



SDMtoolbox User Guide

By Jason L. Brown, Southern Illinois University

Last updated: 2/9/2020

The SDMtoolbox consists of a series of python scripts designed automate complicated ArcGIS analyses that are often too difficult/tedious for the average GIS user to execute efficiently. This tutorial gives a basic introduction to use SDMtoolbox for ecology and evolution studies. A large set of the tools here are designed to complement MaxEnt species distribution models (SDMs) or to improve the predictive performance of MaxEnt models. The MaxEnt program uses maximum entropy modelling of species' geographic distribution (Phillips et al. 2006) and has become one of the most prevalent methods due to its high predicted performance, computational efficiency and ease of use. Many of the analyses are also available for use on other SDM methods (see: SDM Tools → 1. Universal Tools).

If you use the SDMtoolbox please remember to cite it. By citing this you encourage the development of more tools.

Please use the following reference:

Brown, J.L. (2014) SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*



Getting Started

Downloading

The latest version of the toolbox is available for download at: www.sdmtoolbox.org. This software requires ArcMap 10.1 or 10.2 with an active Spatial Analyst license (www.ESRI.com). This toolbox is programmed specifically for ArcMap 10.1 (and above) and due to a series of improvements in this version, it is not backwards compatible with older releases of ArcMap (*i.e.* ArcMap 9.2). This software consists of an ArcGIS toolbox and associated python scripts.

Installation

1. Download, unzip 'SDMtoolbox.zip' and place all files in a folder on your harddrive where you want the toolbox to be stored (e.g ...\\Documents\\ArcGIStoolboxes)
2. Close all ArcGIS programs
3. Open **ArcCatalog 10.1** (a program in the ESRI ArcGIS programs, see icon below)



Above: ArcCatalog Icon

4. Once open, activate the ArcToolbox window (if not already visible) by clicking the red box below

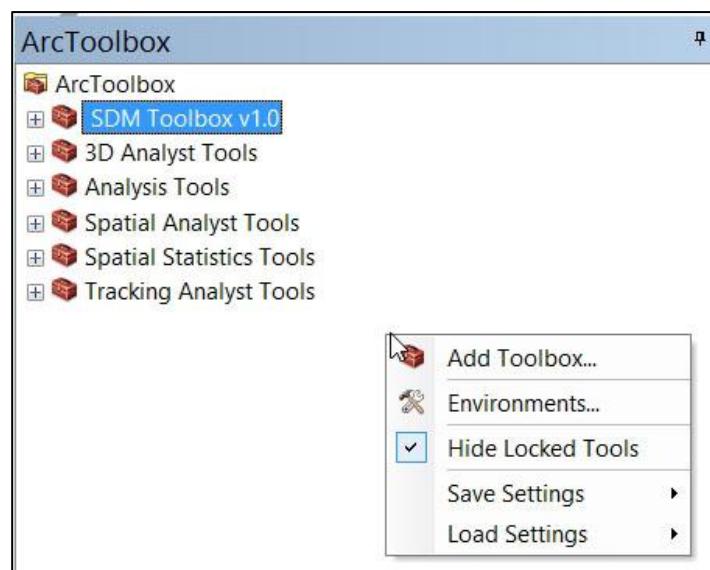


5. Then right-click ArcToolbox Window (see image below) → select 'Add Toolbox.' Then go to location of downloaded toolbox and select 'SDMToolbox v1.0.tbx'

6. The toolbox should appear inside the ArcToolbox (see highlighted toolbox to right). If there, now the toolbox is almost installed.

7. To finish installation simply close the ArcCatalog program.

8. Now start ArcMap and begin using the SDMtoolbox!





How to update or uninstall SDMtoolbox

1. Close all ArcGIS programs
2. Open **ArcCatalog 10** (a program in the ESRI ArcGIS programs, see icon below)

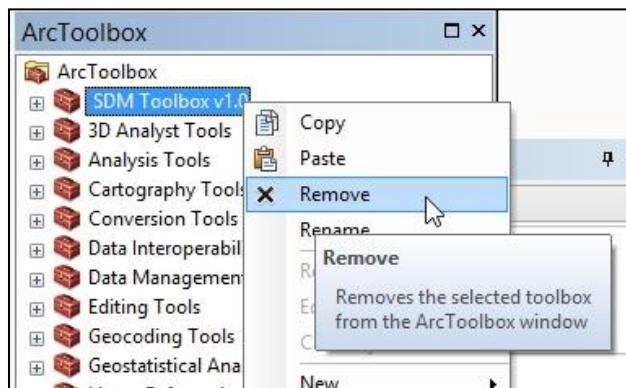


Above: ArcCatalog Icon

3. Once open, activate the ArcToolbox window (if not already visible) by clicking the red box below



4. Then right-click SDMtoolbox in ArcToolbox Window (see Image to right) and select 'Remove'.
5. If only uninstalling, then you are done. If installing a new version of SDMtoolbox then right-click ArcToolbox Window (see image below) → select 'Add Toolbox.' Then go to location of downloaded toolbox and select 'SDMToolbox v1.X.tbx'
6. The toolbox should appear inside the ArcToolbox (see highlighted toolbox to right). If there, now the toolbox is almost installed.
7. To finish installation simply close the ArcCatalog program.
8. Now start ArcMap and begin using the SDMtoolbox!



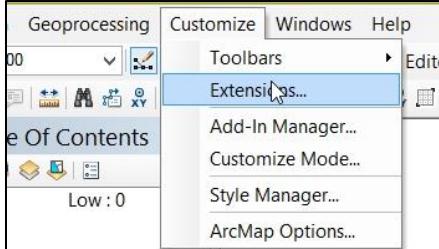
A First Run

Many things will cause the SDMtoolbox to *not* run that have *nothing* to do with the toolbox itself. Upon first use, the following steps should be performed:

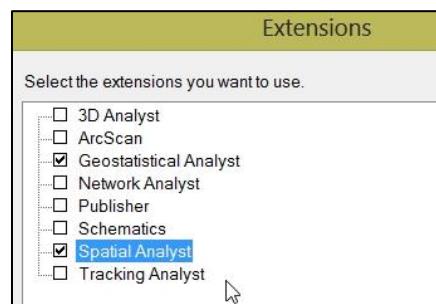
1. Open ArcMap10.1 and activate the ArcToolbox window (if not already visible). Bottom image on this page is the Arctoolbox window. If not visible, click toolbox icon (see image below).



2. Ensure Spatial Analyst is Enabled in ArcMap



A. Go to: Customize → Extensions



B. Check the box next to: *Spatial Analyst* (and *Geostatistical Analyst* if available)

3. Install all relevant ArcMap 10.1 or 10.2 Patches and Service Packs.

The base release of ArcMap 10.1/2 has some serious bugs that cause many analyses to fail. The only way to fix these bugs is the install all relevant patches and service packs.

- A. To search for these go to:

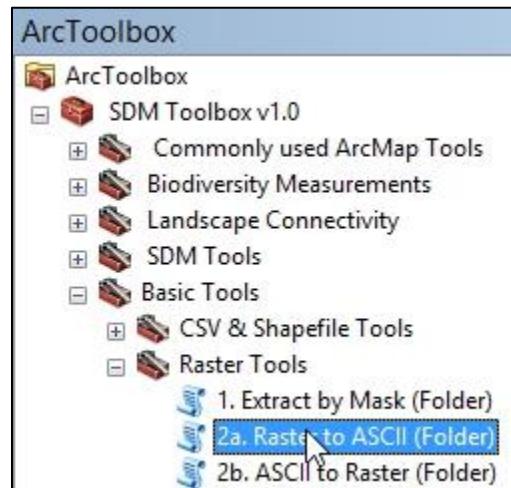
<http://support.esri.com/en/downloads/patches-servicepacks/list?productId=160&productVersions=10.1&categoryTypes=5&categoryTypes=8>

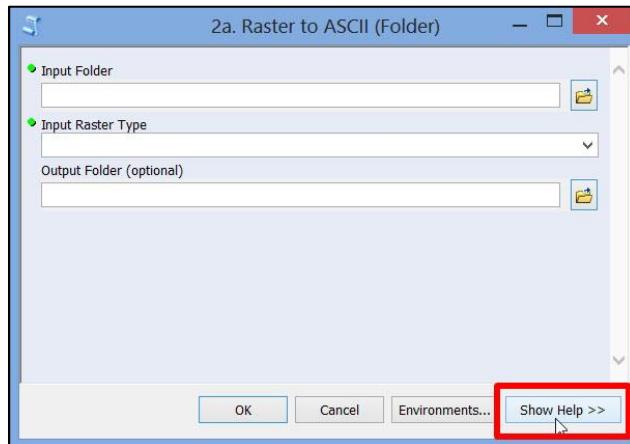
Or: <http://support.esri.com/en/downloads/patches-servicepacks>

- B. Or use the ESRI's 'Patch Finder for Windows' to identify patches, then search for suggested patches/service packs (<http://downloads2.esri.com/Support/downloads/other/PatchFinder.exe>)

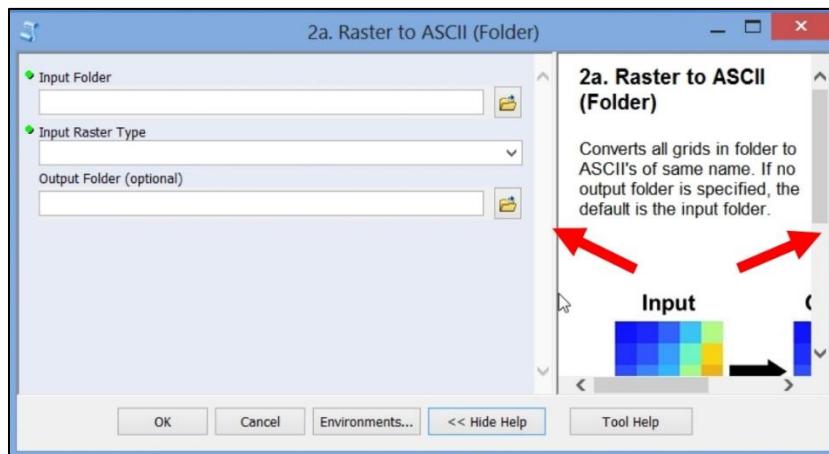
4. Enable 'Tool Help' and resize windows.

Open the 'Basic Tools → Raster Tools → 2a. Raster to ASCII (Folder)' tool. Double-Click the tool (see image to right)

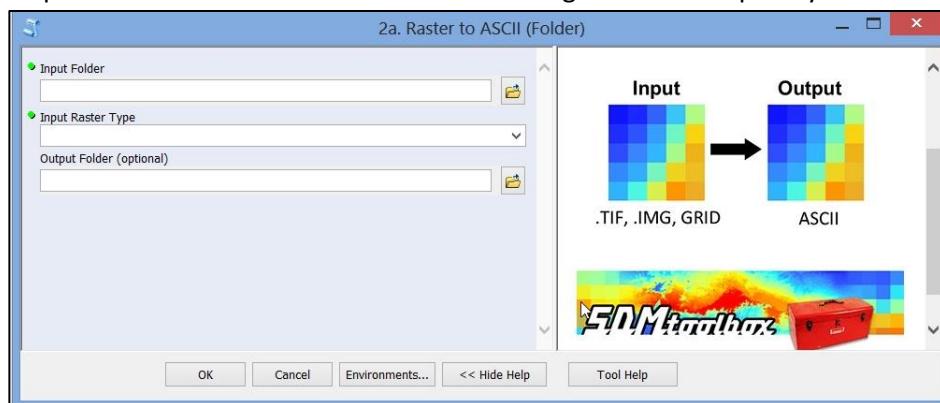




- A. In the newly opened tool click 'Show Help' (red box above)



- B. Resize the window by dragging the two windows (depicted by red arrows) so that the final tool window is as below. Note the SDMtoolbox Icon should be fully visible. Resizing the help window to this size will ensure that all diagrams are completely visible.





Running a SDM in MaxEnt: from Start to Finish

Below is a brief overview of *my view of the best practices* of correlative species distribution modeling and how SDMtoolbox will facilitate achieving them. This overview focuses on modeling in MaxEnt, but many steps are applicable to all types of distribution modeling. For overview of major assumptions and other considerations, see table at the end of this document.

Species distribution modelling (SDM) occurs in two phases: 1) **Data compilation** and 2) **Model creation, calibration, and validation**

Data compilation

This step includes collecting occurrence records of the focal species and environment data for its habitats.

Occurrence Data

The single most important component of any SDM is the input occurrence records. Extra care should go into selecting, and then processing, these points. The quality, distribution and number of points are directly related to the accuracy of the model. Use as many high-quality locality points as possible (e.g. GPS data collected with confident taxonomic identification) and try to collect occurrence records that are evenly sampled throughout the species' range and avoid biases in the sampling method (e.g., sampling only from road transects). It is better to have only a limited number of points that satisfy the above conditions than many points of vague credence (e.g. be skeptical of points downloaded from internet databases, particularly those that are georeferenced from locality info) (Chan et al. 2011).

Environment Data

The environment data provide the landscape-level data to quantify the focal species' ecological tolerances. Include variables that are likely to be directly relevant to the species being modeled. However do not add all available climate data without regard to the redundancy of the data. Many environmental variables are tightly correlated making some redundant, this makes interpreting the influences of each variable in the model difficult. If not included in your model, consider the effects of the following items on the present distribution of your species: fire history, glaciations, contagious diseases, anthropogenic factors, recent geological changes, the species' movement potential through the landscape or biotic interactions.

1. Preparing Worldclim Climate Data for use in MaxEnt Analyses

1A. Preparing Worldclim Climate Data: Clip the raster to area of species' extent

Tools: Extract by Mask (Folder) and Raster to ASCII

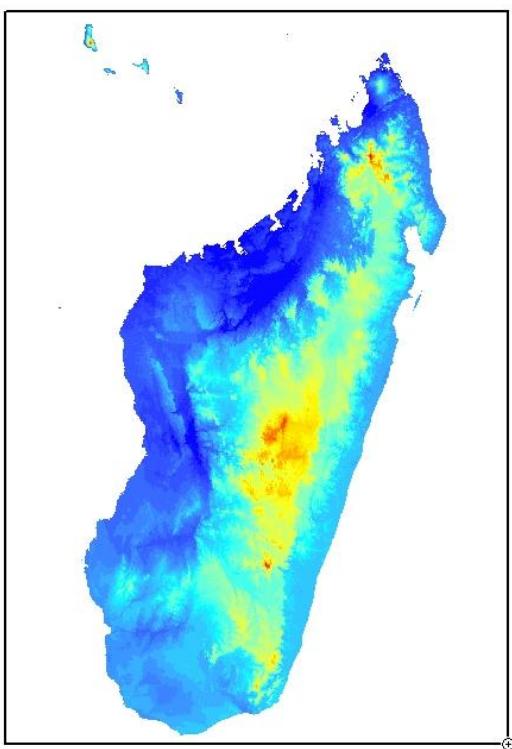
ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Download ESRI grid climate data (e.g. the 30 arc-second bioclim) from worldclim.org
3. Open one of the newly downloaded layers in ArcMap

4. There are several ways to define the area to clip the climate data into. One of the easiest ways is to simply zoom the display window the desired extent (Image below, where I wanted to reduce climate data to Madagascar). Other ways include defining the max-min XY coordinates (a bounding box) or using another GIS layer as a template (such as, a country's boundary). Select one of these methods and continue. Note the area should encompass an area about 50-100 km (ca. 0.5-1 degree) greater than total distribution of **all** your focal species. We will then use bias files to limit background selection of each species to meaningful areas within this area.



Zoom tool



Zoom to desired area

5. Double-click the 'Basic Tools → Raster Tools → 1. Extract by Mask (Folder)
6. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input folder containing the full extent Worldclim data, input raster should be ESRI grid format
2. Check the box depicting that the rasters are ESRI grid format.
3. Select output folder location
4. Select either a 'mask' or select 'extent' and choose the appropriate method for defining the extent (here I used 'Same as Display').
5. Execute tool



1B. Preparing Worldclim Climate Data: Convert rasters to ASCII

Tool: Raster to ASCII

ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the ‘Basic Tools → Raster Tools → 2a. Raster to ASCII (Folder)
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input folder containing the clipped Worldclim data output from previous tool
2. Select raster type “Tiff(.tif)”
3. Select output folder location
4. The climate data are ready for use in MaxEnt

1C. Preparing Worldclim Climate Data: Define projection of ASCII climate data

Tool: 3d. Define Projection as WGS84 or ArcMaps’s Define Projection

IF input data are coordinates and WGS84 then:

ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the ‘Raster Tools → 3d. Define Projection as WGS84 (folder)’
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input folder with clipped ASCII worldclim data.
2. Execute tool

IF input data are another projection, use ArcMap’s tool: “Define Projection”.

2. Optional Step. Which variables should I use? Testing Autocorrelations of Environmental Data

If you are interested interpreting how each input environmental variable contributes to your species distribution model, then you need to reduce autocorrelation of your input environmental data by removing highly correlated variables. It is widely known that many climate variables are highly correlated with each other. While including all these will not affect the predictive quality of your MaxEnt model, it does seriously limit any inference of the contribution of any correlated variables (*i.e.* the MaxEnt outputs from ‘Analysis of variable contributions’ and to some degree ‘Jackknifing environmental variables’). This is mainly because when a model is built in MaxEnt, if a highly correlated variable is included in the model, this often excludes all other highly correlated variables from being incorporated. This is because these variables likely would contribute similarly to the models. Since they are not included, they will not be properly represented in the output ‘Analysis of variable contributions’.

Tool: Explore Climate Data: Remove Highly Correlated Variables

ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the ‘Basic Tools → SDM Tools → Universal SDM Tools → Explore Climate Data → Remove Highly Correlated Variables’
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Select all the clipped Worldclim data (‘control+shift’ will allow you to select all items in a folder). Layers that you wish to retain (vs. the other correlated layers) should be first in the list. All correlated layers that occur after will be excluded. For interpreting influence of environmental layers in the SDM, I prefer to place layers that depict metrics frequently used in non-SDM ecology and evolution studies [such as: BIO1 = Annual Mean Temperature, BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)), BIO12 = Annual Precipitation]. Further for simplicity, these layers often best represent the original input climate data (as they directly reflect the actual measurements) and are not derived from several layers or a subset of the data.
2. Maximum correlation allowed. Multiple values can be input separated by semicolon (‘;’). Input a value between 0-1. The absolute value of the correlation coefficients range from 0 to 1. A value of 1 implies that a linear equation describes the relationship between X and Y perfectly. A value of 0 implies that there is no linear correlation between the variables.
3. Input NoData Value. Note this must be the same for all values or else correlations will not be accurate. Since we used only Worldclim here, this should not be an issue (as all values are the same). To check NoData values, import layers into ArcGIS and right click the layer and select ‘Properties’ and then go to the ‘Source’ tab. Alternatively, you can simply open your “.asc” files in a text editor and at the top of header will be the NoData value.
4. Select output folder location. Output will be two tables with the correlation coefficients among all comparisons and a table with the final list of rasters to include in your model.

3. Preparing Occurrence Data for use in MaxEnt Analyses

3A. Preparing Occurrence Data: Import Species Occurrence Records

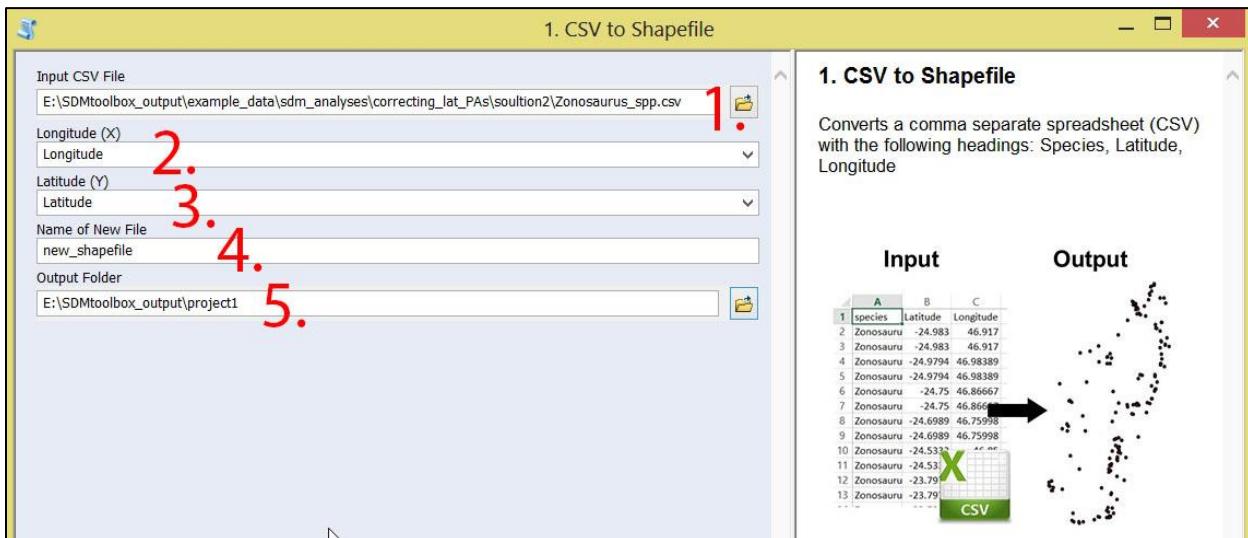
Tool: CSV, TXT, XLS to shapefile

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import CSV, TXT or XLS file with occurrence records. Table of species occurrences containing ONLY: Species ID, Longitude and Latitude. The fields must be in that order. For species ID, do not use non-alphanumeric characters in names (e.g., * : \ / < > | " ? [] ; = + & £ \$), replace spaces with "_", and remove periods.
3. Double-click the 'CSV & Shapefile Tools → 1. CSV to Shapefile'
4. Continue to tool interface instructions (below)

species	Longitude	Latitude
Zonosaurus_aeneus	46.917	-24.983
Zonosaurus_aeneus	46.917	-24.983
Zonosaurus_aeneus	46.98389	-24.97944
Zonosaurus_aeneus	46.98389	-24.97944
Zonosaurus_aeneus	46.86667	-24.75

MaxEnt Input Species Data Format: three columns with species, longitude and latitude--- as ordered here.



CSV to Shapefile tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input CSV file with columns with latitude and longitude
2. Field with longitude
3. Field with latitude
4. Name of new shapefile
5. Select output folder location



3B. Preparing Occurrence Data: Define projection of occurrence points shapefile

Tool: 6b. Define Projection as WGS84 or ArcMaps's Define Projection

ARCGIS STEP-BY-STEP GUIDE:

IF input data are coordinates and WGS84 (if not see below):

1. Double-click the 'CSV & Shapefile Tools → 6b. Define Projection as WGS84'
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input newly imported shapefile points
2. Execute tool

IF input data are another projection, use ArcMap's tool: "Define Projection".

3E. Preparing Occurrence Data: Spatial Rarefying

An Unbiased Sample: a Need for Spatial Rarefying

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 auto-correlated occurrence points. The elimination of spatial clusters of localities is important for model calibrating 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 *et al.* 2012; Boria *et al.* 2014). The *spatially rarefy occurrence data* tool addresses this issue by spatially filtering locality data by a user input distance, reducing occurrence localities to a single point within the 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 km², 10 km² and 30 km² in areas of high, medium and low environmental heterogeneity, respectively. This graduated filtering method is particular useful for studies with limited occurrence points and can maximize the number of spatially independent localities.

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

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

-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.

Tool: Spatially Rarefy Occurrence Data

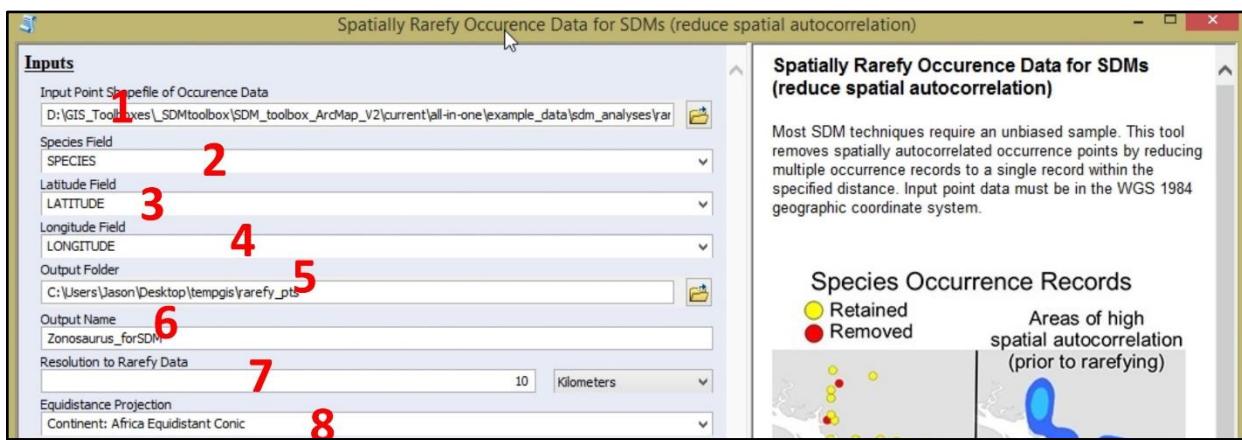
ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the 'SDM Tools → 1. Universal Tools → Spatially Rarefy Occurrence Data for SDMs (reduce spatial autocorrelation)' tool
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input shapefile of species occurrence data, here select 'occurrences.shp'
2. Shapefile field corresponding to species identity, here select 'SPECIES'

3. Shapefile field corresponding to latitude, here select 'LATITUDE'
4. Shapefile field corresponding to longitude, here select 'LONGITUDE'
5. Select output folder location. This should be a new empty folder. If not empty this can cause the analysis to fail, particularly if temporary files from a previous analysis were not properly removed (e.g. this can happen if another SDMtoolbox analysis is terminated early).
6. Name of output file. Note that "_rarefied_points" will be appended to file name.
7. The spatial resolution to rarefy the data. Here use the default settings. Note that this value will NOT actually be used here because you will be executing the multi-distance occurrence data rarefying.
8. Please select proper equidistance projection.



Spatially Rarefy Occurrence Data tool interface

4. Creation of Bias Files

Background Selection via Bias Files

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 collection sampling biases 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 exist 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 favor a model that fail 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 et al. 2013).

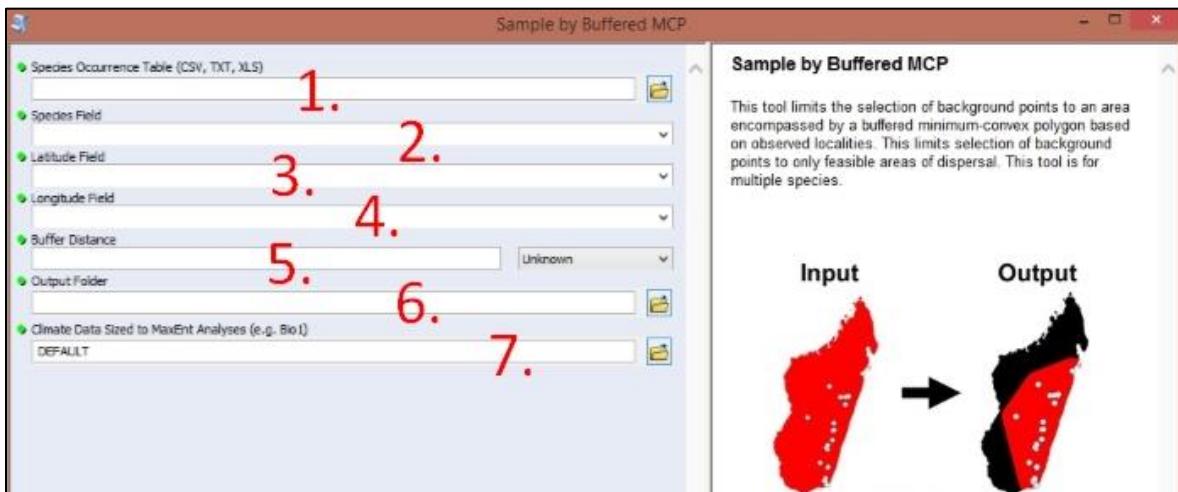
To circumvent this problem, many researchers have begun using background point and pseudo-absence selection methods that are more regional. SDMtoolbox contains three tools to facilitate more sophisticated background selection for use in Maxent. The *Sample by Distance from Obs. Pts.* tool (see: SDM Tools →2. MaxEnt Tools → Background Selection via Bias Files) uses a common method that samples backgrounds within a maximum radial distance of known occurrences (see Thuiller et al. 2009). The Sample by buffered MCP tool restricts background selection with a buffered minimum-convex polygons based on known occurrences (see following guide).

- 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 pseudoabsences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3, 327–338.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudoabsence data. *Ecological Applications*, 19, 181-197.
- Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M. B. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373.

Tool: Background Selection: Sample by Buffered MCP, Sample by Distance from Obs. Pts. or Sample by Buffered Local Adaptive Convex-Hull

ARCGIS STEP-BY-STEP GUIDE:

1. Decide if you want to restrict background sampling to: i) a buffered minimum convex polygon based on observation localities, ii) a radial distance from all occurrence points or an intermediate between those two iii) using the tool: *Sample by Buffered Local Adaptive Convex-Hull*.
2. Double-click the corresponding tool 'SDM Tools → 2. MaxEnt Tools → Background Selection via Bias Files → Background Selection: Sample by Buffered MCP' or 'Background Selection: Sample by Distance from Obs. Pts.' or 'Background Selection: Sample by Buffered Local Adaptive Convex-Hull'
3. Continue to tool interface instructions (Buffered MCP below)



Background Selection: Sample by Buffered MCP tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. CSV file of species occurrences (from previous steps)
2. Field corresponding to species identity
3. File field corresponding to longitude
4. File field corresponding to latitude
5. The distance outside of minimum-convex-polygon included in background selection.
6. Select output folder location. This should be a new empty folder. If not empty this can cause the analysis to fail, particularly if temporary files from a previous analysis were not properly removed (e.g. this can happen if another SDMtoolbox analysis is terminated early).
7. Climate data sized to extent of MaxEnt Modeling. Here use the 'Bio_1.asc' layer. Select one of your climate files sized to your modeling extent (e.g. Bio1.asc). This file will be used to match the bias file to proper extent and resolution (no change will be made to this file).



Model calibration and validation

5. Spatial Jackknifing and independent tests of parameters

Why use SDMtoolbox for MaxEnt modeling:

I. Spatial Jackknifing

Spatial jackknifing (or geographically structured k-fold cross-validation) tests evaluation performance of spatially segregated spatially independent localities. SDMtoolbox automatically generates all the GIS files necessary to spatially jackknife your MaxEnt Models. The script splits the landscape into 3-5 regions based on spatial clustering of occurrence points (e.g. if 3: A,B,C). Models are calibrated with k-1 spatial groups and then evaluated with the withheld group. For example if k=3, models would be run with following three subgroups:

1. Model is calibrated with localities and background points from region AB and then evaluated with points from region C
2. Model is calibrated with localities and background points from region AC and then evaluated with points from region B
3. Model is calibrated with localities and background points from region BC and then evaluated with points from region A

II. Independent Tests of Model Feature Classes and Regularization Parameters

Equally important, this tool allows for testing different combinations of five model feature class types (FC) and regularization multiplier(s) (RM) to optimize your MaxEnt model performance. For example, if a RM was input (here 5), this tool kit would run MaxEnt models on the following parameters for each species:

1. RM: 5 & FC: Linear, 2. RM: 5 & FC: Linear and Quadratic, 3. RM: 5 & FC: Hinge, 4. RM: 5 & FC: Linear, Quadratic and Hinge, 5. RM: 5 & FC: Linear, Quadratic, Hinge, Product and Threshold

III. Automatic Model Selection

Finally, the script chooses the best model by evaluating each model's: 1. omission rates (OR)*, 2. AUC**, and 3. model feature class complexity. It does this in order, choosing the model with the lowest omission rates on the test data. If many models have the identical low OR, then it selects the model with the highest AUC. Lastly if several models have the same low OR and high AUC, it will choose the model with simplest feature class parameters in the following order: 1. linear; 2. linear and quadratic; 3. hinge; 4. linear, quadratic, and hinge; and 5. linear, quadratic, hinge, product, and threshold.

Once the best model is selected, SDMtoolbox will run the final model using all the occurrence points. If desired, at this stage models will be projected into other climates, environmental variables will be jackknifed to measure importance, and response curves will be created.



*For each iteration, OR is weighted by the number of points in the evaluation subgroup. This is necessary because spatial groups may not have identical number of points. The weighing gives equal contribution to all points included in model evaluation.

**AUC is calculated from the total study area in the input bias file (if k=3, then all groups: ABC)

For info and justification for each of these methods see:

- Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling, 275:73-77.
- Radosavljevic, A. and R. P. Anderson. 2014. Making better Maxent models of species distributions: complexity, overfitting, and evaluation. Journal of Biogeography. 41:629-643
- Shcheglovitova, M. and R. P. Anderson. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecological Modelling, 269:9-17.

Tool: Run MaxEnt: Spatially Jackknife

ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the 'SDM Tools → 2. MaxEnt Tools → Modeling in MaxEnt → Run MaxEnt: Spatially Jackknife' tool
2. Continue to tool interface instructions (below)

SDMTOOLBOX STEP-BY-STEP GUIDE:

IMPORTANT NOTE 1: none of the input and output file names/file paths can have spaces in them. If there are any spaces, the output batch scripts will likely fail to work properly.

IMPORTANT NOTE 2: Upon first use of this tool, due to its unique syntax, you need to specify the location of the menu file. If you do not do this, the menu presented by ArcGIS will not make complete sense. For a detailed overview of how to do this (it will take 30 seconds to do), go to:
<http://www.sdmtoolbox.org/menu-fix-spatial-jackknife>

1. Folder with the MaxEnt program. Downloadable from:
<http://www.cs.princeton.edu/~schapire/maxent/>
2. CSV file output from spatial rarefied occurrence data (output from step 3D)
3. CSV field corresponding to species identity
4. CSV field corresponding to longitude
5. CSV field corresponding to latitude
6. Folder with clipped ASCII climate data from step 3C.
7. Select environmental layers from this folder that depict categorical data.
8. Select environmental layers to exclude from analyses (not list in output from step 2).
9. Folder containing bias files from step 4.
10. Output folder for GIS, python scripts and MaxEnt batch files. Output will include all GIS files necessary to run models.
11. For spatial jackknifing this format output will be Logistic, regardless of selection
12. For spatial jackknifing this format output will be ASCII, regardless of selection

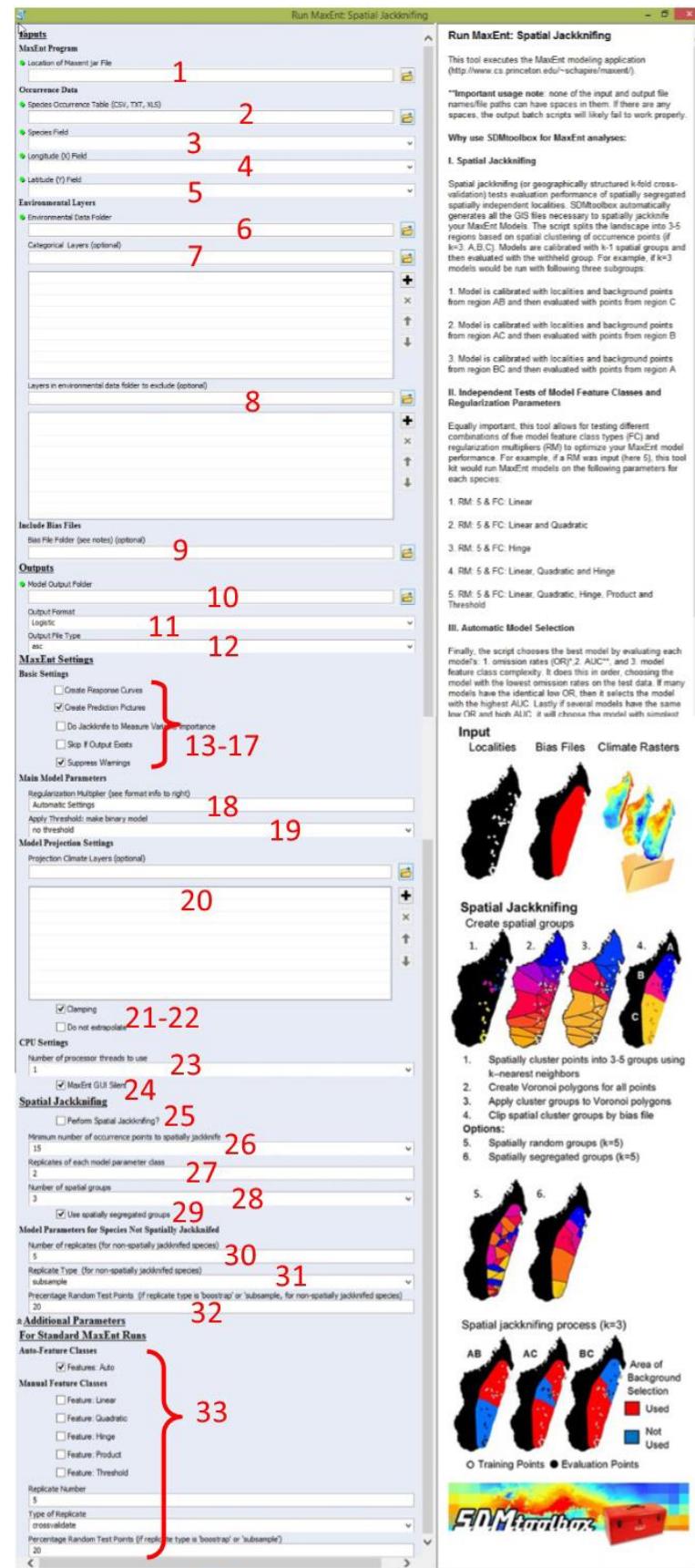
13. Create graphs showing how predicted relative probability of occurrence depends on the value of each environmental layer
14. Create an image of each output model
15. Measure variable importance by jackknifing the variables. Each variable is excluded in turn and a model created with the remaining variables. Then a model is created using each variable in isolation.
16. This will skip the model if an output exists.
17. This will suppress any warnings encountered during modeling. All warnings are always written to the log files.
18. This is the breadth of the model. A higher number gives a more spread-out distribution. Input many values separated by semi-colons.

For example:

0.5; 1; 1.5; 2; 2.5; 3; 4; 5

The default value is 1.

Remember the more values input will produce more SDMs created and require more computation time. For each regularization multiplier (RM), 15 models will be run (5 feature class groups and 3-5 spatial jackknife groups, $5 \times 3 = 15$ to $5 \times 5 = 25$). This number is multiplied for each replicate and each species





modeled. Thus, if you have 2 species, 5 RMs and 2 replicates; this would result in 300 or 500 models run for 3 and 5 spatial jackknife groups, respectively (2 species *2 replicates* 5 RM * 15-25 models per run)

19. Apply a threshold to make binary model. This will generate a binary model in addition to the continuous model. In none is supplied, SDMtoolbox will use '10 percent minimum training presence' to calculate omission rates. If you prefer another threshold, please select it here.
20. Projection Climate Layers. Folders containing environmental data for projecting the MaxEnt models (often are future or past climates). To select multiple folders at once hold to shift. The layers MUST match the input environment layers, e.g. if Bio27 is used to build the model then the projection folder must contain an analog variable and have the identical name. Here resolution and spatial extent do not need to match input environmental layers.
21. Apply clamping when projecting
22. If checked, will not predict areas of climate space outside of limits encountered during training
23. Number of CPUs to use for modeling
24. This will *not* display the MaxEnt GUI when running models- preferred
25. Check this box to perform all the analyses described into the information window. If not checked this will run the modeling as if executed from the MaxEnt GUI (no spatial jackknifing or independent evaluation of RV or feature classes).
26. This is the minimum number of points to execute spatial jackknifing. If *below* this value, the models will be trained and evaluated using either cross-validation, bootstrapping or sub-sampling (as specified below in steps: 30-32). Each non-spatially jackknifed group is optimized with independent tests of different combinations of the five model feature class types (FC) and input regularization multiplier (RM) values.
27. Replicates of each model parameter class in spatial jackknife runs
28. Number of groups to subdivide the landscape into. Higher the number the more models run, but also the more training points included in each model run.
29. If selected:
Groups will be spatially segregated and numbers of occurrences within groups may not be equal. This analysis is more focused on natural spatial groups. This method is best if projecting models into other climates (i.e. current or past) and is particularly useful for training and evaluating model performance in potentially non-analogous climates.
If not selected:
Spatial jackknife groups will be spatially randomly and numbers of occurrences within groups will be equal (+/- 1, due to unequal group sizes for some combinations of occurrences records and group number). This method is best if not projecting models into other climates.
30. Replicates of each model parameter class for species with too few points to spatially jackknife
31. Replicate Type. If replicates are >1, then multiple runs are performed by this type:

Crossvalidate: MaxEnt makes k number of folds of your occurrence data to train and test the data. Here you are not able to tell MaxEnt how many replicates you would like to run or the percentage of occurrence data you would like withheld for model validation (test occurrences). Optimal if you have a large number of species occurrences.

Bootstrap: Replicates samples sets are chosen by sampling with replacement

Subsample: Replicate sample sets are chosen by removing the random test percentage (input in the following window) without replacement, the variables not included are then used for model evaluation

32. If replicate type is 'bootstrap' or 'subsample', input the percentage of points used for subsampling.
33. Additional parameters not used in spatial jackknifing.
34. Run the tool.
35. After files are created, to execute models go to output folder. Click the batch file "Step1_Optimize_MaxEnt_Model_Parameters.bat" this will run all models and summary stats. Know this may take several hours, or even days (if you have many species and RV), to finish.
36. Once all the models from Step 1 are run, the "Step2_Run_Optimized_MaxEnt_Models.bat" will be populated with the best model parameters. Run this file to get final models.
37. To see model ranks for each species, open the corresponding folder and open the "species_name_SUMSTATS_RANKED_MODELS.csv." Here the best model is the first row. Feature number corresponds to feature class group with: 1=linear; 2=linear & quadratic; 3=hinge; 4=linear, quadratic, and hinge; and 5=linear, quadratic, hinge, product, and threshold.

Final remarks

Great job—only a small proportion of all published papers using SDMs/ENMs address the following best practices of modeling:

1. Used species-specific regional background sampling
2. Spatially rarefied occurrence data
3. Spatially jackknifed SDMs to calibrate model parameters
4. Independently evaluated feature class parameters and regularization multiplier(s)
5. Reduced correlation of input climate variables for interpreting influence on model (optional here)

Now consider the following SDMtoolbox tools for further analyses:

If projecting models into future or past climates:

1. Limit Dispersal in Future SDMs
2. Overprediction Correction: Clip Models by Buffered MCPs
3. Distribution Changes Between Binary SDMs:
 - a. Centroid Changes (Lines)
 - b. Distribution Changes Between Binary SDMs

To create better models:

4. Correcting Latitudinal Background Selection Biases



5. Gaussian kernel density of sampling localities

To assess landscape connectivity:

6. Among all sites or between shared haplotypes
7. Create friction layers

To measure spatial biodiversity patterns of many SDMs:

8. Calculate species richness and endemism (weighted endemism & corrected weighted endemism)

Now go publish your results!

Summary of some basic considerations when generating SDMs

Table modified from: Alvarado-Serrano, D. F. and Knowles, L. L. (2014), Ecological niche models in phylogeographic studies: applications, advances and precautions. Molecular Ecology Resources, 14: 233–248. Please see paper for more details.

Assumptions that may affect SDMs	Specific considerations
Data compilation occurrence records: Are species presences (and absence) records representative of the actual distribution?	<ul style="list-style-type: none"> ▪ effects of species' natural history ▪ geographic/environmental bias ▪ intraspecific variability ▪ positional uncertainty ▪ sample size ▪ sampling bias (e.g. towards more accessible areas) ▪ taxonomic accuracy (e.g. subspecies or races) ▪ temporal coverage in relation to environmental data
Data compilation Environmental variables: Do environmental variables accurately capture the association between species subsistence and the environment at the relevant scale?	<ul style="list-style-type: none"> ▪ data quality and biases ▪ effect on species distribution (direct vs. indirect) ▪ resolution in space and time ▪ spatial autocorrelation ▪ spatial extent ▪ temporal coverage and stability ▪ type (categorical vs. continuous)
Model generation and calibration Is the modelling algorithm appropriate given the data available and research question?	<ul style="list-style-type: none"> ▪ algorithm assumptions ▪ algorithm performance ▪ under different scenarios ▪ input data type (e.g. presences only vs. presence/absences) ▪ output generated (e.g. presence/absence vs. continuous prediction) ▪ sensitivity to model parameters
Model generation and calibration Is the model appropriately calibrated for the data available and research question?	<ul style="list-style-type: none"> ▪ model complexity ▪ model selection procedure ▪ setting of model parameters ▪ variable selection strategy
Model validation Is validation performed on truly independent data and under appropriate settings?	<ul style="list-style-type: none"> ▪ assumptions/limitations of accuracy measurement ▪ importance of use of multiple metrics ▪ sensitivity to model parameters ▪ threshold transformation of continuous predictions
Model projection Is the species environment relationship likely to be maintained in space and/or time?	<ul style="list-style-type: none"> ▪ availability of validation data in projected regions ▪ likelihood of niche shifts ▪ model uncertainty ▪ model transferability ▪ risks of interpolation and extrapolation