Methods in Ecology and Evolution
British Ecological Society

Methods in Ecology and Evolution 2014, **5,** 694–700 doi: 10.1111/2041-210X.12200

APPLICATION

SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses

Jason L. Brown*†

Department of Biology, Duke University, 125 Science Drive, Durham, NC 27705, USA

Summary

1. Species distribution models (SDMs) are broadly used in ecological and evolutionary studies. Almost all SDM methods require extensive data preparation in a geographic information system (GIS) prior to model building. Often, this step is cumbersome and, if not properly done, can lead to poorly parameterized models or in some cases, if too difficult, prevents the realization of SDMs. Further, for many studies, the creation of SDMs is not the final result and the post-modelling processing can be equally arduous as other steps.

2. SDMtoolbox is designed to facilitate many complicated pre- and post-processing steps commonly required for species distribution modelling and other geospatial analyses. SDMtoolbox consists of 59 Python script-based GIS tools developed and compiled into a single interface.

3. A large set of the tools were created to complement SDMs generated in Maxent or to improve the predictive performance of SDMs created in Maxent. However, SDMtoolbox is not limited to analyses of Maxent models, and many tools are also available for additional analyses or general geospatial processing: for example, assessing landscape connectivity of haplotype networks (using least-cost corridors or least-cost paths); correcting SDM over-prediction; quantifying distributional changes between current and future SDMs; or for calculating several biodiversity metrics, such as corrected weighted endemism.

4. SDMtoolbox is a free comprehensive python-based toolbox for macroecology, landscape genetic and evolutionary studies to be used in ArcGIS 10.1 (or higher) with the Spatial Analyst extension. The toolkit simplifies many GIS analyses required for species distribution modelling and other analyses, alleviating the need for repetitive and time-consuming climate data pre-processing and post-SDM analyses.

Key-words: ArcGIS, geographic information systems, least-cost corridors, corrected weighted endemism, ecological niche models, Maxent bias files, spatially rarefy occurrences, spatial jackknifing

Species distribution models (SDMs) and associated analyses are broadly used in ecological and evolutionary studies. Almost all SDM methods require extensive data preparation in a geographic information system (GIS) prior to model building. Often, properly formatting occurrence localities and environmental data are cumbersome and, if not properly done, can lead to poorly parameterized models. Further, for many studies, the creation of SDMs is not the final result and the post-modelling processing can be equally arduous. With this in mind, I used an open-source programming language, Python (version 2.7; Python Software Foundation, Wilmington, Delaware, USA), to produce 59 novel scripts for use in macroecology, landscape genetics, landscape ecology and evolutionary studies. Most scripts reduce the need for repetitive actions, and associated errors, by processing all the files in the same way as specified by the user. In lieu of this, it is paramount that users of these scripts carefully consider all input parameters – carelessly parameterized automated analyses are likely to be *only* a more efficient path to mediocre results. These scripts require ArcGIS 10.1 (or higher) and a Spatial Analyst extension licence (ESRI 2012). ArcGIS is a windows-based program, but also can be run on other systems using OS emulators (i.e. Parallels for Mac OS). The structure of ArcGIS facilitates the compilation of many python scripts into a single toolkit that can be easily distributed and incorporated into the software. This facet makes these python scripts accessible to non-GIS specialists and allows them to be linked to other ArcGIS tools to further automate workflows. Here, many of the scripts were created to facilitate data preparation for species distribution modelling (e.g. spatially rarefying occurrence localities, preparation of climate data, and creation of background bias files) and post-modelling analyses of species distribution models (e.g. assessment of the distributional changes between time periods) – because of this, I named this toolkit ‘SDMtoolbox’. The toolbox is not limited to facilitating species distribution modelling and is arranged into five headings: i) commonly used

*Correspondence author. E-mail: sdmtoolbox.help@gmail.com
† Present address: Department of Biology, The City College of New York, 160 Convent Ave, New York, NY 10030, USA

ArcMap Tools, ii) biodiversity measurements, iii) landscape connectivity, iv) species distribution modelling tools and v) basic batch processing tools subdivided into subgroups by input data type: a) raster tools and b) shapefile and table tools.

Commonly used ArcMap tools

This is a convenient grouping of 23 existing Spatial Analyst tools within ArcGIS 10 that are commonly used for geospatial analyses in ecology and evolution studies (i.e. *raster calculator* and *reclassify*). For legacy ArcGIS users, this familiar grouping of functions is almost identical to the tools contained in the Spatial Analysis Toolbar (now absent from ArcMap 10). Because these tools are well documented within the software and are not novel, they are not discussed in detail herein.

Biodiversity measurements

This suite of tools calculate species richness, weighted endemism and corrected weighted endemism and was inspired by the classic ArcView 3 extension ‘Endemicity Tools’ (provided by N. Danho; Crisp *et al.* 2001). Species richness is the sum of species per cell. Weighted endemism is the sum of the reciprocal of the total number of cells each species in a grid cell is found in and emphasizes areas that have a high proportion of species with restricted ranges. Corrected weighted endemism is the weighted endemism divided by the total number of species in a cell and emphasizes areas that have a high proportion of species with restricted ranges, but are not necessarily areas that are species rich. These tools input either binary SDMs or point data of species occurrences (Fig. 1c,d) and output as a raster grid, square grid shapefile or tessellated hexagon shapefile (Table 1a).

Landscape connectivity

These tools measure landscape connectivity among populations. One prevalent method to do this is to estimate least-cost paths (LCPs) among sites (e.g. Ray 2005; McRae & Beier 2007; Etherington 2011). When a single LCP is only considered, often this oversimplifies landscape processes. By using categories of cost paths that include paths with slightly more costly path lengths (relative to the LCP), researchers can better depict habitat heterogeneity and its varying roles in dispersal (also see McRae & Beier 2007 for an alternative method). In previous research, we developed a method that creates dispersal networks that utilize least-cost corridors for visualizing haplotype networks across geographic space (Chan, Brown & Yoder 2011; Fig. 1e), and here I automate the process. For researchers interested in calculating the least-cost corridors among all sites, a separate tool (*least-cost corridors and least-cost paths among all sites*) has been included (Table 1b).

Species distribution modelling tools

Many of the python scripts contained in SDMtoolbox were initially written for species distribution modelling in Maxent (being specific to the naming syntax, input and output formats). However, many of the tools can be applied to most other types of species distribution models, and when necessary, these tools have been modified to handle SDMs from other methods (contained in the ‘Universal Tools’ heading). Because the universal SDM tools require increased flexibility in input/output formats, often they require additional input information to clarify input/output file associations. For all universal tools with corresponding Maxent tools, the core analyses and processes are identical.

MODELLING IN MAXENT: SPATIAL JACKKNIFING

Spatial jackknifing (or geographically structured k-fold cross-validation) tests evaluation performance of spatially segregated localities. The *run Maxent: spatial jackknife* tool splits the landscape into 3–5 regions based on Voronoi polygons and spatial clustering of occurrence points. Models are calibrated with all permutations of the groups using occurrence points and background data from k−1 spatial groups and then evaluated with the withheld group (Table 1g). This tool also facilitates testing different combinations of the five model feature class types and regularization multipliers to optimize model performance (Shcheglovitova & Anderson 2013; Radosavljevic & Anderson 2014).

SPATIALLY RAREFYING OCCURRENCE DATA

Most SDM methods require input occurrence data to be spatially independent to perform well. However, it is common for researchers to introduce environmental biases into their SDMs from spatially autocorrelated occurrence points. The elimination of spatial clusters of localities is important for model calibration and evaluation. When spatial clusters of localities exist, often models are over-fit towards environmental biases (reducing the model’s ability to predict spatially independent data) and model performance values are inflated (Veloz 2009; Hijmans 2012; Boria *et al.* 2014). The *spatially rarefy occurrence data* tool addresses this issue by reducing occurrence localities to a single point within the user-specified Euclidian distance. This tool also allows users to spatially rarefy their data at several distances according to habitat, topographic or climate heterogeneity (Table 1d). For example, occurrence localities could be spatially filtered at 5, 10 and 30 km^2 in areas of high, medium and low environmental heterogeneity, respectively. This graduated filtering method is particularly useful for studies with limited occurrence points and can maximize the number of spatially independent localities.

QUANTIFYING DISTRIBUTIONAL CHANGES

An increasing use of SDMs is to attempt to predict distributional changes under climate change conditions (both future and past). Here, two tools were created that help summarize distributional changes between two time periods. The *distribution changes between binary SDMs* tool calculate the distributional

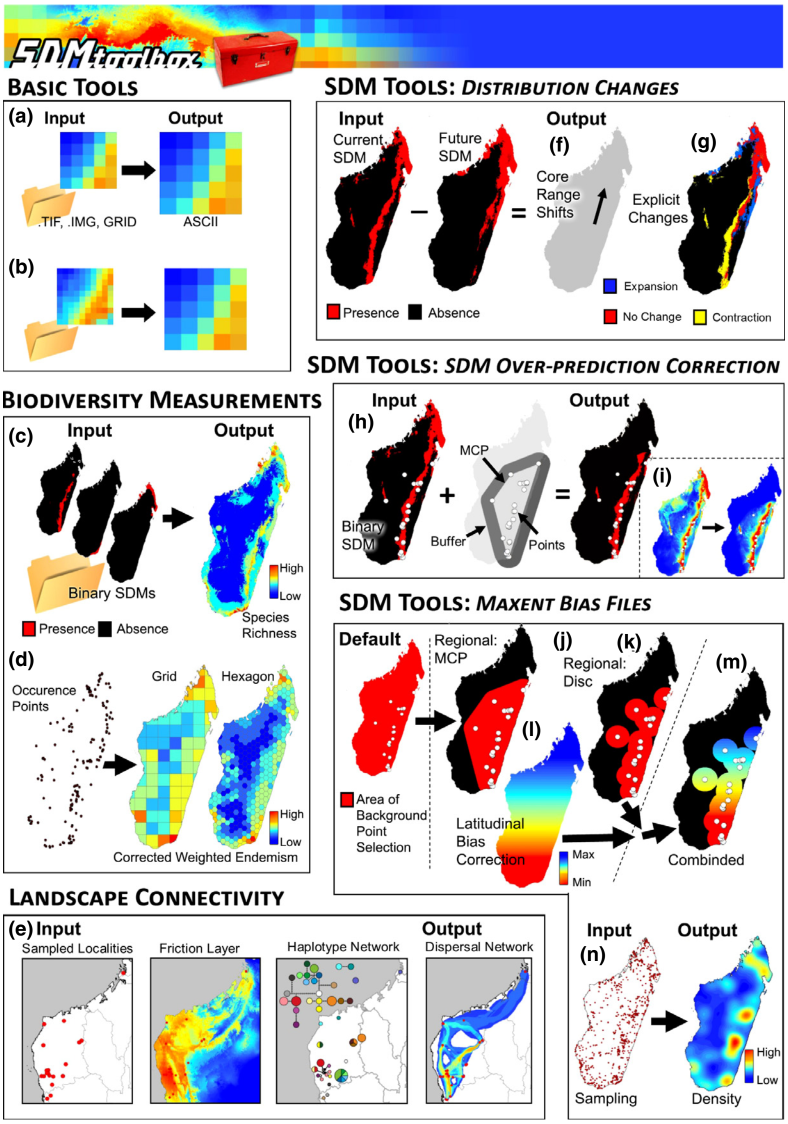

Fig. 1. Illustrative overview of SDMtoolbox. Basic Tools. SDMtoolbox contains 19 basic tools for converting and batch processing shapefile and raster data. Two of these tools are (a) batch conversion to from raster (i.e. ESRI grid) to ASCII rasters (.asc) and (b) batch resampling rasters to different resolutions. Biodiversity measurements. The toolbox also contains tools for calculating (c) species richness, weighted endemism, and (d) corrected weighted endemism from (c) binary species distribution models and (d) point occurrence data. Output from these analyses can be either raster grids, a gridded shapefile or a tessellated hexagon shapefile. Landscape connectivity. Several tools calculate least-cost corridors (and paths) among all individuals and (e) within shared haplotypes to depict a species' dispersal network (a spatial translation of a haplotype network depicting suitable areas of historic dispersal between contemporary populations). SDM Tools: Distribution Changes. Four tools are included that (f) measure the magnitude and direction of distributional changes in a species distribution and (g) quantify the explicit changes between two binary SDMs of different time periods. SDM Tools: SDM Over-Prediction Correction. Over-prediction can be an issue in even the highest quality SDMs and in some studies (i.e. of those gear towards conservation) – SDMs are being used to depict species' actual distributions, and areas of over-prediction need to be removed. SDMtoolbox contains several tools for clipping (h) binary and (i) continuous SDMs by calculating a buffered minimum convex polygon of observed localities and removing areas outside. SDM Tools: Maxent Bias Files can control where background points are selected and the density of background sampling throughout the landscape, avoiding habitat greatly outside of a species' known occurrence or accounting for both regional sampling biases and latitudinal biases associated with coordinate data. Two methods for sampling more regional backgrounds are (j) buffered minimum convex polygons and (k) maximum distance from observed localities (a.k.a disc selection). All SDMs created in a geographic coordinate system are biased in their selection of background and unique observed localities towards the poles. The level of bias depends on the breadth of latitudes an analysis covers. One solution (of two) (i) corrects the bias sampling problem by correcting how pseudo-absence values and unique occurrence localities are selected by creating (m) a latitudinal bias file. This file can be merged with the regional background selection bias file to get the combined benefits of both. (n) The last bias file minimizes the effect of sample biases. This tool produces a bias grid that up-weights presence-only data points with fewer neighbours in the geographic landscape, creating a Gaussian kernel density of sampling localities.

changes between two binary SDMs (e.g. current and future SDMs). Output consists of a table depicting contraction, expansion and stable areas in the species' distribution (in km^2; Fig. 1g). The second tool (*centroid changes*) also calculates the distributional changes between two binary SDMs (e.g. current and future SDMs); however, this analysis aims to summarize

Table 1. SDMtoolbox's functions. Each tool is documented much more extensively within the toolbox

Tool, Tool Group: Subgroup	Function	No. of Tools
A. **Biodiversity Measurements**		
Biodiversity measurements	Calculate species richness, weighted endemism and corrected weighted endemism from point data or binary SDMs. Results are output as rasters grids, vector grids or hexagons	6
B. **Landscape Connectivity**		
Least-cost corridors and least-cost paths among shared haplotypes	Creates a raster of the sum of least-cost corridors and a polyline shapefile of least-cost paths between populations that share haplotypes	1
Least-cost corridors and least-cost paths among all populations	Creates a raster of the sum of least-cost corridors and a polyline shapefile of least-cost paths between all populations	1
C. **SDM Tools: Distribution changes between binary SDMs and over-prediction Correction**[1]		
Distribution changes between binary SDMs	Calculates the distributional changes between two binary SDMs (e.g. current and future SDMs). Output is a table depicting the km^2 of range contraction, range expansion and no change in the species distribution	2
Centroid changes (lines)	This analysis is focused on summarizing the core distributional shifts in many species' ranges. This analysis reduces each species' distribution to a single central point (a centroid) and creates a vector file depicting magnitude and direction of change through time	2
Over-prediction correction: clip models by buffered minimum convex polygons	To limit over-prediction of SDMs, two tools were created that clip SDMs by a buffered minimum convex polygon (MCP) generated from the input point data of each species following the approach of Kremen *et al.* (2008). This method produces models that represent suitable habitat within an area of known occurrence (based on a buffered MCP), excluding suitable habitat greatly outside of the observed range	4
Limit dispersal in future SDMs	This analysis limits the dispersal of a species in future climates to a specified distance	2
D. **SDM Tools: Universal Tools– Pre- and Post-SDM tools**		
Create friction layer: invert SDMs	This tool creates a friction layer from a SDM by inverting the model (x inverted = 1-x; e.g. SDM raster values of 1, 0·7, and 0·2 would be changed to 0, 0·3 and 0·9, respectively)	1
Explore climate data: correlation and summary stats	Calculates summary statistics (mean, maximum, minimum and standard deviation) for each input raster layer. This tool also calculates correlation and covariance coefficients of each input raster to all others (output is in the form of a matrix)	1
Explore climate data: remove highly correlated variables	This tool evaluates the correlations among all input environment data and then removes layers that are correlated at the user-specified level	1
Spatially rarefy occurrence data	Reduces occurrence localities to a single point within a specified distance(s)	1
Calculate climate or topographic heterogeneity for rarefying data	These tools calculate the climatic or topographic variation within a specified area and are meant to be used with the multi-distance option in *spatially rarefy occurrence data* tool	3
E. **SDM Tools: Maxent tools– Correcting latitudinal pseudo-absence selection bias**		
Project climate and distribution data into equal-areas projection	Projects all the Maxent input data to an equal-areas projection	4
Create a bias file for coordinate data	Corrects sampling bias by correcting how background values and unique occurrence localities are selected	4
F. **SDM Tools: Maxent tools– Background Selection via bias files**		
Sample by distance from observed point	Samples background points within a maximum radial distance of known occurrences	1
Sample by buffered MCP	Restricts background point selection within a buffered minimum-convex polygon based on known occurrences.	1
Gaussian kernel density of sampling localities	This tool accounts for sampling biases. This method produces a bias grid that up-weights presence-only data points with fewer neighbours in the geographic landscape	1
G. **SDM Tools: Maxent tools– Modelling in Maxent: Spatial Jackknifing**		
Modelling in Maxent: Spatial Jackknifing	This tool splits the landscape into 3–5 regions based on Voronoi polygons and spatial clustering of occurrence points. Models are calibrated with all permutations of the groups using occurrence points and background data from k−1 spatial groups and then evaluated with the withheld group. This tool facilitates testing different combinations of five model feature class types and regularization multipliers to optimize Maxent model performance	1
H. **Basic Tools: Shapefile and Table Tools**		
CSV, TXT or XLS to shapefile	Converts a CSV, TXT or XLS spreadsheet with latitude and longitude to a point shapefile	1
Shapefile to CSV	Converts a point shapefile to CSV spreadsheet with latitude and longitude	1
Randomly select points	Randomly selects a subset of points	1

(continued)

Table 1. (continued)

Tool, Tool Group: Subgroup	Function	No. of Tools
Split shapefile by fields	Splits a shapefile into several by unique field ID	1
Batch project shapfile to any projection (by folder)	Projects a folder of shapefiles to any specified projection	1
Create tessellated hexagons of a region	Creates regular hexagons tessellated across the defined study area	1
Sample raster or feature values to hexagons	This tool samples raster values to tessellated hexagon shapefiles created with the *create tessellated hexagons of a region* tool. It does this by sampling all the raster values contained within each hexagon and summarizes the values based on user input metrics (e.g. taking a mean of values)	1
I. **Basic Tools: Raster tools**		
Batch extract by mask (by folder)	Clips a folder of rasters (such as climate data) to a smaller area defined by a mask, coordinates or map extent	1
Batch raster to ASCII (by folder)	Converts a folder of rasters to ASCII rasters	1
Batch ASCII to raster (by folder)	Converts a folder of ASCII rasters to another raster format	1
Batch raster to raster (by folder)	Converts a folder of rasters (.tif,.img,.bil,.bip,.bmp,.bsq,.dat,.gif,.jpg,.jp2,.png) to another raster format (.tif,.img,.bil,.bip,.bmp,.bsq,.dat,.gif,.jpg,.jp2,.png,.flt)	1
Batch project raster to equal-area projection (by folder)	Projects a folder of WGS 1984 projected rasters to an equal-area projection	1
Batch project raster to any projection (by folder)	Projects a folder of rasters to any specified projection	1
Batch resample grids (by folder)	Resamples a folder of rasters to any resolution	1
Batch sum rasters – any extent (by folder)	Sums all rasters in a folder of any extent	1
Batch sum rasters – all same extent (by folder)	Sums all rasters in a folder of the same extent	1
Apply same colour ramp to all open rasters	Applies the same colour ramp to all open rasters	1
Quick reclassify to binary	Quickly reclassifies a raster to a binary raster based on a cut-off threshold	1
Quick reclassify	Quickly reclassifies a raster to user-specified values. Simply input the cut-off ranges	1
Batch reclassify (by folder)	Quickly reclassifies all rasters in a folder to the same user-specified values	1
Correlation and Summary Stats	Calculates summary statistics (mean, maximum, minimum and standard deviation) for each input raster layer. This tool also calculates correlation and covariance coefficients of each input raster to all other included rasters (output is in the form of a matrix)	1
Zonal statistics of many rasters to single table	Calculates statistics on values of a raster within the zones of another data set. This tool outputs the results of all input rasters to a single table	1

[1]Tools are available under both *SDM Tool* subcategories: *Universal Tools* and *Maxent Tools*.

the core distributional shifts in multiple species' ranges. Moreover, this analysis reduces each species' distribution to a single central point (a.k.a. a centroid) and creates a vector depicting magnitude and direction of change through time (Fig. 1f; Table 1c). The *limit dispersal in future SDMs* tool limits the maximum dispersal distance from current SDMs into future SDMs. This excludes areas that are predicted climatically suitable, but are too distant to be colonized.

CORRECTING MODEL OVER-PREDICTION

To limit over-prediction of SDMs, a common problem with modelling species' distributions (Peterson *et al*. 2011), the *over-prediction correction: clip models by buffered MCPs* tools were created, which clip SDMs by a buffered minimum convex polygon (MCP) generated from the input point data of each species following the approach of Kremen *et al*. (2008). This method reduces SDMs to represent suitable habitat within an area of known occurrence (based on a buffered MCP), excluding suitable habitat greatly outside of the observed range (Fig. 1h,i; Table 1c).

CREATION OF FRICTION LAYERS: INVERT SDM

The use of least-cost paths and along-path distances often dramatically improve the calculation of geographic distances for testing hypotheses (i.e. isolation by distance). However, few studies have access to meaningful friction landscapes (Spear *et al*. 2010). Some researchers generate friction landscapes from classified satellite images where each major habitat type represents a different friction value (i.e. Broquet *et al*. 2006). A primary downfall to using habitat heterogeneity as a friction landscape is the weighing of each habitat class to represent relevant friction values. Proper implementation relies heavily on expert life-history knowledge, and when done, analyses lose some objectivity. For example, Broquet *et al*. (2006) adjusted the friction values until they satisfied prior expectations. More recently, authors have used inverted SDMs as friction landscapes, the method implemented in SDMtoolbox (Wang *et al*. 2008; Chan, Brown & Yoder 2011). This method is a more objective alternative to expert knowledge, reducing potential bias, and high-quality SDMs can be generated for most species (Table 1d).

BIAS LAYERS FOR MAXENT

A subset of python scripts create bias files used to fine-tune background and occurrence point selection in Maxent. Bias files control where background points are selected and the density of background sampling. Proper use of bias files can avoid sampling habitat greatly outside of a species' known occurrence or can account for both collection sampling biases and latitudinal biases associated with coordinate data.

Background points (and similar pseudo-absence points) are meant to be compared with the presence data and help differentiate the environmental conditions under which a species can potentially occur. Typically background points are selected within a large rectilinear area; within this area, there often exists habitat that is environmentally suitable, but was never colonized. When background points are selected within these habitats, this increases commission errors (false positives). As a result, the 'best' performing model tends to be over-fit because selection criterion favours a model that fails to predict the species in the un-colonized climatically suitable habitat (Anderson & Raza 2010; Barbet-Massin *et al.* 2012). The likelihood that suitable unoccupied habitats are included in background sampling increases with Euclidian distance from the species' realized range. Thus, a larger study spatial extent can lead to the selection of a higher proportion of less informative background points (Barbet-Massin *et al.* 2012). Researchers should not avoid studying species with broad distributions or those existing in regions that do not conform well to rectilinear map layouts, rather they simply need to be more selective in the choice of background points in Maxent (and pseudo-absences in other SDM methods)(Barve *et al.* 2011; Merow, Smith & Silander 2013).

To circumvent this problem, many researchers have begun using background point and pseudo-absence selection methods that are more regional. SDMtoolbox contains two tools to facilitate more sophisticated background selection for use in Maxent. The *sample by distance from obs. pts.* tool uses a common method that samples backgrounds within a maximum radial distance of known occurrences (Fig. 1k; see Thuiller *et al.* 2009). The *sample by buffered MCP* tool restricts background selection within a buffered minimum convex polygon based on known occurrences (Fig. 1j; Table 1e).

One limitation of presence-only data SDM methods is the effect of sample selection bias from sampling some areas of the landscape more intensively than others (Phillips *et al.* 2009). Maxent requires an unbiased sampling of occurrence data, and spatial sampling biases can be reduced by using the *Gaussian kernel density of sampling localities* tool. This method produces a bias grid that up-weights presence-only data points with fewer neighbours in the geographic landscape. To do this, the tool creates a Gaussian kernel density of sampling localities (Fig. 1n). Output bias values of 1 reflect no sampling bias, whereas higher values represent increased sampling bias. Depending on the study, the input points could be all sampling localities for a larger taxonomic group or simply the input sampling localities of a focal species. For example, if I were studying a single species of frog from Madagascar, I could use either i) only the occurrence points from that species, or ii) all sampling points from all amphibians in Madagascar. The former focuses on sampling biases in the focal species, where the latter focuses on widespread spatial sampling biases (e.g. sampling only near roads) and likelihood of detection of your species in all surveys.

A last set of bias files address a prevalent issue in species distribution modelling, yet rarely acknowledged, and regard a latitudinal bias in background selection. To understand this bias, the basic difference between a geographic coordinate system and projected coordinate system must be clarified. Briefly, a geographic coordinate system consists of data measured in angles across a globe in the form of latitude and longitude. Different geographic coordinate systems vary in terms of their frames of reference for measuring locations on the surface of the earth; however, all utilize degrees latitude and longitude. A projected coordinate system refers to data in a two-dimensional plane that can be measured in fixed linear units (i.e. feet or metres). All SDMs created in geographic coordinate systems (i.e. decimal degrees) in Maxent, and most pseudo-absence-based SDM methods, are biasing their selection of pseudo-absence/background points and unique occurrence points towards the poles (Elith *et al.* 2011). The level of bias depends on the breadth of latitudes of the study area and the heterogeneity of habitats encompassed in the landscape. For example, this is a bigger issue for broadly distributed temperate species than for species distributed around the equator. This is because the area occupied by these units decreases latitudinally and areas are largest at equator and smallest at poles. This inequality results from convergence of the meridians, lines of longitude, towards the poles. There are two solutions to this problem: i) the first corrects the bias sampling problem by correcting how background values and unique occurrence localities are selected. ii) The second solution fixes the problem by projecting all the data into an equal-areas projection (EAP). The latter is the preferred method because it directly addresses the issue of unequal area sizes; however, for many modellers, this requires considerable effort and can be confusing due to issues associated with selecting the best EAP. The SDM toolbox facilitates both solutions (Fig. 1l,m; Table 1e). See Supplementary Materials for detailed explanation and comparison of these methods.

Basic tools

Most of the basic tools facilitate conversion or importation of data for use in species distribution modelling software. This includes converting '.csv', '.txt' or '.xls' tables with coordinate data into shapefiles (or vice versa). Ten scripts pertain to batch processing of raster files (i.e. preparing climate data for use in Maxent), and collectively, these reduce the number of processing steps required by an order of magnitude or more. Main batch features include ASCII files to raster files (and vice versa; Fig. 1a), projection of rasters to any projection (including equal-areas projections), clipping raster files to a particular extent, re-sampling rasters to lower or higher resolutions (Fig. 1b), reclassifying raster values and the summation of many rasters. Other basic tools facilitate the processing of larger data files, for example, *the randomly selected points* tool can be used to subsample data sets or the *split shapefile by field attributes* tool can split point shapefiles by any field class (i.e. species ID, locality or genetic group; Table 1h). Lastly, the *raster correlations and summary statistics* script calculates summary statistics and correlation coefficients of rasters, a tool to help researchers explore correlations between environmental rasters before running SDMs (Table 1i).

Conclusion

The scripts contained with SDMtoolbox have been developed through years of geospatial research and I provide them hoping they streamline the geospatial analyses of others. The toolkit simplifies many GIS processes required for species distribution modelling and other common geospatial analyses, alleviating the need for repetitive and time-consuming climate data pre-processing and post-SDM analyses. In the light of this, it is important to remind users to carefully parameterize each tool so that the analyses and corresponding species distribution modelling produce meaningful results. The toolbox also allows researchers to use more current ESRI software and reduces the amount of time that would be spent developing common solutions. The latest version of SDMtoolbox, a user guide, and example data are freely available at http://www.sdmtoolbox.org. For questions or suggestions regarding SDMtoolbox, please email sdmtoolbox.help@gmail.com.

Acknowledgements

I would like to thank members of the Yoder laboratory, in particular Christopher Blair, who inspired this project and provided guidance, comic relief and useful comments after its inception. I am grateful to Jason Roberts, Robert P. Anderson and Mariano Soley-Guardia for their assistance. I would like to thank the anonymous creator(s) of Maxent Tools v9.2 for also inspiring this release and for the careful consideration of the editors and four anonymous reviewers. JLB received salary support from Duke University Start-up funds to Anne D. Yoder. During revisions of manuscript, JLB was supported by a co-funded FAPESP (BIOTA, 2013/50297-0), NSF (DOB 1343578), and NASA Dimensions of Biodiversity grant to Ana C. Carnaval.

Data accessibility

No original data were used in this manuscript.

References

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Boria, R.A., Olson, L.E., Goodman, S.M. & Anderson, R.A. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modeling*, **275**, 73–77.

Broquet, T., Ray, N., Petit, E., Fryxell, J.M. & Burel, F. (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology*, **21**, 877–889.

Chan, L.M., Brown, J.L. & Yoder, A.D. (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution*, **59**, 523–537.

Crisp, M.D., Laffan, S., Linder, H.P. & Monro, A. (2001) Endemism in the Australian flora. *Journal of Biogeography*, **28**, 183–198.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

ESRI (2012) *ArcGIS Desktop and Spatial Analyst Extension: Release 10.1*. Environmental Systems Research Institute, Redlands, CA.

Etherington, T.R. (2011) Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution*, **2**, 52–55.

Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J., Thomas, C.D., Beentje, H. *et al.* (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222–226.

McRae, B.H. & Beier, P. (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, **104**, 19885–19890.

Merow, C., Smith, M.J. & Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Peterson, T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, NJ.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting, and evaluation. *Journal of Biogeography*, **41**, 629–643.

Ray, N. (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes*, **5**, 177–180.

Shcheglovitova, M. & Anderson, R.P. (2013) Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecological Modelling*, **269**, 9–17.

Spear, S.F., Balkenhol, N., Fortin, M.-J., McRae, B.H. & Scribner, K. (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.

Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.

Wang, Y.H., Yang, K.C., Bridgman, C.A. & Lin, L.K. (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, **23**, 989–1000.

Received 1 October 2013; accepted 2 April 2014
Handling Editor: Barbara Anderson

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. A case study and the use of the coordinate bias file.

Table S1. Niche-overlap was measured using three different statistics.

Figure S1. Latitudinal area changes in geographic coordinate data.

Figure S2. A case study of latitudinal biases in background and occurrence point selection.