DETERMINING THE MINIMAL BACKGROUND AREA FOR SPECIES DISTRIBUTION MODELS: MinBAR PACKAGE

Xavier Rotllan-Puig1,✉, and Anna Traveset2

1 ASTER Projects. Barri Reboll, 9, 1r. 08694 - Guardiola de Berguedà (Barcelona). Spain  
2 Terrestrial Ecology Laboratory. Global Change Research Group. Institut Mediterrani d’Estudis Avançats (CSIC-UIB). C/ Miquel Marqués, 21. 07190 - Esporles (Mallorca - Illes Balears). Spain

✉ Correspondence: [Xavier Rotllan-Puig <xavier.rotllan.puig@aster-projects.cat>](mailto:xavier.rotllan.puig@aster-projects.cat)

**Key-words**: species distribution model; ecological niche model; geographical background; background area; MaxEnt

# Abstract

One of the crucial choices when modelling species distributions using pseudo-absences approaches is the delineation of the background area to fit the model. We hypothesise that there is a minimum background area around the geographical centre of the species distribution that characterises well enough the range of environmental conditions needed by the species to survive. Thus, fitting the model within this geographical area should be the optimal solution in terms of both quality of the model and execution time. *MinBAR* is an R package that calculates the optimal background area by means of sequentially fitting several concentric species distribution models (SDMs) until a satisfactory model in terms of the included metrics is reached. The version 1.1.1 is implemented for MaxEnt (using either *maxnet* or the original java program) and uses Boyce Index as a metric to assess models performance. Three case studies are presented to test the hypothesis and assess package’s functionality. We show how partial models trained with part of the species distribution often perform equal or better than those fitted on the entire extent. *MinBAR* is a versatile tool that helps modellers to objectively define the optimal solution.

# Introduction

Species distribution modelling (SDM) has become an essential tool in the field of ecology (Peterson et al., 2015) and particularly for biodiversity conservation (Franklin, 2013). Its popularity, among other reasons, is due to the ease of use of software such as MaxEnt (Phillips et al., 2006) or BIOMOD (Thuiller et al., 2009), but also because of the development of the R-programming (R Core Team, 2019) community and the free availability of biodiversity and environmental data in public repositories (Zurell et al., 2020). However, regarding the public biodiversity data, they are often limited to only species presences and with a lack of occurrences in poorly sampled areas. These facts limit the use of some techniques or algorithms and force to make critical assumptions and choices, introducing different levels of uncertainty to model predictions (Jarnevich et al., 2017).

One of the crucial choices when using pseudo-absences approaches is the delineation of the background area to fit the model, also called “landscape of interest” or “study area” (Elith et al., 2011; Raes, 2012), given that increasing such area too far away from the presences might lead to overestimate the accuracy of the SDMs (Lobo et al., 2008). Defining its extent, however, remains a challenge and, as far as we know, a methodology to objectively define it has not yet been developed. Elith et al. (2011), for instance, argued that it has to be defined by the ecologist and limited by geographic boundaries or by how far the species can disperse. More recently, other authors have considered the interactions with other species or the sampling biases in the data set as constraints (Jarnevich et al., 2017). Yet, in many situations it is difficult to accurately define a background area, either owing to limited knowledge of the species biology or to the lack of available data (Anderson and Raza, 2010; Barve et al., 2011). In addition, studies are often performed at a local or regional level (Melo-Merino et al., 2020) and, then, the background area is constrained to an artificial boundary despite the species distribution might be wider (El-Gabbas and Dormann, 2018). Finally, another limitation may appear with species with large geographical ranges, which makes computations highly resource-demanding and time-consuming, especially when the study encompasses a high number of them. Any of these situations usually lead to fit models using only part of the species distribution (i.e. partial models), which might or might not imply a reduction of model performance (El-Gabbas and Dormann, 2018).

In this work, we hypothesize that there is a minimum background area around the geographical centre of the species distribution that characterizes well enough the range of environmental conditions needed by the species to survive. Thus, fitting the SDM within this minimum geographical area should be the best solution in terms of quality of the model (i.e. giving highest, but not overestimated, levels of model’s evaluation metrics), but also optimal regarding the execution time.

# MinBAR overview

*MinBAR* is an R package that aims at (1) defining the minimum background extent necessary to fit SDMs reliable enough to extract ecologically relevant conclusions from them and (2) optimizing the modelling process in terms of computation demands.

## Problem

On the one hand, fitting partial SDMs might reduce their good performance and lead to underestimated predictions of species’ distribution or to biased descriptions of their niches (Sanchez-Fernandez et al., 2011). On the other hand, making model calibrations and predictions of species with a large geographic range can demand a huge amount of computer resources in terms of time and memory.

The idea behind *MinBAR* is to solve these problems by sequentially fitting several concentric SDMs, each with an increased diameter, from the geographical centre of the species distribution to the periphery, until a model which satisfies the user’s needs (i.e. an acceptable compromise between quality and execution time) is reached. We call “buffers” to these concentric SDMs.

## Evaluation metrics

One of the most widely used metric to evaluate the performance of SDMs is the area under the receiver operating characteristic curve (i.e. AUC), although it has received several critiques (e.g. Jimenez-Valverde, 2012; Lobo et al., 2008). In particular, for the purpose of *MinBAR*, AUC is not the best choice because it is highly influenced by the defined background area. For this reason, *MinBAR* uses the Boyce Index (Hirzel et al., 2006), implemented in the R package *ecospat* (Di Cola et al., 2017).

Boyce Index (BI) is a threshold-independent evaluator for SDMs adequate for algorithms that use background data (Di Cola et al., 2017). BI varies between -1 and 1 (positive values indicate consistent predictions; close to zero, predictions not better than random; negative, bad predictions). See Hirzel et al. (2006) and Di Cola et al. (2017) for further details on BI.

In order to evaluate the predictive performance of the models, *MinBAR* includes two metrics. On the one hand, Boyce Index Partial (BI\_part) evaluates the accuracy of predictions within the buffer (i.e. training area). On the other hand, Boyce Index Total (BI\_tot) assesses predictions beyond the training area, across the whole distribution of the species (i.e. model transferability). In addition, in order to compare the environmental variables used for training the model with those across the projected area, a multivariate environmental similarity surface (MESS) analysss (Elith et al., 2010) is run.

## *minba*: The main function

The main function of *MinBAR* is *minba*. In the version 1.1.1, *minba* is implemented for MaxEnt models and the user can choose to use either the R package *maxnet* (Phillips et al., 2017) or the original java program, if installed (Phillips et al., 2006). See Table 1 for a full list of parameters to pass to *minba* as arguments.

After loading presences’ data and explanatory variables, the function calculates the geographical centre of the species distribution, the most distant occurrence and the buffers. The geographical centre corresponds to the mean location for longitude/latitude coordinates dealing with angularity (presences points and variables are previously “reprojected” to WGS84 if needed). It can be found in the literature some criticisms about the geographical component of the centre-periphery hypothesis (e.g. Pironon et al., 2017). This is why in future versions of MinBAR the user will be able to choose among the geographical and the ecological centre of the species. The number of buffers can be set by the user (10 as default) and are not defined by equal distance, but by % of presences equally distributed. This is particularly useful for relatively highly discontinuous distributions (e.g. introduced species), while not affecting more aggregated populations.

Then, *minba* makes *n* models for each buffer and calculates averages. In this step, it crops the variables to the extent of the buffer +5%, and calculates the number of necessary pseudo-absences to cover the 50% of the cells within the buffer (Guevara et al., 2018). The user can decide the proportion of presences set aside for testing the models. As default it uses 70% of the presences to calibrate the model and 30% for evaluation, all of them from within the buffer (Boyce Index Partial). It also makes predictions and evaluations for the whole extent of the species +5% (Boyce Index Total).

The user can choose either (1) to run the models for all the buffers to see if the selected background area is accurate and how the quality of the models evolves, or (2) to stop the process when it reaches certain conditions, which can be defined by the user as well. The latter option is adequate for large species distributions. In this case, the user also has several options, mainly depending on the aim of the study. On the one hand, if the interest is related to the characteristics of the population (e.g. description of the ecological niche), the focus should be in the Boyce Index Partial. On the other hand, if the aim is to project the model in time or space, the focus should fall on the Boyce Index Total. In turn, both approaches have two possibilities: either (a) fixing a minimum BI to stop the process when it is reached, or (b) to automatically stop it when the standard deviation (SD) of the last four calculated buffer’s BI is small. Thus, the user has four arguments (i.e. BI\_part, BI\_tot, SD\_BI\_part and SD\_BI\_tot) to pass to *minba* in order to define how to proceed. BI\_part and BI\_tot accept two possibilities: either *NULL* (default), which deactivates the condition, or a number below 1, which establishes the limit to be reached. Similarly, SD\_BI\_part and SD\_BI\_tot accept *NULL* (default) to deactivate the condition, or a number to establish the minimum SD. Therefore, there are several combinations to choose from. For instance, if all four arguments are *NULL* (default), all buffers are modelled; alternatively, if both BI\_par and BI\_tot are defined as a number, and so are SD\_BI\_part and SD\_BI\_tot, the process stops when the first of them is reached. Any combination of them is allowed.

## Outputs

At the end of the modelling process, *minba* outputs different information in the form of tables and charts to let the user know which is the buffer resulting in the model that best captures the environmental conditions of the species distribution. In addition, other ancillary data is generated during the process for each model (e.g. MESS maps).

It writes out three tables in *csv* files: *selfinfo\_mod*, *info\_mod* and *info\_mod\_means* (all followed by the name of the species). The first two tables are merely informative about how the modelling process has been developed, whereas *info\_mod\_means* shows the means of the *n* models run for each buffer. See Table S1 in Supplementary Material as an example of *info\_mod\_means*. It contains the metrics BI\_par, BI\_tot and the execution time. Additionally, this table also includes rankings of the buffers derived from the three metrics: *rankFinalNoTime* and *rankFinalWithTime*, which rank for the best buffer with and without taking into account the execution time, respectively.

Finally, *minba* draws scatterplots of the two BI to show the evolution of them with the increase of the buffer diameter in kilometres. It also plots the execution time.

# Implementation (Case Studies)

To test the hypothesis on the existence of an optimal background area, we performed a case study with real data at ecozone level. Then, to check how effective is the best buffer identified by *minba* in capturing the real distribution of the species, we performed a second case study with simulation data at regional level. Finally, a third case study was run on data belonging to an archipelago in order to check the functionality on islands.

By default, the function *minba* defines 10 buffers, with 3 model replicates per buffer, and lets the process produce models for all of them. MaxEnt was run using *maxnet* with default parameters, except for the number of background points. The intention of that was to limit interferences in the results as much as possible for all the species. We used 19 climatic variables available from WorldClim (Fick and Hijmans, 2017) and, although MinBAR is scale-independent, it was tested with two different resolutions depending on the study case.

All the R scripts used in these case studies can be found at <https://github.com/xavi-rp/MinBA>. While *MinBAR* v.1.1.1 can be downloaded from CRAN at <https://CRAN.R-project.org/package=MinBAR>. The latest (unstable) version is available at <https://github.com/xavi-rp/MinBAR>.

## Case 1: Real data

We modelled 25 common plant species native from the Western Palaearctic (longitude –14 to 60; latitude 28 to 75 degrees; see the list in Suppl. Mat. Table S2.1). The occurrences were downloaded from GBIF (references in Suppl. Mat. Table S2.2), and the working resolution was 5 arc-minutes. See Figure 1 for a representation of *Fagus sylvatica* presences, the centre of its distribution and the different buffers.

As an example, the output graphs produced for *Fagus sylvatica* can be seen in Figure 2. Both BI\_tot and BI\_part did not notably improve when increasing the buffers after certain level. A similar pattern was seen for almost all the species studied (see all plots in Supplementary Material S3). Actually, the results (Table 2, Figure 3) showed that the best models for most of the species were those fitted with only part of their distribution, both taking into account the execution time (88%) and not doing so (56%). Therefore, this would corroborate our hypothesis that a part of the species distribution close to its geographical centre would produce models with the highest levels of evaluation metrics.

## Case 2: Simulation data

The aim of this case study was to assess whether the best buffer determined by the function *minba* captured the actual distribution of the species or not, and at which degree. To do that, we had to know the entire and true distribution of the species and, for that, the only way was simulating virtual species distributions. If the result of the comparison between the true distribution and the one modelled (the best buffer) was positive, we could accept that *MinBAR* is effective in capturing the real distribution of the species, and at which degree.

We used *MinBAR* to calculate the best buffer for 25 virtual species generated with *virtualspecies* (Leroy et al., 2016) on the Iberian Peninsula. The working resolution was 5 arc-minutes. Then, for each species and best buffer, we obtained a prediction raster of potential distribution using *ENMeval* (Muscarella et al., 2014). Finally, we compared those predictions with the actual distribution (environmental suitability raster) by means of *ENMTools* (Warren et al., 2017) using D and I metrics. Both metrics are defined to measure similarities between two environmental niche models projected to the geographic space. And both range from 0 to 1, meaning no overlap at all and perfectly identical niches respectively. See Warren et al. (2008) for further details on how they are calculated and their similarities and differences.

Figure 4 shows how the best models for most of the species were those fitted with only part of their distribution (92% and 76% with and without execution time respectively), confirming Case 1 results. On the other hand, the comparison of the rasters of probability of presence for *MinBAR*’s best buffers with the actual distribution of the virtual species (Table 3) gave an average of 0.69 (SD = 0.16) and 0.87 (SD = 0.11) for D and I respectively. These results showed how, on average, the potential distribution of the best buffers resulted on a good overlap with the actual distribution of the species, meaning that *MinBAR* is able to capture the real distribution in a notable degree.

## Case 3: Real data on islands

Besides case studies 1 and 2 presented above, and to assess *MinBAR* for species inhabiting on islands, we run this case study 3 with real data from the Balearic Islands (Western Mediterranean).

We modelled the distribution of 10 species on the Balearic Islands (see the list in Suppl. Mat. Table S4.1), although their native distribution also includes other continental areas. The occurrences were downloaded from Bioatles ([http://bioatles.caib.es](http://bioatles.caib.es/)) using the R package *PreSPickR* (Rotllan-Puig, 2018). The function *minba* was run with default parameters (10 buffers, 3 model replicates per buffer and completing the models for all of them). MaxEnt models were fit using *maxnet* (Phillips et al., 2017) with the default parameters, except for the number of background points. The working resolution was 0.5 arc-minutes.

Both BI\_tot and BI\_part for most of the species did not improve very much when increasing the buffers after the first half. See all plots produced by *minba* in Supplementary Material Figures S4.2-11. Table 4 and Figure 5 show that the best models for most of the species were those fit with only part of their distribution (90%).

Therefore, these results confirm that *MinBAR* performs equally well for islands than for continental data. In addition, they show how this tool can help the user to confirm that, even though the area of study might not be the entire distribution of the species, the models can still be reliable enough.

# Conclusions

The package *MinBAR* has been developed, so far, to work with MaxEnt. It includes the Boyce Index as the main evaluator of the models predictive performance. In coming versions, however, it would be interesting to include other threshold-dependent evaluators based on sensitivity and specificity, as well as the option to pass arguments to the MaxEnt function, or to decide the centre from where to start delimiting buffers for modelling (i.e. geographic or environmental). In addition, the inclusion of an index that would take into account at the same time the accuracy in the training area and after transferring to further areas, such as the one described by Duque-Lazo et al. (2016), might also be quite useful for the users. Furthermore, the implementation of other algorithms and modelling techniques would be highly convenient.

In short, delimiting the background area can strongly affect the results and evaluations of SDMs (Acevedo et al., 2012; Phillips et al., 2009; Rodda et al., 2011). The case studies presented here show that the model including the presences from all the species distribution does not always perform the best. However, given the variability in the response of all the species to the modelling process, further studies should be made to fully understand this fact. Nevertheless, we have demonstrated the usefulness of *MinBAR* and its good level of effectiveness in capturing the real distribution of species. Therefore, this tool will be able to help modellers to objectively define an optimal solution regarding the delimitation of the background area.

# Acknowledgements

We are grateful to the R community for continuously contributing to the development of code and packages. And also to GBIF for their work in gathering and making data freely available, as well as all the volunteers and professionals who share the results of their fieldwork.

# Tables

Table 1: List of arguments to pass to the main function minba() and their description

|  |  |
| --- | --- |
| Arguments | Description |
| occ | Data frame or character. Data set with presences (occurrences). A data frame with 3 columns: long, lat and species name (in this order) |
| varbles | Raster\* object. A raster brick of the independent variables, or a directory where the rasters are. It will use all the rasters in the folder. Supported: .tif and .bil |
| wd | Character. A directory to save the results |
| prj | Numeric. Coordinates system (e.g. ‘4326’ is WGS84; check <http://spatialreference.org/>) |
| num\_bands | Numeric. Number of buffers (default is 10) |
| n\_rep | Numeric. Number of replicates (default is 15) |
| occ\_prop\_test | Numeric. Proportion of presences (occurrences) set aside for testing (default is 0.3) |
| maxent\_tool | Character. Either ‘dismo’ or (default) ‘maxnet’ |
| BI\_part | Numeric. Maximum Boyce Index Partial to stop the process if reached |
| BI\_tot | Numeric. Maximum Boyce Index Total to stop the process if reached |
| SD\_BI\_part | Numeric. Minimum SD of the Boyce Index Partial to stop the process if reached (last 3 buffers) |
| SD\_BI\_tot | Numeric. Minimum SD of the Boyce Index Total to stop the process if reached (last 3 buffers) |

Table 2: Best and second best buffer with and without taking into account execution time, for each species in Case 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species | Best Buffer No-Time | 2nd Buffer No-Time | Best Buffer With-Time | 2nd Buffer With-Time |
| pin\_syl | 6 | 8 | 6 | 8 |
| que\_ile | 7 | 8 | 7 | 8 |
| fag\_syl | 10 | 6 | 10 | 4 |
| fra\_exc | 6 | 5 | 6 | 5 |
| que\_pet | 7 | 10 | 7 | 3 |
| que\_rob | 9 | 10 | 9 | 8 |
| que\_pyr | 10 | 8 | 2 | 5 |
| que\_sub | 9 | 7 | 6 | 7 |
| abi\_alb | 9 | 10 | 9 | 10 |
| ace\_pla | 10 | 9 | 3 | 4 |
| aln\_glu | 10 | 1 | 1 | 2 |
| jun\_oxy | 10 | 9 | 10 | 5 |
| arb\_une | 10 | 9 | 9 | 10 |
| cra\_mon | 6 | 5 | 5 | 6 |
| pru\_spi | 7 | 8 | 7 | 8 |
| bux\_sem | 10 | 6 | 5 | 6 |
| cot\_tom | 9 | 8 | 9 | 7 |
| vio\_mir | 9 | 7 | 7 | 9 |
| dip\_eru | 10 | 9 | 3 | 10 |
| cen\_alb | 10 | 8 | 10 | 2 |
| ger\_luc | 6 | 9 | 6 | 2 |
| lin\_alp | 9 | 6 | 9 | 6 |
| pis\_ter | 10 | 8 | 8 | 10 |
| leo\_com | 9 | 5 | 5 | 9 |
| lot\_edu | 10 | 7 | 7 | 3 |

Table 3: Degree of similarity between the probability of presence of the best buffer calculated by MinBAR with the actual distribution of the same (virtual) species. The number in ‘BestBuffer’ represents the buffer number, being 1 the closest to the centre and 10 the furthest. Metrics ‘D’ and ‘I’ represent the level of similarity, being 0 no similarity at all and 1 identical probability of presence

|  |  |  |  |
| --- | --- | --- | --- |
| Species | BestBuffer | D | I |
| VirtSp s01 | 10 | 0.81 | 0.96 |
| VirtSp s02 | 9 | 0.83 | 0.97 |
| VirtSp s03 | 5 | 0.61 | 0.83 |
| VirtSp s04 | 3 | 0.39 | 0.66 |
| VirtSp s05 | 9 | 0.88 | 0.98 |
| VirtSp s06 | 10 | 0.88 | 0.99 |
| VirtSp s07 | 7 | 0.7 | 0.86 |
| VirtSp s08 | 9 | 0.67 | 0.81 |
| VirtSp s09 | 10 | 0.92 | 0.99 |
| VirtSp s10 | 4 | 0.55 | 0.79 |
| VirtSp s11 | 8 | 0.78 | 0.91 |
| VirtSp s12 | 9 | 0.77 | 0.95 |
| VirtSp s13 | 6 | 0.69 | 0.86 |
| VirtSp s14 | 10 | 0.86 | 0.98 |
| VirtSp s15 | 3 | 0.56 | 0.78 |
| VirtSp s16 | 6 | 0.59 | 0.83 |
| VirtSp s17 | 2 | 0.41 | 0.68 |
| VirtSp s18 | 10 | 0.75 | 0.95 |
| VirtSp s19 | 10 | 0.89 | 0.99 |
| VirtSp s20 | 8 | 0.81 | 0.95 |
| VirtSp s21 | 7 | 0.76 | 0.94 |
| VirtSp s22 | 6 | 0.62 | 0.84 |
| VirtSp s23 | 7 | 0.73 | 0.93 |
| VirtSp s24 | 3 | 0.37 | 0.63 |
| VirtSp s25 | 4 | 0.5 | 0.75 |
| mean |  | 0.69 | 0.87 |
| sd |  | 0.16 | 0.11 |

Table 4: Best and second best buffer with and without taking into account execution time, for each species in Case 3. The number represents the buffer number, being 1 the closest to the centre and 10 the furthest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species | Best Buffer No-Time | 2nd Buffer No-Time | Best Buffer With-Time | 2nd Buffer With-Time |
| arb\_une | 7 | 10 | 7 | 5 |
| asp\_aes | 4 | 1 | 4 | 1 |
| cha\_hum | 3 | 5 | 3 | 5 |
| eph\_fra | 1 | 7 | 1 | 7 |
| hel\_sto | 9 | 7 | 9 | 2 |
| jun\_oxy | 3 | 5 | 3 | 5 |
| pis\_len | 10 | 9 | 10 | 2 |
| que\_coc | 8 | 7 | 8 | 7 |
| rha\_ala | 4 | 5 | 4 | 5 |
| vib\_tin | 5 | 4 | 5 | 1 |

# Figures

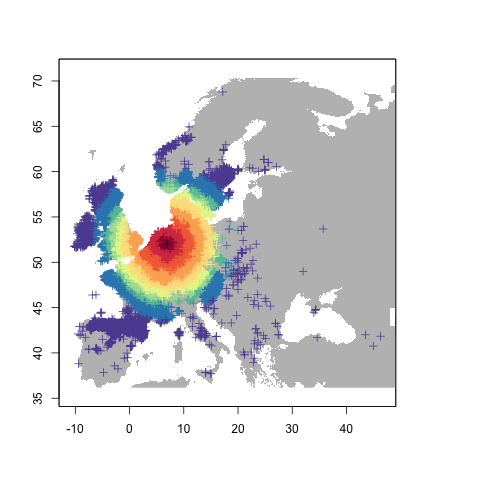


Figure 1: Occurrences of Fagus sylvatica. Black cross represents the geographical centre of the species distribution. Crosses in different colours for each buffer

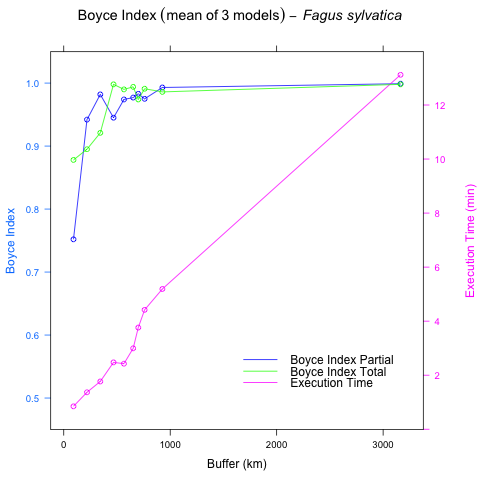


Figure 2: Evolution of Boyce Index Total (green) and Partial (blue) and the execution time in minutes (pink) for Fagus sylvatica. *The origin of the x-axis corresponds to the geographical centre of the species distribution (mean location for longitude/latitude coordinates dealing with angularity). The x-axis increases (in kilometres) with buffers 1 to 10, respectively the closest and the most distant to the centre of species distribution*



Figure 3: Frequencies of best buffer with and without taking into account execution time in Case 1. Buffers range from 1 to 10, corresponding respectively to the closest and the most distant to the centre of species distributions



Figure 4: Frequencies of best buffer with and without taking into account execution time in Case 2. Buffers range from 1 to 10, corresponding respectively to the closest and the most distant to the centre of species distributions



Figure 5: Frequencies of best buffer with and without taking into account execution time in Case 3. Buffers range from 1 to 10, corresponding respectively to the closest and the most distant to the centre of species distributions

# References

Acevedo, P., Jimenez-Valverde, A., Lobo, J.M., Real, R., 2012. Delimiting the geographical background in species distribution modelling. Journal of Biogeography 39, 1383–1390. <https://doi.org/10.1111/j.1365-2699.2012.02713.x>

Anderson, R.P., Raza, A., 2010. The effect of the extent of the study region on gis models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus nephelomys) in venezuela. Journal of Biogeography 37, 1378–1393. <https://doi.org/10.1111/j.1365-2699.2010.02290.x>

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberon, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D’Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R.G., Hordijk, W., Salamin, N., Guisan, A., 2017. Ecospat: An r package to support spatial analyses and modeling of species niches and distributions. Ecography 40, 774–787. <https://doi.org/10.1111/ecog.02671>

Duque-Lazo, J., Gils, H. van, Groen, T.A., Navarro-Cerrillo, R.M., 2016. Transferability of species distribution models: The case of phytophthora cinnamomi in southwest spain and southwest australia. Ecological Modelling 320, 62–70. <https://doi.org/10.1016/j.ecolmodel.2015.09.019>

El-Gabbas, A., Dormann, C.F., 2018. Wrong, but useful: Regional species distribution models may not be improved by range-wide data under biased sampling. Ecology and Evolution 8, 2196–2206. <https://doi.org/10.1002/ece3.3834>

Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution 1, 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of maxent for ecologists. Diversity and Distributions 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. <https://doi.org/10.1002/joc.5086>

Franklin, J., 2013. Species distribution models in conservation biogeography: Developments and challenges. Diversity and Distributions 19, 1217–1223. <https://doi.org/10.1111/ddi.12125>

Guevara, L., Gerstner, B.E., Kass, J.M., Anderson, R.P., 2018. Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. Global Change Biology 24, 1511–1522. <https://doi.org/10.1111/gcb.13992>

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199, 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>

Jarnevich, C.S., Talbert, M., Morisette, J., Aldridge, C., Brown, C.S., Kumar, S., Manier, D., Talbert, C., Holcombe, T., 2017. Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: An example with background selection. Ecological Modelling 363, 48–56. <https://doi.org/10.1016/j.ecolmodel.2017.08.017>

Jimenez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (auc) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography 21, 498–507. <https://doi.org/10.1111/j.1466-8238.2011.00683.x>

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>

Lobo, J.M., Jimenez-Valverde, A., Real, R., 2008. AUC: A misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>

Melo-Merino, S.M., Reyes-Bonilla, H., Lira-Noriega, A., 2020. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. Ecological Modelling 415, 108837. <https://doi.org/https://doi.org/10.1016/j.ecolmodel.2019.108837>

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENMeval: An r package for conducting spatially independent evaluations and estimating optimal model complexity for maxent ecological niche models. Methods in Ecology and Evolution 5, 1198–1205. <https://doi.org/10.1111/2041-210x.12261>

Peterson, A.T., Monica, P., Jorge, S., 2015. Mechanistic and correlative models of ecological niches. European Journal of Ecology 1, 28–38. <https://doi.org/https://doi.org/10.1515/eje-2015-0014>

Phillips, S.J., Anderson, R.P., Dudik, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: An open-source release of maxent. Ecography 40, 887–893. <https://doi.org/10.1111/ecog.03049>

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

Phillips, S.J., Dudik, 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. Ecological Applications 19, 181–197. https://doi.org/10.1890/07-2153.1

Pironon, S., Papuga, G., Villellas, J., Angert, A.L., Garcia, M.B., Thompson, J.D., 2017. Geographic variation in genetic and demographic performance: New insights from an old biogeographical paradigm. Biological Reviews 92, 1877–1909. <https://doi.org/10.1111/brv.12313>

Raes, N., 2012. Partial versus full species distribution models. Natureza & Conservacao 10, 127–138. <https://doi.org/10.4322/natcon.2012.020>

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodda, G.H., Jarnevich, C.S., Reed, R.N., 2011. Challenges in identifying sites climatically matched to the native ranges of animal invaders. PLOS ONE 6, e14670. <https://doi.org/10.1371/journal.pone.0014670>

Rotllan-Puig, X., 2018. PreSPickR: Downloading species presences (occurrences) from public repositories. <https://doi.org/10.13140/RG.2.2.10574.97607/1>

Sanchez-Fernandez, D., Lobo, J.M., Lucia Hernandez-Manrique, O., 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: A case study using iberian diving beetles. Diversity and Distributions 17, 163–171. <https://doi.org/10.1111/j.1472-4642.2010.00716.x>

Thuiller, W., Lafourcade, B., Engler, R., Araujo, M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. Ecography 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>

Warren, D., Matzke, N., Cardillo, M., Dinnage, R., 2017. ENMTools: Analysis of niche evolution using niche and distribution models.

Zurell, D., Franklin, J., Kˆnig, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leit„o, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schrˆder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. Ecography n/a. <https://doi.org/10.1111/ecog.04960>