Received: 1 November 2019 | Accepted: 8 July 2020

DOI: 10.1111/2041-210X.13452

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

NTBOX: An R package with graphical user interface for modelling and evaluating multidimensional ecological niches

Luis Osorio-Olvera[1,3] | Andrés Lira-Noriega[2] | Jorge Soberón[3] | Andrew Townsend Peterson[3] | Manuel Falconi[1] | Rusby G. Contreras-Díaz[1,4] | Enrique Martínez-Meyer[5,6] | Vijay Barve[7] | Narayani Barve[7]

[1]Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, México; [2]CONACyT Research Fellow, Red de Estudios Moleculares Avanzados-Instituto de Ecología, Veracruz, México; [3]Biodiversity Institute, University of Kansas, Lawrence, KS, USA; [4]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio A, Circuito de Posgrados, Ciudad Universitaria, Ciudad de México, México; [5]Instituto de Biología, Universidad Nacional Autónoma de México, México City, México; [6]Centro del Cambio Global y la Sustentabilidad en el Sureste A.C., Villahermosa, Mexico and [7]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

Correspondence
Luis Osorio-Olvera
Email: luismurao@gmail.com

Funding information
This work was partially supported by Consejo Nacional de Ciencia y Tecnología (CONACyT; postdoctoral fellowship number 740751; CVU: 368747 and project FORDECyT 273646), PAPIIT IN116018 and Google Summer of Code 2016 and the NSF grant ABI 1458640.

Handling Editor: Huijie Qiao

Abstract

1. Biodiversity studies rely heavily on estimates of species' distributions often obtained through ecological niche modelling. Numerous software packages exist that allow users to model ecological niches using machine learning and statistical methods. However, no existing package with a graphical user interface allows users to perform model calibration and selection based on convex forms such as ellipsoids, which may match fundamental ecological niche shapes better, incorporating tools for exploring, modelling, and evaluating niches and distributions that are intuitive for both novice and proficient users.
2. Here we describe an R package, NicheToolBox (NTBOX), that allows users to conduct all processing steps involved in ecological niche modelling: downloading and curating occurrence data, obtaining and transforming environmental data layers, selecting environmental variables, exploring relationships between geographic and environmental spaces, calibrating and selecting ellipsoid models, evaluating models using binomial and partial ROC tests, assessing extrapolation risk, and performing geographic information system operations via a graphical user interface. A summary of the entire workflow is produced for use as a stand-alone algorithm or as part of research reports.
3. The method is explained in detail and tested via modelling the threatened feline species *Leopardus wiedii*. Georeferenced occurrence data for this species are queried to display both point occurrences and the IUCN extent of occurrence polygon (IUCN, 2007). This information is used to illustrate tools available for accessing, processing and exploring biodiversity data (e.g. number of occurrences and chronology of collecting) and transforming environmental data (e.g. a summary PCA for 19 bioclimatic layers). Visualizations of three-dimensional ecological niches modelled as minimum volume ellipsoids are developed with ancillary statistics.

This niche model is then projected to geographic space, to represent a corresponding potential suitability map.

4. Using NTBOX allows a fast and straightforward means by which to retrieve and manipulate occurrence and environmental data, which can then be implemented in model calibration, projection and evaluation for assessing distributions of species in geographic space and their corresponding environmental combinations.

KEYWORDS

biodiversity informatics, ecological niche modelling, GIS tools, minimum volume ellipsoid, model evaluation, model selection, model uncertainty, species distribution

1 | INTRODUCTION

The description of species' distributional patterns has become a major focus of biodiversity informatics (Guralnick & Hill, 2009). One of the main developments in this field is in the growing use of correlational ecological niche models (ENMs), which are based on the principle that environmental requirements of a species, or its ecological niche, determine the sites where it can survive (Grinnell, 1917). These models are widely used in macroecological and biogeographic studies (Guisan & Zimmermann, 2000; Raxworthy et al., 2003). Applications often include the analysis of potential biological invasions, prediction of disease outbreaks and understanding effects of climate change on species distributions, as well as in phylogenetic frameworks to infer evolution of ecological niches, among many others (Peterson et al., 2011).

Recently, efforts have been made to develop specialized software to support ecological niche modelling and species distribution modelling: NicheA (Qiao et al., 2016), kuenm (Cobos, Peterson, Barve, & Osorio-Olvera, 2019), Wallace (Kass et al., 2018), ENMeval (Muscarella et al., 2014), biomod2 (Thuiller, Lafourcade, Engler, & Araújo, 2009), dismo (Hijmans, Phillips, Leathwick, & Elith, 2019), sdm (Naimi & Araújo, 2016), SSDM (Schmitt, Pouteau, Justeau, de Boissieu, & Birnbaum, 2017), SDMToolbox (Brown, Bennett, & French, 2017), sklearn (Pedregosa et al., 2011) and others. These efforts include non-open and open source packages that can be run either inside a graphical user interface (GUI; i.e. SDMToolbox, Wallace) or via commands in a programming language like R and python (i.e. dismo; biomod2, sklearn). Although these packages have a great diversity of functions to do ecological niche modelling exercises, few of them have the characteristic of being a GUI and at the same time provide the functions to perform analysis via scripts (i.e. sdm, SSDM).

NicheToolBox (NTBOX) is a software package with functions that can be used as a part of a script or via its GUI; it incorporates routines and tools for modelling correlational ecological niches with information that is straightforward to obtain from online sources about species' occurrences (e.g. Global Biodiversity Information Facility, GBIF; https://www.gbif.org) or environmental landscapes (e.g. CHELSA, Bio-Oracle). Alternatively, it can use data provided by the user, while allowing users to keep track of each step of the procedure by saving the workflow. NTBOX is written in the R language (http://www.r-project.org/), which has become one of the most widely used languages for data analysis in biodiversity informatics. NTBOX provides a detailed log of the modelling process, greatly facilitating replication of analyses. The workflow includes at least five steps: (a) obtaining georeferenced occurrences of species, (b) getting environmental information for modelling (usually in raster format), (c) data curation and spatial filtering of occurrences, (d) model calibration and evaluation and (e) reporting the workflow.

One of the deepest ideas coming from Hutchinson (1957) is that a 'duality' exists between geographic and environmental spaces (Colwell & Rangel, 2009). The entire enterprise of ecological niche modelling is based on this duality, yet software oriented to help a user to explore it is still largely lacking (Qiao et al., 2016, is an exception). Covering this gap is one of the main goals of NTBOX (see Section 5 in Appendix S1). Moreover, despite the broad use of ENMs, no other platform currently implements a model calibration and selection protocol for ellipsoid models with options in a GUI to facilitate visualization of selected models in environmental space.

Several reasons exist for using ellipsoids to model niches. First, convex shapes have been proposed as good models of fundamental niches (Jiménez, Soberón, Christen, & Soto, 2019), since it is reasonable to expect that if a species is able to tolerate extremes of a variable, it should also tolerate intermediate values (Drake, 2015). Second, the structure of fitness in niche space has been hypothesized to be approximately ellipsoidal, with some empirical support for this idea (Maguire Jr., 1973; Osorio-Olvera, Yañez-Arenas, Martínez-Meyer, & Peterson, 2020). Third, ellipsoids are simple models as they require only three parameters for defining them: (a) a niche-centroid, which is the point in ecological space where fitness has been argued to have maximum value (Martínez-Meyer, Díaz-Porras, Peterson, & Yáñez-Arenas, 2013); (b) a shape matrix, which measures how dependent two niche axes are and how they change together; and (c) the proportion of observations to be included in the ellipsoid (Van Aelst & Rousseeuw, 2009). Also, experience shows that simple ellipsoids may offer good models of swarms of points in multivariate space (Farber & Kadmon, 2003). A potential weakness of this approach is the implicit assumption that performance curves have symmetric forms, which may not be a general pattern across species (Holt, 2020), even if they are convex.

2 | PACKAGE DESCRIPTION

Among the basic operations that the NTBOX package supports are: (a) obtaining environmental data through WorldClim (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), CHELSA (Karger et al., 2016, 2017), Bio-Oracle (Assis et al., 2018; Tyberghein et al., 2012) and ENVIREM (Title & Bemmels, 2018); (b) filtering and visualizing species' occurrence data either provided by the user or retrievable from online sources (e.g. GBIF); (c) specifying a model calibration area, either using a pre-defined polygon or by creating one dynamically on-the-fly on a map provided in the GUI; (d) calculating bivariate correlations among environmental variables to choose the least redundant ones; (e) conducting principal component analysis on environmental variables and transferring them to explicit spatio-temporal scenarios; (f) estimating, visualizing and manipulating ecological niches based on minimum-volume ellipsoids (Van Aelst & Rousseeuw, 2009), BIOCLIM (Booth, Nix, Busby, & Hutchinson, 2014; Busby, 1991) and Maxent (Phillips, Anderson, & Schapire, 2006); (g) analysing group membership of observations using *k*-means clustering; (h) evaluating models with statistics based on partial ROC (Peterson, Papes, & Soberon, 2008) and multiple metrics based on the confusion matrix (Fielding & Bell, 1997); (i) a model-selection protocol for ellipsoid models based on statistical significance and model predictive performance (see Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, & Jiménez-García, 2019); (j) map binarization based on several thresholding methodologies; (k) environmental dissimilarity analysis to evaluate extrapolation risk in model transfers, including mobility-oriented parity (MOP; Owens et al., 2013), multivariate environmental similarity surfaces (MESS; Elith, Kearney, & Phillips, 2010) and extrapolation detection tool (ExDet; Mesgaran, Cousens, & Webber, 2014); (l) geographic information system (GIS) tools to crop and mask raster layers and export them on different raster formats; and (m) saving and exporting the workflow report as HTML files at every moment during the process (Figure 1). See Table 1 for a detailed list of NTBOX native functions; users can find help about how to perform each of the above analysis steps inside the GUI in the package reference guide (Appendix S1, which is also compiled as a vignette when installing the package) and also can find code to reproduce examples for each NTBOX function in https://luismurao.github.io/ntbox.

One of the most important characteristics of NTBOX is generation of ecological niche models based on *n*-dimensional minimum volume

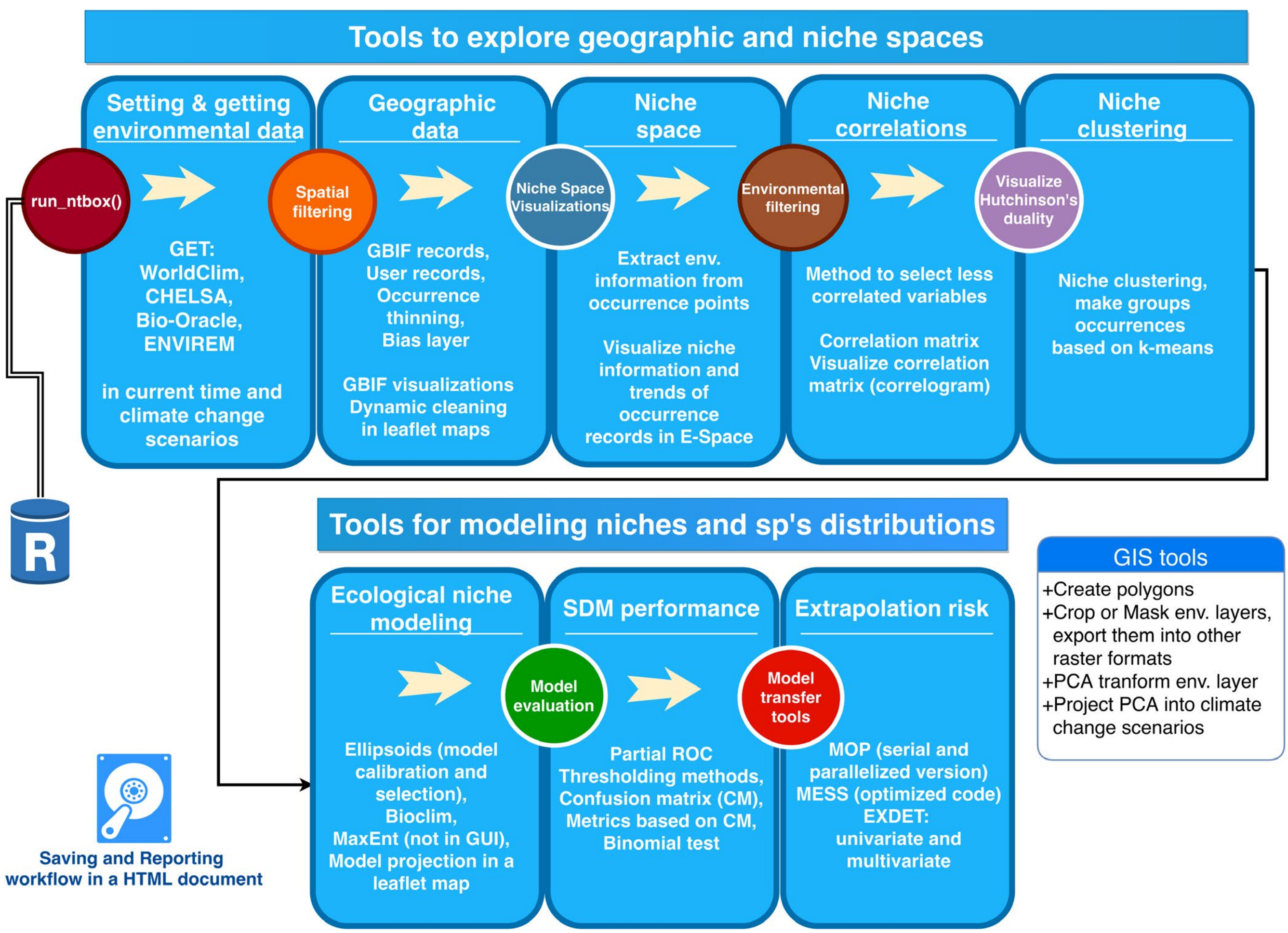

FIGURE 1 Workflow and main functionalities of the NicheToolBox (NTBOX) GUI. The package has tools to explore geographic and environmental spaces: (i) obtaining environmental data; (ii) geographic data; (iii) niche space; (iv) niche correlations for environmental filtering; (v) niche clustering (visualize the Hutchison's duality of the groups). NTBOX also has tools for modelling niches and species' distributions: (vi) ecological niche modelling; (vii) SDM performance (model evaluation) and (viii) extrapolation risk analysis

TABLE 1 List of NTBOX native functions. Note that all functions are accessible from the command-line interface. See package manual (Appendix S1) for a complete guide to how to use them in the GUI environment

Function	Description	Modelling process	In GUI of version 0.4.6.0
get_chelsa	Get current and future bioclimatic layers from CHELSA	Environmental data acquisition (AppSettings tab)	Yes
get_envirem_elev	Get elevation data for current and pass periods (Mid-Holocene and Last Glacial Maximum) from the ENVIREM database	Environmental data acquisition (AppSettings tab)	Yes
get_envirem_clim	Get bioclimate data for current and pass periods (from the ENVIREM database)	Environmental data acquisition (AppSettings tab)	Yes
get_bio_oracle	Get bioclimatic layers from Bio-Oracle for current and future scenarios	Environmental data acquisition (AppSettings tab)	Yes
searh_gbif_data	Function to search GBIF data; produces a leaflet map with metadata information of the occurrence points	Geographic data acquisition (data > GBIF data tab).	Yes
occs_history	Function to visualize GIBIF data using googleVis visualizations	Data exploration and geographic filtering (data > GBIF data > GBIF visualizations tab)	Yes
clean_dup	Function to clean duplicated records given a threshold distance	Data curation and geographic filtering (Data tab)	Yes
biaslayer	Function to create a bias layer for Maxent	Model fitting	No
correlation_finder	Function to find out strong correlations between the environmental variables	Environmental filtering (Niche correlations tab)	Yes
cov_center	Function to compute the minimum volume covariance matrix of an ellipsoid niche model	Niche space, Niche clustering and Model fitting (ENM tab)	Yes
ellipsoid_cluster_plot_3d	Plot cluster data in 3D by modelling them as an Ellipsoid	Niche clustering	Yes
sample_envbg	Function to generate random environmental background data	Ecological niche modelling, model calibration	No
ellipsoid_selection	Model selection protocol for ellipsoid models	Ecological niche modelling (ENM tab)	Yes
inEllipsoid	Determine if a point is inside or outside an ellipsoid	Ecological niche modelling (ENM tab)	No, but used inside the ellipsoid_selection function
ellipsoid_omr	Compute omission rates and partial ROC of ellipsoid models in environmental space	Ecological niche modelling	No, but used inside the ellipsoid calibration and selection process.
swd_format	Prepares data in Samples With Data format (SWD) for Maxent	Ecological niche modelling	No
maxent_call	Call Maxent from R; allows the user to introduce all the arguments that can be passed to Maxent	Ecological niche modelling	No
binomial_test	Binomial significance test for ecological niche models	Model evaluation (SDM performance tab)	Yes
omission_rate	Compute omission rate of a model	Model evaluation (SDM performance)	Yes
confu_mat_optim	Function to find the cut-off threshold that optimizes the confusion matrix	Model evaluation (SDM performance tab)	Yes
bin_model	Binarize a model using a threshold	Model evaluation (SDM performance tab)	Yes
mop	Mobility-Oriented Parity	Extrapolation risk assessment of ecological niche models (Extrapolation risk tab)	Yes
exdet_univar	ExDet univariate: NT1 metric, univariate extrapolation risk analysis for model transfer	Extrapolation risk assessment of ecological niche models (Extrapolation risk tab)	Yes
exdet_multvar	ExDet multivariate: NT2 metric, multivariate extrapolation risk analysis for model transfer	Extrapolation risk assessment of ecological niche models (Extrapolation risk tab)	Yes

(Continues)

TABLE 1 (Continued)

Function	Description	Modelling process	In GUI of version 0.4.6.0
ntb_mess	Multivariate Environmental Similarity Surfaces (MESS); this is an optimized version equivalent to MESS from DISMO package	Extrapolation risk assessment of ecological niche models (Extrapolation risk tab)	Yes
spca	PCA transformation of environmental layers	Data transformation (GIS tools tab)	Yes

ellipsoids (MVEs). It uses Mahalanobis distances to the ellipsoid centroid under the idea that the maximum fitness occurs at the centroid (Osorio-Olvera, Soberón, & Falconi, 2019). This is an attractive alternative to frequently used algorithms for generating correlative niche models (i.e. BIOCLIM, Maxent, GARP) and that lack assumptions about fitness. Mahalanobis distances, on the other hand, show a relationship to abundance (Osorio-Olvera, Yañez-Arenas, et al., 2020; Ureña-Aranda et al., 2015). This approach provides a means by which to define ecological niches based on their shape (i.e. it is widely assumed, through theoretical and experimental work, that fundamental niches have convex forms) and internal structure (i.e. how suitability is distributed within the niche; Jiménez et al., 2019; Maguire Jr., 1973; Osorio-Olvera et al., 2019). NTBOX allows the user to perform a model calibration and selection protocol for ellipsoids based on partial ROC and omission rates; to this end, NTBOX determines if testing points lie inside (correctly predicted presence) or outside of the ellipsoid model (Etherington, 2019; Van Aelst & Rousseeuw, 2009); then, for statistical significance, it performs a partial ROC test using an environmental background provided by the user; model performance is computed as omission rates for both training and testing data (Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, et al., 2019; Osorio-Olvera, Yañez-Arenas, et al., 2020). This evaluation method in environmental space makes the process of model calibration and selection faster than methods that need to create a map of the prediction first and then transform it into a binary map to compute omission rates.

3 | EXAMPLE

We demonstrate the use of NTBOX by modelling the potential distribution of *Leopardus wiedii*, a near-threatened small cat that lives in the Neotropics (Figure 2). Here we modelled the ecological niche of *L. wiedii* using ellipsoids and showed the performance and speed of NTBOX model calibration and selection protocol for MVEs by using environmental information of America at three different spatial resolutions (10′, 5′ and 2.5′). We describe in general terms the functions used at each step of the workflow presented in Figure 1, and provide the code to reproduce the complete example in Appendix S2.

First, in NTBOX, the user can choose from different environmental data layers which will be downloaded for the modelling process; here, we used the function getData from the RASTER package to download WorldClim's (https://worldclim.org) bioclimatic layers at 10′, 5′ and 2.5′ resolutions. Second, we searched 5,000 occurrences for *L. wiedii* from the GBIF database by applying the function search_gbif and explored the provenance and date of collection of these points (Figure 2). The number of georeferenced points that we found before data curation and cleaning was 859. After eliminating wrongly georeferenced localities, filtering occurrences for years ≥1950 and removing spatial duplicates (with the function clean_dup), we had 379 records. We extracted the environmental information of these records for each spatial resolution to look for and remove environmental duplicates. This second process of data curation left us with 231, 274 and 306 records for layers at 10′, 5′ and 2.5′ respectively.

Then, each occurrence dataset was split randomly into training and testing data in a proportion of 70:30. For each dataset, we estimated correlations among environmental variables and eliminated the correlated ones using the function correlation_finder with a threshold of $r \leq 0.95$. The number of environmental variables kept after this process was 14.

Finally, we calibrated MVE models with the function ellisposid_selection, and selected best models by applying the following criteria: (a) a significant value of partial ROC test ($p \leq 0.5$); (b) omission rate values for training and testing data ≤5%; and (c) models that presented the maximum value of AUC regarding the three spatial resolutions. To compute the partial ROC test and AUC, we used 10,000, 15,000 and 30,000 background points for the 10′, 5′ and 2.5′ layers respectively. Ellipsoid models were built in 3, 4 and 5 dimensions, using all possible combinations of the 14 least correlated variables, which generated 3,367 models for each spatial resolution (Table 2 and see Appendix S2 for details). The main arguments of the ellisposid_selection function are: (a) environmental training data (env_train); (b) environmental testing data (env_test); (c) a vector with the names of environmental variables to be fitted (env_vars); (d) the proportion of training points to be included in the MVE (level); (e) the environmental background (env_bg), which is a data.frame used to estimate the partial ROC and AUC test (this allows the algorithm to estimate statistical significance and performance with no need for writing raster models); (f) omission rate criteria for selecting models (omr_criteria); (g) a logical argument that indicates if models will be run in parallel (parallel); and (h) a logical argument to specify whether or not to return the value of partial ROC. Users can find details about how to use all the functions described in this example at https://luismurao.github.io/ntbox/.

In Table 2, we show the performance and the time to run the calibration and selection process of MVEs. Best models had omission rates lower than 5% and an AUC ranging from ~0.85 to 0.87 (some models had bigger AUCs but they had omission rates >5%; see Appendix S2). The time to run the 3,367 models for each resolution ranged from ~2 to 4 min (computer specifications: core i9 3.6 GHz, 64 GB RAM, Windows 10). Differences in running times relate to the number of background points used to compute the partial ROC and AUC test. The code to replicate this example is given in Appendix S2, and Figure 2 shows results

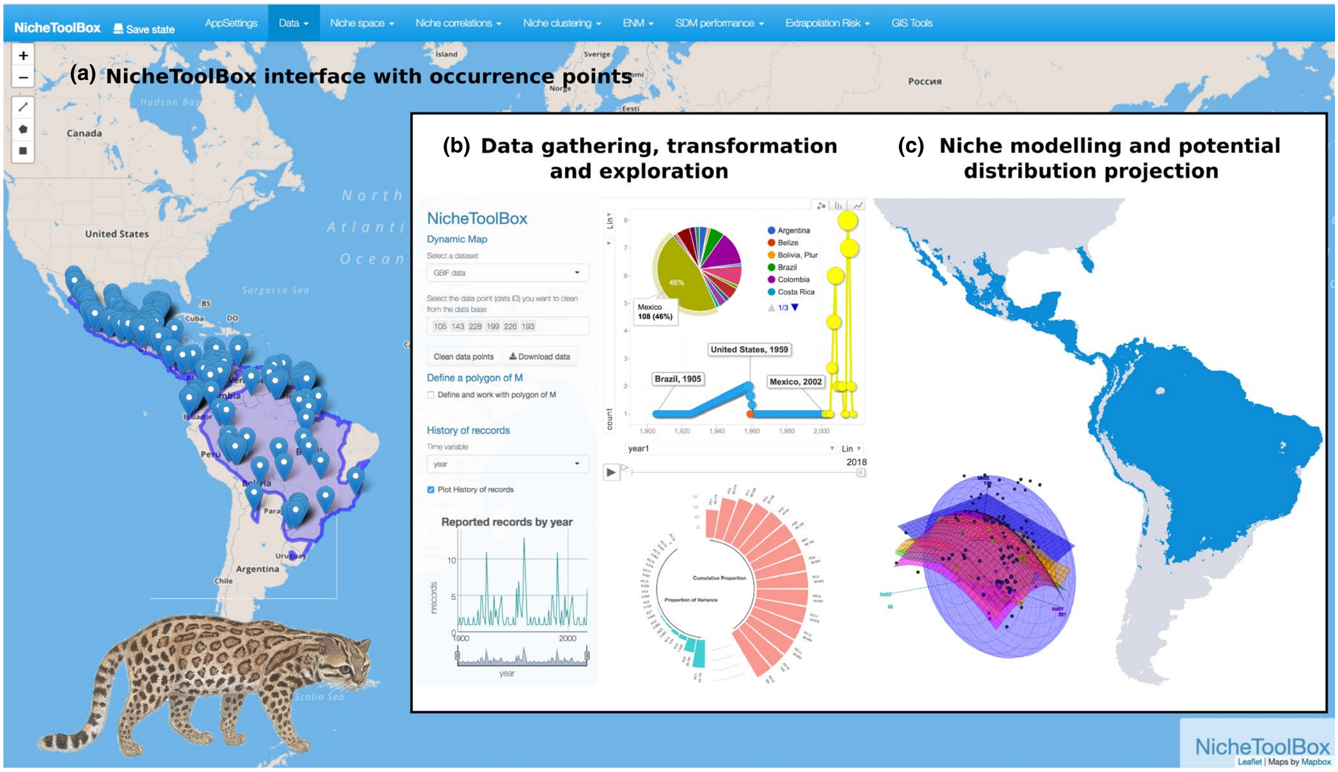

FIGURE 2 NicheToolBox (NTBOX) interface and examples of basic tools. (i) NTBOX map from the Data tab showing occurrences of *Leopardus wiedii* and its imported IUCN polygon (IUCN, 2007). (ii) Representation of some of the tools and processes available in accessing and exploring biodiversity data (e.g. number of occurrences and chronology of collecting) and transformation of environmental data (e.g. summary of PCA on 19 bioclimatic layers). (iii) 3D minimum-volume ellipsoid ecological niche model for *L. wiedii* chosen as best in the model selection process with additive and generalized lineal model trends and corresponding potential suitability map in geographic space for layers of 2.5′. Drawing of *L. wiedii* on the map was modified from (Sánchez et al., 2015)

TABLE 2 List of the best minimum-volume ellipsoid models for *Leopardus wiedii* using the model calibration and selection protocol implemented in the NTBOX R package. We tested performance and speed of the algorithm using three different spatial resolutions (Res). The number of calibrated models for each resolution was 3,367; here, we show those models that had an omission rate (OR) ≤0.05, a significant value of partial ROC ($p < 0.01$) and the highest value of AUC

Best model variables (bios)	OR train	OR test	*p*-val pROC	AUC	Res	Num. models	Time to run
4, 12, 19	0.049	0.043	0.000	0.863	10.0 min	3,367	1.77 min
3, 4, 12, 19	0.047	0.049	0.000	0.865	5.0 min	3,367	2.80 min
4, 9, 13, 15, 18	0.047	0.033	0.000	0.849	2.5 min	3,367	3.98 min

as they are displayed in the GUI, with the corresponding binary distribution map of the best model at a 2.5′ spatial resolution after imposing a 10th-percentile threshold on the suitability raster.

4 | CONCLUSIONS

Here we present NTBOX, an R package that supports functionalities for doing tasks related to ecological niche modelling, ranging from getting occurrence and environmental data and performing common GIS tasks, to evaluating ENMs and SDMs (see Figure 1; Table 1). These analyses can be run either via the GUI or from the R command line, thus allowing users to share and replicate the protocols generated from the different processes according to requirements from different users, such as larger-scale analyses or in collaborative schemes with users with different degrees of programming skills. Although the ENM part of the workflow focuses on modelling niches as ellipsoids, the package allows users to run other algorithms as well, including Maxent with all its parameters (currently available only in the command-line interface; see Table 1). One strength of the ellipsoid calibration and selection protocol presented here is speed and performance (Table 2). The example shown here is just for demonstrating the use of NTBOX; however, it is important to bear in mind that the workflow for modelling species' niches may take into account other factors such as sampling biases in occurrence records, the occurrence thinning method for trying to avoid problems related to spatial autocorrelation, the relationship between the time when an observation was made and the temporal resolution of modelling layers, among other things (Peterson et al., 2011).

The aim of NTBOX is to supply tools for conducting analyses related to understanding species' niches and geographic distributions. Future modules of the package will allow users to run dynamic distribution models to estimate, for example, the dispersal dynamics of invasive species (Osorio-Olvera & Soberon, 2020). As NTBOX is a project born in the open-source community (the source code has been available freely on GitHub since 2016), anyone can contribute to it via its GitHub repository (https://github.com/luismurao/ntbox); we expect that users wishing to contribute could help by providing code to run other modelling algorithms and statistical tools or by reporting bugs.

ACKNOWLEDGEMENTS

Many of the ideas and improvements to the software were realized after several rounds of analysis and on different projects with researchers and students during a research stay by L.O.-O. at the Biodiversity Institute, University of Kansas, as well as during courses taught at the Instituto de Ecología, A.C. and UNAM. L.O.-O. acknowledges the Posgrado en Ciencias Biológicas, UNAM. We also thank Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) for providing web support and space on their servers (shiny.conabio.gob.mx:3838/nichetoolb2/). J.S., R.G.C.-D. and L.O.-O. thank Blitzi Soberon for providing them moral encouragement. We thank Ángela Nava-Bolaños for her comments on a version of the manuscript.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

L.O.-O. conceived the project and coded the software; A.L.-N. and L.O.-O. lead the writing of the manuscript; J.S., A.T.P., A.L.-N., M.F., R.G.C.-D., E.M.-M., V.B. and N.B. gave suggestions to improve the package. All authors tested and verified the software to give final approval for publication.

DATA AVAILABILITY STATEMENT

Obtaining NicheToolBox: The full code of the R package NicheToolBox (NTBOX), along with instructions on how to install and use it, is available in the corresponding author's main GitHub repository: https://github.com/luismurao/ntbox. The source code is licensed under GPL-3. NTBOX package is publicly available on Zenodo at https://doi.org/10.5281/zenodo.3937910 (Osorio-Olvera, Lira-Noriega, et al., 2020).

ORCID

Luis Osorio-Olvera https://orcid.org/0000-0003-0701-5398
Andrés Lira-Noriega https://orcid.org/0000-0002-3219-0019
Jorge Soberón https://orcid.org/0000-0003-2160-4148
Andrew Townsend Peterson https://orcid.org/0000-0003-0243-2379
Manuel Falconi https://orcid.org/0000-0002-4296-7258
Rusby G. Contreras-Díaz https://orcid.org/0000-0002-0569-8984
Enrique Martínez-Meyer https://orcid.org/0000-0003-1184-9264
Vijay Barve https://orcid.org/0000-0002-4852-2567
Narayani Barve https://orcid.org/0000-0002-7893-8774

REFERENCES

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284. https://doi.org/10.1111/geb.12693

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, *20*(1), 1–9. https://doi.org/10.1111/ddi.12144

Brown, J. L., Bennett, J. R., & French, C. M. (2017). SDMtoolbox 2.0: The next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, *5*, e4095. https://doi.org/10.7717/peerj.4095

Busby, J. R. (1991). BIOCLIM – A bioclimate analysis and prediction system. In C. R. Margules & M. P. Austin (Eds.), *Nature conservation: Cost-effective biological surveys and data analysis* (pp. 64–68). Canberra, Australia: CSIRO.

Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019). kuenm: An R package for detailed development of ecological niche models using Maxent. *PeerJ*, *7*, e6281. https://doi.org/10.7717/peerj.6281

Cobos, M. E., Peterson, A. T., Osorio-Olvera, L., & Jiménez-García, D. (2019). An exhaustive analysis of heuristic methods for variable selection in ecological niche modeling and species distribution modeling. *Ecological Informatics*, *53*, 100983. https://doi.org/10.1016/j.ecoinf.2019.100983

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences USA*, *106*(2), 19651–19658. https://doi.org/10.1073/pnas.0901650106

Drake, J. M. (2015). Range bagging: A new method for ecological niche modelling from presence-only data. *Interface*, *12*(107), 20150086. https://doi.org/10.1098/rsif.2015.0086

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Etherington, T. R. (2019). Mahalanobis distances and ecological niche modelling: Correcting a chi-squared probability error. *PeerJ*, *7*, e6678. https://doi.org/10.7717/peerj.6678

Farber, O., & Kadmon, R. (2003). Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, *160*(1–2), 115–130. https://doi.org/10.1016/S0304-3800(02)00327-7

Fielding, A. H., & Bell, J. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49. https://doi.org/10.1017/S0376892997000088

Grinnell, J. (1917). The niche-relationships of the California Thrasher. *The Auk*, *34*, 427–433. https://doi.org/10.2307/4072271

Guisan, A., & Zimmermann, N. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Guralnick, R., & Hill, A. (2009). Biodiversity informatics: Automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics*, *25*, 421–428. https://doi.org/10.1093/bioinformatics/btn659

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2019). *Package 'dismo'*. R package version 1.1-4. CRAN. Retrieved from http://cran.r-project.org/web/packages/dismo/index.html

Holt, R. D. (2020). Reflections on niches and numbers. *Ecography*, *43*(3), 387–390. https://doi.org/10.1111/ecog.04828

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*, 415–427.

IUCN. (2007). *The IUCN Red List of threatened species. Leopardus wiedii*. Cambrige, UK: IUCN, NatureServe and IUCN (International Union for Conservation of Nature).

Jiménez, L., Soberón, J., Christen, J. A., & Soto, D. (2019). On the problem of modeling a fundamental niche from occurrence data. *Ecological Modelling*, *397*, 74–83. https://doi.org/10.1016/J.ECOLMODEL.2019.01.020

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., … Kessler, M. (2016). *CHELSA climatologies at high resolution for the earth's land surface areas (Version 1.1)*. World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.1594/WDCC/CHELSA_V1_1

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., … Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122. https://doi.org/10.1038/sdata.2017.122

Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C., & Anderson, R. P. (2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution*, *9*, 1151–1156. https://doi.org/10.1111/2041-210X.12945

Maguire Jr., B. (1973). Niche response structure and the analytical potentials of its relationship to the habitat. *The American Naturalist*, *107*(954), 213–246. https://doi.org/10.1086/282827

Martínez-Meyer, E., Díaz-Porras, D., Peterson, A. T., & Yáñez-Arenas, C. (2012). Ecological niche structure and rangewide abundance patterns of species. *Biology Letters*, *9*(1), 20120637. https://doi.org/10.1098/rsbl.2012.0637

Mesgaran, M. B., Cousens, R. D., & Webber, B. L. (2014). Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, *20*(10), 1147–1159. https://doi.org/10.1111/ddi.12209

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*(11), 1198–1205. https://doi.org/10.1111/2041-210X.12261

Naimi, B., & Araújo, M. B. (2016). sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*, 368–375. https://doi.org/10.1111/ecog.01881

Osorio-Olvera, L., Lira-Noriega, A., Soberón, J., Peterson, A. T., Falconi, M., Contreras-Díaz, R. G., … Barve, N. (2020). luismurao/ntbox: ntbox first release (Version v1.0.0). *Zenodo*, https://doi.org/10.5281/zenodo.3937910

Osorio-Olvera, L., & Soberon, J. (2020). *bam: Dynamic distribution models using the BAM framework*. R package version 0.1.0, 1–20. Retrieved from https://github.com/luismurao/bam

Osorio-Olvera, L., Soberón, J., & Falconi, M. (2019). On population abundance and niche structure. *Ecography*, *42*, 1415–1425. https://doi.org/10.1111/ecog.04442

Osorio-Olvera, L., Yañez-Arenas, C., Martínez-Meyer, E., & Peterson, A. T. (2020). Relationships between population densities and niche-centroid distances in North American birds. *Ecology Letters*, *23*, 555–564. https://doi.org/10.1111/ele.13453

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., … Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, *263*, 10–18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peterson, A. T., Papes, M., & Soberon, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*(1), 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions (MPB-49)* (Vol. 56). Princeton, NJ: Princeton University Press.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Qiao, H., Peterson, A. T., Campbell, L. P., Soberón, J., Ji, L., & Escobar, L. E. (2016). NicheA: Creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography*, *39*(8), 805–813. https://doi.org/10.1111/ecog.01961

Raxworthy, C. J., Martínez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M., & Peterson, A. T. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, *426*, 837–841. https://doi.org/10.1038/nature02205

Sánchez, O., Pineda, M. A., Benítez, H., González, B., Berlanga, H., & Rivera-Téllez, E. (2015). *Guía de identificación para las aves y mamíferos silvestres de mayor comercio en México protegidos por la CITES* (SEMARNAT-CONABIO, Ed.) (2nd ed.). Ciudad de México: SEMARNAT-CONABIO.

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F., & Birnbaum, P. (2017). SSDM: An r package to predict distribution of species richness and composition based on stacked species distribution models. *Methods in Ecology and Evolution*, *8*(12), 1795–1803. https://doi.org/10.1111/2041-210X.12841

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*(2), 291–307. https://doi.org/10.1111/ecog.02880

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*(2), 272–281. https://doi.org/10.1111/j.1466-8238.2011.00656.x

Ureña-Aranda, C., Rojas-Soto, O., Martínez-Meyer, E., Yáñez-Arenas, C., Landgrave Ramírez, R., & Espinosa de los Monteros, A. (2015). Using range-wide abundance modeling to identify key conservation areas for the micro-endemic Bolson Tortoise (*Gopherus flavomarginatus*). *PLoS ONE*, *10*(6), 1–14. https://doi.org/10.1371/journal.pone.0131452

Van Aelst, S., & Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 71–82. https://doi.org/10.1002/wics.19

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Osorio-Olvera L, Lira-Noriega A, Soberón J, et al. NTBOX: An R package with graphical user interface for modelling and evaluating multidimensional ecological niches. *Methods Ecol Evol*. 2020;11:1199–1206. https://doi.org/10.1111/2041-210X.13452