

## rangemap: AN R PACKAGE TO EXPLORE SPECIES GEOGRAPHIC RANGES

MARLON E. COBOS<sup>1\*</sup>, VIJAY BARVE<sup>2</sup>, NARAYANI BARVE<sup>2</sup>, ALBERTO JIMÉNEZ-VALVERDE<sup>3</sup>,  
CLAUDIA NUÑEZ-PENICHER<sup>1</sup>

<sup>1</sup>*Department of Ecology and Evolutionary Biology & Biodiversity Institute, University of Kansas,  
Lawrence, Kansas 66045, USA*

<sup>2</sup>*Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA*

<sup>3</sup>*Universidad de Alcalá, Departamento de Ciencias de la Vida, 28805, Alcalá de Henares,  
Madrid, Spain*

\*Corresponding author: Marlon E. Cobos, Email: [manubio13@gmail.com](mailto:manubio13@gmail.com)

*Abstract.* Data exploration is a critical step in understanding patterns and biases in information about species' geographic distributions. We present rangemap, an R package that implements tools to explore species' ranges based on simple analyses and visualizations. The rangemap package uses species occurrence coordinates, spatial polygons, and raster layers as input data. Its analysis tools help to generate simple spatial polygons summarizing ranges based on distinct approaches, including spatial buffers, convex and concave (alpha) hulls, trend-surface analysis, and raster reclassification. Visualization tools included in the package help to produce simple, high-quality representations of occurrence data and figures summarizing resulting ranges in geographic and environmental spaces. Functions that create ranges also allow generating extents of occurrence (using convex hulls) and areas of occupancy according to IUCN criteria. A broad community of researchers and students could find in rangemap an interesting means by which to explore species' geographic distributions.

*Key words:* area of occupancy, buffer, concave hull, convex hull, extent of occurrence, suitable areas, trend-surface analysis

Biodiversity conservation and research in biogeography, macroecology, disease ecology, and other fields, rely heavily on information about species' geographic distributions. Understanding species' ranges is, however, a challenging task, because where a species is located depends on the existence of suitable environmental conditions, adequate dispersal ability, and appropriate biotic interactions (Soberón and Peterson 2005). Multiple analyses can be done to characterize species' distributions (e.g., correlative models, dispersal simulations, mechanistic models, etc.), and the quality of the results from such processes depend on many factors, including the algorithm used, knowledge of the species' ecology and natural history, and quality and quantity of information available (Clobert et al. 2012; Radosavljevic and Anderson 2014; Qiao et al. 2015).

Species' occurrence data are among the most numerous types of biological data available online thanks to recent technological developments and initiatives for data archiving and sharing, like the

Global Biodiversity Information Facility, iNaturalist, iDigBio, and others (Hardisty and Roberts 2013; Peterson et al. 2015). The existence of these data has facilitated use of diverse tools to model and simulate distributions and distributional dynamics, which has improved understanding of species' ranges (Franklin 2010; Peterson et al. 2011). Using these tools is data demanding, and requires considerable effort (Anderson 2015), including intensive processes of data exploration. However, simple tools to facilitate such explorations of species distributional data are still scattered (i.e., multiple tools from distinct GIS and statistical software may be required to perform such analyses).

Initial explorations are critical steps when working with distributional data (Cobos et al. 2018). These processes consist of a series of analyses and visualizations that allow researchers to identify patterns and recognize potential errors (Sillero et al. 2021). Among the most common steps for data exploration, plotting records on a map offers a simple

Table 1. Description of the main functions included in rangemap. More detailed documentation of these functions and other helper functions used to run analyses can be seen at<sup>1</sup>.

Function	Group	Description
rangemap_buff	Analysis	Generates a range by buffering species' occurrences using a user-defined distance.
rangemap_boundaries	Analysis	Creates a range by selecting all features of a spatial polygon layer in which the species is known to occur. Individual polygons are selected considering geographic occurrences and/or by manually defining their names.
rangemap_hull	Analysis	Generates ranges by creating a convex hull (Eddy 1977) or concave hull (Park and Oh 2012) polygons based on occurrence data. Polygons can be split based on geographic clustering using hierarchical (Everitt 1974) or k-means (Hartigan and Wong 1979) algorithms. Final polygons can be buffered if needed.
rangemap_enm	Analysis	Creates ranges by thresholding a continuous raster layer resulting from ecological niche modeling or species distribution modeling exercises. The threshold value is a user-specified level of omission or a specific value present in the continuous raster layer (Nenzén and Araújo 2011).
rangemap_tsa	Analysis	Generates range polygons using species' occurrences and trend surface analyses. Trend surface analysis is a method based on low-order polynomials of spatial coordinates for estimating a regular grid of points from scattered observations (Legendre and Legendre 1998). This method assumes that all cells not occupied by occurrences are absences; hence its use depends on the quality of data and the completeness of sampling in the region of interest.
rangemap_explore	Visualization	Creates simple figures to visualize occurrence data on top of a country map. All countries with at least one record are shown in the plot.
rangemap_plot	Visualization	Produces plots of ranges resulting from analysis functions. Species' ranges, extents of occurrence, and occurrences can be plotted on the same map if needed. Other aspects of a map can be added: legend, north arrow, scale bar, and axis values.
ranges_emaps	Visualization	Plots one or more ranges of a species on one or more raster layers of environmental variables.
ranges_espace	Visualization	Generates a three-dimensional plot, in environmental space, of ranges created using distinct algorithms.

visualization that helps to recognize potential biases. Creating simple range areas, derived from buffers or other types of polygons, also helps researchers to appreciate certain biases in sampling, geographic outliers, and areas related to species' distributions. Visualizations that integrate distributional data (occurrence data and/or simple range areas) and environmental information (in geographic or environmental spaces) give a general idea of conditions under which species have been detected and could exist. Once a better understanding of the patterns of data has been gained by performing these explorations, more robust analyses to characterize species' distributions can be performed. This is particularly relevant in conservation projects in which the quality of the data used and the modeling exercises determine the effectiveness of decisions that derive from the results of such applications.

Thanks to the relatively recent development of specialized packages, R (R Core Team 2021) is rapidly becoming an excellent alternative for analyzing spatial patterns in biodiversity data. Taking advantage of these specialized packages and the versatility of R, we developed “rangemap,” a new R package to explore species' distributions using

simple algorithms. This set of tools offers handy and robust open-source options to generate and visualize ranges, which otherwise will require users to perform numerous analyses. The functions in rangemap facilitate automation of exploratory analyses, which can be helpful when working with large numbers of species.

#### SOFTWARE DESCRIPTION

The rangemap R package contains tools that help researchers to explore species' ranges based on occurrence data and simple algorithms. The main tools in this package help to perform analyses to generate hypotheses of species ranges (Table 1). The functions in rangemap can be separated into two groups: (1) analysis functions, and (2) visualization functions.

#### *Data required*

The main input for analysis functions is a set of occurrence data (geographic coordinates) in the form of a “data.frame”. Other inputs vary depending on the function used, but they are common types of data used in geographic analyses (i.e., spatial polygons

<sup>1</sup> <https://cloud.r-project.org/web/packages/rangemap/rangemap.pdf>.

and raster layers). Examples of all types of data required to use the analysis functions are included in the package to help users to understand the structure and classes of such data.

#### *Analysis toolset*

Analysis functions are the ones used to generate estimates of species' ranges. The five analysis functions use different approaches: buffers, feature selection in polygon layers, convex hulls, concave (alpha) hulls, trend surface analyses, and raster classification (see details in Table 1). Before creating spatial polygons of species ranges, these functions perform simple steps of data cleaning: (1) erasing duplicates; (2) deleting occurrences lacking coordinates; and (3) excluding records falling outside of a region of interest. By default, the regions of interest are spatial polygons representing country borders (using data included in the package `mapproj`; Bivand and Lewin-Koh 2021), but other spatial polygons can be provided by users if smaller geographic regions need to be considered or if working in marine areas.

Diverse arguments in the functions allow users to explore different parameterizations in the analyses (e.g., buffer distance, geographic projection, algorithm, occurrence clustering, etc.). Other arguments allow users to produce spatial polygons to represent the extent of occurrence (using convex hulls) and the area of occupancy of species according to IUCN criteria (IUCN Standards and Petitions Committee 2019). These functions also permit saving results directly in a directory, which could help to prevent memory issues if multiple analyses are performed, and the user does not want results to be stored in memory (see ways to use such arguments in the code provided to reproduce examples). All results obtained with analysis functions are returned in objects of class *S4*, together with information describing how the range polygons were created. Users can save these results in multiple formats (according to the class of each of the results) if needed, or results can be used in further analyses.

#### *Visualization tools*

Four functions of `rangemap` help to visualize occurrence data and ranges generated with analysis functions (Table 1). To start, users can explore the occurrences to be used and roughly identify potential problems with the data (`"rangemap_explore"`). After

obtaining spatial polygons of species' ranges, another function (`rangemap_plot`) can be used to produce maps of ranges, occurrences, and, if present in results, extents of occurrence. The generic maps that are obtained can also be modified using arguments from the plotting function that control some graphic and map attributes (e.g., color, legend, north arrow, scale bar, etc.).

Two special visualization tools in `rangemap` (`"ranges_emaps"` and `"ranges_espace"`; Table 1) allow users to represent species' ranges considering environmental information. These functions help to visualize the environmental implications of using distinct approaches to produce range estimates. Environmental conditions are represented in geography using raster layers, and in environmental space using three-dimensional views of variable values present in records and ranges. For representations in environmental space, principal components derived from variables can also be used. To simplify comparisons of ranges in environmental space, ellipsoids (Norris et al. 2006; Nuñez-Penichet et al. 2021), created based on environmental conditions within each range, are used instead of all points.

#### EXAMPLE APPLICATION

The example consists of a step-by-step guide to explore, create, and plot species' ranges using `rangemap`. All code to reproduce this example is presented in the Supplementary Material (File S1). The occurrence data, spatial polygons, and raster layers used in these examples can be obtained using the code provided. Most of the data used are included in the `rangemap` package; this information can be loaded using the code provided and following other available guides (see Software availability).

#### *Data for examples*

Occurrence data for four species were used in our example application: (1) *Dasyopus kappleri* (greater long-nosed armadillo), (2) *Peltophryne fustiger* (western giant toad), (3) *Peltophryne empusa* (Cuban small-eared toad), and (4) *Amblyomma americanum* (lone star tick). Other data used in our examples are in raster format; one of these layers is the result of an ecological niche modeling exercise with the species *A. americanum* (Raghavan et al. 2019); the other raster layers represent bioclimatic variables (Hijmans et al. 2005): `BIO_5` = maximum temperature of

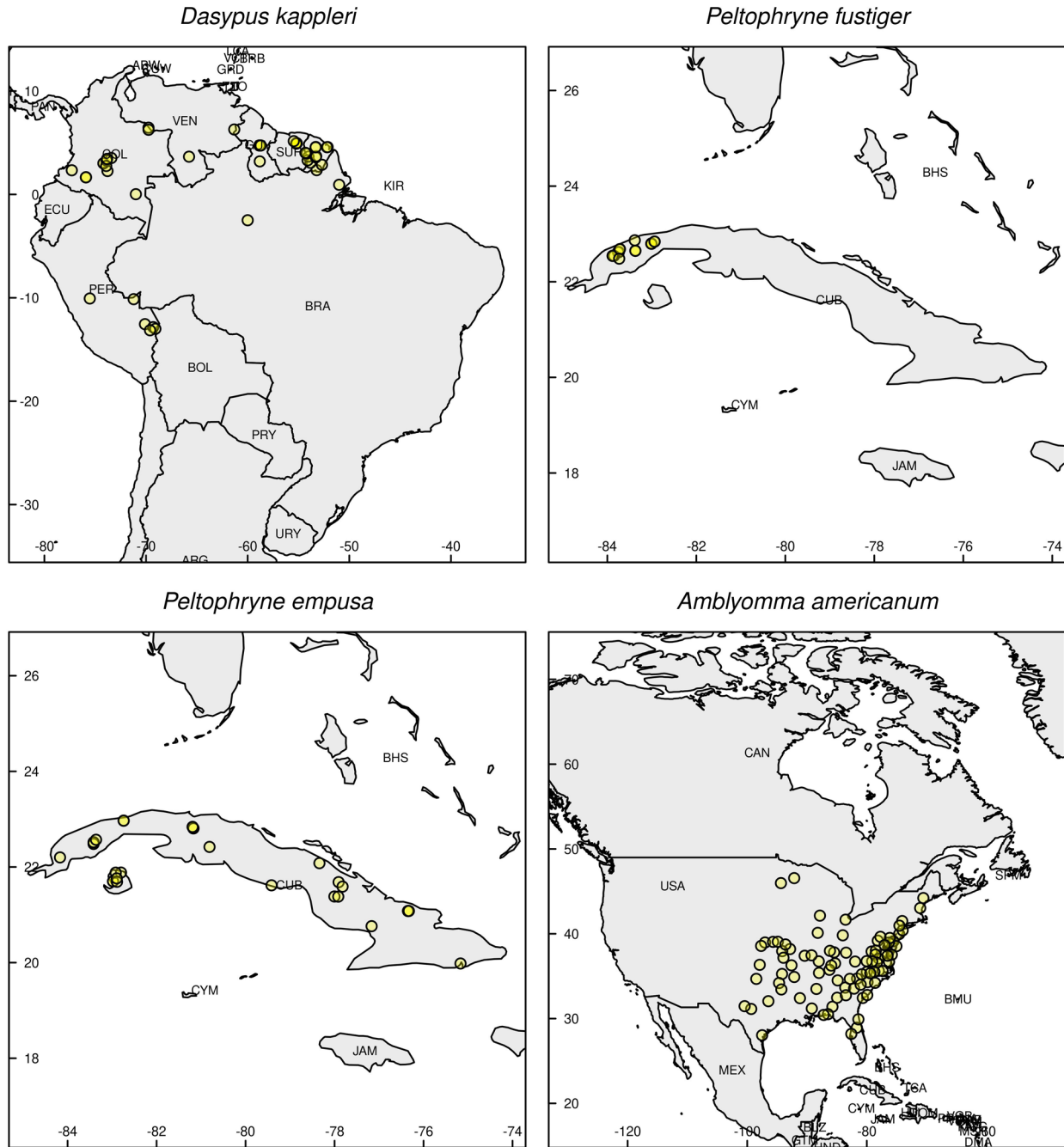


Figure 1. Examples of figures that can be produced to perform initial explorations of species occurrence data using the function “rangemap\_explore.”

warmest month; BIO\_6 = minimum temperature of coldest month; BIO\_13 = precipitation of wettest month; and BIO\_14 = precipitation of driest month. The spatial polygons used in analyses are prepared using data from the maptools package (Bivand and Lewin-Koh 2021). Layers to represent distinct administrative boundaries were downloaded using the function that performs this analysis (information available online<sup>2</sup>).

<sup>2</sup> <https://gadm.org/data.html>.

#### Exploring data and generating range estimates

We developed plots to explore the occurrence data using the function “rangemap\_explore” (Fig. 1). We generated polygons to represent ranges using approaches based on buffers, polygon selection, convex hulls, concave hulls, and trend-surface analysis. To demonstrate the use of ecological niche modeling outputs, we used the data for *A. americanum*. When ranges were created based on buffers or convex or concave hulls, we used the



following buffer distances: 300 km for *D. kappleri*; 50 km for *P. fustiger*; 30 km for *P. empusa*; and 350 km for *A. americanum*. Ranges generated by feature selection used the following administrative boundaries: countries for *D. kappleri*; municipalities for *P. fustiger*; provinces for *P. empusa*; and states for *A. americanum*. For the example case using the species *D. kappleri* and convex hulls, we used a distance of 1500 km to separate data in hierarchical clusters.

### Visualization of ranges

After the creation of range estimates, we produced figures to represent results in simple maps using the function “rangemap\_plot” (Fig. 2; S1–S5). To exemplify how ranges can be explored considering environmental conditions, we used range estimates created for *D. kappleri* and *A. americanum* and the functions “ranges\_emaps” and “ranges\_espace” (Fig. 3–4). Figures that represent ranges in environmental space were produced using three environmental variables and three principal components derived from four environmental variables (see section Data for examples).

### DISCUSSION

The tools included in the rangemap package allowed exploration of occurrence data (Fig. 1) and distinct options to generate spatial polygons (Fig. 2) that help to understand species’ ranges. The plots created with these tools showed the results from analyses and helped visualize environmental characteristics corresponding to species’ ranges (Fig. 3–4). Geographic patterns like clustering and density of records were easily identified using such explorations (see, e.g., species *A. americanum*; Fig. 1–2). Lack of sampling in certain areas of the regions of interest was also clear for *D. kappleri* (Fig. 1–2). The recognition of patterns like clustering, disjunction, and lack of sampling, can also help to identify potential errors in occurrence records, which are rather common in this type of data.

Using different algorithms to generate range estimates has marked effects on the results that can be obtained (e.g., *D. kappleri* ranges deriving from buffers and feature selection; Fig. 2). Changing some parameters in the tools used also affected the outcome of analyses; for instance, allowing the algorithm to recognize hierarchical clusters of occurrence records based on a distance, showed a discontinuity

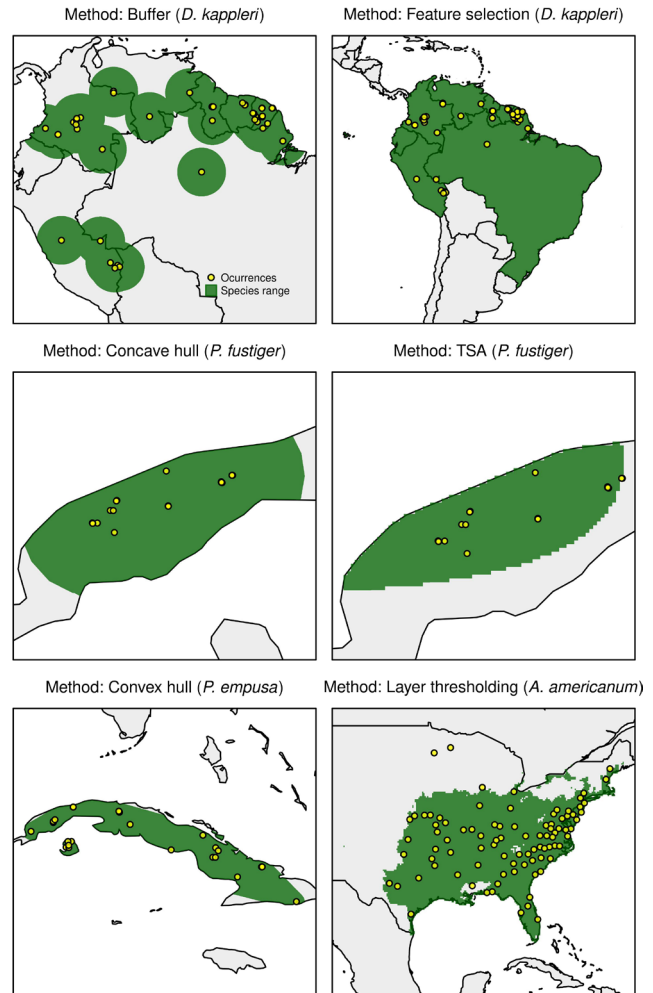


Figure 2. Examples of simple range estimates generated based on buffers, feature selection of country polygons, concave hulls, a trend surface analysis (TSA), convex hulls, and thresholding of a continuous layer resulting from an ecological niche modeling exercise.

in the range of *D. kappleri* (Fig. S3; see also Meyer et al. 2017 for applications of disjunct hulls). We recommend using multiple algorithms for range creation and different parameterizations of these tools to perform more complete explorations of the distributional data available. One of the parameters that could be of special importance is the one that sets the distance for buffers in some of our tools. Although, in our examples, distances were used for purposes of demonstration only, we encourage users to pick values of distance based on appropriate considerations (e.g., dispersal abilities, home range, etc.).

Consideration of environmental dimensions adds valuable information when exploring species’ distributions. Visualizing environmental conditions in plots and ranges on maps (Fig. 3) could help to

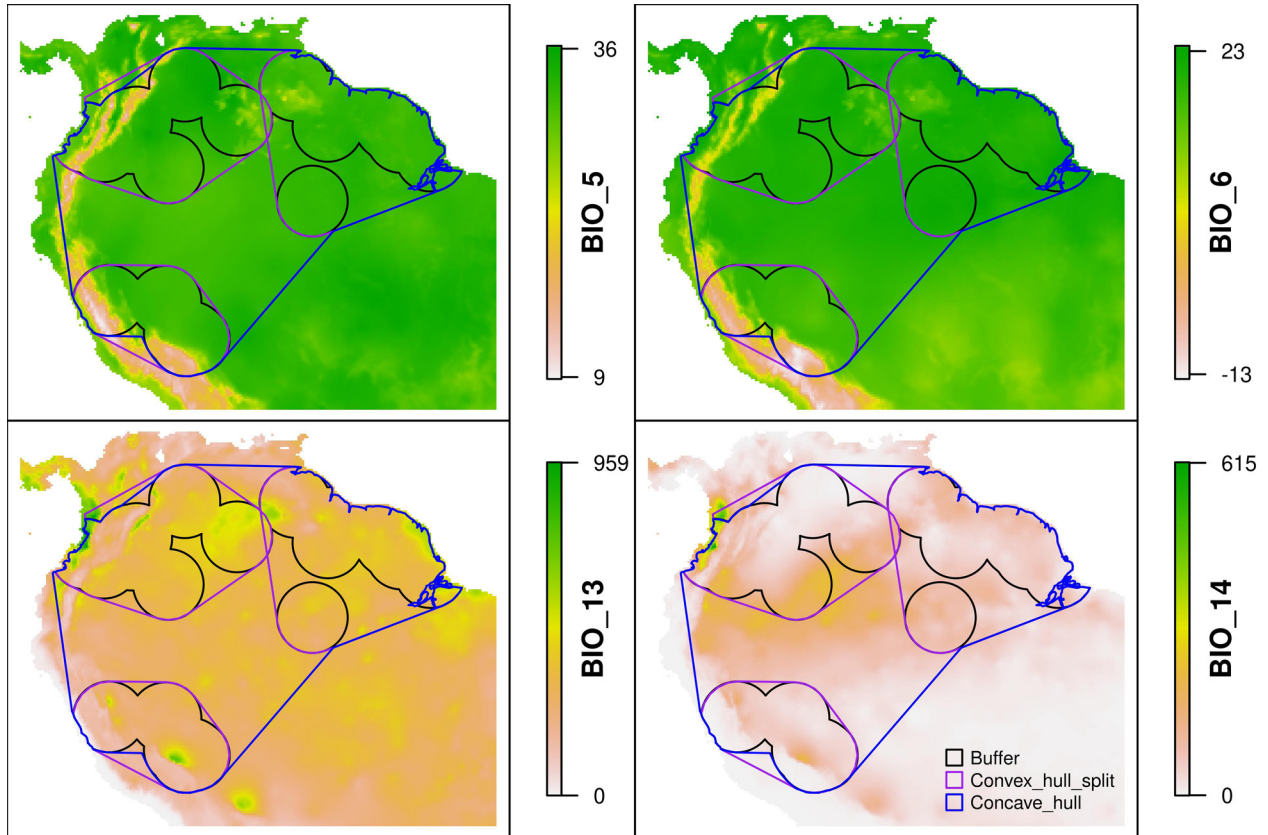


Figure 3. Comparison of ranges of the species *Dasypus kappleri* created with distinct algorithms. Ranges are shown on top of four bioclimatic variables. BIO\_5 = max temperature of warmest month; BIO\_6 = min temperature of coldest month; BIO\_13 = precipitation of wettest month; BIO\_14 = precipitation of driest month.

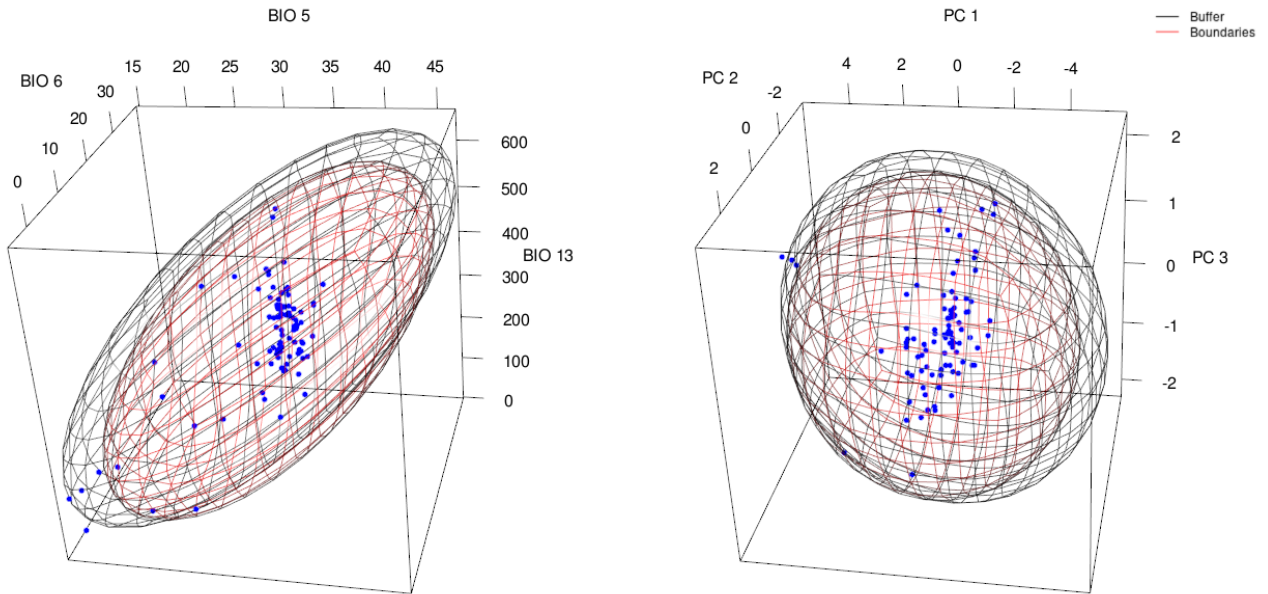


Figure 4. Visualization of distinct hypotheses of ranges for *Amblyomma americanum* in environmental space. Environmental space is represented by three raw environmental dimensions (left) and by the three first principal components (PC) derived from such dimensions (right): BIO\_5 = maximum temperature of warmest month, BIO\_6 = minimum temperature of coldest month, BIO\_13 = precipitation of wettest month.

identify potential biogeographic barriers derived from changes in such conditions in the continuous geographic region of interest. Plots produced to represent ranges in environmental space help to visualize effects of considering results from one type of analysis or another. For instance, using buffers of 350 km to create a range for *A. americanum* translates to wider limits in terms of temperature as compared to a range deriving from selecting the states where the species has been detected (Fig. 4). This type of visualization also helps to start understanding conditions used by the species (the occurrences) compared with conditions available nearby. In fact, some of the methods used to generate range estimates with our tools have also been proposed to create areas for model calibration in ENM or SDM exercises (e.g., Acevedo et al. 2012; Gonzalez et al. 2021). This is, generating calibration areas as spatial polygons obtained from buffering points, creating convex or concave hulls, or selecting polygons from layers representing ecoregions, is not uncommon in the ENM/SDM literature. However, these areas should be delimited based on more biologically-relevant considerations as they represent the regions that have been accessible to the species of interest for relevant periods of time (the M from the BAM diagram; Barve et al. 2011). For that reason, we recommend caution if the intention is to use our tools to generate such areas, as other, more advanced tools have been developed for this purpose (Machado-Stredel et al. 2021).

As previously established, the tools presented here aim to help users to perform initial explorations of species' ranges. Results from tools that use ecological niche modeling outputs could be interpreted as areas with suitable conditions for species (Franklin 2010; Peterson et al. 2011). However, we recommend caution in interpreting results, as no other ecological processes are considered in creation of species ranges. Another consideration when using our tools is that buffers for polygons are based on distances measured on spatial objects converted to the Azimuthal Equidistant projection, centered on the geographic centroid of the occurrence data. When working with very large areas, distances far from the center will lose precision (Snyder 1987).

In sum, the rangemap package offers handy options to explore species' distributions with minimal data requirements. After appropriate considerations of the limitations of the analyses performed to generate

range estimates, the information deriving from the toolset presented here could be useful in diverse applications. Researchers working on projects dealing with questions related to biogeographic patterns and conservation planning, for instance, could find in rangemap a friendly tool to perform initial but necessary steps to explore and filter occurrence data. Although the number of analyses allowed in this package is currently limited, future versions will provide options to explore occurrence data and distribution ranges considering environmental dimensions in more detail. The rangemap package and its dependencies (Table S1) are on CRAN (see Software availability).

#### ACKNOWLEDGMENTS

We thank M. Weber, the members of the KUENM working group at the University of Kansas Biodiversity Institute, and students from the Google Code-In program, for their help in beta-testing rangemap. Students from the Google Code-In program also helped to develop the logo of the package. This R package was developed during and with support from the 2018 Google Summer of Code program; R Organization, project “Species range maps in R.”

#### COMPETING INTERESTS

The authors have declared that no competing interests exist.

#### SOFTWARE AVAILABILITY

rangemap is available on CRAN<sup>3</sup>. Other detailed guides to using rangemap can be seen at<sup>4</sup>.

#### DATA AVAILABILITY

All data used to produce the examples presented here can be obtained with the code provided as part of the example application.

#### SUPPLEMENTARY MATERIAL

All supplementary materials can be accessed at<sup>5</sup>.

#### LITERATURE CITED

Acevedo, P., A. Jiménez-Valverde, J. M. Lobo, and R. Real. 2012. Delimiting the geographical background in species distribution modelling. *J. Biogeogr.* 39:1383–1390.

<sup>3</sup> <https://CRAN.R-project.org/package=rangemap>.

<sup>4</sup> <https://marloncobos.github.io/rangemap/>.

<sup>5</sup> <https://doi.org/10.6084/m9.figshare.16624228.v1>.

- Anderson, R. P. 2015. El modelado de nichos y distribuciones: No es simplemente “clic, clic, clic.” *Biogeografía* 8:4–27.
- 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. *Ecol. Model.* 222:1810–1819.
- Bivand, R., and N. Lewin-Koh. 2021. *maptools: Tools for handling spatial objects*. R package<sup>6</sup>.
- Clobert, J., M. Baguette, T. G. Benton, and J. Bullock (eds). 2012. *Dispersal Ecology and Evolution*. Oxford University Press, Croydon, UK.
- Cobos, M. E., L. Jiménez, C. Nuñez-Penichet, D. Romero-Alvarez, and M. Simões. 2018. Sample data and training modules for cleaning biodiversity information. *Biodivers. Inform.* 13:49–50.
- Eddy, W. F. 1977. A new convex hull algorithm for planar sets. *ACM Trans. Math. Softw.* 3:398–403.
- Everitt, B. 1974. *Cluster Analysis*. Heinemann Educational for Social Science Research Council, London.
- Franklin, J. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge, UK.
- Gonzalez, V. H., M. E. Cobos, J. Jaramillo, and R. Ospina. 2021. Climate change will reduce the potential distribution ranges of Colombia’s most valuable pollinators. *Perspect. Ecol. Conserv.* 19:195–206.
- Hardisty, A., and D. Roberts. 2013. A decadal view of biodiversity informatics: Challenges and priorities. *BMC Ecol.* 13:16.
- Hartigan, J. A., and M. A. Wong. 1979. A k-means clustering algorithm. *J. Royal Stat. Soc. C* 28:100–108.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965–1978.
- IUCN Standards and Petitions Committee. 2019. *Guidelines for using the IUCN Red List categories and criteria*. Version 14<sup>7</sup>.
- Legendre, P., and L. F. J. Legendre. 1998. *Numerical Ecology*. 2<sup>nd</sup> ed. Elsevier, Amsterdam.
- Machado-Stredel, F., M. E. Cobos, and A. T. Peterson. 2021. A simulation-based method for identifying accessible areas as calibration areas for ecological niche models and species distribution models. *Front. Biogeogr.* 13:e48814.
- Meyer, L., J. A. F. Diniz-Filho, and L. G. Lohmann. 2017. A comparison of hull methods for estimating species ranges and richness maps. *Plant Ecol. Divers.* 10:389–401. Taylor & Francis.
- Nenzén, H. K., and M. B. Araújo. 2011. Choice of threshold alters projections of species range shifts under climate change. *Ecol. Model.* 222:3346–3354.
- Norris, J. R., S. T. Jackson, and J. L. Betancourt. 2006. Classification tree and minimum-volume ellipsoid analyses of the distribution of ponderosa pine in the western USA. *J. Biogeogr.* 33:342–360.
- Nuñez-Penichet, C., M. E. Cobos, and J. Soberón. 2021. Non-overlapping climatic niches and biogeographic barriers explain disjunct distributions of continental *Urania* moths. *Front. Biogeogr.* 13:e52142.
- Park, J.-S., and S.-J. Oh. 2012. A new concave hull algorithm and concaveness measure for n-dimensional datasets. *J. Inf. Sci. Eng.* 28:587–600.
- Peterson, A. T., J. Soberón, and L. Krishtalka. 2015. A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC Ecol.* 15:15.
- 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*. Princeton University Press, Princeton.
- Qiao, H., J. Soberón, and A. T. Peterson. 2015. No silver bullets in correlative ecological niche modelling: Insights from testing among many potential algorithms for niche estimation. *Methods Ecol. Evol.* 6:1126–1136.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Radosavljevic, A., and R. P. Anderson. 2014. Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *J. Biogeogr.* 41:629–643.
- Raghavan, R. K., A. T. Peterson, M. E. Cobos, R. Ganta, and D. Foley. 2019. Current and future distribution of the Lone Star Tick, *Amblyomma americanum* (L.) (Acari: Ixodidae) in North America. *PLoS ONE* 14:e0209082.
- 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. *Ecol. Model.* 456:109671.
- Snyder, J. P. 1987. *Map Projections: A Working Manual*. U.S. Geological Survey, Washington, D.C.
- Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species’ distributional areas. *Biodivers. Inform.* 2:1–10.

<sup>6</sup> <https://CRAN.R-project.org/package=maptools>.

<sup>7</sup> <http://www.iucnredlist.org/documents/RedListGuidelines.pdf>.