



How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels

Bart Steen^{a,b,*}, Olivier Broennimann^{a,c}, Luigi Maiorano^{b,1}, Antoine Guisan^{a,c,1}

^a Institute of Earth Surface Dynamics, University of Lausanne, CH-1015 Lausanne Switzerland

^b Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

^c Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne Switzerland

ARTICLE INFO

Original content: [Stability \(similarity\) values of output maps generated using different background point sampling strategies](#)

Keywords:

SDMs
Ecological niche models
Background points
Pseudo-absences
Environmental space
Model stability
Performance metrics
Invasive alien species

ABSTRACT

Species distribution models (SDMs) have become central tools in ecology and biogeography. Although they can be fitted with different types of species data (e.g. presence-absence, abundance), the most common approach, based on data from large species repositories, is to use simple occurrences (i.e. presence-only) combined with background points (BP; also called pseudo-absences). But how should we sample these background points, and how does this choice affect SDMs? In most studies so far, BP were sampled randomly in geographic space, yet theory rather suggests, if a species is at equilibrium, that it is better to sample them in a stratified way in environmental space. However, this potential improvement of SDM predictions has never been tested. Furthermore, a typical assumption behind SDMs is that the modelled species are at equilibrium with their environment. But how do these models perform when species are in disequilibrium, as is the case for most invasive species? To answer these questions, we selected 30 different species (10 insects, 10 mammals and 10 plants; for each group 5 were invasive and 5 were considered at equilibrium) and for each we calibrated SDMs with different types of background selections: random in environmental space, random-stratified in environmental space, random in geographic space, and random-stratified in geographic space. For each SDM we assessed both predictive performance using standard metrics and their stability using a new approach that compares the model's habitat suitability projection with those of a SDM calibrated with virtual occurrence data generated from the most suitable areas. Finally, we compared the predictive performance of species distribution models of invasive alien (disequilibrium) species versus native (equilibrium) species by comparing model stability and performance metrics of the two groups. We found that sampling BP in a stratified-random way in environmental space yields the highest performance metrics, and that sampling fully randomly in environmental space yields the most stable models. This has implications for the use of SDMs in conservation, as the classical and frequently used fully random in geographic space BP are found to produce both less accurate and less stable models. Our results indicate that the best approach is to use stratified random in environmental space BP sampling if accuracy is essential, and fully random in environmental space BP sampling if model stability is essential.

1. Introduction

Species Distribution Models (SDMs), also known as Ecological Niche Models (ENMs), constitute a widespread method for ecological, biogeographic, and conservation studies (Franklin 2010; Peterson et al., 2011; Guisan et al. 2013; Guisan et al. 2017; Araujo et al. 2019), often being particularly useful when dealing with invasive alien species (IAS) (Gallien et al. 2012; Slodowicz et al. 2018; Vicente et al. 2019). SDMs take geographic observations of species and georeferenced

environmental datasets in a study area to statistically identify the conditions in which the species can occur (i.e., a part or all of the species' realized environmental niche), and map the occurrence probability of the species based on these conditions back in geographic space (Colwell and Rangel 2009). The usefulness of SDMs depends both on model accuracy, i.e., how well a model is able to make predictions, and model stability, i.e., the degree of model sensitivity to stochasticity (Duan et al. 2014; Grimett et al. 2020; Mateo et al. 2010). Therefore, consistency of geographical predictions is a pivotal factor to research.

* Corresponding author at: Institute of Earth Surface Dynamics, University of Lausanne, CH-1015 Lausanne Switzerland.

E-mail address: bart-steen@unil.ch (B. Steen).

¹ Shared last authorship

<https://doi.org/10.1016/j.ecolmodel.2024.110754>

Received 7 December 2023; Received in revised form 8 May 2024; Accepted 9 May 2024

Available online 17 May 2024

0304-3800/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The vast majority of SDMs use presence-only data, likely because occurrence data in national or global databases (e.g., the Global Biodiversity Information Facility ([GBIF]); Anderson et al. 2016, Fletcher et al. 2019, Chapman et al. 2020) have become numerous and easily accessible (Chauvier et al. 2021; Valavi et al. 2021, Nolan et al. 2022). In contrast, absence data are more rarely available and hard to generate, especially in the case of animal species (Gormley et al. 2011; Lobo et al. 2010). Presence-background SDMs function by contrasting the species occurrences to points sampled (often randomly) in the study area (Valavi et al. 2022), ideally along the main environmental gradients used to define the species niches (Barve et al. 2011; but see Iturbide et al. 2018 for other possibilities). This set of points representing the environment are called background points (BP; often also pseudo-absences, but see Senay et al. 2013) and are most typically sampled considering a geographically random scheme (Liu et al. 2019; Chauvier et al. 2021; Barber et al. 2022). Therefore, presence-background SDMs make the implicit assumption that BP sampled in geographic space represents the available environment (measured as all environmental combinations within the study area) appropriately (Barbet-Massin et al. 2012; Wisz and Guisan 2009; Liu et al. 2019, Cengic et al. 2020, Valavi et al. 2022).

However, since SDMs fit response curves along environmental predictors, this random selection of BP should theoretically be conducted in environmental rather than geographical space and in a stratified way (Fraginière et al. 2022). This is also supported by the fact that, in order to produce accurate and robust SDM results, species occurrence data should optimally be sampled in the field in a random-stratified in environmental space manner (Hirzel and Guisan 2002).

Whilst there have been studies that used background points selected in environmental space, this was usually done in a variety of stratified fashions, arranging the background points along environmental gradients (e.g. Fraginière et al. 2022). It has also recently been shown that geographically constraining background point selections to areas that are geographically or environmentally close to the occurrence points can strongly influence SDM predictive accuracy (Whitford et al. 2024; Schartel and Cao 2024). However, it was rarely done fully randomly in environmental space. Here, we present the first study that also uses a fully random selection of background points in environmental space, allowing us to test the four possible types of selection based on geographic versus environmental spaces, and stratified or not, and thus whether the implicit assumption that the geographic space reflects the environmental space is acceptable or not when fitting SDMs. As Hutchinson's niche-biotope duality tells us that the geographic space does not reflect the environmental space (Colwell and Rangel 2009; Guisan et al. 2014), the question is whether it can still give a good approximation.

Another major assumption made when using SDMs is that the modelled species is at equilibrium with its environment. This is already not true for all native species (Normand et al. 2011; Svenning and Sandel 2013), but it is almost never true in the case of invasive alien species (IAS), as their spread leads them to progressively occupy new habitats and territories (Gallien et al. 2012; Petitpierre et al. 2012). Therefore, statistical species-environment relationships usually only represent a snapshot "pseudo-equilibrium" in time (Guisan & Zimmermann 2000). The question is, then, how much a disequilibrium causes problems to SDM fitting and predictions (Svenning and Sandel 2013)? Therefore, the method of sampling background points may be even more important when a species is not at equilibrium with the environment.

Here, we address this second dimension of (dis)equilibrium with the environment together with our assessment of background point selection, by using species with different putative degrees of equilibrium. More specifically, we make the simplified assumption that well-established native species, i.e., those for which we know that they have a dynamic that has not changed recently, are likely at equilibrium inside their home range and fill their realized niche. In this study, we want to answer the two following research questions around the two assumptions previously exposed:

- (1) Which method of BP sampling produces the best-performing SDMs for invasive and equilibrium species in terms of model accuracy and stability?
- (2) Do the optimal BP sampling strategy differ between native and invasive alien species?

2. Materials and methods

2.1. Species selection and study area

In our analyses, we considered both invasive alien species (IAS), i.e., species not at equilibrium with the environment, and species which we assume to be at the equilibrium (see Table 1). Both invasive and equilibrium species included 5 insect species, 5 plant species and 5 vertebrate species, making a total of 30 target species.

All equilibrium species are mountain species, which are naturally distributed along steep elevational and environmental gradients, thus making their distribution constrained in geographic space (i.e., where mountains are). Also, this limited our study area to mountain biomes, which was advantageous for computational considerations. A final reason to select non-generalist mountain species for equilibrium studies is because whilst many of these species might physiologically be able to occur at lower altitudes, as demonstrated by growing them in botanical gardens (Vetaas 2002), competition with other species drives them away from there. For instance, the Alpine aster (*Aster alpinus*) usually occurs at elevation of over 2000 m because it gets outcompeted at lower altitudes (Lyu and Alexander, 2022). High-altitude species therefore are likely to occupy their entire realized niche, and therefore they are at (pseudo-)equilibrium with their realized environment in mountains.

We selected 15 IAS from the EU list of Union Concern (https://ec.europa.eu/environment/nature/invasivealien/list/index_en.htm) and the factsheets on invasive alien species in Switzerland provided by the

Table 1

Study species and matching rationale to insert them into the studies and number of occurrences (in the study area, after removal of non-native range occurrences for equilibrium species and thinning). Note that the invasive alien species all at least occur in mountain areas in their home range.

Category	Species group	Species name	Number of occurrences
Invasive	Plants	<i>Asclepias syriaca</i>	1342
		<i>Heracleum mantegazzianum</i>	833
		<i>Impatiens glandulifera</i>	769
		<i>Lespedeza cuneata</i>	236
		<i>Lupinus polyphyllus</i>	820
	Vertebrates	<i>Branta canadensis</i>	920
		<i>Myocastor coypus</i>	400
		<i>Ondatra zibethicus</i>	281
		<i>Sciurus niger</i>	686
		<i>Tamias sibiricus</i>	77
	Insects	<i>Aedes albopictus</i>	95
		<i>Cameraria ohridella</i>	225
		<i>Harmonia axyridis</i>	1377
		<i>Ips typographus</i>	77
		<i>Popillia japonica</i>	794
Equilibrium	Plants	<i>Alnus alnobetula</i>	164
		<i>Aster alpinus</i>	86
		<i>Helictrotrichon pratense</i>	70
		<i>Leontopodium nivale</i>	423
		<i>Ranunculus glacialis</i>	107
	Vertebrates	<i>Lagopus muta</i>	64
		<i>Lepus timidus</i>	286
		<i>Marmota marmota</i>	123
		<i>Nucifraga caryocatactes</i>	433
		<i>Tetrao urogallus</i>	445
	Insects	<i>Boloria pales</i>	75
		<i>Colias phicomone</i>	87
		<i>Erebia meolans</i>	140
		<i>Parnassius apollo</i>	212
		<i>Parnassius mnemosyne</i>	93

Swiss Federal Office for the Environment (FOEN) (OFEV, O. fédéral de l'environnement (n.d.)), [invasive_alien_species_in_switzerland_factsheets.pdf](#). The Swiss lists were chosen because of the high abundance of high-altitude species and extensive work on the effect of non-native flora and fauna on biodiversity. It was ensured that these species at least occur in mountains in their home range and that most of their occurrence data (see Section 2.4) were inside the study area (see Section 2.1). The list of species is given in Table 1 (see Appendix 1, table S5 for the list along with the rationale to include the species).

We defined our study area using the biomes from Olson et al. (2001) where our species occur (Fig. 1): temperate broadleaf and mixed forests, temperate coniferous forests, temperate grasslands, savannahs and shrublands, montane grasslands and shrublands, boreal forests/taiga and tundra. We excluded biomes where the presence of our species was marginal (see Appendix 3, Table S31 for occurrence points omitted per species). For an additional analysis on the potential effect of the omitted occurrence points on the model outcomes, showing that the environmental values of the occurrence points off the biomes are represented into the environmental values of the background points and of the occurrence points included in the study, see Appendix 3 (Figs. S32 and S33).

2.2. Algorithm choice and settings

MaxEnt (which stands for 'maximum entropy'), a type of Poisson point process regression model (Renner and Warton 2013), is among the most widespread algorithms used to fit SDMs (Elith et al. 2011; Gomes et al. 2018). It also draws upon presence-only (and background) data, which is much more widely available than data on presence-absence of species (Fletcher et al. 2019; Guillera-Aroita et al. 2015). To avoid model overfitting, the ENMeval package version 2.0.3 (Kass et al. 2021) was used to find the most optimal model settings. MaxEnt uses several mathematical function types (linear, quadratic, hinge, product, threshold) and the Regularization Multiplier (RM), a controller of model complexity, to identify the relationship between data occurrence and the underlying environmental variables (Phillips et al. 2009). ENMeval calibrates a MaxEnt model by pointing out the optimal settings of the function types and RM. Since ecological interactions are never accurately described by only one type of function, e.g., only linear functions, we only chose three combinations of functions: 1) linear, quadratic and hinge (LQH), 2) linear, quadratic and product (LQP), and 3) linear, quadratic product and hinge (LQPH). The RM was varied from 1 to 5. Threshold functions were never used because hinge functions have been shown to be a better replacement (Elith et al. 2011).

2.3. Environmental variables

Bioclimatic variables were downloaded from CHELSA, version 1.2 (Karger et al., 2017), at a resolution of 0.0083°. This version was chosen because the variables are composed of the average values from the time series of 1979 – 2013 and species responses to climate change often lag behind the actual climate at a given time (Menendez et al. 2006). Though this may vary among species groups, we make this assumption in this study, for the sake of comparability. An average for a time period is therefore more informative than choosing the values at one specific time point. The Ecospat R package (Di Cola et al. 2017; Broennimann et al. 2023) was used to investigate correlation among the variables (i.e., collinearity) and to select the best set of (partly) uncorrelated variables to insert in the models. A correlation limit of 0.70 between variables was used (Dormann et al. 2013). In the end, seven variables were used: mean annual temperature, diurnal temperature range, total temperature range, annual precipitation, precipitation seasonality, precipitation of driest quarter, precipitation of coldest quarter.

2.4. Species occurrence data

We downloaded data from the Global Biodiversity Information Facility (GBIF; Anderson et al. 2016) on 30 species (GBIF.org, 2022a,b,c, GBIF.org, 2023). Only occurrence records from the year 2000 or later were selected, in order to match the CHELSA version 1.2 timeline. The delay from 1979–2013 to 2000–2022 is assumed to make up for the delayed species response (see Section 2.3). To properly focus on our question of equilibrium state, only the native range of species is modeled. Conversely, IAS are expected to be at disequilibrium with the environment (Gallien et al. 2012). For these species, all the known range (in the study area) was modeled in this study. This is firstly to research the effect of this disequilibrium on model stability, and secondly because more is often known about those species in their invaded range than in their native range (Poland et al. 2021).

Many species that are not listed as invasive are still naturalized alien in some countries, which might mean that they are not at equilibrium in part of their range. Therefore, the DASCO R package (Hseebens 2022) was used to identify the native distribution range of the species at equilibrium and all occurrence records from outside that native range were removed. Then, occurrence records that were outside the study area were removed. For an additional analysis on the potential effect of this on niche truncation and model's ability to accurately characterize the niche, see supplementary material, Appendix 3, Figs. S32 and S33. The data were subsequently 'cleaned'. This implied removal of unlikely coordinates, occurrences with coordinate uncertainty over 50 m, observations with zero abundance and/or an occurrence status of "absent" (in many languages) and visual inspection of the cleaned data set. This was done using the scrubr R package ([scrubr source : R/scrubr-package.R; Downloading and cleaning GBIF data with R | R-bloggers](#)). For species with more than 20'000 occurrence records, 20'000 occurrence points were selected at random, for reasons of computing efficiency. This number of occurrence points was considered big enough to represent the species' distribution range effectively, thus we lost no information in this step. Finally, the occurrence data for each species were thinned (Steen et al. 2021, Inman et al. 2021) as follows. First, a Minimum Convex Polygon (MCP) was drawn around the (remaining) points. This is a technique to draw the borders of the home range of the species. Within this MCP, as many points as there are occurrence data were randomly drawn, following Chiocchio et al. (2021). This process was repeated 50 times and each time, the average nearest neighbor distance between the randomly drawn points was recorded. The average of these averages was calculated, and the species occurrence data was thinned by this distance using the spThin R package (Aiello-Lammens et al., 2015). This method matches the clustering of the occurrence records to clustering expected for a random distribution, as it returns the number of points expected in the species' home range and therefore limits sampling bias of occurrence points. The downloaded occurrence data were then thinned to 1 point per grid cell of our environmental rasters (Steen et al. 2021) because our version of MaxEnt only accepts one occurrence point per grid cell. To limit the risk of sample size influencing our results, we kept only species that had at least 50 occurrence records remaining for further analyses, to be well above the limit of 30 occurrences under which no algorithm predicts consistently (Wisz et al. 2008).

Originally, many more species were considered to be included in the study, however, this method of thinning often removed too many records.

2.5. Background points sampling

We sampled 10,000 background points for each sampling strategy. It is generally acknowledged that this is a sufficient background point sample size for SDMs built using MaxEnt (Barbet-Massin et al. 2012). Each of the 4 methods of background point sampling – fully random in environmental space (randenv), stratified random in environmental space (stratenv), fully random in geographic space (randgeo) and

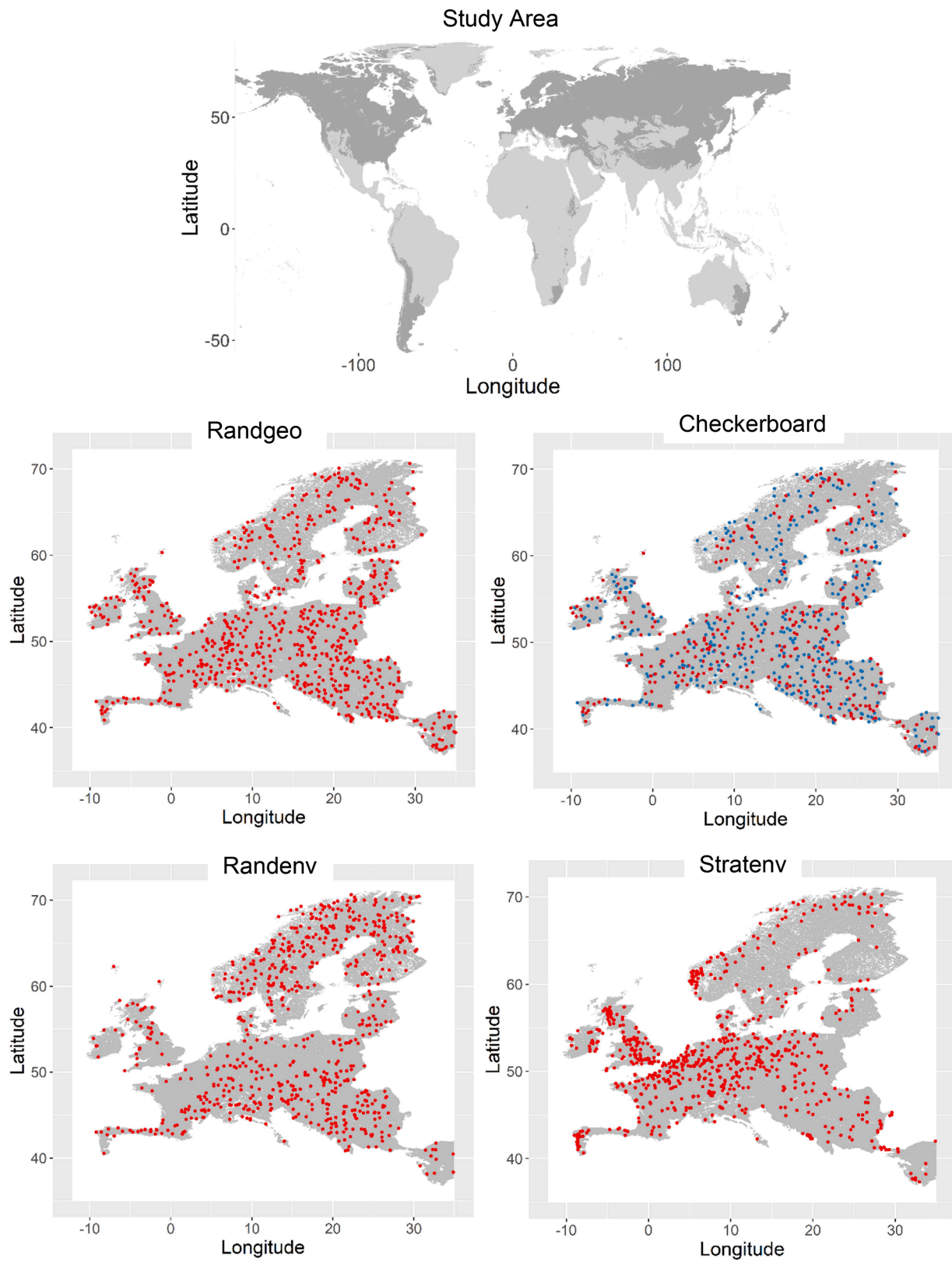


Fig. 1. Study area and comparison of methods of background creation. The background point datasets are shown for Europe only, in order to facilitate visual comparisons: Study area (plotted in dark gray); Fully random in geographic space (randgeo); Checkerboard treatment. The different bins of the data are shown in red and blue. Note that the grid size of the checkerboard blocks have been enlarged to 100 km in order to show the checkered division of the BP data across the bins; Fully random in environmental space (randenv); Stratified-random in environmental space (stratenv). Note we did not display the stratified random in geographic space method here because the checkerboard models also use the random in geographic space background points (in B). Please print in color.

stratified random in geographic space (presented by checkerboard 1 (cb1) and checkerboard 2 (cb2)) – is described below. The background points were sampled from the study area described under Section 2.1. See Fig. 2 for a schematic explanation. All the background sample sizes were downweighted to the number of occurrences.

2.5.1. Fully random in environmental space (randenv)

The simplest approach would be to draw all points randomly in environmental space at once, but when the available environment of the study area is very restricted (as here with mountains), it becomes difficult to make such random points match actual conditions in the field. For this reason, we had to develop a specific procedure. First, a PCA analysis of all the environmental variables was performed using the “PCA” function of the R package “FactoMineR” (Husson et al. 2020). The first two PCA axes explain 74.4% of the variation. Then, a two-dimensional kernel density estimate was performed on the two axes using the “kde2D” function of the R package “MASS” (Venables et al. 2002). Based on this, we divided the environmental dataset into 10 quantiles, each of which contained roughly 10% of all the values in the PCA axes, see Appendix 1, supplementary Fig. S3. Then, the values of the axes were rounded to 1 decimal. Finally, 1000 background points were sampled randomly across all 10 quantiles, resulting in a final dataset of 10,000 BP, as follows: First, random numbers ranging between the minimum and the maximum value of the first PCA axis were generated. The same was done for the second PCA axis. Then, to perform the random sampling, both random numbers were matched to the values of both axes and when there was a match, the corresponding longitude and latitude were recorded into the randenv background point dataset. When performing the random sampling, two conditions had to be met: 1) the randomly sampled values on both the PC1 axis and the PC2 axis had to match the values in the matching cell and 2) no duplicate values of either PC1 or PC2 were allowed. The latter condition was used in order to prevent oversampling of more common values. Thus, we made sure that the final dataset of background points was one that at least approached a representative, fully random sample of all the environmental conditions available. For additional details, see Appendix 1, supplementary Fig. S3.

2.5.2. Stratified random in environmental space (stratenv)

The 7 CHELSA raster layers were each divided into 3 equal environmental strata as follows. First, the difference between the minimum and maximum value in a layer was calculated. Then, all the cells with values that fell within a one-third quantile of the group were selected, thus making three strata per layer. Each of these strata was assigned a unique numeric code which was not included in any of the strata of other layers. Therefore, each matching pixel across raster layers has seven different numeric codes. Then, a new raster layer was created in which

all the numeric codes were summed, creating a unique new code for each combination of strata. Finally, the *sampleStratified* function of the raster package was used to sample across this new layer and create the stratenv background points. This function makes an effort to sample across all the cells in the summed layer that have the same code. Therefore, all strata were sampled (more or less) equally. Since, in environmental space, the number of the strata (and thus their resolution) should not affect the results, we used only one here (i.e. 3 strata per climatic layer).

2.5.3. Fully random in geographic space

This option is the most commonly used and therefore many options exist to conduct it. Here, the *randomPoints* function of the *dismo* (Hijmans et al. 2019) package was used to generate randgeo background points. This function samples fully randomly across the rows and columns of the geographic raster variables (omitting cells with no data), with weights on the rows, so not too many cells can be selected near the poles. Also, no cell can be selected more than once.

2.5.4. Stratified random in geographic space

The checkerboard method of the *ENMeval* package (Kass et al. 2021) was used to divide the randgeo background data (as well as the occurrence data) according to square “checkerboard” blocks (Steen et al. 2019). As, in geographic space (unlike in environmental space), the grain of the checkerboard could influence the result, we tested two different block sizes. The size of the checkerboard blocks was set at 5 km x 5 km to create the first set of bins (checkerboard 1 (cb1)) in one dataset and at 10 km x 10 km to create the second set (checkerboard 2 (cb2)). All the analyses described under Section 2.7 were done for all species, with both sets.

2.5.5. Background points comparison

The datasets of 10'000 background points used for each treatment are shown in Fig. 1. In order to investigate if the background point (BP) datasets are, as one would expect, spatially different from each other, we calculated the Moran's I for each dataset. See Fig. 2 for a finer illustration of the comparison between the classic fully random sampling in geographic space and the stratified random sampling in environmental space.

2.6. Optimal model selection and general modeling framework

The thinned occurrence datasets for the target invasive (non-equilibrium) and equilibrium species were randomly divided into two bins: one that contained 80% of their occurrence data and one that contained 20%, a common choice of partition sizes for SDMs (Cosentino et al. 2023). Only the 80% bin was used for the selection of the optimal

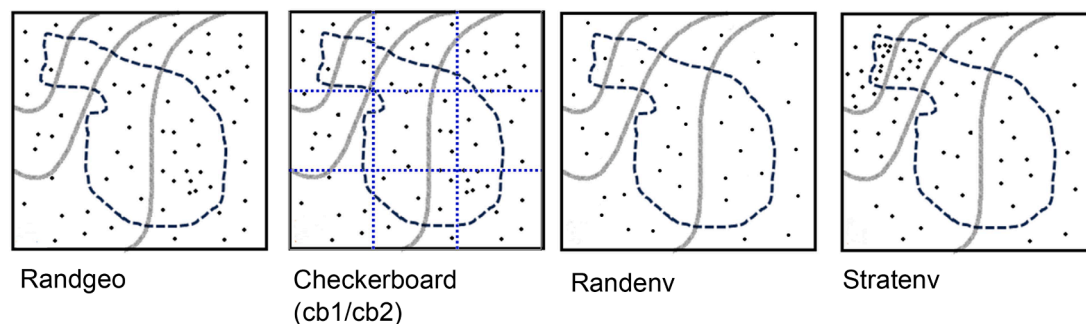


Fig. 2. Geographic representation of the classic fully random in geographic space (randgeo) background sampling, the stratified random in geographic space (i.e., checkerboard; cb1/cb2) approach, the fully random in environmental space (randenv) approach, where the number of BP is proportional to the size of the strata and the stratified random in environmental space (stratenv) background sampling, where the same number of points are sampled in any combination of environmental variables. The dots represent sampling points, plain lines represent isolines of some gradually varying environmental variable (e.g., temperature changing along elevation) and the dashed line polygon represents a patch of some landscape feature (e.g. a forest). Modified from Hirzel & Guisan (2002).

models. For each of these 80% bins, all pertaining to a different target species, 50 maxent models were run using the ENMeval package, with the MaxEnt.jar model method selected and using the aforementioned settings (see Section 2.2). Before running each maxent model, the 80% bin of data was randomly split 50–50 across a training bin and a validation bin. We diverted here from the more typically used 80–20 or 70–30 division because the bins used in the checkerboard analyses would split the data roughly evenly across training and validation datasets. The best model settings were selected based on their average value of 3 predictive performance metrics: the continuous Boyce index (CBI; Hirzel et al. 2006), the Area Under the ROC Curve (AUC, Jimenez et al. 2020) and the sum of both (see Fig. 4). The last one was calculated in order to not neglect models that had both (relatively) high Boyce (i.e., good calibration in predictions) and (relatively) high AUC (i.e., good discrimination in predictions). This is necessary, because it is possible for a model to have a high AUC and a low Boyce index or vice versa (i.e. balancing discrimination versus calibration in the evaluation; Guisan et al. 2017), and in both cases that would be a bad model from at least one perspective. Before the sum was calculated, the AUC was recalculated to the gini AUC ($2 \cdot \text{AUC} - 1$) ranging between -1 and +1 in order to give it the same “weight” as the Boyce index. This therefore yielded 3 optimal model settings (Boyce index, AUC and the sum of both) for each species. Subsequently, the full 100% of the occurrence data for each species was used to fit a final ‘full’ model using the identified optimal settings (James et al. 2013; Guisan et al. 2017) and again splitting the occurrence data 50–50 across the training and the validation bin in order to ensure that the checkerboard analyses are comparable to the other three treatments. This full model was finally projected in geographic space to obtain a habitat suitability map. This procedure was repeated for each species and all 4 methods of background selection. Note that only one dataset of background points for each method was used for the creation of SDMs in this entire study.

2.7. Model similarity analysis: a measure for model stability

To assess model stability, we used a form of virtual simulations based on the habitat suitability maps of the species generated from SDMs at the previous step. We refer to these SDMs and the matching distribution maps as being of generation 1. For this, we artificially considered the generation 1 prediction maps as “true”, meaning a perfect presentation of habitat suitability for the species (as in Thibaud et al. 2014). New occurrence points were then generated from each of these true habitat suitability maps, using the values of the map as probability of occurrence, i.e., the higher the habitat suitability was, the higher the likelihood that an occurrence point got generated there. For each species, 7 new occurrence datasets were created: ones of 10, 30, 50, 100, 300, 500 and 1000 occurrence points. For each of these datasets, the model fitting and selection procedures described in Section 2.6 were performed again, minus the splitting up of data into bins of 80% and 20%, i.e., the generation 2 data were divided 50–50, used to train 50 model replicates to identify optimal settings, then re-run with 100% of the data using those optimal settings. We underline that the virtual occurrence datasets were never resampled, i.e., the same occurrence record dataset was used for the 50 MaxEnt runs. These are the models of generation 2. The resulting habitat suitability maps were then compared to the original habitat suitability maps for each species as follows: first, the raster datasets were converted to vectors, and then their similarity was determined using the cor() function in the R language. Our measure of model stability is consequently the measure of spatial similarity between the original map and the map created using the virtual occurrences, i.e., through describing spatial autocorrelation. An overview of the study design is given in Fig. 3. See Appendix 1, Fig. S4 for an idealized presentation of maximum model stability and Appendix 1, Fig. S1 and Table S2 for the effect of the number of generated virtual points on stability.

Many models, especially generation 2 models with low numbers of generated occurrences (10, 30, sometimes 50) yielded no performance

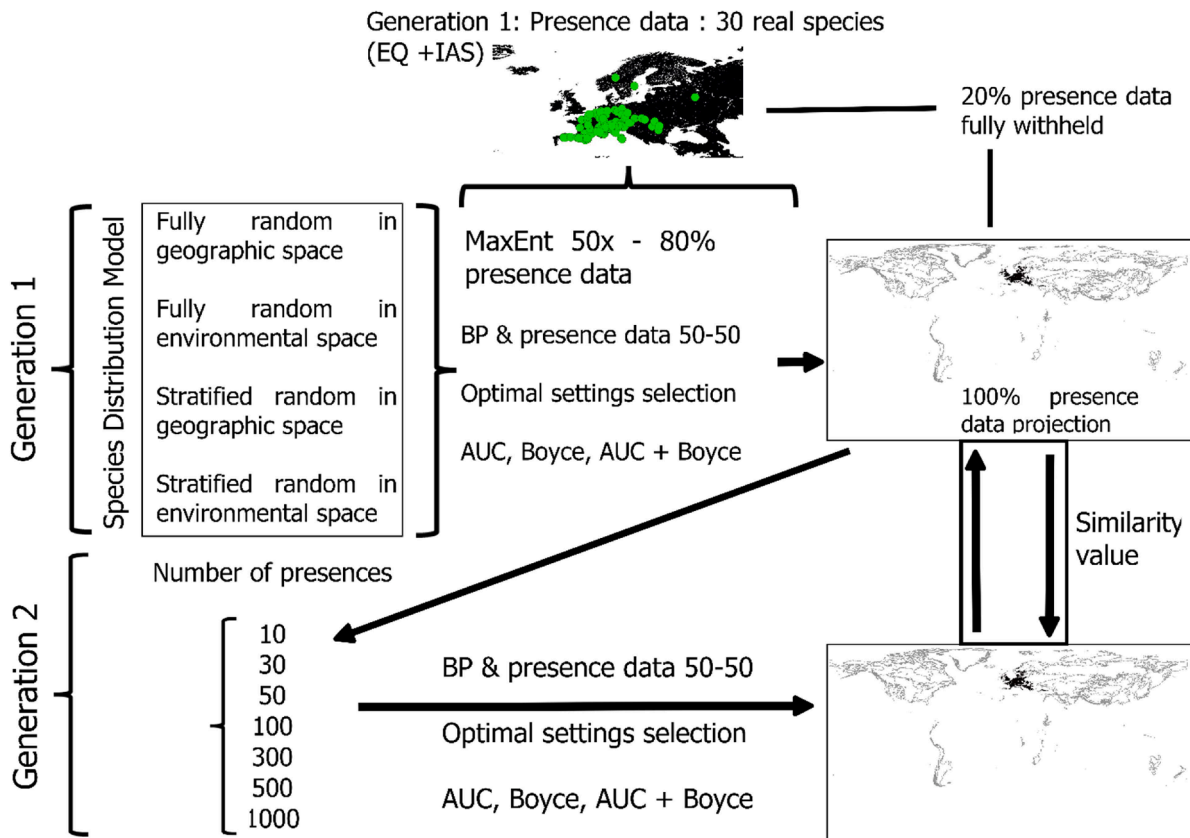


Fig. 3. Workflow and design of the study. Please print in color.

metrics (see Appendix 1, supplementary Table S4), which is logical, given the very low sample size and 50–50 random split of data. In such a case, the model was not included in the similarity assessments, because no optimal settings could be selected. As a result, the amount of similarity values differed across background point creation strategies: there were 1032 for stratenv, 887 for randenv, 931 for the classic randgeo, 885 for checkerboard 1 (cb1) and 920 for checkerboard 2 (cb2). We compared all the values of the strategies with Welch *t*-tests. Since we tested no additional explanatory variables, the *t*-tests were not corrected for multiple comparisons.

In order to address the different sample sizes, we also did an additional analysis in which we only used Welch *t*-tests after random elimination of values until the two compared strategies had the same number of values. This method was repeated 100 times over, with a new random sample of eliminated values each time. The Welch *t*-test was chosen because no equal variance could be assumed among both samples.

3. Results

The moran's I was $-3.967e-05$ for the randgeo dataset, $3.573e-05$ for the randenv background points and 0.001 for stratenv background points. Note that the same BP dataset was used for randgeo, cb1 and cb2; it was merely split into different bins for the latter two.

3.1. Model performance metrics comparison

Models calibrated for species at the equilibrium yielded significantly higher AUC and AUC+CBI in generation 2 for the stratenv and CBI treatments than invasive species models. In generation 1, no significant differences were found when comparing invasive species models to equilibrium species models.

When comparing performance metrics between background point selection strategies (regardless of level of species equilibrium with the environment), stratenv yielded higher AUC values than any other strategy in generation 1 models. No additional significant effects were found in this generation. In generation 2, more significant differences were found. Stratenv had higher AUC than any other strategy, and randenv had a higher AUC than all the geographically sampled BP datasets. The same pattern was observed for AUC+CBI. Stratenv also had higher Boyce index than all other treatments, except for randenv, with which there was no significant difference.

In generation 1, when comparing AUC, CBI and AUC+CBI between IAS and EQ species, within each background point selection strategy, no significant differences in performance metrics were found. When comparing results for equilibrium species and invasive alien species within each BP sampling strategy, significant differences in average AUC and AUC+CBI measures were found in generation 2 for stratenv background points, with a slightly higher average for equilibrium species. The same pattern was found in cb1 treatments. See Fig. 4 for a graphical presentation of these results and Appendix 1, Table S1 for the same results in tabular format.

3.2. Model stability

3.2.1. Effects of background points on model stability

When comparing all of the model stability values (i.e., the measure of similarity between a generation 2 model and the matching generation 1 model) between two treatments (i.e., not using the random elimination method), the results of the Welch *t*-tests indicate that the randenv strategy yields far higher model similarity (i.e., has far higher model stability) than any other strategy, for both invasive species and equilibrium species. Stratenv yields higher similarity values than the classic randgeo (Tables 2 and 3) and both checkerboard treatments. When randomly eliminating models, the pattern changes slightly. Randenv is still by far the highest in terms of model similarity and stratenv still yields significantly higher results than randgeo, but not higher than

either checkerboard treatment anymore (see Table 4).

3.2.2. Effects of equilibrium species vs invasive alien species on model stability

The results were analyzed by both Welch *t*-tests of all the values and again by random elimination, just as described under Section 2.7. Only the stratenv resulted in higher average model stability for equilibrium species than for invasive alien species, but this result was not found to be statistically significant. The other four background treatments (i.e. including two options for the stratified in geographic space: cb1 and cb2) demonstrated the opposite effect, but this time with significant results, except for cb2 (Fig. 5). The random elimination method however yielded significantly higher similarity values for equilibrium species in the stratenv treatment in 100% of the cases. The strategies randenv, randgeo and cb1 also yielded 100% significant results, but with a higher average for invasive species. For cb2, no difference was found in similarity values for invasive species and equilibrium species. See Appendix 1, Table S3 for the same results in tabular format.

3.3. Model stability vs performance metrics

The stratified random in environmental space background points models yielded the highest performance metrics, particularly AUC, for both generation 1 and generation 2 models, and the fully random in environmental space treatment yielded the highest model similarity (see Fig. 6). See Appendix 1, Fig. S2 for the same analyses on the CBI.

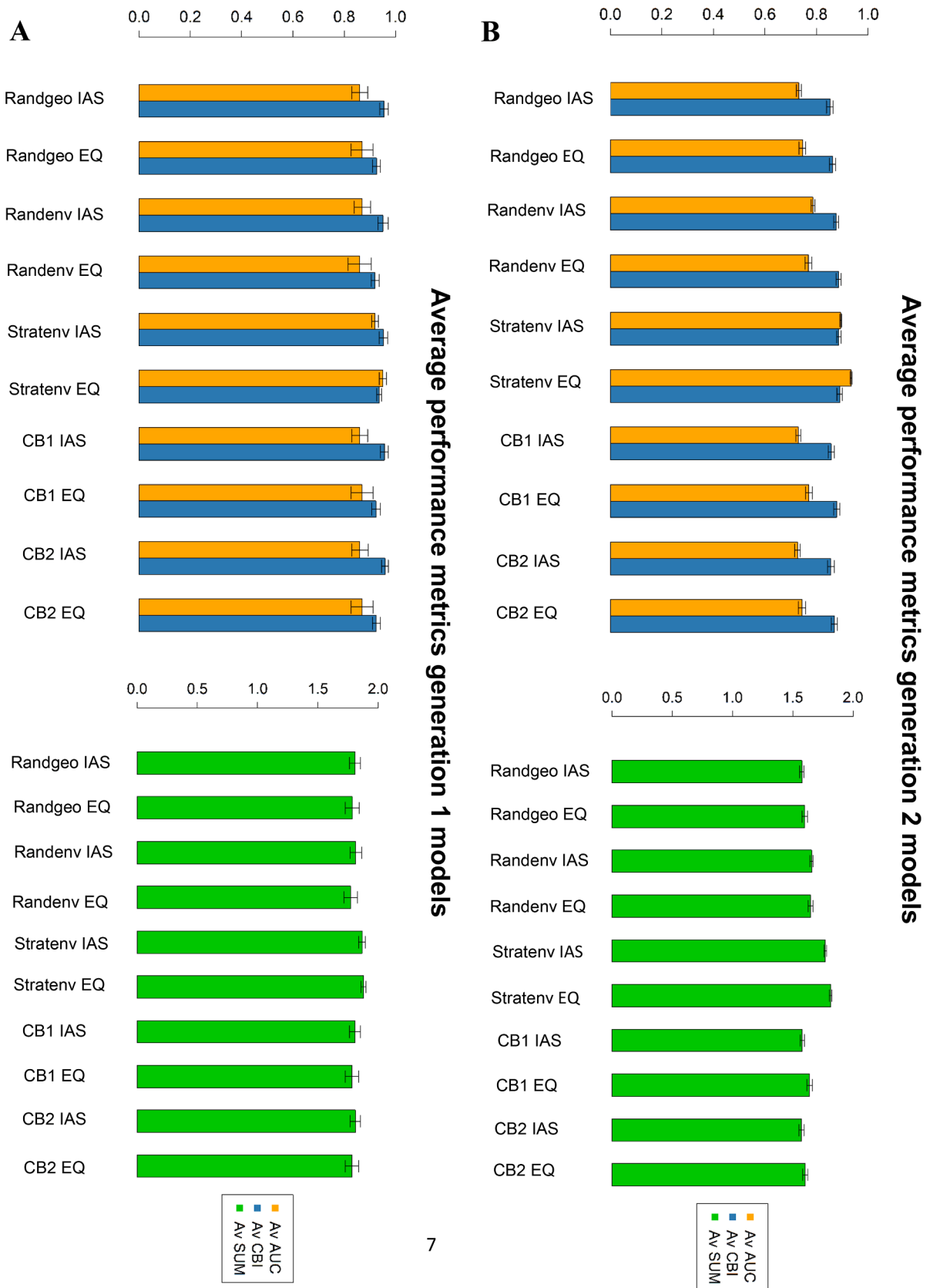
4. Discussion

This research focuses on two questions: (1) Which method of background points sampling produces the best-performing SDMs for invasive and equilibrium species in terms of model accuracy and stability? (2) Do these background point sampling strategies differ between native and invasive alien species?

4.1. What drives model stability?

Our results indicate that the classic randgeo sampling of background points produces less stable models (i.e., models with lower similarity values) than both strategies of sampling in environmental space. This is logical and intuitive because the environmental space is used to model the environmental niche (Guisan et al. 2017). In addition, it is also optimal to sample species observations in the field based on environmental space (Hirzel and Guisan 2002). The finding likewise concurs with previous findings that randgeo sampling can cause prediction biases (Botella et al. 2020; Freeman et al. 2022). However, sampling background points in geographic space is still much more common than sampling in environmental space (examples of the latter are e.g. Fragnière et al. 2022, Bazzichetto et al. 2023).

However, the randenv method of background point selection produces by far the most stable (measured by similarity) models for both invasive alien species and species at equilibrium, though the model stability is significantly higher for invasive alien species. This is consistent with the results of Bazzichetto et al. 2023, who used virtual species to show that sampling background points uniformly in environmental space (which is intended to also be fully random in environmental space) produces models that most closely approximate reality for species with wide environmental tolerances, which, in the real world, would include invasive alien species. Therefore, the superior model similarity we found for invasive alien species using this background sampling strategy could be explained by the fact that the stratenv data points are very good at capturing the extremes of the environment, whilst the randenv background points are more continuously spread along the core parts of the environmental gradients. The latter probably informs the models better, particularly the models on invasive alien species, because these are often not present in the



7

Fig. 4. Average performance metrics with error bars for invasive species (IAS) and equilibrium species (EQ) for models of generation 1 (A) and generation 2 (B). Please print in color.

Table 2

Comparison of model similarity (i.e., as a measure of model stability) per background sampling method across all species for method 1 (in which all models were kept.). Stratenv = stratified random in environmental space, randenv = fully random in environmental space and randgeo = fully random in geographic space, cb1 = checkerboard 5 km, cb2 = checkerboard 10 km.

	Stratenv	Randgeo	Randenv	cb1	cb2
Std error	0.0067	0.00748	0.0030	0.0083	0.0075
Average	0.4802	0.4420	0.8113	0.4534	0.4500

Table 3

P-values of *t*-tests comparing similarity score yielded for models per background point treatments for method 1.

	Stratenv	Randgeo	Randenv	cb1	cb2
Stratenv		0.00014	< 2.2e-16	0.0120	0.0028
Randgeo	0.00014		< 2.2e-16	0.3065	0.4482
Randenv	< 2.2e-16	< 2.2e-16		< 2.2e-16	< 2.2e-16
cb1	0.01202	0.3065	< 2.2e-16		0.7634
cb2	0.0028	0.4482	< 2.2e-16	0.7634	

Table 4

results of background point sampling strategy comparison when randomly eliminating values, to ensure equal sample size (i.e., method 2).

Strategy comparison	Stratenv	Randgeo	Randenv	cb1	cb2
Stratenv		100%	100%	0%	0%
Randgeo	100%		100%	100%	0%
Randenv	100%	100%		100%	100%
cb1	0%	100%	100%		0%
cb2	0%	0%	100%	0%	

extremes of the environment (e.g. there are very few IAS at high elevations (Fuentes-Lillo et al. 2021; Barros et al. 2022)). Therefore, the distribution of invasive alien species and the randenv BP are probably similar. This could be further reinforced by the fact that citizen science datasets like GBIF are subject to human sampling bias (Chauvier et al. 2021) and therefore contain less observations in high mountains, even for alpine species.

4.2. What drives model accuracy?

On the side of performance metrics, the stratenv background method yields higher AUC and AUC+CBI (our combined index) for equilibrium species than any other background treatment, particularly in generation 2 models. This is almost entirely due to higher AUC and the effect is noticeable in all equilibrium species models. Still, the ability of the stratenv models to yield higher AUC indicates that they are better at discerning suitable area from unsuitable area. Namely, a possible

explanation that they yield higher AUC but not higher Boyce is that the stratenv models include fewer false positives than the other methods, likely due to their ability to classify the absences of the species.

Our results therefore show that although randenv background treatment yields by far the most stable models, it is less good at modeling realistic species distributions than the stratenv treatment. This is likely because the randenv treatment distributes its background points more heavily in environmental values that are common, rather than in the extremes. Both presence points and randenv BP find themselves often in the most common environmental values. Hence, it is often easier for randenv models to identify relationships between the presence points and the environment and therefore create stable models, but they are not necessarily the right relationships. The latter are more likely to be identified by the stratenv background model treatments.

4.3. Do optimal BP selection strategies differ between equilibrium and non-equilibrium species?

When answering question 2 (if different BP selection strategies produce the most accurate and stable models for EQ species and IAS), we would expect that EQ species would yield higher performance metrics and more stable models, since SDMs assume that the species is at equilibrium with the environment and that the species occupies its whole realized niche. In this study, it is assumed that the latter condition is satisfied for EQ species, but not for IAS. Stratified sampling in environmental space was the only method where SDMs trained on occurrence data of equilibrium species yielded both higher performance metrics and higher stability values than for non-equilibrium species.

This might be because the stratenv background points are agglomerated into areas where environmental gradients are steep, such as in mountains. Previous studies indicate that model performance is better when background point sampling biases match the bias of the occurrence records (Botella et al. 2020; Freeman et al. 2022; Schartel & Cao (2024)). This may explain why the stratenv background point sampling method, unlike all other methods, yielded higher model similarity for equilibrium species than for invasive alien species: the distribution pattern of these background points matches the distribution pattern of the species. This agrees with the observation that modeling species not at equilibrium with the environment can lead to inaccurate predictions (Freeman et al. 2022). It is therefore possible that the stratenv sampling works better for equilibrium species than for alien species because the stratenv background points are much more abundant around mountains. Hence, the models may be more stable not because the species are at equilibrium, but because they are primarily mountain species. In addition, many of the equilibrium species chosen in this study have relatively narrow distributions, meaning that stratenv background points, which are agglomerated where the environment is heterogeneous, can provide the model with a more complete view of the habitat preferences of narrow-ranged species. This may indicate that the methods have great conservation applications, as narrow-ranged species are both vulnerable

Model similarity of EQ species vs model similarity of IAS across background sampling methods

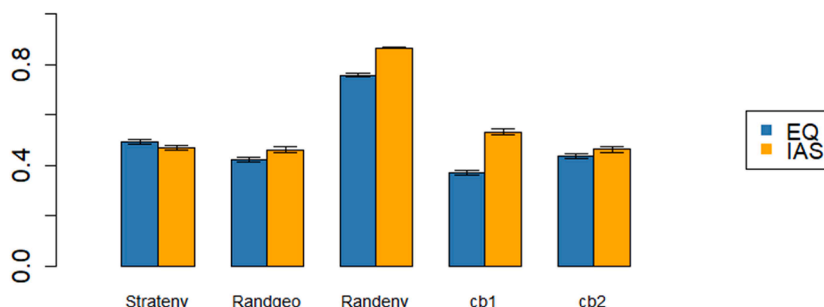


Fig. 5. Comparison of model similarity for equilibrium species and invasive alien species per background sampling method. Please print in color.

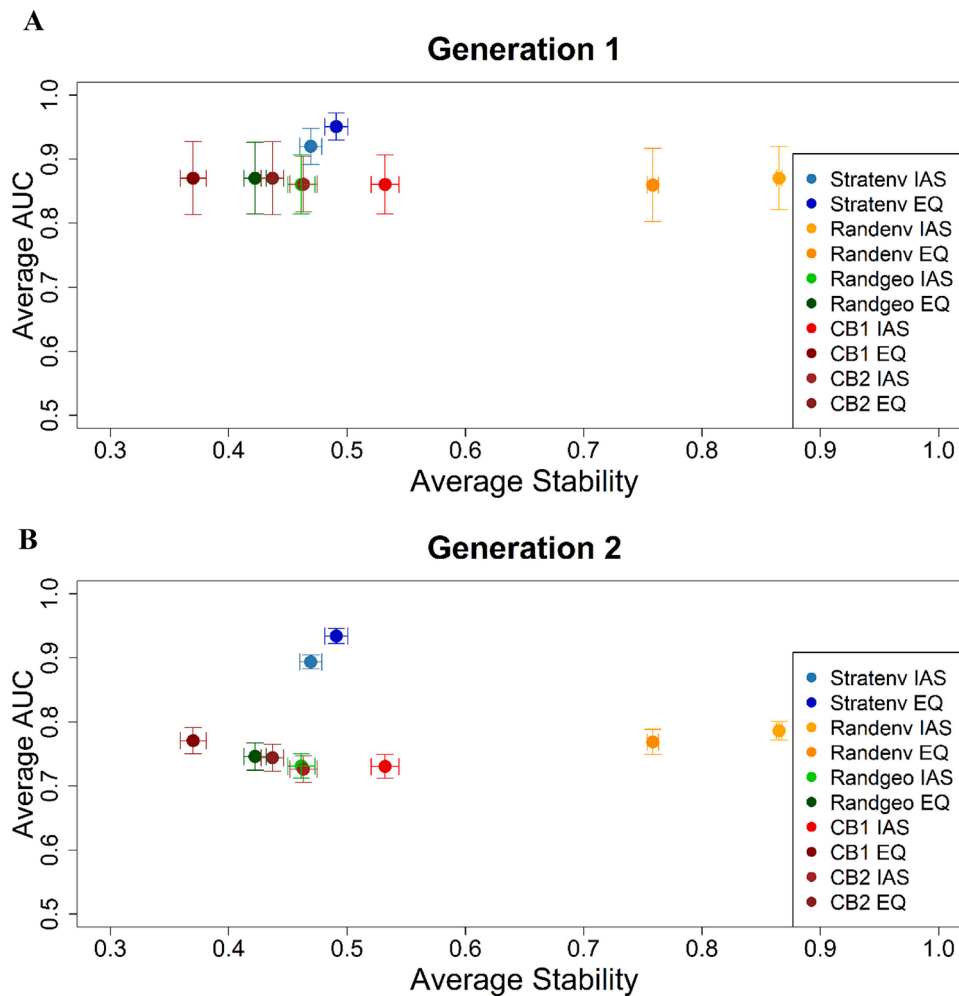


Fig. 6. Average model similarity vs average AUC for real species models of generation 1 (the models that were run using real occurrence data) (A) and generation 2 (the models that were run using virtual occurrence data generated from the output suitability maps of generation 1 models) (B). Please print in color.

to climate change and notoriously difficult to model due to low sample size (Hernández-Lambrano et al. 2020). However, environmental strata are very strongly correlated to topography and elevation, so likely, the random stratified sampling in environmental space informs SDMs better on the margins of the environmental tolerance of each species. Still, we cannot refute the possibility that stratenv sampling is better suited to equilibrium species, regardless of whether these occur primarily in mountains or not.

An additional criticism might be that the spatial autocorrelation of the stratenv BP dataset would lead to errors in model training and predictions. Since this has been proven to be true for occurrence datasets (Veloz 2009; Miller et al. 2012), it might also be true for BP datasets. However, model training is about statistical correlations between occurrences and environmental values, it takes place in environmental space, not in geographic space. Therefore, we do not see how model errors resulting from spatial autocorrelation of the stratenv BP models could artificially inflate the stability values. In addition, our sampling of the stratenv BP data should be (most) uniform along environmental gradients, not biased.

The better stability and performance results for invasive alien species than for equilibrium species yielded by all sampling strategies except stratenv can be explained by their distributions being spread relatively evenly over the study area, just like the distributions of the selected species. The species used for this study are widespread, at an advanced stage of invasion. Sampling strategies that cover the more common environmental values therefore inform the model better. It is also

possible that the stratenv background points produce models that are less sensitive to sample size, as the invasive alien species have on average more than twice as many occurrence points than the equilibrium species (see Table 1). A complementary argument is that since the chosen invasive alien species are at an advanced stage of invasion, they could be closer to equilibrium with the environment than initially assumed. This may also explain why the result of stratenv background points yielding higher model similarity results for equilibrium species is significant, but only barely: the invasive alien species are relatively close to equilibrium, as well.

We found that no performance metrics could be generated from many models of generation 2 (i.e., the models built using occurrences generated from the generation 1 habitat suitability maps) that were trained with only 10 or 30 occurrences. This is an additional finding that models trained with small sample size may still yield performance metrics (but see Collart and Guisan 2023) but fail our additional stability test. Our results therefore suggest that it is advisable to not run SDMs with fewer than 50 occurrence records (which has already been shown to be problematic (Adde et al. 2023)), and that testing with performance metrics may not be sufficient to evaluate SDMs. Null models are a possible measure of reliability that may alleviate this concern, as proposed in Collart and Guisan (2023).

4.4. Possible expansions

Further investigations expanding on our methods might yield

interesting results. For instance, different methods of assessing overlap of species' predictions, e.g., using Schoener's D (e.g. Warren et al. 2008) or Kulczynski's coefficient (e.g. Randin et al. 2006) instead of our correlation method used to calculate model stability could be used in future studies. This might help put our findings in the larger context of other studies, e.g. on calculating niche and distribution overlaps (Warren et al. 2008), or to compare model predictions in transferability assessments (Randin et al. 2006). Another potentially promising method to assess the effect of spatial autocorrelation on models would be to use Dutilleul's modified *t*-test, which is especially designed to account for the degree of spatial autocorrelation in statistical inference (Dutilleul et al. 1993). This test would thus be mostly ideal when one is interested in comparing two samples in the presence of spatial autocorrelation.

Secondly, limiting the SDM study area to the home range of the respective species might provide valuable insights, as large home ranges may artificially inflate AUC (Lobo, 2008). In this study, we have not done so, as the comparability of the models, particularly of the widespread IAS and the narrow-ranged EQ species, would be compromised.

Finally, alternate methods of thinning occurrence data might strongly influence performance metrics. We have already presented an alternate thinning method and the effect on model stability and performance metrics for two species in the supplementary materials (appendix 3), however environmental thinning of occurrence data has been shown to improve model performance metrics (Varela et al. 2014).

In conclusion, from our findings, we suggest sampling background points stratified randomly in environmental space when dealing with 1) low sample sizes, 2) species at equilibrium with the environment and 3) species with narrow ranges. Conversely, it is better to sample the background points randomly in environmental space when species are widespread and have a broad environmental niche. In addition, our results show that stratified random in environmental space background sampling yields both more accurate and more stable models than the classic fully random in geographic space sampling, whilst fully random in environmental space background sampling yields by far the most stable, but not necessarily the most correct models. These findings should however be confirmed by further studies.

CRediT authorship contribution statement

Bart Steen: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Olivier Broennimann:** Software, Methodology. **Luigi Maiorano:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Antoine Guisan:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization.

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 authors do not have permission to share data. [Stability \(similarity\) values of output maps generated using different background point sampling strategies](#)

Acknowledgments

We give thanks to the Swiss Confederation and to La Sapienza University of Rome for awarding the doctoral grants that made this research possible, as well as to the Institute of Earth Surface Dynamics of the University of Lausanne for providing additional financial support. BS and AG also benefitted from the Swiss national Valpar.CH project for additional support. In addition, we thank the University of Lausanne for

granting access to their High Performance Computing (HPC) cluster. Finally, we thank both of our anonymous reviewers for their swift responses, thoroughness, and insights, which have allowed us to greatly improve the scientific work presented here.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2024.110754](https://doi.org/10.1016/j.ecolmodel.2024.110754).

References

- Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38 (5), 541–545. <https://doi.org/10.1111/ecog.01132>.
- Anderson, R.P., Araújo, M.B., Guisan, A., Lobo, J.M., Martinez-Meyer, E., Peterson, A.T., Soberón, J., 2016. Are species occurrence data in global online repositories fit for modeling species distributions. The Case of the Global Biodiversity Information Facility (GBIF). Final report of the Task Force On GBIF Data Fitness For Modelling. Global Biodiversity Information Facility (GBIF), Secretary Office, Aarhus, Denmark.
- Araujo, 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 (1) <https://doi.org/10.1126/sciadv.aat4858> eaat4858 ARTN.
- Barber, R.A., Ball, S.G., Morris, R.K.A., Gilbert, F., 2022. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Divers. Distribut.* 28 (1), 128–141. <https://doi.org/10.1111/ddi.13442>.
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting Pseudo-Absences for Species Distribution Models: How, Where and How Many? <https://doi.org/10.1111/j.2041-210X.2011.00172.x>.
- Barros, A., Haider, S., Mullerova, J., Alexander, J.M., Alvarez, M.A., Aschero, V., Daehler, C., Peyre, G., Ratier Backes, A., Arévalo, J.R., Cavieres, L., Dar, P., Fuentes-Lillo, E., Liedtke, R., McDougall, K., Milbau, A., Morgan, J.W., Naylor, B.J., Nuñez, M.A., Pauchard, A., Rashid, I., Reshi, Z.A., Rew, L., Sandoya, V., Seipel, T., Vorstenbosch, T., Vítková, M., Neville, W., Wedegartner, R.E.M., Zong, S., Lembrechts, J.J., 2022. The role of roads and trails for facilitating mountain plant invasions. In: Barros, A., Shackleton, R., Rew, L., Pauchard, A. (Eds.), *Tourism, Recreation and Biological Invasions*. CABI publishing, London, pp. 14–26.
- Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Modell.* 222 (11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>.
- Bazzichetto, M., Lenoir, J., Da Re, D., Tordoni, E., Rocchini, D., Malavasi, M., Barták, V., Sperandii, M.G., 2023. Sampling strategy matters to accurately estimate response curves' parameters in species distribution models. *Glob. Ecol. Biogeogr.* 32 (10), 1717–1729. <https://doi.org/10.1111/geb.13725>.
- Botella, C., Joly, A., Monestiez, P., Bonnet, P., Munoz, F., 2020. Bias in presence-only niche models related to sampling effort and species niches: lessons for background point selection. *PLoS One* 15 (5), e0232078. <https://doi.org/10.1371/journal.pone.0232078>.
- Broennimann, O., DiCola, V., Guisan, A., 2023. *ecospat*: spatial ecology miscellaneous methods. R package version 4.0.0. < <http://www.unil.ch/ecospat/home/menuguid/ecospat-resources/tools.html>. >
- Chapman, A.D., Belbin, L., Zermoglio, P.F., Wiczorek, J., Morris, P.J., Nicholls, M., Schigel, D., 2020. Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. *Biodivers. Inf. Sci. Stand.* 4 <https://doi.org/10.3897/biss.4.50889>.
- Chauvier, Y., Zimmermann, N.E., Poggiato, G., Bystrova, D., Brun, P., Thuiller, W., 2021. Novel methods to correct for observer and sampling bias in presence-only species distribution models. *Glob. Ecol. Biogeogr.* 30 (11), 2312–2325. <https://doi.org/10.1111/geb.13383>.
- Chiochetto, A., Arntzen, J.W., Martínez-Solano, I., de Vries, W., Bisconti, R., Pezzarossa, A., Maiorano, L., Canestrelli, D., 2021. Reconstructing hotspots of genetic diversity from glacial refugia and subsequent dispersal in Italian common toads (*Bufo bufo*). *Sci. Rep.* 11 (1), 1–14. <https://doi.org/10.1038/s41598-020-79046-y>. 2021 11:1.
- Collart, F., Guisan, A., 2023. Small to train, small to test: dealing with low sample size in model evaluation. *Ecol. Inform.* 75, 102106.
- Colwell, R.K., Rangel, T.F., 2009. Hutchinson's duality: the once and future niche. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19651–19658. <https://doi.org/10.1073/pnas.0901650106>.
- Cosentino, F., Seamark, E.C.J., Van Cakenberghe, V., Maiorano, L., 2023. Not only climate: the importance of biotic interactions in shaping species distributions at macro scales. *Ecol. Evol.* 13 (3), e9855. <https://doi.org/10.1002/ece3.9855>.
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., d'Amén, 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 (6), 774–787.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013.

- Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 36 (1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Downloading and cleaning GBIF data with R | R-bloggers (n.d.). Retrieved October 6, 2022, from <https://www.r-bloggers.com/2021/03/downloading-and-cleaning-gbif-data-with-r/>.
- Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Fan, W.-Y., Wang, Z.-G., 2014. The predictive performance and stability of six species distribution models. *PLoS One* 9 (11), e112764. <https://doi.org/10.1371/journal.pone.0112764>.
- Dutilleul, P., Clifford, P., Richardson, S., Hemon, D., 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49 (1), 305–314. <https://doi.org/10.2307/2532625>.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distribut.* 17 (1), 43–57. <https://doi.org/10.1111/J.1472-4642.2010.00725.X>.
- Fletcher, R.J., Hefley, T.J., Robertson, E.P., Zuckerman, B., McCleery, R.A., Dorazio, R.M., 2019. A practical guide for combining data to model species distributions. *Ecology* 100 (6), e02710. <https://doi.org/10.1002/ecy.2710>. ARTN.
- Fraginière, Y., Gremaud, J., Pesenti, E., Bétrisey, S., Petitpierre, B., Guisan, A., Kozłowski, G., 2022. Mapping habitats sensitive to overgrazing in the Swiss Northern Alps using ecological niche modeling. *Biol. Conserv.* 274, 109742 <https://doi.org/10.1016/j.biocon.2022.109742>.
- Franklin, J., 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.
- Freeman, M.S., Dick, J.T.A., Reid, N., 2022. Dealing with non-equilibrium bias and survey effort in presence-only invasive Species Distribution Models (iSDM): predicting the range of muntjac deer in Britain and Ireland. *Ecol. Inform.* 69, 101683 <https://doi.org/10.1016/j.ecoinf.2022.101683>.
- Fuentes-Lillo, E., Lembrechts, J.J., Cavieres, L.A., Jiménez, A., Haider, S., Barros, A., Pauchard, A., 2021. Anthropogenic factors overrule local abiotic variables in determining non-native plant invasions in mountains. *Biol. Invas.* 23 (12), 3671–3686. <https://doi.org/10.1007/s10530-021-02602-8>.
- Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E., Thuiller, W., 2012. Invasive species distribution models – how violating the equilibrium assumption can create new insights. *Glob. Ecol. Biogeogr.* 21 (11), 1126–1136. <https://doi.org/10.1111/j.1466-8238.2012.00768.x>.
- GBIF.org, 2022a. GBIF Occurrence Download. <https://doi.org/10.15468/dl.tcf6fqe>.
- GBIF.org, 2022b. GBIF Occurrence Download. <https://doi.org/10.15468/dl.ns4jzr>.
- GBIF.org, 2022c. GBIF Occurrence Download. <https://doi.org/10.15468/dl.mpxxws>.
- GBIF.org, 2023. GBIF Occurrence Download. <https://doi.org/10.15468/dl.cbxxn3>.
- Gomes, V.H.F., Ijff, S.D., Raes, N., Amaral, L.L., Salomão, R.P., de Souza Coelho, L., de Almeida Matos, F.D., Castilho, C.V., de Andrade Lima Filho, D., López, D.C., Guevara, J.E., Magnusson, W.E., Phillips, O.L., Wittmann, F., de Jesus Veiga Carim, M., Martins, M.P., Irueme, M.V., Sabatier, D., Molino, J.-F., ter Steege, H., 2018. Species Distribution Modelling: contrasting presence-only models with plot abundance data. *Sci. Rep.* 8 (1), 1 <https://doi.org/10.1038/s41598-017-18927-1>. Article.
- Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S.L., Scroggie, M.P., Woodford, L., 2011. Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *J. Appl. Ecol.* 48 (1), 25–34. <https://doi.org/10.1111/j.1365-2664.2010.01911.x>.
- 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. Modell.* 431, 109194 <https://doi.org/10.1016/j.ecolmodel.2020.109194>.
- Guillera-Aroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Glob. Ecol. Biogeogr.* 24 (3), 276–292. <https://doi.org/10.1111/Geb.12268>.
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., Kueffer, C., 2014. Unifying niche shift studies: insights from biological invasions. *Trend. Ecol. Evol. (Amst.)* 29 (5), 260–269. <https://doi.org/10.1016/j.tree.2014.02.009>.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16 (12), 1424–1435. <https://doi.org/10.1111/Ele.12189>.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat Suitability and Distribution Models, With Applications in R*. Cambridge University Press, Cambridge.
- Hernández-Lambrano, R.E., Carbonell, R., Sánchez-Agudo, J.Á., 2020. Making the most of scarce data: mapping distribution range and variation in population abundance of a threatened narrow-range endemic plant. *J. Nat. Conserv.* 57, 125889 <https://doi.org/10.1016/j.jnc.2020.125889>.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2019. Package ‘dismo’. R Package Version 1.1-4. CRAN. Retrieved from <https://cran.r-project.org/web/packages/dismo/index.html>.
- Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Modell.* 157 (2), 331–341. [https://doi.org/10.1016/S0304-3800\(02\)00203-X](https://doi.org/10.1016/S0304-3800(02)00203-X).
- 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 (2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>.
- hseebens, 2022. Hseebens/DASCO workflow [HTML]. https://github.com/hseebens/DASCOworkflow/blob/30d6467a6f3259c5a0f622fe1d6a8ef4dbf5b159/run_DASCO_workflow.R. Original work published 2021.
- Husson, F., Josse, J., Le, S., Mazet, J., 2020. Package ‘FactoMineR’. Multivar. Explorat. Data Analy. Data Min. Available on CRAN <https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>.
- 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.gioplacha.2018.03.008>.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*. Springer.
- Jimenez, L., Soberon, 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. *Method. Ecol. Evol.* 11 (12), 1571–1586. <https://doi.org/10.1111/2041-210x.13479>.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Krefl, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth’s land surface areas. *Sci. Data* 4 (1), 170122. <https://doi.org/10.1038/sdata.2017.122>.
- Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guarda, M., Anderson, R.P., 2021. ENMeval 2.0: redesigned for customizable and reproducible modeling of species’ niches and distributions. *Method. Ecol. Evol.* 12 (9), 1602–1608. <https://doi.org/10.1111/2041-210X.13628>.
- Liu, C.R., 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 (3), 535–548. <https://doi.org/10.1111/ecog.03188>.
- Lobo, J.M., Jimenez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*. 33 (1), 103–114. <https://doi.org/10.1111/j.1600-0587.2009.06039.x>.
- AUC: a misleading measure of the performance of predictive distribution models—Lobo—2008—Global Ecology and Biogeography—Wiley Online Library. (n.d.). Retrieved February 28, 2024, from <https://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2007.00358.x>.
- Lyu, S., Alexander, J.M., 2022. Competition contributes to both warm and cool range edges. *Nat. Commun.* 13 (1), 1 <https://doi.org/10.1038/s41467-022-30013-3>.
- Mateo, R.G., Felicísimo, Á.M., Muñoz, J., 2010. Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. *J. Veget. Sci.* 21 (5), 908–922. <https://doi.org/10.1111/j.1654-1103.2010.01198.x>.
- Menéndez, R., Megías, A.G., Hill, J.K., Braschler, B., Willis, S.G., Collingham, Y., Fox, R., Roy, D.B., Thomas, C.D., 2006. Species richness changes lag behind climate change. *Proceed. Roy. Soc. B: Biol. Sci.* 273 (1593), 1465–1470. <https://doi.org/10.1098/rspb.2006.3484>.
- Miller, J.A., 2012. Species distribution models: spatial autocorrelation and non-stationarity. *Progr. Phys. Geogr.: Earth Environ.* 36 (5), 681–692. <https://doi.org/10.1177/0309133312442522>.
- Normand, S., Ricklefs, R.E., Skov, F., Bladt, J., Tackenberg, O., Svenning, J.C., 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. *Proceed. Roy. Soc. B-Biol. Sci.* 278 (1725), 3644–3653. <https://doi.org/10.1098/rspb.2010.2769>.
- OFEV, O. fédéral de l’environnement. (n.d.). Espèces exotiques en Suisse. Retrieved April 7, 2023, from <https://www.bafu.admin.ch/bafu/fr/home/themen/thema-biodiversitaet/biodiversitaet-publikationen/publikationen-biodiversitaet/gebietfremde-arten-in-der-schweiz.html>.
- 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 (11), 933–938. <https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R., Martínez-Meyer, E., Nakamura, M., Araujo, M.P., 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A., 2012. Climatic niche shifts are rare among terrestrial plant invaders. *Science* (1979) 335 (6074), 1344–1348. <https://doi.org/10.1126/science.1215933>.
- 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. *Ecolog. Applic.* 19 (1), 181–197. <https://doi.org/10.1890/07-2153.1>.
- Poland, T.M., Patel-Weyand, T., Finch, D.M., Miniati, C.F., Hayes, D.C., Lopez, V.M. (Eds.), 2021. *Invasive Species in Forests and Rangelands of the United States: A Comprehensive Science Synthesis for the United States Forest Sector*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-45367-1>.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? *J. Biogeogr.* 33 (10), 1689–1703. <https://doi.org/10.1111/j.1365-2699.2006.01466.x>.
- Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. *Biometrics* 69 (1), 274–281. <https://doi.org/10.1111/j.1541-0420.2012.01824.x>.
- Schartel, T.E., Cao, Y., 2024. Background selection complexity influences Maxent predictive performance in freshwater systems. *Ecol. Modell.* 488, 110592. <https://doi.org/10.1016/j.ecolmodel.2024.110592>.
- scrubr source : R/scrubr-package.R (n.d.). Retrieved October 6, 2022, from <https://rdrr.io/cran/scrubr/src/R/scrubr-package.R>.
- Senay, S.D., Worner, S.P., Ikeda, T., 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PLoS One* 8 (8), e71218. <https://doi.org/10.1371/journal.pone.0071218>.

- Slodowicz, D., Descombes, P., Kikodze, D., Broennimann, O., Müller-Schärer, H., 2018. Areas of high conservation value at risk by plant invaders in Georgia under climate change. *Ecol. Evol.* 8 (9), 4431–4442. <https://doi.org/10.1002/ece3.4005>.
- Steen, B., Cardoso, A.C., Tsiamis, K., Nieto, K., Engel, J., Gervasini, E., 2019. Modelling hot spot areas for the invasive alien plant *Elodea nuttallii* in the EU. *Manag. Biol. Invas.* 10 (1), 151–170. <https://doi.org/10.3391/MBI.2019.10.1.10>.
- Steen, V.A., Tingley, M.W., Paton, P.W.C., Elphick, C.S., 2021. Spatial thinning and class balancing: key choices lead to variation in the performance of species distribution models with citizen science data. *Method. Ecol. Evol.* 12 (2), 216–226. <https://doi.org/10.1111/2041-210x.13525>.
- Svenning, J.C., Sandel, B., 2013. Disequilibrium vegetation dynamics under future climate change. *Am. J. Bot.* 100 (7), 1266–1286. <https://doi.org/10.3732/Ajb.1200469>.
- Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C., Guisan, A., 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Method. Ecol. Evol.* 5 (9), 947–955. <https://doi.org/10.1111/2041-210x.12203>.
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2021. Modelling species presence-only data with random forests. *Ecography* 44 (12), 1731–1742. <https://doi.org/10.1111/ecog.05615>.
- 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 (1), e01486. <https://doi.org/10.1002/ecm.1486>. ARTN.
- 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 (11), 1084–1091. <https://doi.org/10.1111/j.1600-0587.2013.00441.x>.
- Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* 36 (12), 2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics With S*, 4th edition. Springer, New York. ISBN 0-387-95457-0. <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Vetaas, O.R., 2002. Realized and potential climate niches: a comparison of four *Rhododendron* tree species. *J. Biogeogr.* 29 (4), 545–554.
- Vicente, J.R., Vaz, A.S., Queiroz, A.I., Buchadas, A.R., Guisan, A., Kueffer, C., Marchante, E., Cabral, J.A., Nesper, M., Broennimann, O., Godoy, O., Alves, P., Castro-Díez, P., Henriques, R., Honrado, J.P., 2019. Alien plant species: environmental risks in agricultural and agro-forest landscapes under climate change. *Clim. Change Manag.* 215–234. https://doi.org/10.1007/978-3-319-75004-0_13/FIGURES/1.
- Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* (N.Y.) 62 (11), 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>.
- 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, 110604.
- Wisn, 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 (8) <https://doi.org/10.1186/1472-6785-9-8> online.
- Wisn, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Distribut, N.P. S., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distribut.* 14 (5), 763–773. <https://doi.org/10.1111/j.1472-6462.2008.00482.x>.