

SPATIALLY ADAPTIVE BACKGROUND DELIMITATION STRATEGIES FOR SPECIES DISTRIBUTION MODELS WITH UNEVENLY SURVEYED OCCURRENCES

A. MÁRCIA BARBOSA

Spatial Biology Lab, Porto University, Portugal

Abstract. The quality of species distribution models (SDM) and ecological niche models (ENM) depends critically on how the background or pseudo-absence region is defined. Traditional background delimitation strategies include convex, concave or alpha hulls, buffers, biogeographic regions, and/or administrative units. However, in the frequent cases when species occurrence data are unevenly sampled – due to spatial variations in accessibility, survey effort or other biases –, traditional backgrounds may either include large unsurveyed areas, or exclude ecologically relevant (though insufficiently surveyed) portions of the species’ distribution range. Such mis-specification can reduce model predictive accuracy, increase extrapolation error, and worsen biodiversity knowledge gaps. This paper presents a function in the `fuzzySim` R package that implements a spatially adaptive approach to delineating SDM backgrounds, in a way that reflects the spatial distribution of the available occurrence records, which is largely shaped by survey effort. The method uses data-driven, differently-sized buffers to delimit regions that reflect local differences in occurrence data availability, with the aim of being neither overly conservative nor overly permissive. The approach is illustrated using a real-world species occurrence dataset with spatially varying degrees of survey bias. Spatially adaptive buffers yield more realistic, ecologically interpretable and justifiable background regions for modelling, which better address known biodiversity data constraints. They avoid placing pseudo-absence or unoccupied background points in unsurveyed regions, retaining isolated occurrence points without inflating the modelling area. This approach can thus help reduce the SDM uncertainty that arises from unevenly sampled biodiversity data, contributing toward the identification, visualization, and mitigation of the effects of knowledge gaps.

Key words: ecological niche modeling, background selection, pseudo-absence selection, model calibration extent, sampling bias

INTRODUCTION

Species Distribution Models (SDMs) and Ecological Niche Models (ENMs) are widely used tools for inferring where species are likely to occur based on environmental correlates of occurrence data (Elith & Leathwick, 2009; Franklin, 2010; Peterson et al., 2011; Sillero et al., 2021; Kass et al., 2025). A critical, yet often under-appreciated, component of SDMs is the definition of the background region from which background or pseudo-absence points are drawn. This background should represent the set of environmental conditions available to the species, to provide the contrast against which occurrence locations are evaluated. The delineation of this region substantially influences model predictive performance, particularly under extrapolation scenarios. Poorly chosen background extents may

lead to spatial and environmental bias, over- or under-prediction, and misleading inferences about species’ ecological requirements and potential distribution (Anderson & Raza, 2010; Barve et al., 2011; Acevedo et al., 2012).

Real-world occurrence datasets, particularly at the scale of species’ geographical distributions, typically integrate a variety of data sources, including scientific literature, biological collections, grey literature, and citizen science records. Such datasets are inevitably produced with spatially uneven sampling effort, yielding clusters of well-surveyed areas as well as large under-surveyed regions where only scattered records exist (Boakes et al., 2010; Meyer et al., 2015). Such unevenness arises from varying accessibility, observer bias, historical data collection practices, logistic constraints, and differential

funding or interest. In these circumstances, conventional background delimitation strategies – such as convex or concave hulls, fixed buffers around occurrence points, administrative or biogeographic borders – often fail to capture a uniformly meaningful region, where survey was reasonably well distributed. These approaches often lead to the inclusion of vast areas with deficient survey effort, where pseudo-absence or unoccupied background points are unreliable; or to the exclusion of ecologically relevant areas, because species occurrence records there are overly sparse to be representative of the species' environmental choices (Barbosa et al. 2013). Those approaches treat all occurrence records equally, although some records represent well-surveyed areas, while others are isolated and reflect opportunistic rather than well-distributed sampling (Phillips et al., 2009; Lobo et al., 2010).

These limitations exacerbate biodiversity knowledge gaps (Hortal et al., 2015). Mis-specified background areas can increase extrapolation into poorly sampled environments, distort estimates of environmental suitability, and reduce the confidence and effectiveness of model-based conservation decisions. Theory provides criteria for optimal background delimitation, such as including all areas accessible to the species (Soberón & Peterson, 2005; Owens et al., 2013), and capturing environmental variation and sampling effort homogeneously. However, real-world datasets can rarely satisfy those ideals.

A range of practical background delimitation strategies have been proposed in the SDM/ENM literature. A relatively common approach in the past was the convex or a concave hull around occurrences (Figure 1 A-B), which could be expanded by a buffer to allow for potential dispersal (e.g., Brown et al., 2017). Several other studies defined the background based on administrative or political boundaries, such as countries or provinces, especially when these reflect reliable survey extents or management units (Broennimann et al., 2007; Barbosa et al., 2003, 2009; Mateo et al., 2010; Lara-Galván et al., 2020, 2023). More natural options include the species' expert-drawn distribution range map (Figure 1C), or the set of ecoregions where the species occurs (Figure 1D; Sillero et al., 2021). Geographically-based options include the spatial trend of the occurrences (Acevedo et al., 2012; Figure 1E), or buffers with a fixed user-defined radius around each occurrence (Figure 1F), to capture areas that are likely within the species' dispersal reach (VanDerWal et al., 2009; Fourcade et al., 2014; Gallardo et al., 2024).

Some works have advocated for delimiting the background according to the accessible area concept of Soberón & Peterson (2005), which emphasizes incorporating

dispersal limitations and biogeographic barriers (Barve et al., 2011). More data-demanding approaches incorporate sampling effort or bias layers to weigh background points (Phillips et al., 2009; Kramer-Schadt et al., 2013; Fourcade et al., 2014; Stolar & Nielsen, 2015; Rocchini et al., 2023), or employ environmental similarity metrics to restrict background regions to conditions relevant for model transferability (e.g. Owens et al., 2013). This diversity of strategies highlights both the importance and the difficulty of defining biologically and statistically meaningful background regions for SDMs and ENMs. Moreover, a key challenge often remains insufficiently addressed: that a reliable background region should include not only areas that are within the species' movement ability, but also where the species was likely surveyed, so that pseudoabsences or unoccupied background points more adequately reflect areas not selected by the species.

This paper addresses this challenge by proposing, and implementing in a free and open-source software tool, a set of spatially adaptive background delimitation algorithms. The approach acknowledges that occurrence data often form clusters of varying extent and density, many of which likely reflect regions of higher survey effort rather than natural species distribution structure. It then accounts for how isolated or well-connected the occurrence points are. Key features include: (a) defining clusters of occurrences located near one another, of which the most numerous and extensive likely reflect well-surveyed areas for the target species; (b) computing buffers with a radius tailored accordingly for each cluster, rather than using a uniform-width buffer; (c) down-weighting or reducing buffer extent for more isolated points, which are more likely to reflect poorly surveyed areas; and (d) avoiding the loss of valuable information that would result from altogether excluding areas with only scarce and/or remote, isolated observations.

Spatially adaptive background delimitation can be a practical way to minimize the impacts of uneven survey bias on model quality, besides helping to identify and address biodiversity knowledge shortfalls. This can be a useful tool for biodiversity informatics, facilitating more reliable predictions from species occurrence datasets that are inevitably uneven and incomplete. Rather than attempting to demonstrate gains in model accuracy – a task that is highly dataset-dependent and hampered by the very sampling biases that motivate this method, which affect both model building and evaluation –, this paper illustrates the advantages of adaptive background delimitation through visual empirical examples, based on a real-world species occurrence dataset.

SOFTWARE TOOL DESCRIPTION

The implemented algorithms for delineating spatially adaptive background regions include three different approaches:

1) Inverse distance. This involves computing, for each species occurrence point, the sum of its pairwise distances to every other point, as a measure of how isolated the point is. These distances are then normalized to vary between 0 and 1, and inverted to instead represent proximity. Finally, a buffer is drawn around each point, whose radius is this sum of normalized proximities to the other points. Occurrence points with fewer and/or farther other points in their vicinity thus get smaller buffers around them, thus avoiding background areas in more scarcely surveyed regions. An exponential distance decay function is optionally used, to make farther points generate progressively smaller buffers.

2) Cluster mean distance. This starts by defining clusters of occurrence points within a given (user-specified, or default 100-km) distance of one another. This is done by drawing a circular buffer of half this radius around each point, and then merging the intersecting buffers into groups. Then, around each cluster, another buffer is drawn whose radius is the mean pairwise distance between the points in that cluster. This is a measure of the spatial spread of the cluster, and therefore the spatial reach of both the species and the surveyed region that the cluster reflects. For single-point clusters, the mean distance of zero is converted to a very small number, so that the pixel including the point is kept, but the region around it is not included in the background.

3) Cluster width. This consists in grouping occurrence points into clusters as in the previous approach, but then using, as the cluster buffer radius, the width of the smallest rectangle that can contain the points in each cluster. This is a more conservative measure of the spatial spread of the cluster, and it can be more suitable for elongated clusters. As above, for single-point clusters, the width of zero is converted to a very small number, so that the point is kept, but the region around it is not included in the background.

For all these methods, intersecting buffers are aggregated in the end. Additionally, a multiplier can be used to increase or decrease the output buffer radius around the points or clusters, if the user deems it more suitable (after expert-guided visual inspection) under the available knowledge of their target species' ecology, biogeographic history, and survey patterns. For the cluster-based methods, there is an additional option to weigh the radius of the buffer around each cluster proportionally to the number of points that it contains. If used, clusters with fewer

points (possibly indicating more scarcely surveyed areas) get proportionally smaller buffers than the mean distances or the widths of those clusters.

The proposed procedures are implemented and documented in the `getRegion` function of the `fuzzySim` R package (Barbosa, 2015; Barbosa, 2026), version 4.50 or above, written in the R programming language (R Core Team, 2025). The source code is available at <https://github.com/AMBarbosa/fuzzySim>, and bug reports can be sent under the “Issues” tab. Spatial operations within `getRegion` are computed with the `terra` R package (Hijmans, 2025), using functions such as `distance`, `width`, `buffer` and `aggregate`, which use modern geodesic algorithms that account for the curvature of the Earth. The output polygons are produced as R objects of class `SpatVector`, which can be handled directly with package `terra`, or converted for other spatial packages (e.g. with function `st_as_sf` of package `sf`). They can also be exported to any spatial vector file format, such as GeoPackage, GeoJSON, or ESRI Shapefile. The R script to reproduce the analysis and figures in this paper is provided as Supplementary Material, and it can easily be adapted (by changing just one line of code) to produce adaptive buffers for any species of the user's choice.

EXAMPLE APPLICATION

All background delimitation procedures are illustrated here with an empirical species occurrence dataset of the Eurasian otter *Lutra lutra* (GBIF.org, 2025) (Figures 1 and 2). This dataset reflects the wide spatial variations in survey intensity that are common in real-world biodiversity data, particularly at the scale of most species' geographical distributions, while still being small enough (using only records from the year 2025) for the entire example analysis to be easily reproduced in any computer.

Traditional background delimitation strategies either include large under-surveyed areas (Figure 1), or (if restricted to well-surveyed areas, or areas containing more points) can leave out a considerable number of occurrences. In contrast, the adaptive background delimitation strategies proposed here produce areas that are larger where survey was more widespread (as documented by the number and extent of occurrence records there), and smaller around more isolated records (Figure 2), which generally indicate more opportunistic and incomplete surveys. With these algorithms, all presence records are included, but “unoccupied” background is avoided in regions where survey is not demonstrated. These strategies can simply be run as follows:

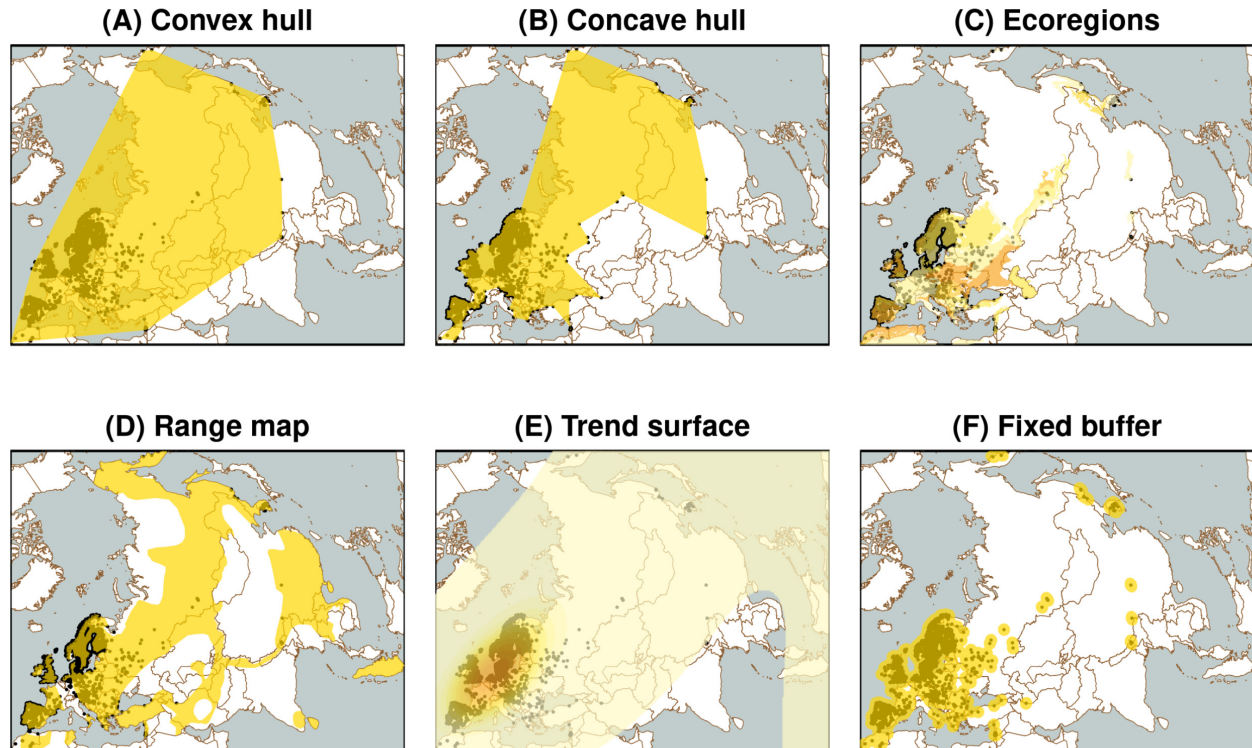


Figure 1. Regions derived from several conventional background delimitation strategies (yellow-shaded areas) applied to an empirical dataset of species occurrences (black dots). All tend to include relatively large areas where survey effort was visibly uneven.

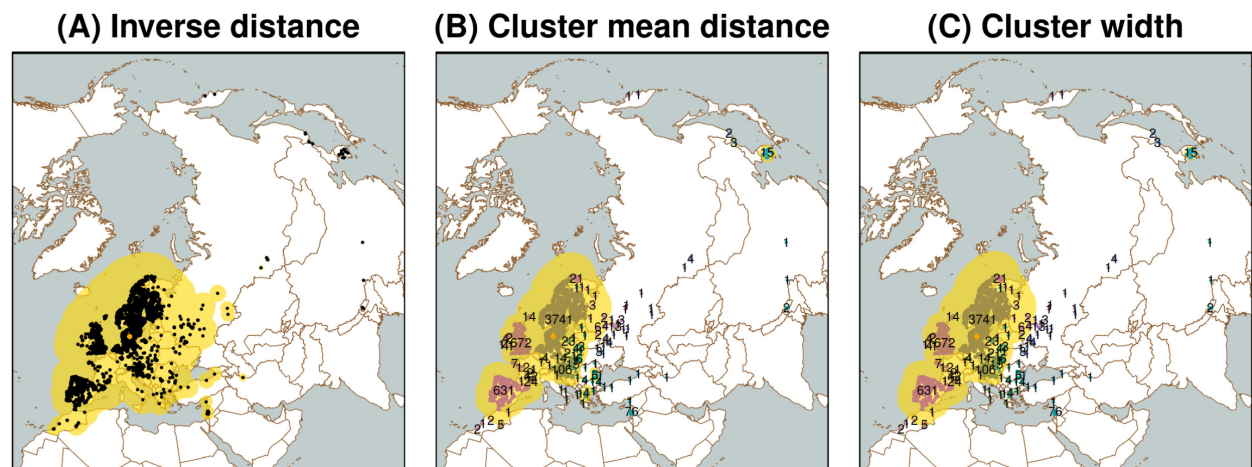


Figure 2. Spatially adaptive background regions (yellow-shaded areas) estimated with three different algorithms for an empirical dataset of species occurrences (dots). Numbers and colours in the two right-hand plots show clusters of occurrences within a given distance of one another (in this case 100 km). See text for details.

```
reg_invdist <-
fuzzySim::getRegion(occurrence_points,
  type = "inv_dist")

reg_clustmeandist <-
fuzzySim::getRegion(occurrence_points,
  type = "clust_mean_dist")
```

```
reg_clustwidth <-
fuzzySim::getRegion(occurrence_points,
  type = "clust_width")
```

As explained under *Software tool description* above, additional arguments can be added to further control the size of the buffers, such as `dist_mult`, `width_mult` and `weight`. Users should run `?getRegion` to see the

help file of the function, where all these parameters are described in detail; and try out the examples with values other than the default, to see their effect on the output and judge whether they could produce more appropriate background regions for their dataset (see *Discussion*). Further arguments can be added e.g. with plotting parameters such as titles, colours and opacity levels. See the full script in Supplementary Material for examples.

To accurately depict the distances between occurrences, the maps shown here (Figures 1 and 2) are in Azimuthal Equidistant projection centered on the aggregated occurrence points. However, the buffers were computed on the original occurrence points in unprojected longitude and latitude coordinates (in degrees), using a geodesic algorithm (implemented in package *terra*) that accounts for the spherical nature of these coordinates, and hence the curvature of the Earth. This produces more accurate results, particularly over long distances, which are common at the scale of species distributions. If input coordinates are in a projected system, the function defaults instead to a straight-line (rather than arc) distance measure, but it emits a note about the consequent inaccuracy.

The procedures leading to background delimitation determine which data are provided to the models and preceding analyses, therefore affecting variable selection and model computation: likely unsurveyed areas should not be included in either. However, prediction or projection of model outputs does not need to be limited to these areas. The models can predict species' potential distributions across any region, within and beyond the modelled background, provided that extrapolation is properly assessed (see Supplementary Material for a worked example).

DISCUSSION

As is apparent when comparing Figures 1 and 2, the spatially adaptive buffers implemented in the *getRegion* function can produce model training regions that, upon expert visual inspection, provide more sensible, defensible and justifiable background areas than other methods previously available – especially regarding the models' key assumption that the background is (at least to some degree) surveyed for the target species. These regions are better aligned with both biogeographic knowledge and biodiversity data availability, avoiding the pitfalls of overly inclusive, overly restrictive, or overly data-demanding approaches for background delimitation. By explicitly accounting for the spatial structure and uneven distribution of species occurrence data, these approaches provide a pragmatic compromise between ecological realism and data availability, minimizing overestimates of the area that is both accessible to, and surveyed for, the target species.

Within the defined modelling region, pseudo-absence or background points can be further selected based on a range of different strategies (e.g., Barbet-Massin et al., 2012; Kramer-Schadt et al., 2013; Stolar & Nielsen, 2015; Botella et al., 2021; Da Re et al., 2023; Balej et al., 2025). The *fuzzySim* package also facilitates the use of several such strategies, e.g. through the *selectAbsences* and the *biasLayer* functions, as is documented in their help file examples. But first, it is essential to delimit minimally adequately surveyed regions within which this selection can be made, which is now implemented in the *getRegion* function.

Among the proposed strategies, inverse distance (Figure 2A) is usually computationally faster, although it may clog the R memory if there are too many points for which to compute the pairwise distances. It has the advantage of requiring fewer user-defined parameters than the strategies involving clusters. It may work best when the frequency of recorded occurrences decreases progressively towards the edge(s) of the core surveyed area. Cluster-based strategies (Figure 2 B-C) work generally well regardless of the survey pattern, but they usually take longer to compute, and they require user-defined parameters such as the distance within which points should be included in the same cluster. The default clustering distance in the currently implemented function is 100 km, but this can be changed depending on the mobility and survey patterns of the target species. Compared to cluster mean distance, the “cluster width” strategy can be more appropriate when occurrence clusters extend across relatively narrow, elongated areas, where the mean pairwise distance may generate overly large buffer sizes that may not reflect evenly surveyed areas. When in doubt, users should try out different values of the function parameters, visually judge the results (taking into account the biogeography and survey patterns of the target species and region), and ideally perform a sensitivity analysis using an appropriate, unbiased set of validation data whenever possible.

As the spatial extent of a species' accessibility and survey effort is rarely accurately known, background delimitation cannot be objectively optimized from occurrence data alone. Therefore, rather than prescribing a single optimal background extent, the framework proposed here offers a set of different spatially adaptive strategies, reflecting the fact that unevenly surveyed data rarely contain sufficient information to objectively infer the “true” background which is both accessible and surveyed for any given species. The choice among these strategies should be guided by visual inspection of the resulting background regions, overlaid on the available occurrences, and interpreted in light of existing knowledge on the ecology,

biogeography and survey of the target species. Such expert-guided assessment is a standard diagnostic step in SDM and ENM workflows, and it provides a transparent basis for selecting a background region when formal optimization is not possible.

Users can employ additional criteria to further edit the obtained regions, by removing or adding parts (e.g., using *terra* package functions like *erase* or *+*, respectively), whenever more detailed knowledge is available on local patterns of species occurrence, movement, and/or survey effort. But whether or not this is the case, the proposed strategies allow an objective delimitation of sensible background regions purely from the available occurrence points, which can reflect survey effort variations across the species' occurrence area. These strategies thus address biodiversity knowledge shortfalls, tackling local and regional data gaps of spatially varying intensity, with no occurrence data loss and with improved confidence that the selected background generally reflects areas surveyed for the species.

FUTURE DEVELOPMENTS

Potential future improvements to the method include increases in computational efficiency, particularly for datasets with many thousands of occurrence records; and consideration of geographic and anthropogenic barriers, such as mountain chains or large water bodies (for terrestrial species), land areas (for aquatic species), densely humanized areas, or large infrastructures like dams or motorways. This may avoid the inclusion of physically separated (albeit relatively near) occurrences in the same proximity buffer.

ACKNOWLEDGMENTS

The author thanks all those involved in obtaining, formatting and uploading species occurrence data to biodiversity databases, and all those producing and maintaining the free software tools which the presented tool builds on.

CONFLICT OF INTEREST STATEMENT

The author has declared that no competing interests exist.

DATA AVAILABILITY

All data used in the examples are referenced and publicly available. A brief illustrated guide on how to install the required packages and reproduce the examples, along with all data used, is provided as Supplementary Material. Detailed examples of the function use are also described in the package help files.

SOFTWARE AVAILABILITY

The *fuzzySim* package, which contains the *get-Region* function, is implemented in the R programming language (R Core Team, 2025) and is freely available under a GNU General Public License. The stable release can be downloaded from the Comprehensive R Archive Network (<https://cran.r-project.org/package=fuzzySim>). Development versions, issue tracking, and source code are maintained on R-Forge (<https://fuzzysim.r-forge.r-project.org/>) and GitHub (<https://github.com/AMBarbosa/fuzzy-Sim>). The function is fully documented within the package, and usage examples are provided in the help files. The package is platform-independent and has been tested on Linux, Windows, and MacOS operating systems.

REFERENCES

- Acevedo, P., A. Jiménez-Valverde, J. M. Lobo, and R. Real. 2012. Delimiting the geographical background in species distribution modelling. *Journal of Biogeography* 39:1383–1390.
- Anderson, R. P., and A. Raza. 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.
- Balej, P., V. Moudrý, D. Prajzlerová, L. Gábor, N. Sillero, D. Rocchini, and P. Šimová. 2025. Species-observer link and kernel density estimation of background points allow for sampling bias correction in bird species distribution models. *Ecography* 2025:e08202.
- Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3:327–338.
- Barbosa, A. M. 2015. *fuzzySim*: applying fuzzy logic to binary similarity indices in ecology. *Methods in Ecology and Evolution* 6:853–858.
- Barbosa, A. M. 2026. *Fuzzy Similarity in Species Distributions*. R package version 4.50. URL: <http://fuzzysim.r-forge.r-project.org/>.
- Barbosa, A. M., R. Real, J. Olivero, and J. M. Vargas. 2003. Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. *Biological Conservation* 114:377–387.
- Barbosa, A. M., R. Real, and J. Mario Vargas. 2009. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecological Modelling* 220:747–754.
- Barbosa, A. M., M. Pautasso, and D. Figueiredo. 2013. Species-people correlations and the need to account for survey effort in biodiversity analyses. *Diversity and Distributions* 19:1188–1197.

- Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222:1810–1819.
- Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O'Connor, and G. M. Mace. 2010. Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLOS Biology* 8:e1000385. Public Library of Science.
- Botella, C., A. Joly, P. Bonnet, F. Munoz, and P. Monestiez. 2021. Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data. *Methods in Ecology and Evolution* 12:933–945.
- Broennimann, O., U. A. Treier, H. Müller-Schärer, W. Thuiller, A. T. Peterson, and A. Guisan. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10:701–709.
- Brown, J. L., J. R. Bennett, and C. M. French. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ* 5:e4095. PeerJ Inc.
- Da Re, D., E. Tordoni, J. Lenoir, J. J. Lembrechts, S. O. Vanwambeke, D. Rocchini, and M. Bazzichetto. 2023. USE it: Uniformly sampling pseudo-absences within the environmental space for applications in habitat suitability models. *Methods in Ecology and Evolution* 14:2873–2887.
- Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697. Annual Reviews.
- Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLOS ONE* 9:e97122. Public Library of Science.
- Franklin, J. 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge.
- Gallardo, B., S. Bacher, A. M. Barbosa, L. Gallien, P. González-Moreno, V. Martínez-Bolea, C. Sorte, G. Vimmercati, and M. Vilà. 2024. Risks posed by invasive species to the provision of ecosystem services in Europe. *Nat Commun* 15:1–12. Nature Publishing Group.
- GBIF.org (31 December 2025) GBIF Occurrence Download <https://doi.org/10.15468/dl.68pm8z>
- Hijmans R (2025). *terra: Spatial Data Analysis*. R package version 1.8-91, <https://rspatial.org/>.
- Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 46:523–549. Annual Reviews.
- Kass, J. M., A. B. Smith, D. L. Warren, S. Vignali, S. Schmitt, M. E. Aiello-Lammens, E. Arlé, A. M. Barbosa, O. Broennimann, M. E. Cobos, M. Guéguen, A. Guisan, C. Merow, B. Naimi, M. P. Nobis, I. Ondo, L. Osorio-Olvera, H. L. Owens, G. E. Pinilla-Buitrago, A. Sánchez-Tapia, W. Thuiller, R. Valavi, S. J. E. Velazco, A. Zizka, and D. Zurell. 2025. Achieving higher standards in species distribution modeling by leveraging the diversity of available software. *Ecography* 2025:e07346.
- Kramer-Schadt, S., J. Niedballa, J. D. Pilgrim, B. Schröder, J. Lindenborn, V. Reinfelder, M. Stillfried, I. Heckmann, A. K. Scharf, D. M. Augeri, S. M. Cheyne, A. J. Hearn, J. Ross, D. W. Macdonald, J. Mathai, J. Eaton, A. J. Marshall, G. Semiadi, R. Rustam, H. Bernard, R. Alfred, H. Samejima, J. W. Duckworth, C. Breitenmoser-Wuersten, J. L. Belant, H. Hofer, and A. Wilting. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions* 19:1366–1379.
- Lara-Galván, J. L., M. M.-S. Martín, J. F. Martínez-Montoya, J. J. Sigala-Rodríguez, J. A. Bañuelos-Alamillo, and M. Barbosa. 2023. Assessing the distribution of the West Mexican Coral Snake, *Micrurus distans* Kennicott, 1860, in Zacatecas State, Mexico, using modelling based on multiple data sources. *Herpetology Notes* 16:827–836.
- Lara-Galván, J. L., J. F. Martínez-Montoya, J. J. Sigala-Rodríguez, C. E. Esparza-Estrada, O. C. Rosas-Rosas, L. Ávila-Herrera, and A. M. Barbosa. 2020. Rattlesnake (*Crotalus* spp.) distribution and diversity in Zacatecas, Mexico. *ZooKeys* 1005:103–132. Pensoft Publishers.
- Lobo, J. M., A. Jiménez-Valverde, and J. Hortal. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103–114.
- Mateo, R. G., Á. M. Felicísimo, and J. Muñoz. 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. *Journal of Vegetation Science* 21:908–922.
- Meyer, C., H. Kreft, R. Guralnick, and W. Jetz. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat Commun* 6:8221. Nature Publishing Group.
- Owens, H. L., L. P. Campbell, L. L. Dornak, E. E. Saupe, N. Barve, J. Soberón, K. Ingenloff, A. Lira-Noriega, C. M. Hensz, C. E. Myers, and A. T. Peterson. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* 263:10–18. Elsevier.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011.

- Ecological Niches and Geographic Distributions. University Press, Princeton and Oxford.
- Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19:181–197.
- R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Áustria.
- Rocchini, D., E. Tordoni, E. Marchetto, M. Marcantonio, A. M. Barbosa, M. Bazzichetto, C. Beierkuhnlein, E. Castelnovo, R. C. Gatti, A. Chiarucci, L. Chieffallo, D. Da Re, M. Di Musciano, G. M. Foody, L. Gabor, C. X. Garzon-Lopez, A. Guisan, T. Hattab, J. Hortal, W. E. Kunin, F. Jordán, J. Lenoir, S. Mirri, V. Moudrý, B. Naimi, J. Nowosad, F. M. Sabatini, A. H. Schweiger, P. Šímová, G. Tassarolo, P. Zanini, and M. Malavasi. 2023. A quixotic view of spatial bias in modelling the distribution of species and their diversity. *npj biodiversity* 2:1–11. Nature Publishing Group.
- Sillero, N., S. Arenas-Castro, U. Enriquez-Urzelai, C. G. Vale, D. Sousa-Guedes, F. Martínez-Freiría, R. Real, and A. M. Barbosa. 2021. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecological Modelling* 456:109671. Elsevier.
- Soberón, J., and A. T. Peterson. 2005. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics* 2.
- Stolar, J., and S. E. Nielsen. 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Diversity and Distributions* 21:595–608.
- VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* 220:589–594.