to the environment. These responses are multiplied together, and their product defines the final suitability value across the environmental space. The highest suitability values correspond to areas where all Gaussian responses reach their peak. Additionally, species are allowed to exhibit any niche breadth (standard deviation from the mean), ensuring flexibility in the range of ecological scenarios represented. This approach assumes that species respond to climate changes without lags, i.e., reaching equilibrium as environmental changes occur (Araújo and Pearson, 2005; Munguía et al., 2012). The species generation process ensures uniqueness, as the probability of identical parameter combinations (such as centroid location, range size, mean environmental conditions, and niche breadth) is nearly zero.

One challenge of using two temporal datasets is that a species may be present in the current environment but go extinct in the future, or

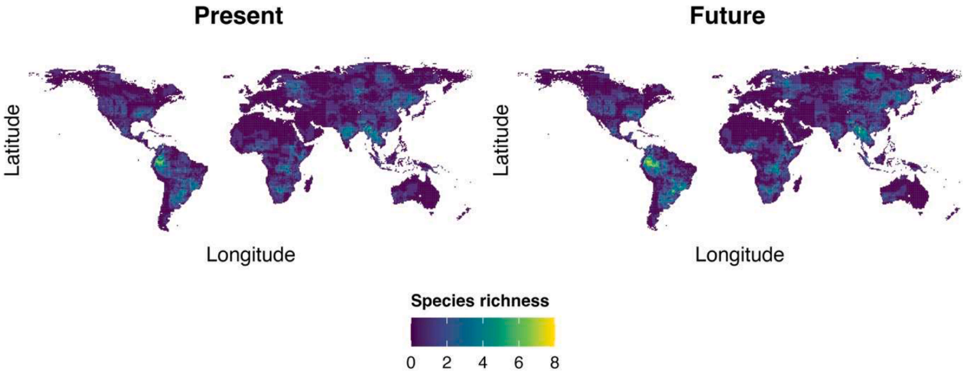


Fig. 1. Distribution of virtual species richness simulated for this work for present and future times. For visualization purposes, for each virtual species we considered pixels above suitability values of 0.8 as presence after scaling (all species had suitability from 0 to 1), below that threshold pixels were transformed into absences. Binary maps were then summed up in a global map for each time period.

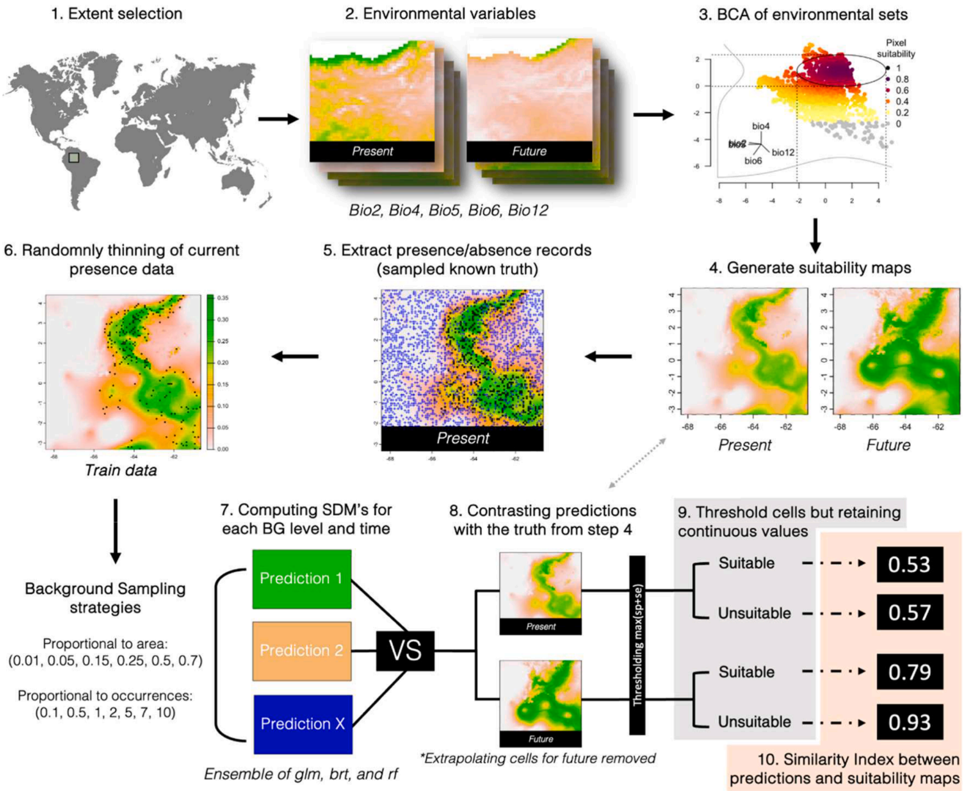


Fig. 2. Virtual species generation process, sampling, and model evaluation workflow. 1) Extent selection - The spatial extent for species distribution is determined following a global bird species richness template and the range of such extent calculated based on the latitude according to the specified formula above. 2) Environmental variables for the present and future time periods are cropped to the selected extent. 3) Between Component Analysis - A Principal Components Analysis (PCA) is first conducted on combined environmental variables for both time periods. A BCA is performed on the PCA results to differentiate between the two environmental data sets. Gaussian response curves are computed for the first two BCA axes to simulate species' environmental preferences. 4) Environmental Suitability Calculation - The Gaussian response curves of the two axes are then multiplied among them to generate the final environmental suitability map, representing the niche of the species. With this, areas where both axes are high obtain as well the highest possible values of environmental suitability. 5) Presence and absence sampling - Presence and absence points are sampled probabilistically based on the suitability scores, ensuring ecologically realistic distributions of occurrence data. 6) Downgrading presence points - A random 70 % of the sample presence points are removed, reducing the data to a manageable size for computational efficiency. 7) SDM development - Species distribution models are computed for each species. Multiple models are created using varying proportions of backgrounds while maintaining a consistent number of presence points. Models are developed for the present and predicted/projected for both time periods (present and future) using three modelling methods. 8) Model prediction and projection - The predictions for the present and projections for the future periods are compared against the habitat suitability maps (the known truth) from step 4 using Spearman correlation. 9) We divide the results of our predictions and projections into suitable and unsuitable areas based on the threshold from the present that maximizes the sum of the specificity + sensitivity. Once suitability was binarized, model accuracy was assessed for each suitability class based on the original continuous values. 10) Model evaluation – Spearman correlation values are computed by comparing the model predictions with the known truth suitability across the delimited suitable and unsuitable areas, providing a quantitative measure of model performance across different background sampling levels and time periods.

conversely, previously restricted environments may become more widespread, enabling species expansion. To ensure reliable evaluation of niche similarity and maintain a minimum niche overlap across both periods, we implemented an iterative loop. For each defined range size, twenty combinations of means (representing the center of the response curve of the species) and standard deviations (niche breadths) were tested in the environmental space until one combination met the criteria of a minimum persistence threshold of 0.2. This threshold means that for a species to become included in our analysis, the species has to persist in both times with at least a maximum value of the suitability raster of 0.2. If a generated species for a particular environment could not find a combination that achieved a maximum suitability score higher than 0.2 in either period within twenty attempts, it was excluded from further analysis. This approach ensured the availability of enough high- and low-suitability pixels in both time periods to evaluate niche similarities effectively.

2.2. Occurrence data

Occurrence data were sampled from a probabilistically transformed presence-absence map representing the truth. To generate this map, 80 % of the cells with a suitability value of 0.8, 70 % of cells with a suitability of 0.7, and so on were considered as presence. From the presence-absence map, we randomly subsampled 40 % to reduce computation times while thinning the data. The extracted presence data along with background points served as train data (sampled known truth), while the suitability maps themselves represented the absolute known truth.

After extracting occurrence data, species prevalence was then calculated as the proportion of presence points relative to the total number of occurrences (presences / presences + absences). In our approach, sample prevalence (the proportion of sampled presence-absence points) is expected to closely reflect species prevalence (the proportion of the study area occupied). This is because species prevalence remains unchanged as long as the sampling process maintains the original class proportions. When randomly subsampling 40 % of the data, each class should be sampled at approximately the same rate (40 %), thereby preserving the original prevalence, regardless of its absolute value. We also categorized prevalence values into five classes; 0–0.1, 0.2–0.3, 0.4–0.5, 0.6–0.7, and 0.8–0.9, corresponding to ranges 0–0.19, 0.20–0.39, 0.40–0.59, 0.60–0.79, and 0.80–0.99, respectively.

2.3. Climatic data

The climatic variables used to construct the environmental space for the virtual species included mean diurnal range (Bio2), temperature seasonality (Bio4), maximum temperature of warmest month (Bio5), minimum temperature of coldest month (Bio6) and annual precipitation (Bio12). These variables were obtained from the WorldClim dataset (version 2.1) (Fick and Hijmans, 2017) at a spatial resolution of 5 min representing long-term averages of climate conditions for the period 1970–2000. These variables were selected to capture a broad range of critical climatic gradients known to influence species distributions globally (Epele et al., 2024; Harsch and HilleRisLambers, 2016; MacPherson et al., 2018).

Future environmental conditions were derived from the ACCESS-CM2 global circulation model (GCM), under a Shared Socio-economic Pathway (SSP) 2.45. This scenario represents intermediate challenges in mitigation and adaptation and covers the period 2021–2040, providing a working example of projection of near future climate change conditions (e.g. Meinshausen et al., 2020).

2.4. Species distribution models

Following the generation of virtual species distributions (hereafter termed “known truth”), a training sample for Species Distribution Models (SDMs) was obtained by randomly downgrading the known truth, removing 70 % of the points. As only presence data from the virtual species were used, we tested two different background generation strategies.

1. Background proportional to presence points: This strategy involved generating background points with a size relative to the number of presence points at seven different ratios: 0.1, 0.5, 1, 2, 5, 7, and 10. For instance, a ratio of 10 represented ten times more background points than the number of presence points.
2. Background proportional to area size: This approach generated background points as a proportion of the area, tested at six levels: 0.01, 0.05, 0.15, 0.25, 0.5, and 0.7. The highest level (0.7) represented 70 % of the area sampled as background points. For highly prevalent species, this was sometimes unachievable due to overlap between presence and background points, leading to fewer observations at this level.

SDMs were developed for each virtual species using background records generated with both strategies across all respective levels over the species-specific extent. For both cases, background could not be sampled from a pixel coinciding with a presence location, deviating slightly from the strict definition of background. The *sdm* R package (Naimi and Araújo, 2016) was used to fit the models, focusing on the subset of three climatic variables (bio5, bio6 and bio12). For this set, we calculated the variance inflation factor (VIF) to detect potential collinearity among variables using the vifstep function from *usdm* R package (Naimi et al., 2014). All variables remained under 5, signalling that no collinearity is found for this set of variables. This subset of variables was selected to capture key thermal and hydrological niches while reducing redundancy and computational complexity. Also, this approach was intended to emulate a realistic scenario in species distribution modelling (SDM) where researchers often lack complete knowledge about all the environmental factors influencing a species’ distribution. By withholding two of the original variables, the modelling process was designed to be "blind" to the full range of climatic factors affecting the virtual species, thereby introducing an element of uncertainty and reducing the risk of overfitting to the artificial data.

We fitted an ensemble of three widely used SDM methods: “Generalized Linear Models” (GLM; McCullagh and Nelder, 1989), Boosted Regression Trees (BRT; Elith et al., 2008) and Random Forest (RF; Breiman, 2001). 1000 trees were employed for BRT and RF. All models were replicated five times subsampling 30 % of the data as test. To address uncertainty associated with varying predictions of individual models, we derived a consensus by combining the outputs of these models (Araújo and New, 2007), following best practices in SDM studies (Araújo et al., 2019).

2.5. Model evaluation

After computing the SDMs, we evaluated their performance by comparing the predictions and projections of each consensus model output against the known truth (suitability maps) for each species across all background levels. The degree of similarity between model outputs and the suitability maps was quantified using the Spearman correlation coefficient, which measures pixel-by-pixel similarity. Values close to 1 refer to strong alignment between the model’s predictions and the known suitability values, indicating high accuracy. We used Spearman over other commonly used metrics such as Schoener’s D for several key reasons: The first one, Spearman showed to be less sensitive to prevalence according to post-hoc analysis (Supp. Fig. 3), which provided robustness to the later interpretation. Second, our primary focus was not on capturing absolute differences between observed and predicted suitability values, as Schoener’s D is designed to do, but rather on assessing how well the predicted suitability ranks align with observed suitability. Spearman’s correlation is more suitable for this purpose, as it evaluates the consistency of relative rankings between observed and predicted values, rather than their absolute agreement. To ensure accurate model evaluations for future predictions and minimize extrapolation effects, we excluded future pixels with environmental values beyond the range observed during the training phase in the present time. This environmental similarity across both time periods was assessed with function MESS from 'modEvA' R package (Márcia Barbosa et al., 2013).

To obtain a clearer assessment of accuracy, we categorised the suitability maps for both present and future time periods into two classes: suitable and unsuitable areas. These classifications were determined using the threshold from current ensembles that maximizes [specificity + sensitivity]. Areas with suitability values above this threshold were labelled as suitable, while those below were labelled as unsuitable. This categorisation allowed us to assess the performance of the models across suitable and unsuitable areas, separately. Therefore, four groups of performance values were calculated for each model: Suitable-present, Suitable-future, Unsuitable-present, Unsuitable-future. For each suitability class, the evaluation using the Spearman correlation coefficient was conducted with an equal number of pixels to ensure comparability across classes. This approach provided a detailed understanding of models' performance in accurately predicting suitable and unsuitable areas under both current and future conditions, enabling a comprehensive assessment of the SDMs' ability to reflect the known truth.

2.6. *Statistical analysis*

Following the evaluation of the models, we performed a Kruskal-Wallis test to assess differences in models' performance across the various background proportions (Table S1). We used Kruskal-Wallis as some proportions did not have a normal distribution. To disclose pairwise comparisons, we employed a Dunn's Kruskal-Wallis test with the "bonferroni" method (Table S2).

2.7. *Explaining model performance*

To investigate the factors influencing model performance, we built a regression model to explain the Spearman values and identify the most important variables. A Random Forest algorithm was chosen for this analysis, as initial exploratory data suggested non-linear relationships between the dependent variable and the predictors. Random Forest models are particularly well-suited for capturing complex non-linear interactions (Fratello and Tagliaferri, 2019; Marchese Robinson et al., 2017; Scornet et al., 2015). The analysis was structured as follows. The dependent variable was the Spearman similarity values, representing the degree of alignment between model predictions and the known truth. The independent variables included area size (the spatial extent used for each species), number of presence records used to train the models, number of background records (determined based on a ratio relative to either the number of presence points or the area size), niche position, measured using outlying mean index (OMI; Dolédec et al., 2000; Karasiewicz et al., 2017) representing the centroid of a species' environmental niche within the multivariate environmental space, niche breadth (tolerance; a measure of the environmental range a species occupies), niche marginality (the distance between the species' niche centroid and the centroid of the overall environmental space), environmental similarity (a metric comparing the environmental conditions between present and future scenarios), and species prevalence (the proportion of the study area classified as suitable habitat in both the present and future). Niche related variables were calculated using the 'ade4' R package (Chessel et al., 2004; Dray and Dufour, 2007).

3. Results

After analyzing the distribution of 280 virtual species, we observed no substantial differences in model performance between the two background sampling strategies (Fig. 3). Sampling background records proportional to the area size or to the number of occurrences provided, on average, similar results and patterns. The patterns were consistent for both time periods (current and future) and for both suitable and unsuitable areas, with only marginal differences between suitable and present times (Fig. 3A and 3C) and unsuitable future times (Fig. 3F and 3H).

When background records were sampled proportional to the number

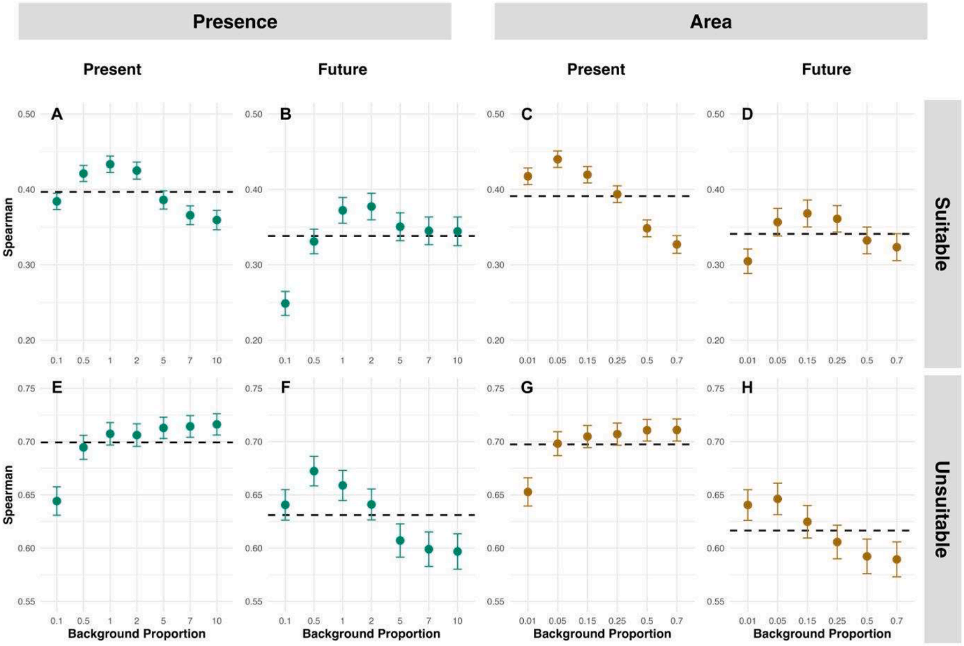


Fig. 3. Differences in model performance across background sampling strategies in current and future times, evaluated for suitable (upper row) and unsuitable areas (lower row) under present and future climate scenarios. Sampling strategies include 'Presence' (background points proportional to the occurrence points) and 'Area' (background points proportional to the area size). Black dashed lines show the mean of all background proportions combined for each suitability-time combination.

of occurrences (Fig. 3), background proportion 1 consistently exhibited the highest average performance under suitable areas (Fig. 3A). In future projections for suitable areas (Fig. 3B), only significant differences were found among proportion 0.1 with the rest of the proportions (Table S2), while only significant differences in proportion of background records appeared between 0.1 and 5, 7, 10 in the present-unsuitable and 0.5 with 10 in future-unsuitable scenarios (Fig. 3E and 3F, respectively).

When background records were sampled proportional to the area size (Fig. 3), the present-suitable combination demonstrated a response to varying background proportions (Fig. 3C). In this case, the best performing background proportion was 0.05 (sampling 5 % of the area as background points). Only for future scenarios under suitable areas proportion of 0.1 had significant differences with 0.15 (Fig. 3D). Under unsuitable areas (Fig. 3G and H), no pairwise comparison showed significant differences after Bonferroni’s correction (Table S2).

Overall, background selection appears to enhance prediction accuracy particularly in suitable areas under present conditions (Fig. 3) where differences in background proportions become statistically significant (Table S2), and selecting one proportion over another meaningfully impacts model performance. For future scenarios or within unsuitable areas, selecting one proportion over another showed no substantial effects on model performance (Fig. 3 and Table S2), with the only difference appearing when comparing the lowest proportions to the highest ones.

The optimal proportions can be observed, on average, with a 1:1 ratio of background records when sampling proportionally to presences, and with approximately 5 % of the study area sampled when proportional to the region. Future scenario projections generally performed slightly worse than current predictions. Notably, the projections demonstrated higher accuracy in unsuitable areas and achieved better performance metrics compared to those in suitable areas.

Nevertheless, disclosing the effect of background across the different prevalence groups provides valuable insights for each time-suitability combination (Fig. 4), which helps to understand how prevalence modulates the effect of background on model performance.

Across currently suitable areas (Fig. 4A and 4C), species with lower prevalence tended to achieve the highest model performances, indicating that narrowly distributed species are more effectively modelled under current climate conditions. In contrast, under other suitability contexts—including areas that are currently unsuitable or unsuitable in the future—species with intermediate prevalence consistently exhibited the best model performance (Fig. 4). Notably, for future scenarios in suitable areas (Fig. 4B and 4D), species with a prevalence of 0–0.1 and 0.2–0.3 achieved the highest scores. In all scenarios involving unsuitable areas, both current and future, the prevalence category of 0.2–0.3 consistently yielded the best performance.

These results highlight that prevalence should be carefully considered when parametrizing background settings, as prevalence levels respond differently to background ratios. While highly prevalent species tend to produce the poorest-performing models, it is not always the least prevalent species that yield the highest performance. Interestingly, only within a specific time-suitability combination (present-suitable) do scarcely represented species achieve the best scores.

If modelling results are disclosed by algorithm (Figures S1-S2), each behaved slightly different. BRT performed better in present-suitable conditions, GLM, however, seemed to have a particular better performance for unsuitable areas under future conditions. RF seemed to be the most sensitive to prevalence (Figure S2), followed by BRT and GLM.

Regarding variable importance (Fig. 5), niche-related factors, especially niche position (OMI), emerged as important predictors for explaining variance in Spearman correlation across all time periods and suitability bins. However, background proportion stood out particularly when predicting present-suitable conditions, reinforcing the finding that the choice of background proportions plays a relatively minor role in future scenarios or unsuitable areas.

On the whole, the variance explained by the explanatory models ranged between 74 % and 92 %. This suggests that the variables included in the random forest regression effectively account for differences in Spearman values. Current scenarios consistently exhibited slightly higher explained variance than future scenarios, while unsuitable areas outperformed their suitable counterparts in this regard.

4. Discussion

We provide evidence that background proportions in Species Distribution Models (SDMs) are primarily relevant for current predictions,

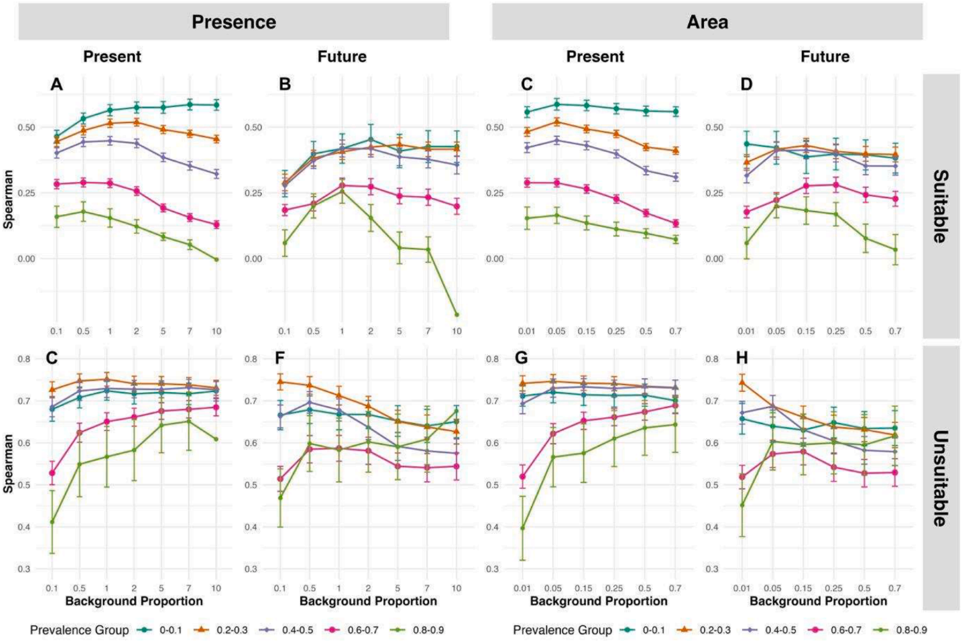


Fig. 4. Impact of background proportions on model performance across different prevalence levels and each sampling strategy analyzed. This figure illustrates the relationship between the Spearman correlation index (model performance) and species background proportion for suitable (A-B) and unsuitable areas (E-H) under current and future scenarios. Each dot represents the average accuracy score for 280 virtual species and the standard error.

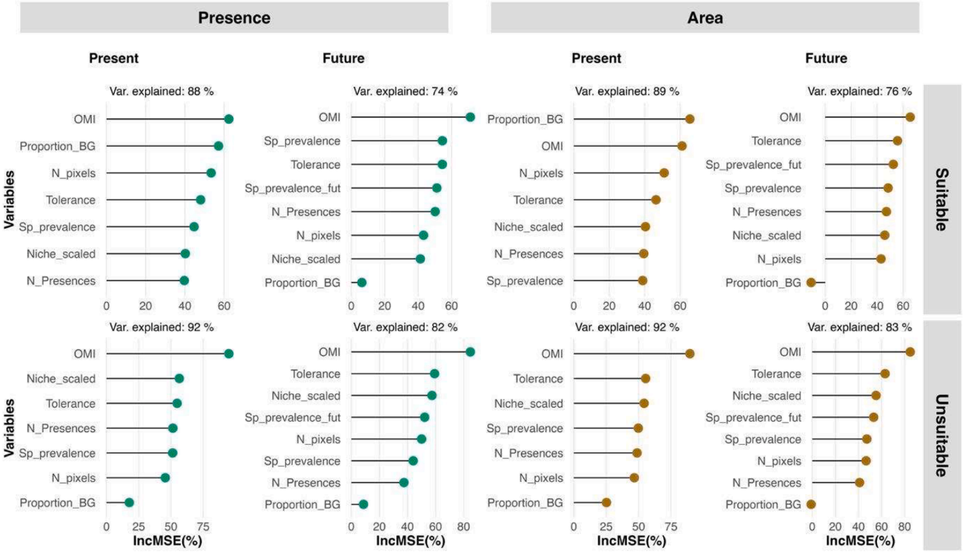


Fig. 5. Variable importance and explained variance for models using presence-based background sampling. This figure presents the importance of variables, measured as the Increase in Mean Standard Error (IncMSE), in explaining model performance depending on the background generation strategy selected. The percentage of variance explained by the model is also included for both present and future times and suitable (top row) and unsuitable areas (bottom row). Proportion_BG; Background proportion. OMI; Outlying Mean Index (niche position). N_pixels; Number of available cells or pixels in the study area. Tolerance; Niche breadth. N_presences; Number of occurrences used. Sp_prevalence; Species prevalence in the present. Sp_prevalence_fut; Species prevalence in the future. Niche_scaled; Niche marginality.

with a negligible influence on future projections under climate change scenarios. These findings partially align with Whitford et al. (2024), who explored how background prevalence affects model reliability in the present albeit for a smaller set of virtual species. Additionally, our analysis reveals that the influence of varying numbers of background records differs between suitable and unsuitable areas. This distinction is crucial for understanding how numbers of background records impact model performance and interpretation. For future projections, Santini et al. (2021) explored how background records might influence predictive capacity in forecasts. However, their study did not identify specific configurations of background settings that consistently improved performance. To the best of our knowledge, our study is the first to provide clear and systematic evidence on how varying numbers of background records affect SDM performance, addressing both current predictions and future projections.

4.1. *Model evaluation*

Model evaluation can be used to verify how well a model fits the training data or a non-independent test sample, such as a random sample from the data, or it can be used to assess how well models predict an independent test data (Araújo and Guisan, 2006). Regardless of whether the focus is on an 'internal' or an 'external' evaluation, a metric is needed to measure the agreement between observations and predictions. For SDMs, performance is frequently assessed using discrimination metrics like the widely known Area Under the Curve of the Receiving Operating Characteristic (AUC-ROC; Hanley and McNeil, 1983). Despite its popularity, AUC-ROC has faced significant criticism particularly when applied to presence-background data (Jiménez and Soberón, 2020; Jiménez-Valverde, 2022; Lobo et al., 2008); a criticism that while focusing often on AUC-ROC, should extend to any metric using presence and absence to evaluate models fitted with presence only data. Similarly, the Boyce index (Boyce et al., 2002; Hirzel et al., 2006), designed as a calibration metric specifically for presence-only data, can be problematic for prevalent species (species with widespread distributions in the study area), as it penalises large areas with high predicted probability.

Recognizing that the choice of the evaluation metric can lead to different interpretations of model performance, we opted to evaluate performance with a direct evaluation approach. Specifically, we compared the suitability maps of each virtual species (the known truth) with the predicted suitability maps, which is a different exercise compared to building an SDM without an underlying assumption that a true distribution exists. This approach reduces the number of assumptions inherent to the evaluation metrics available, such as the existence of balanced presence-absence proportions, and provides a more realistic reflection of SDM under varying background treatments.

It is important to note that, on average, model performance was low, particularly in the future projections. Predicted unsuitable areas exhibited higher accuracy compared to suitable ones. This moderate overall performance could be attributed to the thinning process carried out for the occurrence data, which resulted in a reduced set of data points that might have not fully reflect the actual distribution of the species. Sparseness of data, especially if they exacerbate environmental biases, are known to significantly affect model predictive capacity in SDMs (e.g., Araújo et al., 2009). To minimize the loss of rare environmental conditions, future studies might consider applying thinning directly within environmental space. However, such limitations are commonplace in real world scenarios, where species distributions are often under sampled or poorly characterized.

4.2. *Background generation strategy*

Strategies for determining the numbers of background records (i.e. proportional to the area or proportional to the number of occurrences) to use in SDM fitting generated extremely similar performances. We believe that the average performance is likely more influenced by data quality itself (e.g. how well the randomly sampled data represents the true distribution) rather than by specific background proportion. This observation does not imply that background is not relevant; on the contrary, our results underscore its importance as a predictor at least for present-suitable conditions (Fig. 5). However, selecting an optimum background proportion within a range established by data quality, could further improve the performance of the SDMs. Data quality essentially sets the upper and lower bounds of potential performance, while background setting can help achieve the maximum performance values determined by the quality of the data. This observation is particularly useful for modellers, who often have limited control over data quality but can optimize background parameters, such as the background proportion to improve model performance.

Contrary to our expectations, we observed minimal differences in average performance across background generation strategies, despite the varying ranges of background records. For example, we anticipated that a background range of 50–4000 records would yield different average performances compared to a range of 150–15,000 records for the same species and area. However, this was not the case. One possible explanation is that averaging all the species together, instead of observing species-specific responses, is obscuring a subtle relationship between species, background numbers, and model performance. Another possibility is that we have yet to fully account for the relationship between area size and the number of occurrences (Brotons et al., 2004; Segurado and Araújo, 2004), which could play a role in determining the optimal background proportion. Steen et al. (2024) also explored background sampling strategies. They did not test these specific approaches, but found that sampling in environmental space led to more stable models but used a fixed background sample size. Future research could examine how varying this number in the environmental space impacts model performance, which remains unknown. Interestingly, other studies explored the "how" question introduced by Barbet-Massin et al. (2012) from different perspectives (Fernandez et al., 2022; Schartel and Cao, 2024), but using a fixed number of background data in their multiple tests, which still requires to assume and unknown number of background to use regardless of the methodology used to sample the points.

Additionally, an unexamined aspect in this study is habitat heterogeneity, which could introduce variability into the selection of the optimal number of background records to adequately sample environmental conditions, introducing noise in the relationship between background records and performance.

4.3. Optimal proportion

Determining the optimal proportion of background points has been a prominent topic of debate in the SDM literature (e.g., Barbet-Massin et al., 2012; Tong et al., 2023; Whitford et al., 2024). Despite its importance, only a limited number of studies have systematically explored the variance in model performance relative to the number of background points. In this context, the findings of Whitford et al. (2024), derived from a smaller set of virtual species, align closely with our findings. Although the studies are not directly comparable, they also found that model accuracy-measured as the percentage of overlap between IUCN range maps and predictions was highest when the number of background points matched the number of occurrences. This corresponds to our observed optimal performance at proportions ranging from 0.5–1 when sampling background points proportional to the size of occurrence data (Fig. 3), a common proportion as well in logistic regression (Hosmer and Lemeshow, 2000). However, this study found a consistent pattern of background proportion optimization only for virtual species under current conditions, with no clear pattern emerging for real-world species. This discrepancy is interesting because it might indicate that virtual species simulations, and associated interpretations, may have limited applicability in real-world conditions. Factors such as biotic interactions and historical contingencies, which can disrupt species-environmental equilibrium (Araújo and Pearson, 2005; Munguía et al., 2012), are often not accounted for when fitting SDMs, potentially reducing their effectiveness in natural settings that are highly constrained by these dynamics (Peterson et al., 2011)

Further empirical evidence for the complexity of background point optimization comes from Liu et al. (2019), who conducted a comprehensive analysis using 1800 virtual species with varying prevalence levels. Their study identified four distinct response patterns in model performance as the number of background points increased: increasing, decreasing, arch-shaped, and horizontal. Our findings align with their observation that model performance is not universally improved by larger background samples, and that optimal quantities often depend on both the modelling technique and species characteristics. Their work also demonstrated that ensemble models consistently outperformed individual models, particularly with a small presence sample size. In our case, the general response of the individual models (i.e. how each algorithm reacted to background levels) did not particularly resemble the response pattern of ensemble models, suggesting an emerging and novel response from ensemble models. These findings are particularly relevant in light of Santini et al.'s (2021) comprehensive assessment of SDM reliability. Their review revealed that 85 % of recent SDM studies used presence data with background points, yet nearly half (48.7 %) did not report their background point sampling strategy. Furthermore, their virtual species simulations demonstrated that sample size was one of the main predictors of actual model accuracy, though it had limited influence on estimated accuracy. This disconnection between estimated and actual performance underscores the importance of careful background point selection and highlights potential risks in relying solely on traditional validation metrics when optimizing background point quantities.

4.4. Suggestions for improving performance according to their prevalence

SDMs provide predictions of probability of occurrence or habitat suitability, typically ranging from 0 to 1. To achieve reliable results, these models require sufficient contrast in the data. A key challenge in modelling generalist or widespread species, as observed for the highly prevalent species (Fig. 4), is the lack of contrast between areas where the species occur and where they do not. If a priori this information is noisy, this limitation makes it difficult for algorithms to discriminate suitable habitats from unsuitable ones. A potential underlying mechanism for decreasing SDM performance with larger background proportions, could be attributed to the decrease in contrast of background observations and those points associated with species presence due to increasing spatial autocorrelation. To address this problem, we suggest that modellers strive to use a reduced set of background points (e.g. 1:1 ratio or inferior), especially if there is a suspicion that they are dealing with a prevalent species in their study area, which are inherently more difficult to model (Segurado and Araújo 2004).

For prevalent species, we suspect that three different approaches could improve model performance, although all three present certain challenges. The first would be reducing species prevalence by increasing the extent of the study area, or at least to avoid truncating the extent to geopolitical or artificial boundaries if there is not a proper ecological explanation for it. By doing this, we would be including also areas where the species has access to but does not occupy, helping the algorithm to learn the contrast in occupancy. The second one would imply focusing background sampling on areas that are potentially or demonstrably unsuitable. It might not be possible, but more ecological thought addressing these gaps might be beneficial before running SDMs. The third one would be to reduce the number of background points. These three approaches would enhance predictive accuracy by helping algorithms to better distinguish suitable cells from unsuitable ones, as it would reduce noise in suitable areas and increment the presences' signal in suitable areas. However, true prevalence or unsuitable areas are usually unknown in presence-background models, and therefore, the application of these suggestions might be limited.

For species with narrow niches, background areas outside the species' range are more likely to represent biologically meaningful absences—areas where the species could potentially occur but are absent due to reduced fitness. However, for species with high prevalence, determining whether a background truly represents an absence is more complex. As illustrated in Fig. 4 (comparison between suitable and unsuitable areas for prevalent species), increasing the background for such species leads to two simultaneous outcomes:

1) The additional background reduces the signal of occurrences by wrongly occupying cells in potentially suitable areas, creating a negative correlation between background size and performance.
2) The representation of truly unsuitable areas improves, increasing contrast and thereby enhancing predictive accuracy.

The latter effect appears stronger for unsuitable areas, resulting in greater absolute changes in performance between the smallest and largest background proportions. This phenomenon may lead to the conclusion that larger background sample sizes are universally better (Barbet-Massin et al., 2012; Liu et al., 2019; Morán-Ordóñez et al., 2017; Phillips and Dudík, 2008). However, such conclusions often overlook the differential impact of background size on suitable and unsuitable areas.

To mitigate these issues and if possible, researchers should carefully sample backgrounds from meaningful unsuitable areas—locations where the species could potentially occur but fails to thrive due to environmental, biological, or historical factors. Incorporating such considerations into the background sampling process will likely improve model contrast and accuracy, especially for challenging scenarios involving prevalent species.

4.5. Why do unsuitable areas perform better than suitable ones?

The reasons behind the superior performance of unsuitable areas compared to suitable ones are not entirely clear, but a plausible explanation lies in the impact of background occurrence data on model performance. Background points generated in suitable areas, where the species could potentially occur, are likely to introduce noise. This noise arises because such background points overlap in areas that could hold presences, reducing the contrast between presence and absence data, which is critical for model accuracy. In contrast, increasing the number of background points in unsuitable areas tends to enhance model performance, as these areas are solely represented by absence-like conditions without the confounding presence of species data.

This difference may explain why unsuitable areas often yield better performance metrics than their suitable counterparts. In unsuitable areas, background records consistently provide a clear signal of absence, enabling models to accurately identify unsuitable conditions. On the other hand, in suitable areas, the presence of both occurrence and background points can confuse the model, leading to reduced accuracy in predicting suitability. This overlap likely dilutes the strength of the signal provided by presence data, making it harder for the algorithm to discern true suitability from the background noise.

4.6. Why does background have little effect on future projections?

Our findings suggest that background proportions become irrelevant when future projections are made, a significant observation given the limited knowledge available on this topic until now.

We hypothesize that this could be attributed to changes in the proportions and representation of the environmental conditions within the study area over time. During model training, prevalent environmental conditions (the most common ones) tend to be well characterized by the occurrence data, including both presence and background points. In contrast, the marginal, which are less common, are often underrepresented in the models. This imbalance becomes problematic when marginal conditions in the present shift and become more prevalent in future scenarios, increasing the number of cells that were poorly characterized during training.

Even though strict extrapolation was avoided in our study, the shift in the relative proportions of environmental conditions could still reduce the resolution of the models. Our analysis only assessed the percentage of future cells that remained within the training range, without accounting for changes in the distribution of environmental conditions. If marginal conditions that were poorly characterized in the present become more widespread in the future, the models may struggle to accurately predict suitability in these areas.

This limitation could explain the observed lower performance rates of future projections, particularly for the most sensitive species to changes in the environment, such as those with restricted distributions or low prevalence (Fig. 3). Therefore, environmental shifts within the training range can alter the proportions of accurately characterized cells, leading to models with reduced resolution and diminished performance in future projections. Addressing this issue will require novel strategies for accounting for changes in the representation and distribution of environmental conditions over time.

5. Conclusion

This study provides guidance for SDM practitioners by demonstrating how background proportions influence model performance in the current climate condition (baseline) and offering insights into their limited relevance for future projections. By highlighting the interaction between species prevalence and background proportion, we show that SDMs can be optimised even when data quality is constrained. This research offers one of the first comprehensive evaluations of background effects on future projections, advancing our understanding of how to refine SDMs to support varying biodiversity modelling applications.

CRediT authorship contribution statement

Armand Rausell-Moreno: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Núria Galiana:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Babak Naimi:** Writing – review & editing, Software, Methodology. **Miguel B. Araújo:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

We declare no conflict of interests.

Acknowledgments

Funding was provided by the Spanish Ministry for the Ecological Transition and Demographic Challenge to the CSIC through project 20233TE007 for the preparation of technical documents supporting biodiversity protection against climate change (Art. 24, Law 7/2021). We want to thank also the laboratory for management and analysis of massive biological data from the MNCN, particularly to DLQ and CSIC's Drago team for the help in scaling up computation capacities to allow this kind of research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2025.111177.

Data availability

All the necessary R code is provided as supplementary files.

References

Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D., Thuiller, W., 2011. Climate change threatens European conservation areas. Ecol. Lett. 14, 484–492. https://doi.org/10.1111/j.1461-0248.2011.01610.x.

Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. Sci. Adv 5, 4858–4874.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2006.09.010.

Araújo, M.B., Pearson, R.G., 2005. Equilibrium of species' distributions with climate. Ecography 28, 693–695. https://doi.org/10.1111/j.2005.0906-7590.04253.x.

Araújo, M.B., Pearson, R.G., Thuillier, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Glob. Chang. Biol. 11, 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.001000.x.

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93, 1527–1539. https://doi.org/10.1890/11-1930.1.

Araújo, M.B., Thuiller, W., Yoccoz, N.G., 2009. Reopening the climate envelope reveals macroscale associations with climate in European birds. Proc. Natl Acad. Sci. 106, 34. https://doi.org/10.1073/pnas.0813294106. –34.

Araújo, M.B., Williams, P.H., Fuller, R.J., 2002. Dynamics of extinction and the selection of nature reserves. Proc. R Soc. B 49, 1971–1980. https://doi.org/10.1098/rspb.2002.2121.

Baker, D.J., Maclean, I.M.D., Goodall, M., Gaston, K.J., 2022. Correlations between spatial sampling biases and environmental niches affect species distribution models. Global Ecol. Biogeogr. 31, 1038–1050. https://doi.org/10.1111/geb.13491.

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.

Barker, J.R., MacIsaac, H.J., 2022. Species distribution models: administrative boundary centroid occurrences require careful interpretation. Ecol. Modell. 472. https://doi.org/10.1016/j.ecolmodel.2022.110107.

Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. Ecol. Modell. 157, 281–300.

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

Brotons, L., Thuiller, W., Araújo, M.B., Brotons, A.H.H., Jo, A., Hirzel, M.B., Thuiller, /. W., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27, 4.

Broussin, J., Mouchet, M., Goberville, E., 2024. Generating pseudo-absences in the ecological space improves the biological relevance of response curves in species distribution models. Ecol. Modell. 498, 110865. https://doi.org/10.1016/j.ecolmodel.2024.110865.

Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol. Modell. 210, 478–486. https://doi.org/10.1016/j.ecolmodel.2007.08.010.

Chessel, D., Dufour, A., Thioulouse, J., 2004. The ade4 Package - I: One-Table Methods.

Dolédec, S., Chessel, D., Gimaret-Carpentier, C., 2000. Niche separation in community analysis: a new method. Ecology 81, 2914–2927. https://doi.org/10.1890/0012-9658(2000)081[2914:NSICAA]2.0.CO;2.

Dray, S., Dufour, A.-B., 2007. Journal of Statistical Software The ade4 Package: implementing the Duality Diagram for Ecologists.

Egorova, Y., Reygondeau, G., Cheung, W.W.L., Pakhomov, E.A., 2024. Species distribution models for mesopelagic mesozooplankton community. J. Biogeogr. https://doi.org/10.1111/jbi.15011.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Epele, L.B., Williams-Subiza, E.A., Bird, M.S., Boissezon, A., Boix, D., Demierre, E., Fair, C.G., García, P.E., Gascón, S., Grech, M.G., Greig, H.S., Jeffries, M., Kneitel, J. M., Loskutova, O., Maltchik, L., Manzo, L.M., Mataloni, G., McLean, K., Mlambo, M. C., Oertli, B., Pires, M.M., Sala, J., Scheibler, E.E., Stenert, C., Wu, H., Wissinger, S. A., Batzer, D.P., 2024. A global assessment of environmental and climate influences on wetland macroinvertebrate community structure and function. Glob. Chang. Biol. 30. https://doi.org/10.1111/gcb.17173.

Fernandez, M., Sillero, N., Yesson, C., 2022. To be or not to be: the role of absences in niche modelling for highly mobile species in dynamic marine environments. Ecol. Modell. 471. https://doi.org/10.1016/j.ecolmodel.2022.110040.

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.

Fratello, M., Tagliaferri, R., 2019. Decision Trees and Random Forests. Encyclopedia of Bioinformatics and Computational Biology. Elsevier, pp. 374–383. https://doi.org/10.1016/B978-0-12-809633-8.20337-3.

Glad, A., Mallard, F., 2022. Alpine marmot (Marmota marmota) distribution evolution under climate change: the use of species distribution models at a local scale in the western Pyrenees massif (France). Ecol. Inform. 69. https://doi.org/10.1016/j.ecoinf.2022.101646.

Grimmett, L., Whitsed, R., Horta, A., 2021. Creating virtual species to test species distribution models: the importance of landscape structure, dispersal and population processes. Ecography 44, 753–765. https://doi.org/10.1111/ecog.05555.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., Mcdonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H. P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16, 1424–1435. https://doi.org/10.1111/ele.12189.

Hanberry, B.B., He, H.S., Palik, B.J., 2012. Pseudoabsence generation strategies for species distribution models. PLoS One 7. https://doi.org/10.1371/journal.pone.0044486.

Hanley, J.A., McNeil, B.J., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148, 839–843.

Harsch, M.A., HilleRisLambers, J., 2016. Climate warming and seasonal precipitation change interact to limit species distribution shifts across Western North America. PLoS One 11. https://doi.org/10.1371/journal.pone.0159184.

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. Modell. 199, 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression. Wiley. https://doi.org/10.1002/0471722146.

Hysen, L., Nayeri, D., Cushman, S., Wan, H.Y., 2022. Background sampling for multi-scale ensemble habitat selection modeling: does the number of points matter? Ecol Inform 72. https://doi.org/10.1016/j.ecoinf.2022.101914.

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.

Jiménez-Valverde, A., 2022. The uniform AUC: dealing with the representativeness effect in presence–absence models. Methods Ecol. Evol. 13, 1224–1236. https://doi.org/10.1111/2041-210X.13826.

Karasiewicz, S., Dolédec, S., Lefebvre, S., 2017. Within outlying mean indexes: refining the OMI analysis for the realized niche decomposition. PeerJ 2017. https://doi.org/10.7717/peerj.3364.

Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. virtualspecies, an R package to generate virtual species distributions. Ecography 39, 599–607. https://doi.org/10.1111/ecog.01388.

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.

Lobo, J.M., Jiménez-valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecol. Biogeogr. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Lumbierres, M., Dahal, P.R., Soria, C.D., Di Marco, M., Butchart, S.H.M., Donald, P.F., Rondinini, C., 2022. Area of Habitat maps for the world's terrestrial birds and mammals. Sci. Data 9. https://doi.org/10.1038/s41597-022-01838-w.

Luna, S., Peña-Peniche, A., Mendoza-Alfaro, R., 2024. Species distribution model accuracy is strongly influenced by the choice of calibration area. Biodiversity Informatics.

MacPherson, M.P., Jahn, A.E., Murphy, M.T., Kim, D.H., Cueto, V.R., Tuero, D.T., Hill, E. D., 2018. Follow the rain? Environmental drivers of Tyrannus migration across the New World. Auk 135, 881–894. https://doi.org/10.1642/AUK-17-209.1.

Mainali, K.P., Warren, D.L., Dhileepan, K., Mcconnachie, A., Strathie, L., Hassan, G., Karki, D., Shrestha, B.B., Parmesan, C., 2015. Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. Glob. Chang. Biol. 21, 4464–4480. https://doi.org/10.1111/gcb.13038.

Marchese Robinson, R.L., Palczewska, A., Palczewski, J., Kidley, N., 2017. Comparison of the Predictive Performance and Interpretability of Random Forest and Linear Models on Benchmark Data Sets. J. Chem. Inf. Model. 57, 1773–1792. https://doi.org/10.1021/acs.jcim.6b00753.

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.

Martínez-Fonseca, J.G., Westeen, E.P., Jenness, J., Zahratka, J.L., Chambers, C.L., 2024. Species distribution models predict potential habitat for the endangered New Mexico jumping mouse. J. Wildl. Manage. https://doi.org/10.1002/jwmg.22646.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Routledge. https://doi.org/10.1201/9780203753736.

McCulloch-Jones, E.J., Kraaij, T., Crouch, N., Faulkner, K.T., 2023. Assessing the invasion risk of traded alien ferns using species distribution models. NeoBiota 87, 161–189. https://doi.org/10.3897/neobiota.87.101104.

Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., Van Den Berg, M., Velders, G.J.M., Vollmer, M.K., Wang, R. H.J., 2020. The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geosci. Model. Dev. 13, 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020.

Morán-Ordóñez, A., Lahoz-Monfort, J.J., Elith, J., Wintle, B.A., 2017. Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? Global Ecol. Biogeogr. 26, 371–384. https://doi.org/10.1111/geb.12545.

Munguía, M., Rahbek, C., Rangel, T.F., Diniz-Filho, J.A.F., Araújo, M.B., 2012. Equilibrium of global amphibian species distributions with climate. PLoS One 7. https://doi.org/10.1371/journal.pone.0034420.

Naimi, B., Araújo, M.B., 2016. Sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39, 368–375. https://doi.org/10.1111/ecog.01881.

Naimi, B., Capinha, C., Ribeiro, J., Rahbek, C., Strubbe, D., Reino, L., Araújo, M.B., 2022. Potential for invasion of traded birds under climate and land-cover change. Glob Chang Biol 28, 5654–5666. https://doi.org/10.1111/gcb.16310.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37 (2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 161–175. https://doi.org/10.1111/j.2007.0906-7590.05203.x.

Sanczuk, P., De Lombaerde, E., Haesen, S., Van Meerbeek, K., Van der Veken, B., Hermy, M., Verheyen, K., Vangansbeke, P., De Frenne, P., 2022. Species distribution models and a 60-year-old transplant experiment reveal inhibited forest plant range shifts under climate change. J. Biogeogr. 49, 537–550. https://doi.org/10.1111/jbi.14325.

Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., Huijbregts, M.A.J., 2021. Assessing the reliability of species distribution projections in climate change research. Divers Distrib 27, 1035–1050. https://doi.org/10.1111/ddi.13252.

Schartel, T.E., Cao, Y., 2024. Background selection complexity influences Maxent predictive performance in freshwater systems. Ecol. Modell. 488. https://doi.org/10.1016/j.ecolmodel.2023.110592.

Scornet, E., Biau, G., Vert, J.-P., 2015. Consistency of random forests. Annals Statist. 43. https://doi.org/10.1214/15-AOS1321.

Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. J. Biogeogr. 31, 1555–1568. https://doi.org/10.1111/j.1365-2699.2004.01076.x.

Steen, B., Broennimann, O., Maiorano, L., Guisan, A., 2024. How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecol. Modell. 493. https://doi.org/10.1016/j.ecolmodel.2024.110754.

Stevens, G.C., 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. Am. Nat. 133, 240–256.

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.

Tong, R., Yesson, C., Yu, J., Luo, Y., Zhang, L., 2023. Key factors for species distribution modeling in benthic marine environments. Front. Mar. Sci. 10. https://doi.org/10.3389/fmars.2023.1222382.

Triviño, M., Kujala, H., Araújo, M.B., Cabeza, M., 2018. Planning for the future: identifying conservation priority areas for Iberian birds under climate change. Landsc. Ecol. 33, 659–673. https://doi.org/10.1007/s10980-018-0626-z.

Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M.B., Balaguer, L., Benito-Garzón, M., Cornwell, W., Gianoli, E., van Kleunen, M., Naya, D.E., Nicotra, A.B., Poorter, H., Zavala, M.A., 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecol. Lett. 17, 1351–1364. https://doi.org/10.1111/ele.12348.

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. Modell. 220, 589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010.

Whitford, A.M., Shipley, B.R., McGuire, J.L., 2024. The influence of the number and distribution of background points in presence-background species distribution models. Ecol. Modell. 488. https://doi.org/10.1016/j.ecolmodel.2023.110604.

Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M., Schwartz, M.W., 2009. Using species distribution models to predict new occurrences for rare plants. Divers Distrib. 15, 565–576. https://doi.org/10.1111/j.1472-4642.2009.00567.x.

Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol. 9. https://doi.org/10.1186/1472-6785-9-8.