

SDMtoolbox User Guide

By Jason L. Brown, The City College of New York

Last updated: 10/14/2014

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*



Table of Contents

Getting Started	5
<i>Downloading</i>	5
<i>Installation</i>	5
<i>Uninstalling or updating</i>	6
<i>A First Run</i>	7
Analyses Guide	10
Chapter 1. Biodiversity Measurements	12
Estimate Richness and Endemicity (WE and CWE)	13
<i>ARCgis STEP-BY-STEP GUIDE:</i>	13
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	14
Results	15
Chapter 2. Landscape Connectivity	18
Landscape Connectivity	19
<i>ARCgis STEP-BY-STEP GUIDE:</i>	19
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	21
Results	23
Chapter 3. Species Distribution Modeling Tools	24
<i>Correcting Latitudinal Background Selection Bias</i>	24
Solution 1	25
Tool: 1. Bias File for Coordinate Data (BFCD) in MaxEnt	25
<i>ARCgis STEP-BY-STEP GUIDE:</i>	25
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	26
Tool: 2a. Background Selection: Sample by Buffered MCP	27
<i>ARCgis STEP-BY-STEP GUIDE:</i>	27
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	27
Tool: 3. Clip BFCD by BS Bias File	28
<i>ARCgis STEP-BY-STEP GUIDE:</i>	28
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	28
Solution 2	29
Tool: 1. CSV to EAP. MaxEnt format output (runs both 1a and 1b)	29
<i>ARCgis STEP-BY-STEP GUIDE:</i>	29



SDMTOOLBOX STEP-BY-STEP GUIDE:	29
Tool: 2. Project Climate Data (Raster) to Equal-Area Projection (Folder)	31
ARC GIS STEP-BY-STEP GUIDE:.....	31
SDMTOOLBOX STEP-BY-STEP GUIDE:	31
<i>Distribution Changes Between Binary SDMs</i>.....	32
Tool: Centroid Changes (Lines)	32
ARC GIS STEP-BY-STEP GUIDE:.....	32
SDMTOOLBOX STEP-BY-STEP GUIDE:	33
Tool: Distribution Changes Between Binary SDMs	34
ARC GIS STEP-BY-STEP GUIDE:.....	34
SDMTOOLBOX STEP-BY-STEP GUIDE:	34
Results.....	35
<i>Overprediction Correction: Clip Models by Buffered Minimum Convex Polygons</i>.....	36
Tool: Binary Models (folder)	36
ARC GIS STEP-BY-STEP GUIDE:.....	36
SDMTOOLBOX STEP-BY-STEP GUIDE:	37
Results.....	38
Tool: Continuous Models (folder).....	39
ARC GIS STEP-BY-STEP GUIDE:.....	39
SDMTOOLBOX STEP-BY-STEP GUIDE:	39
Results.....	40
<i>Background Selection via Bias Files</i>	41
Tool: Background Selection: Sample by Buffered MCP	42
ARC GIS STEP-BY-STEP GUIDE:.....	42
SDMTOOLBOX STEP-BY-STEP GUIDE:	43
Tool: Background Selection: Sample by Distance from Obs. Pts.	43
ARC GIS STEP-BY-STEP GUIDE:.....	43
SDMTOOLBOX STEP-BY-STEP GUIDE:	44
<i>Create Friction Layer: Invert SDM</i>.....	44
ARC GIS STEP-BY-STEP GUIDE:.....	45
SDMTOOLBOX STEP-BY-STEP GUIDE:	45
Results.....	46



Spatially Rarefy Occurrence Data	45
<i>ARCgis STEP-BY-STEP GUIDE:</i>	46
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	46
Chapter 4. Basic Tools: Raster, CSV & Shapefile tools	50
Tool: CSV to shapefile	50
<i>ARCgis STEP-BY-STEP GUIDE:</i>	50
<i>SDMTOOLBOX STEP-BY-STEP GUIDE:</i>	50
Chapter 5. Running a SDM in MaxEnt: from start to finish	51
Data Preparation	51
1. Preparing Worldclim Data	51
2. Testing Autocorrelations of Environmental Data.....	54
3. Preparing Occurrence Data.....	55
4. Creation of Bias Files.....	62
Model Creation, Calibration and Evaluation	64
5. Spatial Jackknifing	64

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.

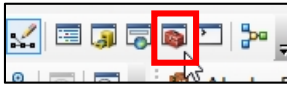
Installation

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\ArcGIS\Toolboxes)
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

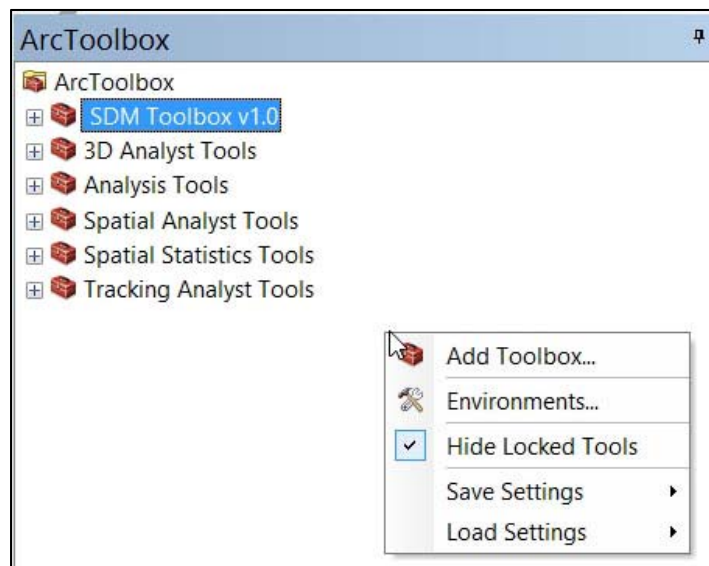


5. Then right-click ArcToolbox Window (see image below) → select 'Add Toolbox.' Then go to location of downloaded toolbox and select 'SDMToolbox v1.0.tbx'

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!



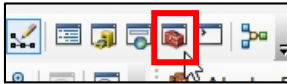
How to update or uninstall SDMtoolbox

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



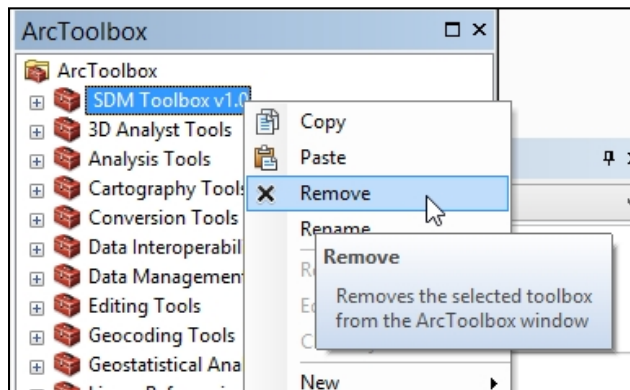
Above: ArcCatalog Icon

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



4. Then right-click SDMtoollbox 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.

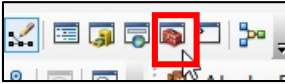
7. To finish installation simply close the ArcCatalog program.

8. Now start ArcMap and begin using the SDMtoobox!

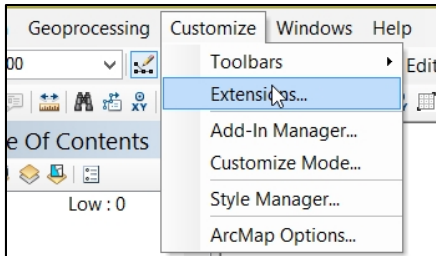
A First Run

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:

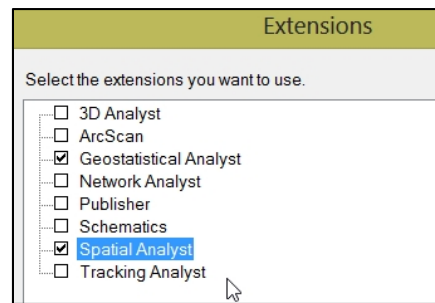
1. **Open ArcMap10.1 and activate the ArcToolbox window (if not already visible). Bottom image on this page is the ArcToolbox window. If not visible, click toolbox icon (see image below).**



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:

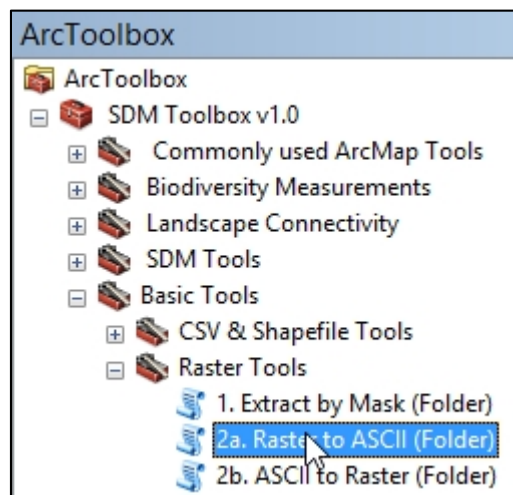
<http://support.esri.com/en/downloads/patches-servicepacks/list?productid=160&productVersions=10.1&categoryTypes=5&categoryTypes=8>

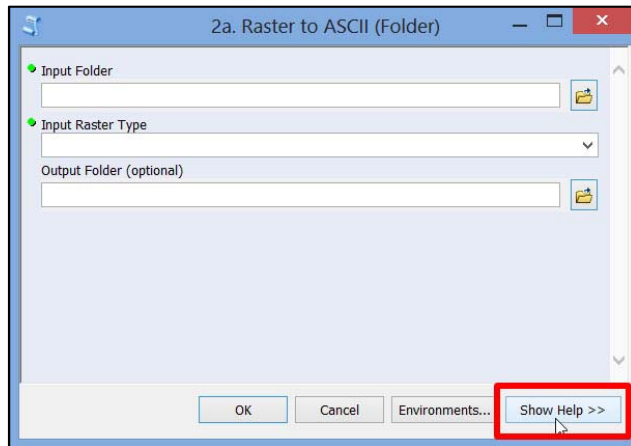
Or: <http://support.esri.com/en/downloads/patches-servicepacks>

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>)

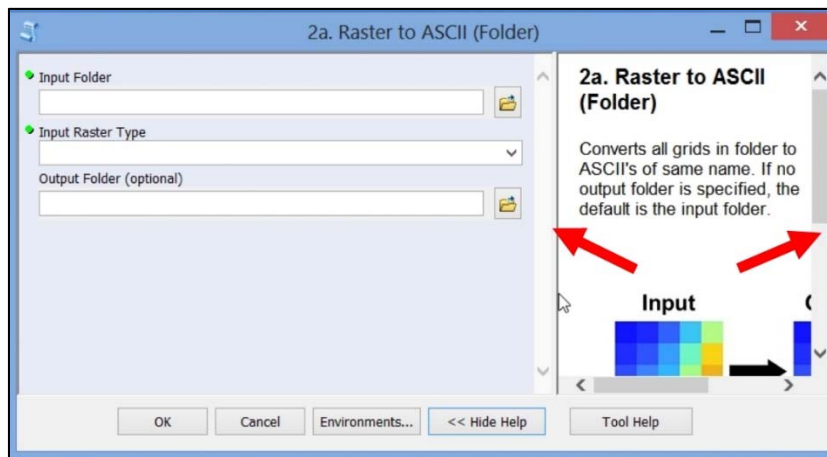
4. **Enable 'Tool Help' and resize windows.**

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.





A note on map projections

Most user reported errors have to do with map projections. As follows is a checklist of things you need to check if you have a script fail. In most cases your data should be projected, with the projection specified in the GIS file.

1. **If you change or define the map projection** after opening the map document. You **NEED** to save all projected files, then close the map document and **open a new map session**. The first file open in your new session needs to be one of the newly projected files—this defines the map datum. This is particularly important if you change the spatial units as the result of projection, e.g. decimal degrees to meters. If you do not do this, often SDMtoolbox scripts will fail or result in inaccurate results.
2. **If the tool requires a CSV of coordinates AND** your analyses are going to be **in another projection**, remember to convert input coordinates (or XYs) to the new, and same, datum prior to running tools or for use in species distribution modeling.

For most, the quickest way to do this for CSV files (and other tables) is to:

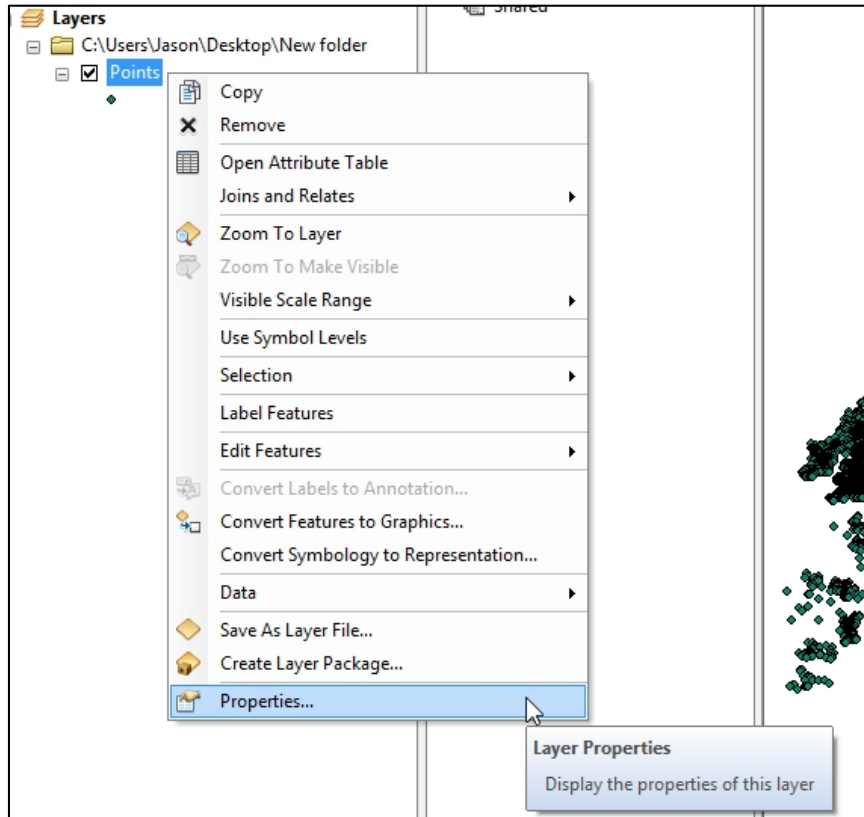
- 1) Import the CSV to a shape file (*Table & Shapefile Tools* → *CSV to shapefile* tool)
- 2) Define input projection of shapefile and then project to the new projection (a. *ArcGIS define projection* and b. *project* tools)
- 3) Add the new coordinates/XYs to Shapefile using the *ArcGIS Add XYs* tool
- 4) Export shapefile table as CSV file (SDMtoolbox the *Table & Shapefile Tools* → *Shapefile to CSV* tool)

If using an equal-areas projection and input data are WGS-1984 coordinates, try the SDMtoolbox tool that will do all the above steps:

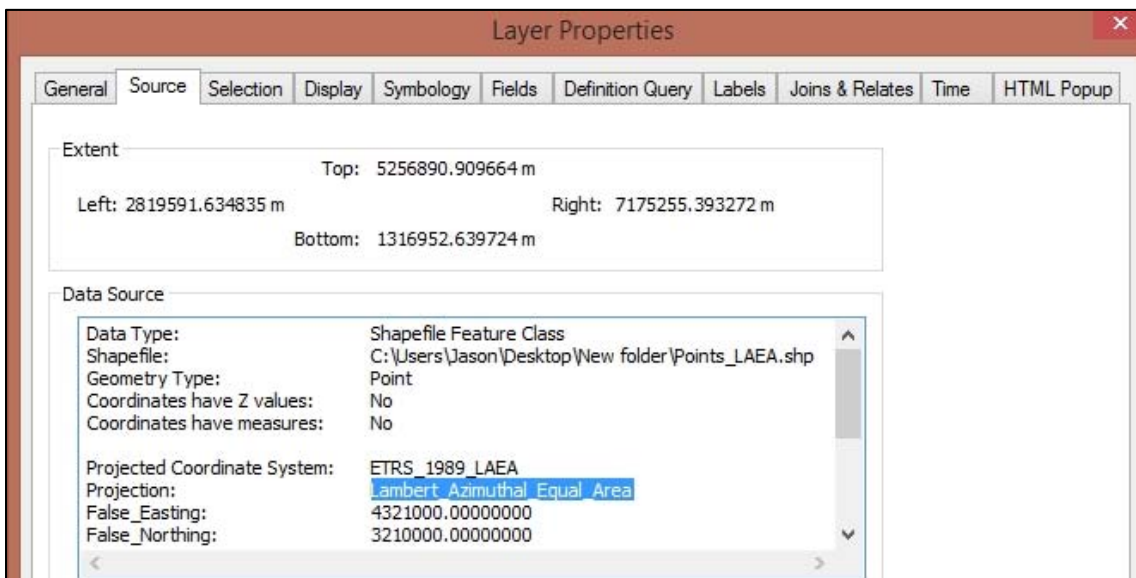
SDM tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection Biases → Solution 2: Project Input Data to Equal-Area Projection (EAP) → 1. CSV to EAP. MaxEnt format output

3. Check that **all your input data are projected and in the same projection**. This is the most common error. Often reference ASCII files are not projected at all.
4. Last, **visualize all GIS files** to make sure they are in fact properly projected (all maps should overlap etc.). Sometimes files say they are projected to the same projection, however in reality something was incorrect.

To check the projection of each GIS file



Import GIS files, then right click each file and select 'Properties'



Select the 'Source' tab and page down until you see the 'Projection'. Here the projection is Lamberts Azimuthal Equal Area



To accompany this guide I have provided example data available at www.sdmtoolbox.org. Download this, the latest version of the SDMtoolbox, and this guide before beginning.

This guide does not cover all the tools in the SDMtoolbox. However, as an overview, the guide covers tools from all the major groups. For many tools lacking a guide I have included example data to execute each analysis. The example data should be contained in a folder respective to the hierarchy of the toolbox. For example, for the tool 'SDM toolbox: Biodiversity Measurements → Input: Point Data → Calculate Richness and Endemicity (WE and CWE)', the example data are in the folder '...example_data\biodiversity_measurements\biodiversity_points'

Lastly, each tool is annotated and instructions should be contained within each tool's help file from within ArcGIS. Follows are the major groups of the toolbox. The guide treats each one as a separate chapter.

Toolbox Guide

Chapter Structure

- 1. Biodiversity Measurements**
- 2. Landscape Connectivity**
- 3. SDM Tools**
- 4. Basic Tools (Both CSV & Shapefile tools and Raster tools)**
- 5. Running a SDM in MaxEnt: from start to finish**



Chapter 1. Biodiversity Measurements

Tool Overview

These tools estimate three common biodiversity metrics: species richness, weighted endemism and corrected weighted endemism. There are two sets of analyses here: analyses that utilize point occurrence data and analyses that use binary SDMs.

As follows are the three diversity metrics:

1. **Species Richness (SR)** is sum of unique species per cell.

$$SR = K \text{ (the total number of species in a grid cell)}$$

2. **Weighted Endemism (WE)**, which is the sum of the reciprocal of the total number of cells each species in a grid cell is found in. A WE emphasizes areas that have a high proportion of animals with restricted ranges.

$$WE = \sum 1/C \text{ (C is the number of grid cells each endemic occurs in)}$$

3. **Corrected Weighted Endemism (CWE)**. The corrected weighted endemism is simply the weighted endemism divided by the total number of species in a cell (Crisp 2001). A CWE emphasizes areas that have a high proportion of animals with restricted ranges, but are not necessarily areas that are species rich.

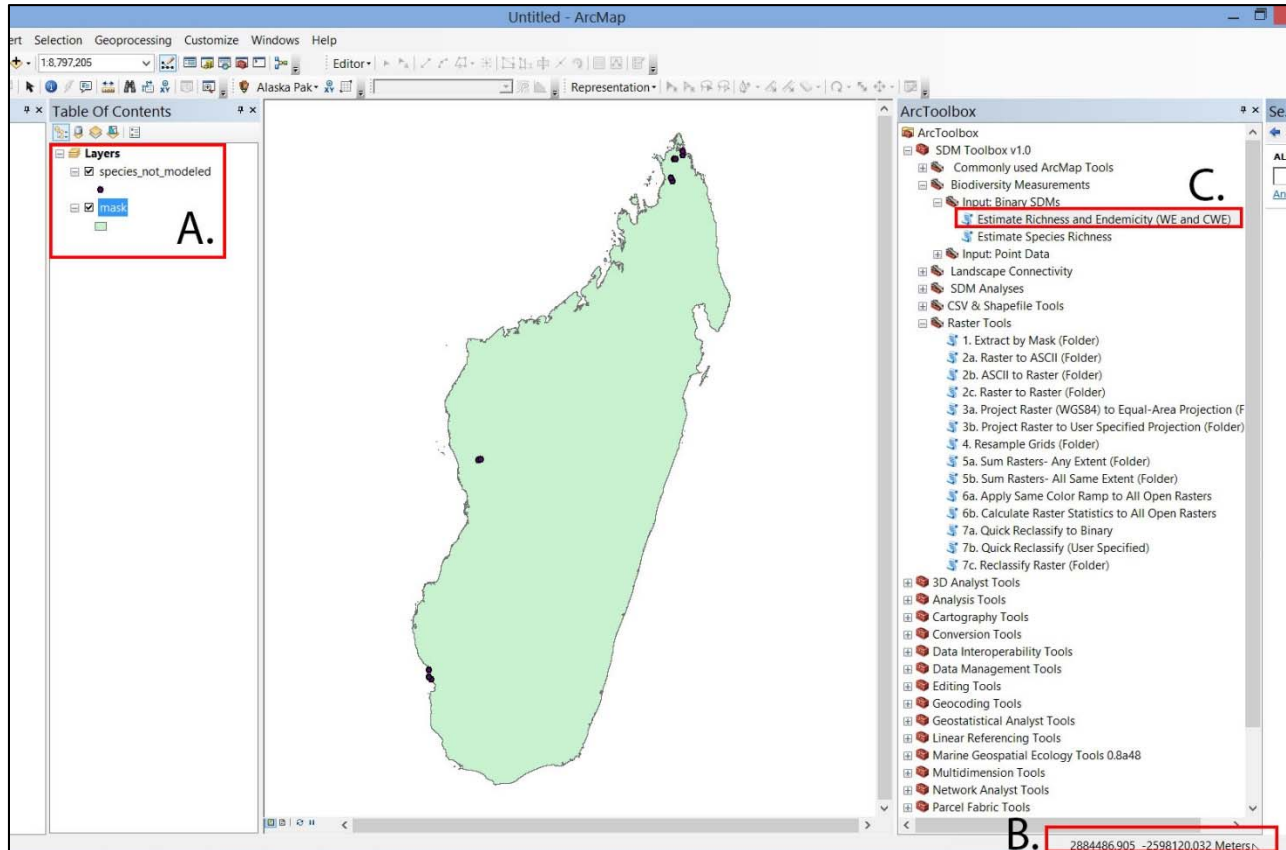
$$CWE = WE/K \text{ (K is the total number of species in a grid cell)}$$

Crisp, M. D., Laffan, S., Linder, H. P., and Monro, A. 2001. Endemism in the Australian flora. *Journal of Biogeography* 28:183-198.

Estimate Richness and Endemicity (WE and CWE)

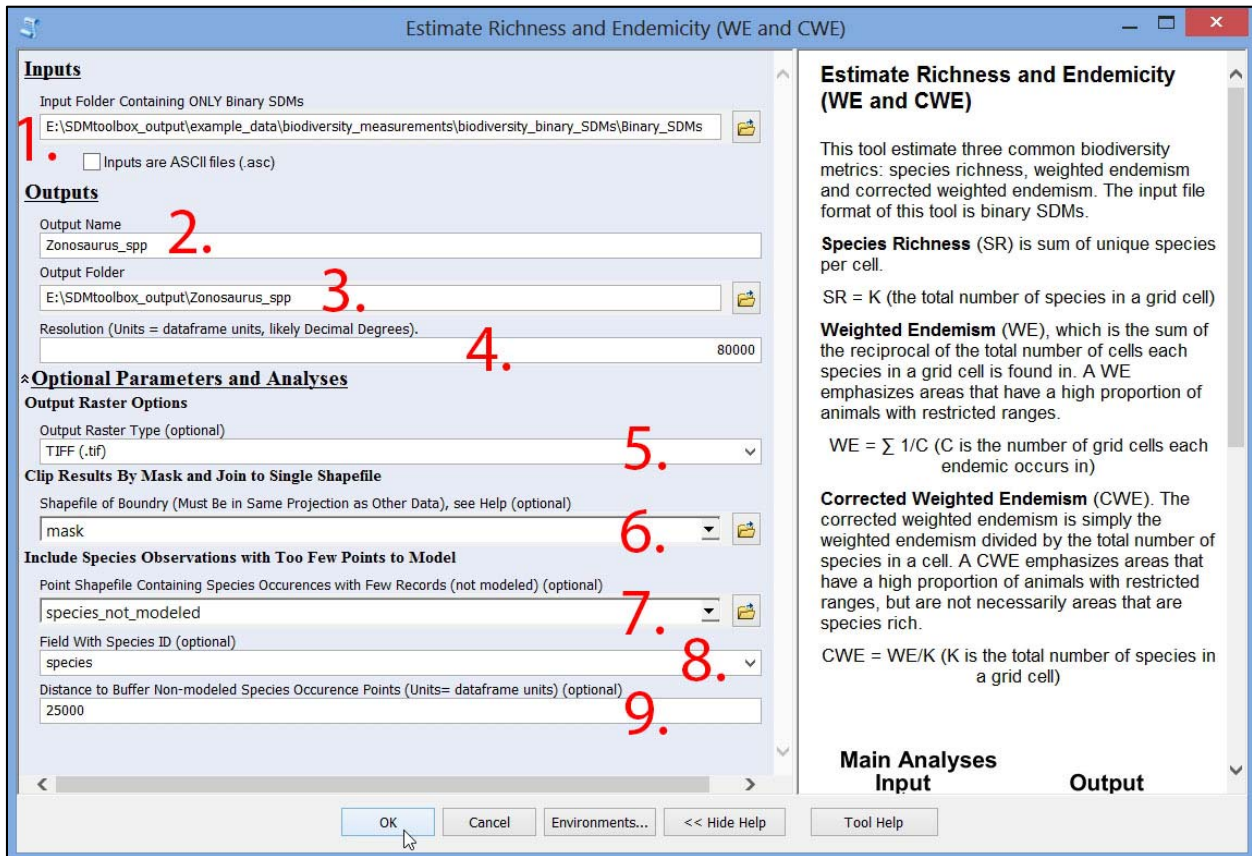
ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'species_not_modeled.shp' and 'mask.shp' (Location: ...\\example_data\biodiversity_measurements\biodiversity_binary_SDMs)



A. Imported data. **B.** Note that these data units are meters **C.** Tool to be used in this guide

3. Next double-click the 'Biodiversity Measurements → Input Binary SDMs → Output: Grid → Estimate Richness and Edemicity (WE and CWE)' tool
4. Continue to tool interface instructions (following page)



Estimate Richness and Endemicity (WE and CWE) tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Select example data folder:
...\\example_data\\biodiversity_measurements\\biodiversity_binary_SDMs\\Binary_SDMs

Leave the 'Inputs are ASCII files (.asc)' unchecked. Here the inputs are TIFF raster files.

2. Input desired output name. Note when this category is highlighted the Help box displays the text in the box to the right. I input 'Uruloke_spp', this means that the output files will be named:
'Uruloke_spp_Estimated_Spp_Rich.tif',
'Uruloke_spp_Est_Spp_Rich_Low_Res.tif',
'Uruloke_spp_WE.tif' and 'Uruloke_spp_CWE.tif'.

Output Name

Name of output file. Note:
 "_Estimated_Spp_Rich",
 "_Est_Spp_Rich_Low_Res", "_WE", "_CWE" will
 be appended to the end of output name as
 relevant.

3. 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).

4. Here I used a resolution of 80,000m (or 80km). The data used here are in meters. In the future, however, your data might be in feet or degrees. The distance value input should be large enough to capture landscape processes, but not too large where regional differences are lost. I recommend starting with a value equivalent to 50 or 100km. Note 1: your actual values will be in the map units (likely meters, feet or decimal degrees). Note 2: 100km = ~0.8983 DD at Equator.
5. I prefer TIFF files as a raster output format because they allow for longer file names (vs. ESRI grid files that are limited to 13 characters) and don't have too many raw parts to each file. There is, however, a slight reduction in performance (vs. ESRI grid files), thus, if processing thousands of rasters this should be taken into consideration.
6. A polygon mask of a country outline. This will clip the edges of pixels by the boundary of this mask. This produces a much more visually pleasing output. All biodiversity metrics will be appended to the shapefile table and can be visualized in the file symbology.

NOTES:

-This **MUST** be in the same projection as other input data (e.g. all projections must be WGS 1984).

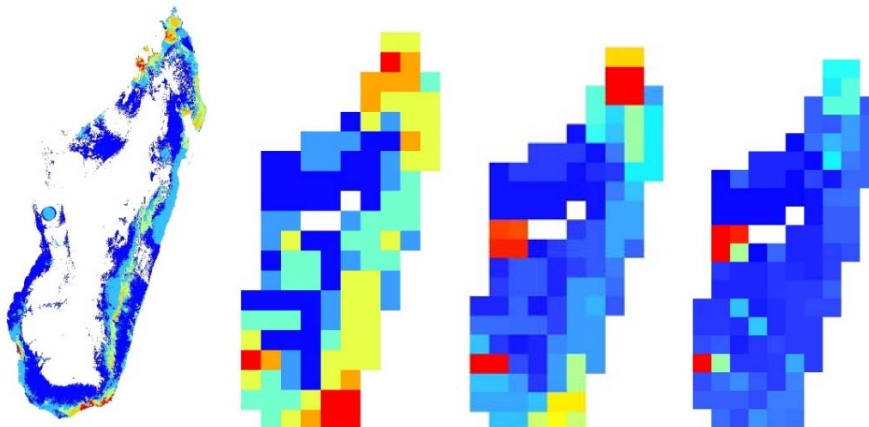
-A good source for clipping mask (country boundaries etc.) is:

<http://www.diva-gis.org/Data>

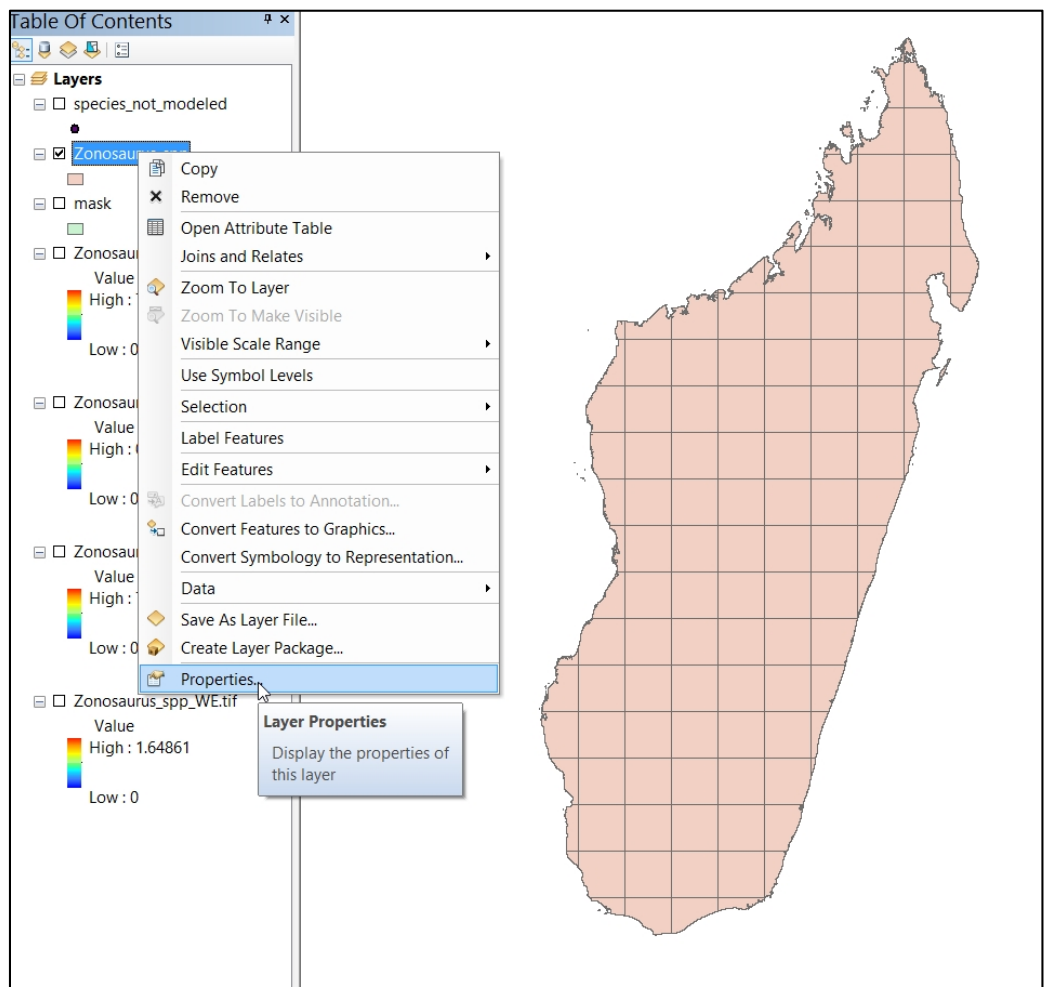
-Again, be sure to project the file to match your input rasters

7. Often due to lacking occurrence data, not all species can be modeled. This feature will include species occurrences that were not modeled to be included in the biodiversity estimates.
8. Select table field corresponding to species ID
9. Buffer distance. The input points will be buffered to this distance. Here I chose 25,000m (25km)--- this means a circle with a 25km radius will be created around each point.

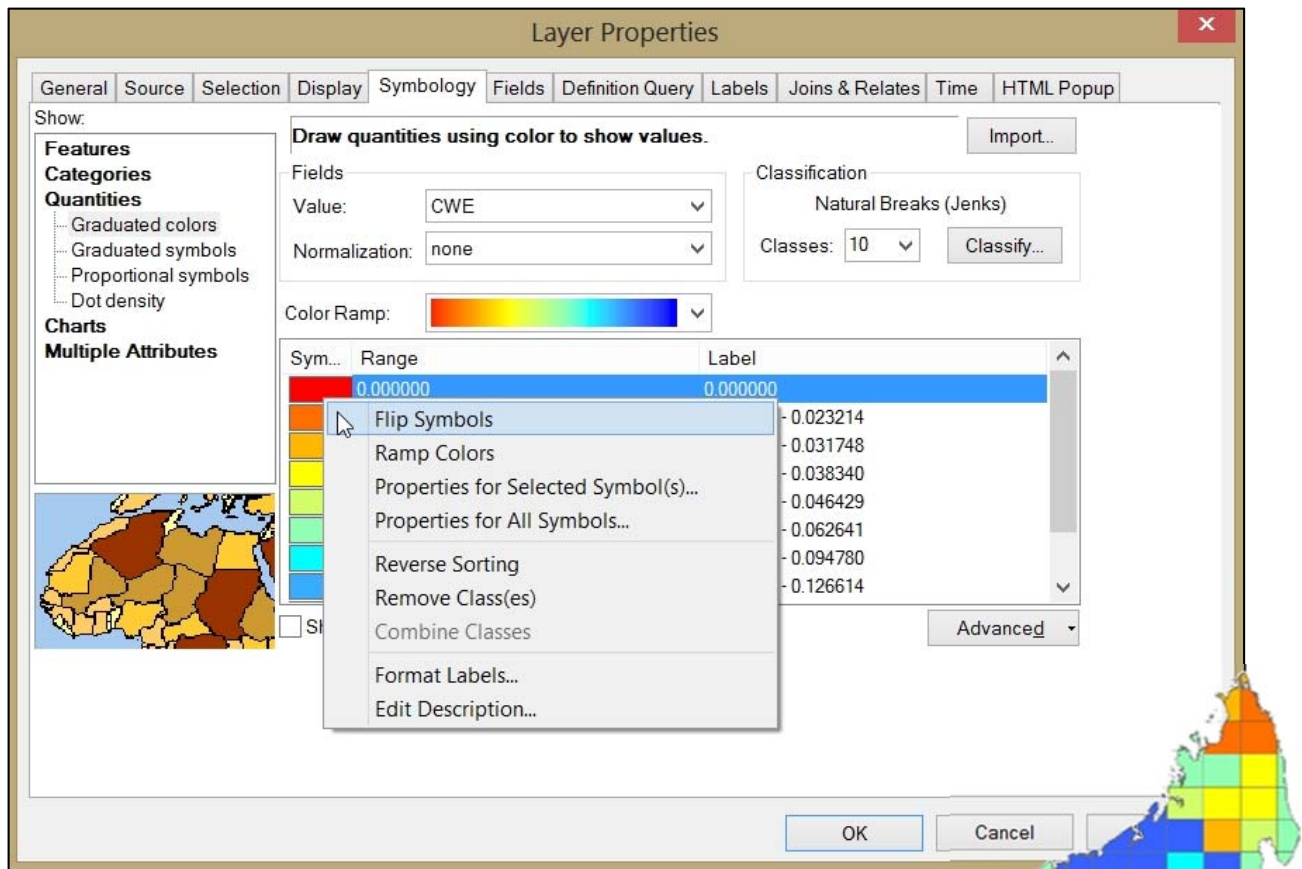
Results



Outputs from analyses. Left to Right: Estimated Spp. Richness, Estimated Spp. Richness (low resolution), WE and CWE

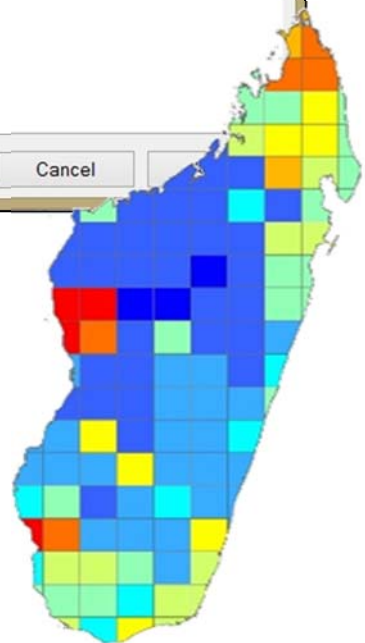


To change the appearance of the output shapefile, right click the file and select 'Properties'



Then select the 'Symbology' tab. In the left column select 'Quantities → Graduated colors', for 'Fields Value' select any of the three biodiversity fields. Here I chose to use 10 classes defined by 'Natural Breaks (Jenks).' To keep consistent with the other color schemes, I used the RGB color ramp and inverted it so that the highest values are red. To do this, right click one of the colors (see above) and select 'Flip Symbols.' Small image to right is the output from these settings.

Extra bit: For a little extra pizzazz, place a digital elevation model below the biodiversity layer (use black to white colorramp) and make the biodiversity layer 10-20% transparent. This gives the map a little texture corresponding to topography.



Chapter 2. Landscape Connectivity

Tool Overview

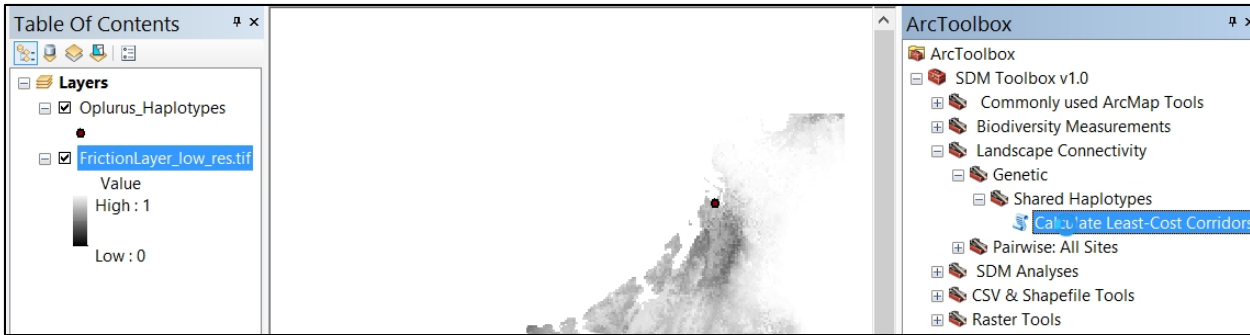
This tool creates a raster of the sum of least-cost corridors and a polyline shapefile of least-cost paths between populations that share haplotypes. Often a single LCP between sites oversimplifies landscape processes. By using categories of cost paths that include paths with slightly more costly path lengths (relative to the LCP), you can better depict habitat heterogeneity and its varying role in dispersal. For each comparison you can classify the lowest cost paths into three categories. Lastly, a density analysis will produce a raster depicting the frequency that LCPs traverse the same path.

For more information, see: Chan LM, Brown JL, Yoder AD (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Mol Phylogenet Evol* 59(2):523-37. doi: 10.1016/j.ympev.2011.01.020.

Landscape Connectivity

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'species_not_modeled.shp' and 'mask.shp' (Location: ...\\example_data\\biodiversity_measurements\\biodiversity_binary_SDMs)
3. Next double-click the 'Landscape Connectivity → Genetic → Shared Haplotypes → Calculated Least-Cost Corridors' tool
4. Continue to the tool interface instructions (following page)



Imported data

Calculate Least-Cost Corridors and Paths
— □ ×

Inputs

Input Points
Oplurus_Haplotypes **1.**

Field with haplotype names
HAPLOTYPE **2.**

Field with unique site ID
SITE_NAME **3.**

Input Friction Layer (results will be sized to this layer)
FrictionLayer_low_res.tif **4.**

Outputs

Output File Name
Oplurus_cuveri **5.**

Output Folder
E:\SDMtoolbox_output\LCCsOutput **6.**

Output Raster Type
Erdas Imagine (.img) **7.**

Create Least-Cost-Path Lines
 Yes **8.**

Optional Parameters

Method of Selecting Least-Cost-Corridors
Percentage of LCP value (recommended) **9.**

LCC Class Percentages (input values are integers, e.g. 1% = 1)

High Cutoff (values between this and the Mid Cutoff = The High Class) **10.**
5

Mid Cutoff (values between this and the Low Cutoff = The Mid Class) **11.**
2

Low Cutoff (values below this = The Low Class) **12.**
1

LCC Class Weights

High Class **13.**
1

Mid Class **14.**
2

Low Class **15.**
5


Calculate Least-Cost Corridors and Paths

This tool creates a raster of least-cost corridors and a polyline shapefile of least-cost paths between populations that share haplotypes. Often a single LCP between sites oversimplifies landscape processes. By using categories of cost paths that include paths with slightly more costly path lengths (relative to the LCP), you can better depict habitat heterogeneity and its varying roles in dispersal. For each comparison you can classify the lowest cost paths into three categories. Lastly, a density analysis will produce a raster depicting the frequency that LCPs traverse the same path.

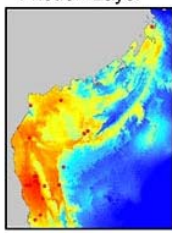
For more information, see: Chan LM, Brown JL, Yoder AD (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Mol Phylogenet Evol* 59(2):523-37. doi: 10.1016/j.ympev.2011.01.020.

Input

Sampled Localities




Friction Layer

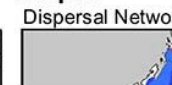


Output

Haplotype Network



Dispersal Network



OK
Cancel
Environments...
<< Hide Help
Tool Help

Calculate Least-Cost Corridors and Paths tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input point shapefile containing sites and haplotypes to be included in least-cost paths and corridors. This shapefile must have two columns: one depicting site names and another depicting haplotype ID.

LATITUDE	LONGITUDE	SITE_NAME	HAPLOTYPE
-16.38	45.345	1432680	C
-14.31	47.915	1569175	A
-16.525	44.49	1660120	A
-16.525	44.49	1660120	C
-16.525	44.49	1660120	F
-19.75	44.6167	APR1031	B
-18.7967	44.8817	HER2558	B
-18.6533	44.6583	HER2725	D
-18.6533	44.6583	HER2725	E

Required geospatial and haplotype data for analysis

Tip. If you need to convert a table to a shapefile for use here. You need two extra columns (latitude and longitude, see table above). Then use the tool 'CSV & Shapefile Tools → 1. CSV to Shapefile'
-Avoid non-alphanumeric characters in both names (e.g., avoid: * : \ / < > | " ? [] ; = + & £ \$, etc.).

2. Table field (part of shapefile) depicting haplotype ID
3. Table field (part of shapefile) depicting site ID. Note this cannot be only a number. If a number recode as alphanumeric ID (e.g. change '9' to '9a')
4. A friction Layer is a raster that depicts the ease of dispersal from each locality through the landscape. In this analysis, the friction layer depicts the output extent and spatial resolution of analysis. If you want a larger extent or a lower spatial resolution, increase both in your friction layer. However, the larger both are—the longer the analysis will take.

Output File Name
Name of output files. Note: "_Dispersal_Network", "_LCPs.shp", "_LCPs_Line_Density" will be appended to the end of in output name as relevant.

Creating a friction layer. One prevalent way to create a friction layer that doesn't suffer from applying weights to habitat types (often associated with contributing to biases in results) is the use of SDMs. An SDMtoolbox tool (location below) will invert a SDM for use as a friction surface. Using this method, areas of high suitability will be converted to areas of low dispersal cost.

Tool Path: SDM Tools → 1. Universal Tools → Create Friction Layer → Invert SDM

5. Input desired output name. Note when this category is highlighted the Help Box displays the text in the box to the right. I input 'Oplurus_cuvieri', this means that the output files will be named: 'Oplurus_cuvieri_Dispersal_Network', 'Oplurus_cuvieri_LCPs.shp', 'Oplurus_cuvieri_LCPs_Line_Density'
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. Output file type. Here I selected the 'Erdas Imagine (.img)' raster format.

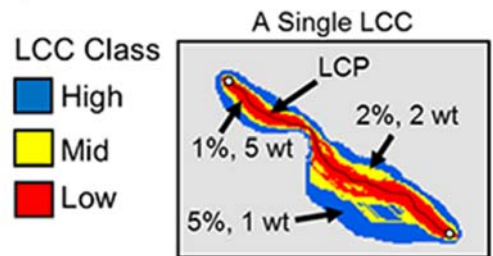
Tip. I prefer TIFF files as output format because they allow for longer file names (vs. ESRI grid files that are limited to 13 characters) and don't have too many raw parts to each file. There is, however, a slight reduction in performance (vs. ESRI grid files), thus, if processing thousands of rasters this should be taken into consideration.

8. Create least-cost path lines. Least-cost path analysis (LCP) allows researchers to find the 'cheapest' way to connect two locations within a cost surface (*i.e.* a friction layer).
9. There are two methods for calculating the least-cost corridors (LCCs):

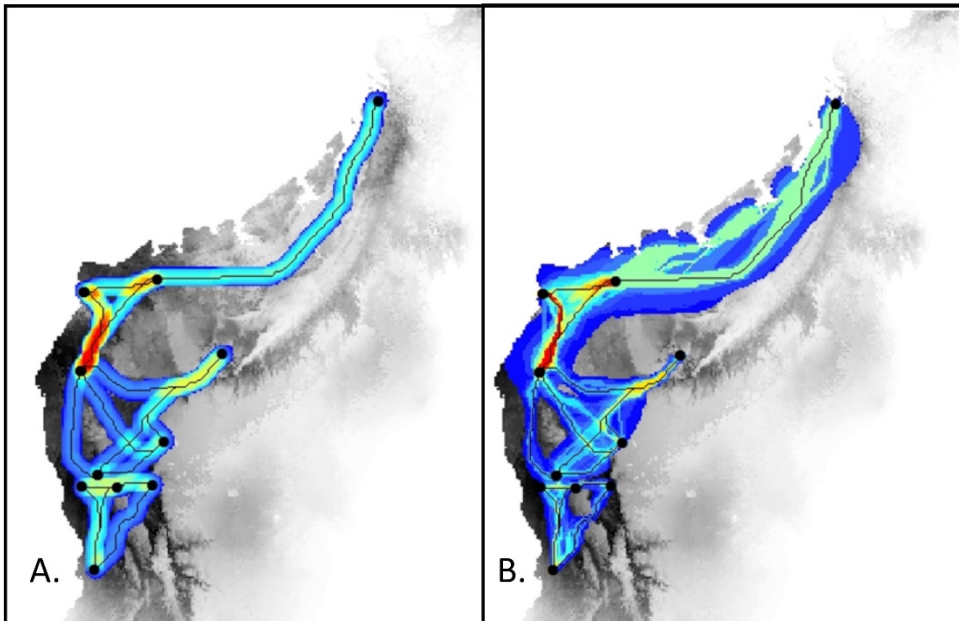
A. Percentage of LCP value is based on the LCP between each site. For example, if the LCP was 5.0 and a 1% LCC class cutoff was selected, LCPs with values between 5.0-5.05 would be included in that class. This formula focuses only on the LCP value and is not affected by larger cost-path values.

B. Percentage of Cost Path Values is based on the range of cost path values across the entire landscape for each comparison. For example, if the least and highest cost paths were 5.0 and 505, respectively, and a 1% LCC class cutoff was selected, LCPs with values between 5.0-10.0 would be included in that class. This formula focuses on the total range of cost path values and is greatly affected by the cost-path values across the landscape.

10. The *Calculate Least-Cost Corridor* tool calculates LCCs in three classes: High, Mid and Low. **High Cutoff.** Cost paths with values between this and the Mid Cutoff will be included in the High Class. The input values are integers of the percentage (e.g. 1% = 1). For overview see figure below.



11. **Mid Cutoff.** Cost paths with values between this and the Low Cutoff will be included in the Mid Class.
12. **Low Cutoff.** Cost paths with values between this and the LCP will be included in the Low Class.
13. LCC weight values. Because all LCCs are summed to create a dispersal network, weighted values are placed on each category of LCCs. The input in this box is the weight applied to the **High LCC class.**
14. Weight applied to the **Mid LCC class.**
15. Weight applied to the **Low LCC class.**



Results

Dispersal Networks. A. Least-cost paths and LCP line densities. Warmer colors depict LCP lines traversed more frequently. B. Haplotype Dispersal Networks. Warmer colors depict cost-paths traversed more frequently and represent likely connections in habitat (due to common ancestry among shared haplotypes).

Chapter 3. Species Distribution Modeling Analyses

- *Correcting Latitudinal Background Selection Bias*
 - Solution 1: Bias Files
 - Solution 2: Project Input Data to Equal Areas Projection
- *Distribution Changes Between Binary SDMs*
 - Centroid Changes (Lines)
 - Distribution Changes Between Binary SDMs
- *Overprediction Correction: Clip Models by Buffered Minimum Convex Polygons*
 - Binary Models (folder)
 - Continuous Models (folder)
- *Background Selection via Bias Files*
 - Sample by Buffered MCPs
 - Sample by Distance from Observed Localities
- *Create Friction Layer: Invert SDM*

Correcting Latitudinal Background Selection Bias

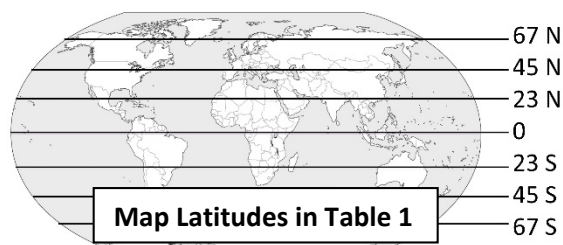
TOOL OVERVIEW

If you are using data that are in a geographic coordinate system (such as, degrees minutes seconds or decimal degrees) for MaxEnt analyses (and most other background and pseudoabsence based SDM methods)--- then **you** are biasing your selection of background points (or pseudoabsence points) and unique observed localities toward the poles. The level of bias depends on the breadth of latitudes your analyses cover. The reason for this is due to the area occupied by these units decreases latitudinally (as values increase, see Table 1), with areas largest at equator and smallest at poles. This inequality results from convergence of the meridians (lines of longitude) towards the poles.

Table 1. Area Covered By a Square Decimal Degree at Different Latitudes

Degrees ²	N/S or E/W at equator	E/W at 23N/S	E/W at 45N/S	E/W at 67N/S
1.0	12392 km ²	11407 km ²	8762 km ²	4842 km ²
% Area of Equator	100%	92.0%	70.7%	39.0%
Relative Sampling Bias (1= none, 2= 2x sampling bias)	1	1.09	1.41	2.85

There are two solutions to this issue. The first solution corrects the bias sampling problem by correcting how background values and unique occurrence localities are selected. The second solution fixes the problem by projecting all the data into an equal-areas projection (EAP). The latter is the preferred method, however for many modelers, this requires





considerable effort and can be confusing due to issues associated with selecting the best EAP. The SDM toolbox facilitates both solutions.

The first set of tools clips a Coordinate Bias File (downloadable from www.sdmtoolbox.org) to the size of your MaxEnt analysis and then calculates a BFCF for that area. The Coordinate Bias File accounts for background sampling biases associated with latitudinal changes in the area encompassed by decimal degree units. This tool converts the Coordinate Bias File to the proportion of area occupied (relative to the largest value present in your analyses) to be used as a bias file in MaxEnt. A value of 2 in the output file depicts a 50% reduction in cell area (vs. cell values of 1). Thus, in absence of a BFCF, the probability of creating a background in the cell with a value of 2 is twice as high as a cell with a value of 1. The BFCF allows equal sampling of background points throughout the landscape in geographic projections.

Info on the downloaded Coordinate Bias File: Units equal percent of area relative to size at equator (where square decimal degrees constitute the largest area). Raw percentage in downloaded file was multiplied by 10 (reducing file size by saving it as an integer file). Thus values 0 –1000 reflect 0.0 to 100.0 % of area occupied by a decimal degree (or any subunit of them) at the equator.

The second set of tools facilitate transforming climate and species occurrence data to EAP for use in MaxEnt (formatting output for direct use).

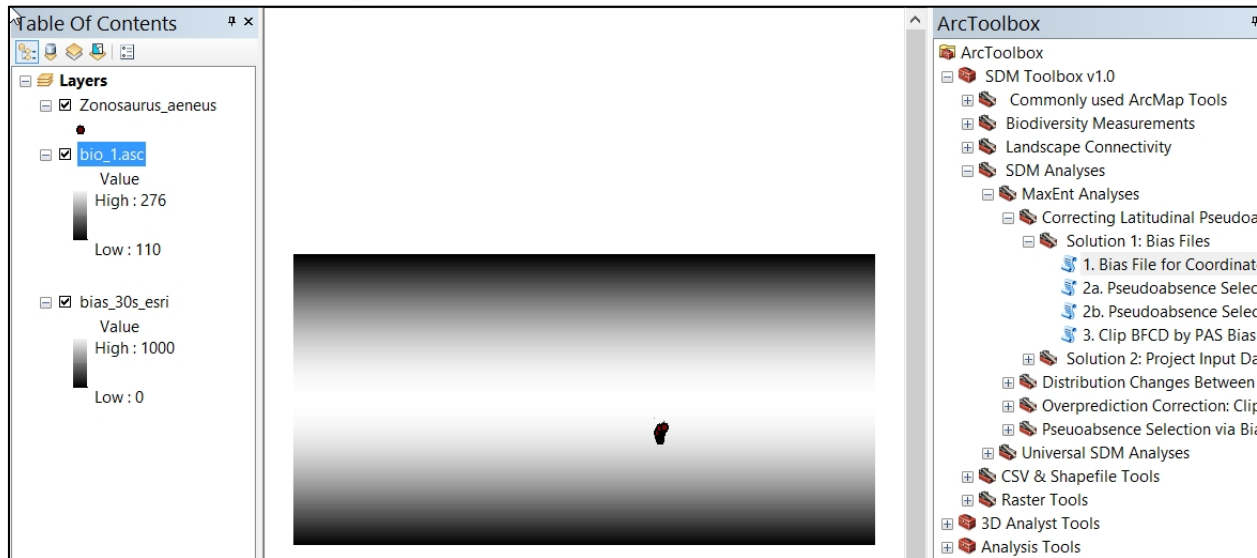
Solution 1

Quick and Easy Solution. Creating a Bias file in MaxEnt to account for biases

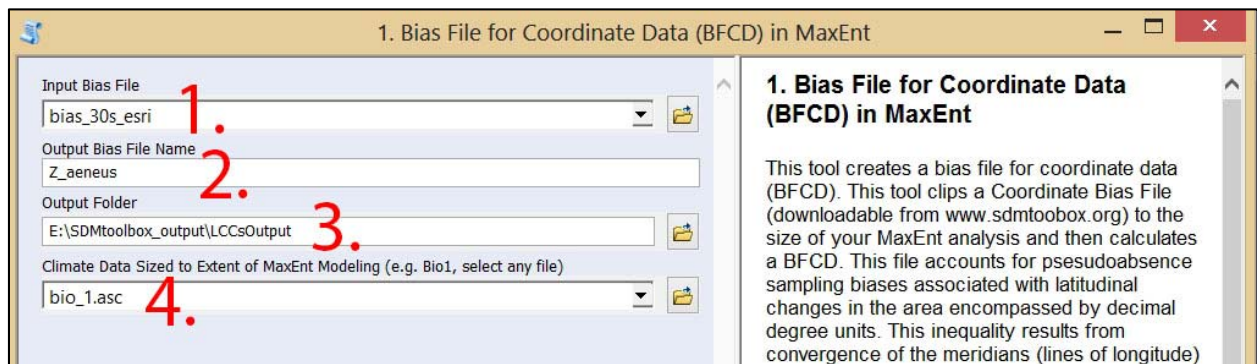
Tool: 1. Bias File for Coordinate Data (BFCF) in MaxEnt

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'Uruloke_eowynae.shp' and 'bio_1.asc'
(Location: ...\\example_data\\sdm_analyses\\correcting_lat_BPs\\soultion1\\)
3. Import the downloaded 30-arc second coordinate bias file
4. Next double-click the 'SDM Tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection → Solution 1: Bias Files → 1. Bias File for Coordinate Data (BFCF)' tool
5. Continue to tool interface instructions (following page)



Imported Data



1. Bias File for Coordinate Data (BFCFD) in MaxEnt tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

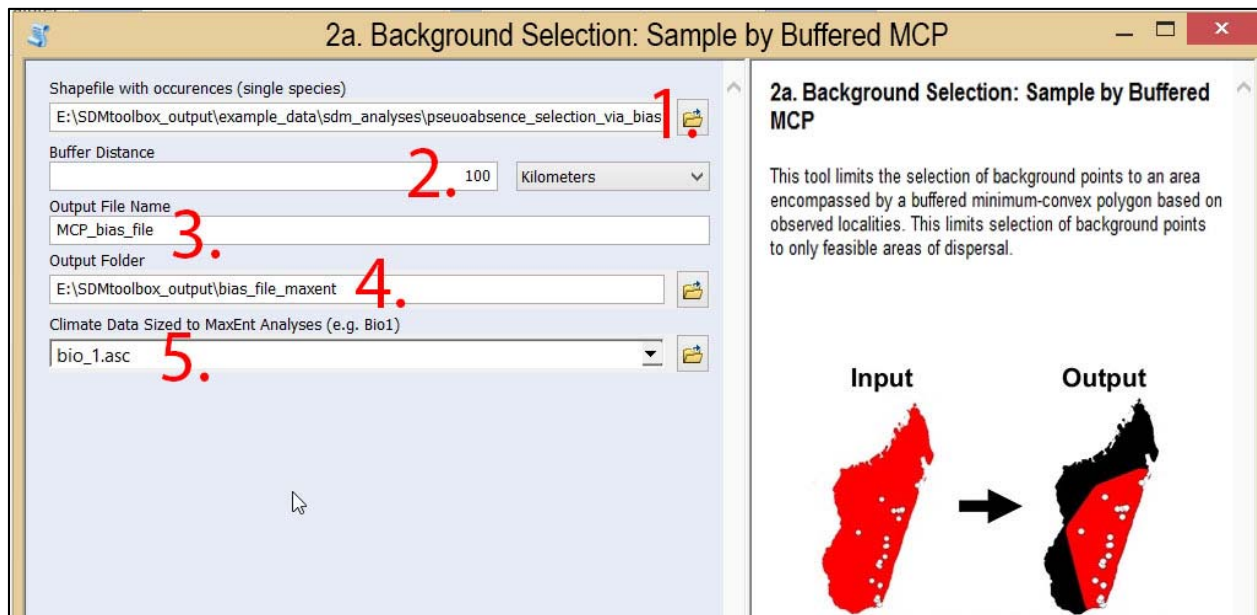
1. Download and unzip Coordinate Bias File from www.sdmtoolbox.org (this must match the resolution of your climate data)
2. Output name. Here I used 'U_eowynae'. Note: '_Coord_Bias_file_for_MaxEnt' will be appended to the end of the output name.
3. Select output folder location. This should be a new 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).
4. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).

Correcting Latitudinal Background Selection Bias: Solution 1

Tool: 2a. B Selection: Sample by Buffered MCP

ARC GIS STEP-BY-STEP GUIDE:

1. This analysis is continued from previous (the same input data are required)
2. Double-click the 'SDM Tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection → Solution 1: Bias Files → 2a. Background Points: Sample by Buffered MCP'
3. Continue to tool interface instructions (below)



2a. Background Selection: Sample by Buffered MCP tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Shapefile of species occurrence.

Tip. If you need to convert a table to a shapefile for use here. You need two extra columns with latitude and longitude. Then use the tool 'CSV & Shapefile Tools → 1. CSV to Shapefile' -Avoid non-alphanumeric characters in both names (e.g., avoid: * : \ / < > | " ? [] ; = + & £ \$, etc.).

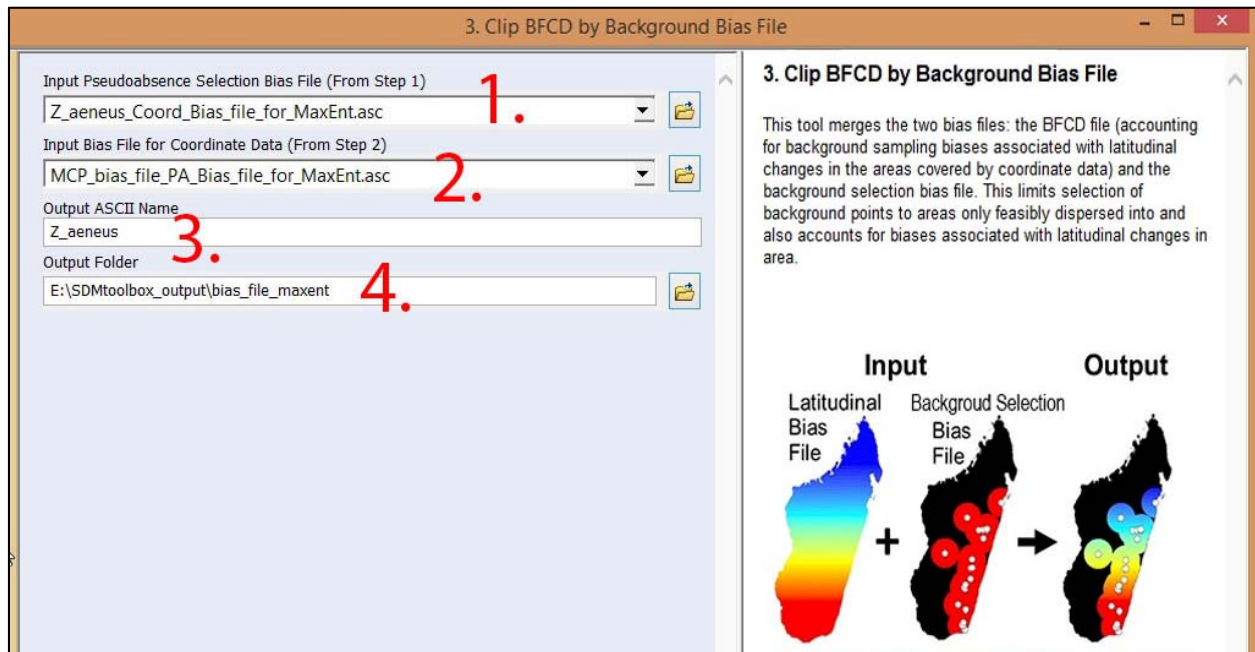
2. The distance outside of minimum-convex-polygon included in background point selection
3. Output name. Note: '_BP_Bias_file_for_MaxEnt' will be appended to the end of the output name.
4. Select output folder location. This should be a new empty folder. If not empty this can cause the analysis to fail, particularly if temporary files from a previous analysis were not properly removed (e.g. this can happen if another SDMtoolbox analysis is terminated early).
5. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).

Correcting Latitudinal Background Selection Bias: Solution 1

Tool: 3. Clip BFCD by Background Selection Bias File

ARCGIS STEP-BY-STEP GUIDE:

1. This analysis is continued from previous (outputs from both steps are required)
2. Next double-click the 'SDM Tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection → Solution 1: Bias Files → 3. Clip BFCD by BS Bias File'
3. Continue to tool interface instructions (below)



3. Clip BFCD by Background Selection Bias File tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Output from tool: 1. Bias File for Coordinate Data (BFCD) in MaxEnt
2. Output from tool:
 - 2a. Background Selection: Sample by Buffered MCP Tool Interface
 - 2b. Background Selection: Sample by Distance from Pts.
3. Output name. Note: '_Both_Bias_file_for_MaxEnt' will be appended to the end of the output name.
4. 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).

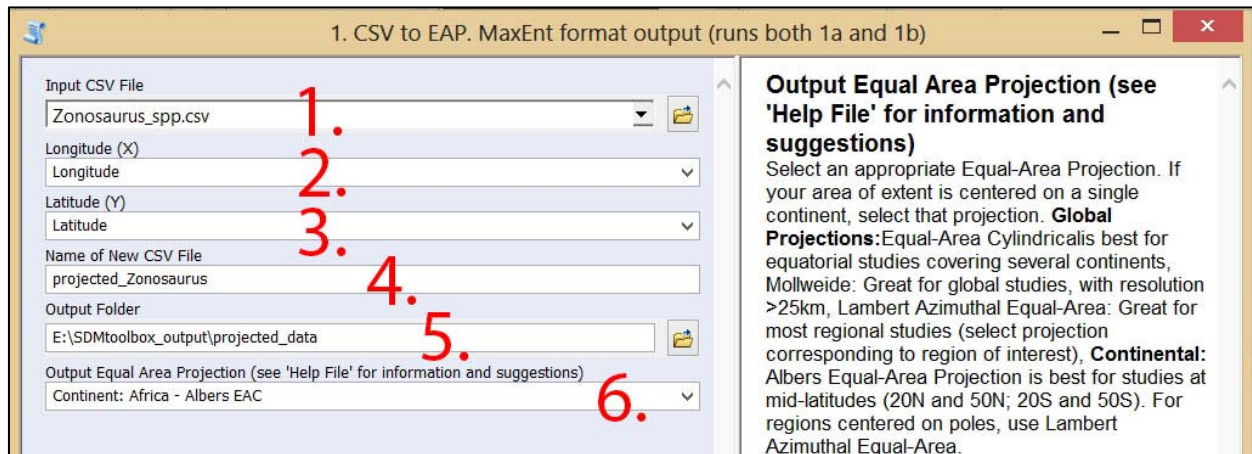
Solution 2

Best practice: Projecting all data into an equal-areas projection

Tool: 1. CSV to EAP. MaxEnt format output (runs both 1a and 1b)

ARCgis STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document.
2. Next double-click the 'SDM Tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection → Solution 2: Project Input Data to Equal-Area Projection (EAP) → 1. CSV to EAP. MaxEnt format output (runs both 1a and 1b)' tool
3. Continue to tool interface instructions (below)



1. CSV to EAP. MaxEnt format output (runs both 1a and 1b) tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Import 'Uruloke_spp.csv'
(Location: ...\example_data\sdm_analyses\correcting_lat_BPs\soultion2\)
This file should match the format required for MaxEnt (see table layout below, saved as '.csv' file). Note this tool will only accept latitude written as: 'LATITUDE', 'latitude', 'Latitude' and longitude written as: 'LONGITUDE', 'latitude', 'Latitude'.
2. Select field corresponding to latitude.
3. Select field corresponding to longitude.
4. Output name for projected CSV file.
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. Select an appropriate Equal-Area Projection. If your area of extent is centered on a single continent, select that projection.

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

Table 3. MaxEnt Input Species Data Format: three columns with species, longitude, latitude.



Help with Equal Area Projections:

Global Projections: Equal-Area Cylindrical is best for equatorial studies covering several continents; Mollweide is great for global studies with resolution >25km; Lambert Azimuthal Equal-Area is great for most regional studies.

Continental: Albers Equal-Area Projection is best for studies at mid-latitudes (20N and 50N; 20S and 50S). For regions centered on poles, use Lambert Azimuthal Equal-Area.

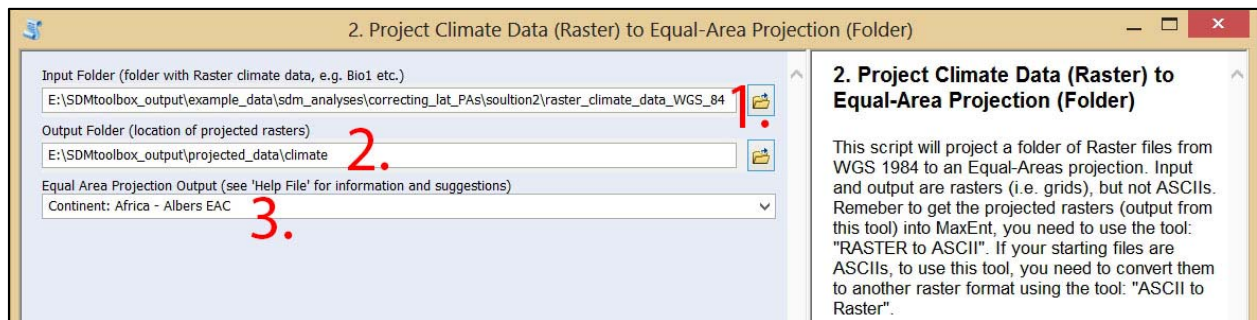
Reminder. Use the same projection for all rasters and CSVs in the study

Correcting Latitudinal Background Selection Bias: Solution 2

Tool: 2. Project Climate Data (Raster) to Equal-Area Projection (Folder)

ARCGIS STEP-BY-STEP GUIDE:

1. Double-click the 'SDM Tools → 2. MaxEnt Tools → Correcting Latitudinal Background Selection → Solution 2: Project Input Data to Equal-Area Projection (EAP) → 2. Project Climate Data (Raster) to Equal-Area Projection (Folder)' tool
2. Continue to tool interface instructions (below)



2. Project Climate Data (Raster) to Equal-Area Projection (Folder) tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input data need to be a non-ASCII raster files and all files should be in a single folder. If ASCII files, first use the '2b. ASCII to Raster (folder)' tool (part of the 'Basic Tools → Raster Tools' group).
2. 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).
3. Select an appropriate Equal-Area Projection. Use the same projection selected for the CSV species file (pg. 26 step 6).
4. Run tool. Then use the '2a. Raster to ASCII (folder)' tool, part of the 'Raster Tools' group, to convert newly projected rasters to ASCII files (input format for use in MaxEnt).

TOOL OVERVIEW

Distribution Changes Between Binary SDMs

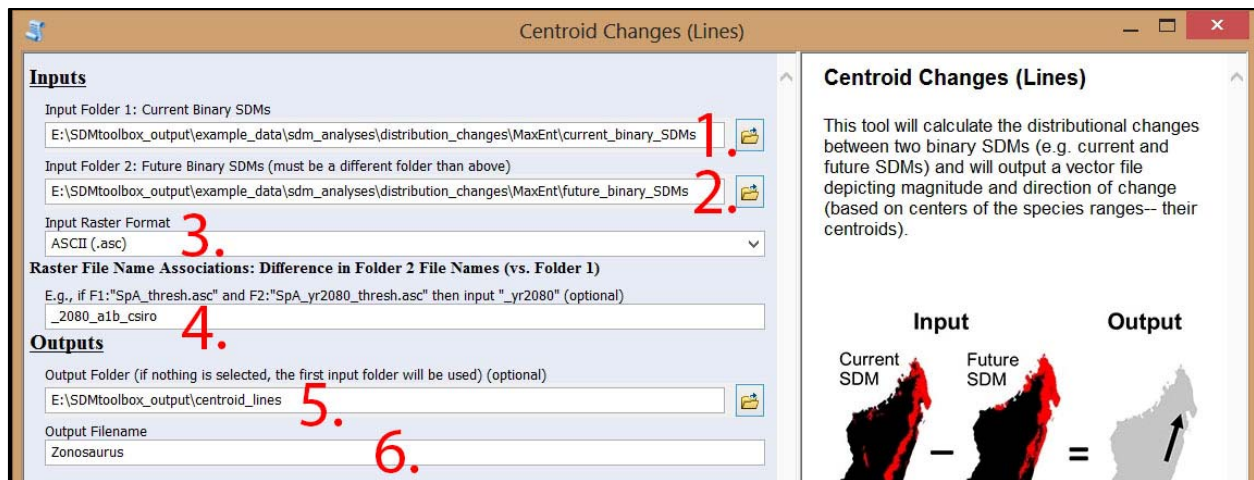
A common use of species distribution models is to predict distributional changes due to climate change. Here I created two tools that help summarize distributional changes. The first tool calculates the distributional changes between two binary SDMs (e.g. current and future SDMs). Output is a table depicting predicted contraction, expansion, and areas of no change in the species' distribution. A second tool also calculates the distributional changes between two binary SDMs (e.g. current and future SDMs), however this analysis is focused on summarizing the core distributional shifts of the ranges of many species. This analysis reduces each species' distribution to a single central point (known as a centroid) and creates a vector file depicting magnitude and direction of predicted change through time.

Distribution Changes Between Binary SDMs

Tool: Centroid Changes (Lines)

ARCGIS STEP-BY-STEP GUIDE:

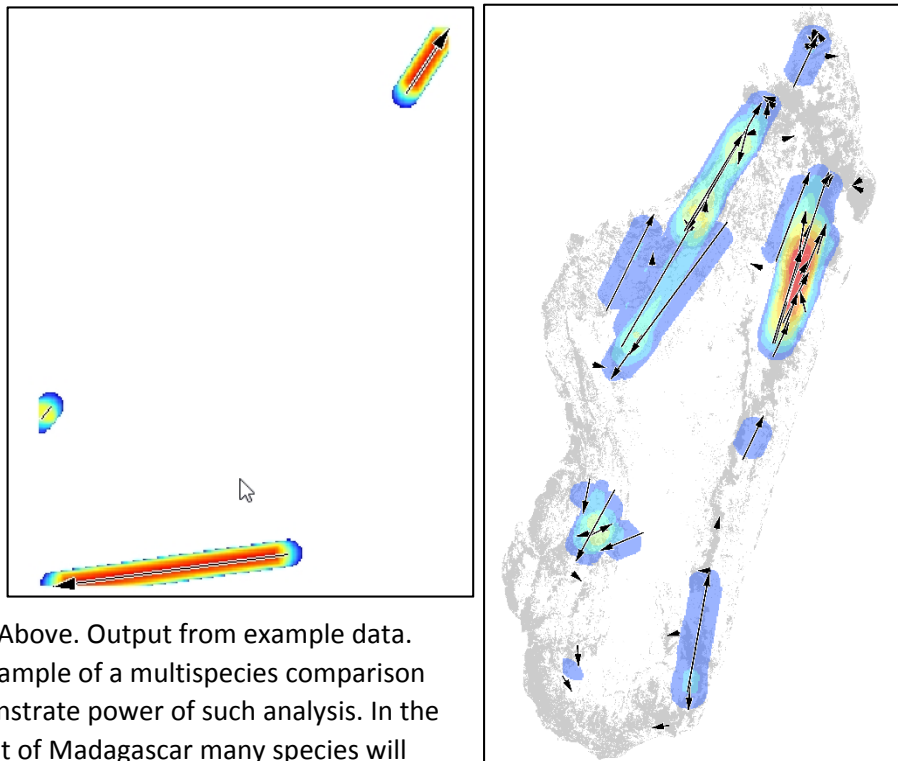
1. Open a fresh ArcMap document.
2. Double-click the 'SDM Tools → 2. MaxEnt Tools → Distribution Changes Between Binary SDMs → Centroid Changes' tool
3. Continue to tool interface instructions (following page)



Centroid Changes (Lines) tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input Folder 1: The folder containing current binary SDMs.
(Location: .../example_data/sdm_analyses/distribtutionchanges/MaxEnt/current_binary_SDMS)
2. Input Folder 2: The folder containing future binary SDMs.
(Location: .../example_data/sdm_analyses/distribtutionchanges/MaxEnt/future_binary_SDMS)
3. Input raster type. Here the example data are 'ASCII (.asc)' raster files.
4. Raster file name associations. Here input '_2080_a1b_csiro'. Because MaxEnt names the future and current SDMs differently, you need to input the difference in file names. For example, if the current SDM for a species A is named 'SpA_threshold.asc' and the future is named 'SpA_yr2080_threshold.asc' then input '_yr2080'. Note: The suffix '_threshold' will automatically be accounted for by the tool. If your SDMs do not have this, then you must use the generic 'Centroid Changes (Lines)' tool [located at: SDM Tools → 1. Universal Tools → Centroid Changes (Lines)], which has more flexibility in input file names.
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. Output name. Note: '_centroid_change' and '_cen_chg_density' will be appended to the end of the output name.



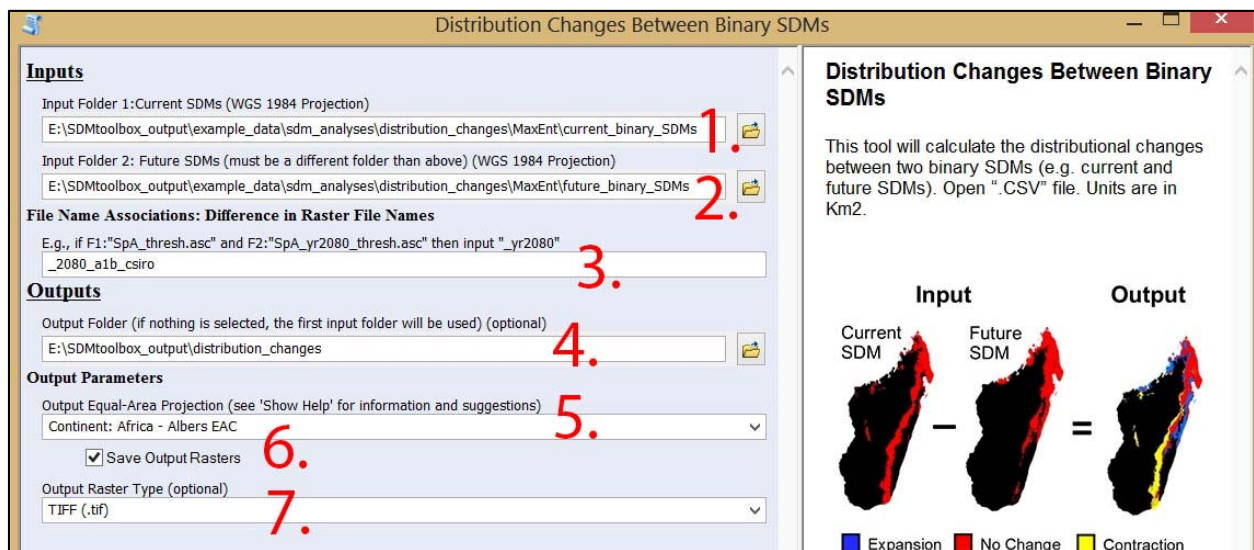
Results: Above. Output from example data. Right. Example of a multispecies comparison to demonstrate power of such analysis. In the northeast of Madagascar many species will respond similarly to climate change

Distribution Changes Between Binary SDMs

Tool: Distribution Changes Between Binary SDMs

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document.
2. Double-click the 'SDM Tools → 2. MaxEnt Tools → Distribution Changes Between Binary SDMs → Distribution Changes Between Binary SDMs' tool
3. Continue to tool interface instructions (following page)



Distribution Changes Between Binary SDMs tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input Folder 1: The folder containing current binary SDMs.
(Location: ...\\example_data\\sdm_analyses\\distribtutionchanges\\MaxEnt\\current_binary_SDMs)
2. Input Folder 2: The folder containing future binary SDMs.
(Location: ...\\example_data\\sdm_analyses\\distribtutionchanges\\MaxEnt\\future_binary_SDMs)
3. Raster file name associations. Here input ' _2080_a1b_csiro'. Because MaxEnt names the future and current SDMs differently, you need to input the difference in file names. For example, if the current SDM for species A is named 'SpA_threshold.asc' and the future is named 'SpA_yr2080_threshold.asc' then input ' _yr2080'. Note: The suffix ' _thresholded' will automatically be accounted for. If your SDMs do not have this then you must use the universal 'Distribution Changes Between Binary SDMs' tool. [located at: SDM Tools→ 1. Universal Tools→ Distribution Changes Between Binary SDMs], which has more flexibility in input file names.
4. 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).

- Select an appropriate Equal-Area Projection. If your area of extent is centered on a single continent, select that projection.

Help with Equal Area Projections:

Global Projections: Equal-Area Cylindrical is best for equatorial studies covering several continents; Mollweide is great for global studies with resolution >25km; Lambert Azimuthal Equal-Area is great for most regional studies.

Continental: Albers Equal-Area Projection is best for studies at mid-latitudes (20N and 50N; 20S and 50S). For regions centered on poles, use Lambert Azimuthal Equal-Area.

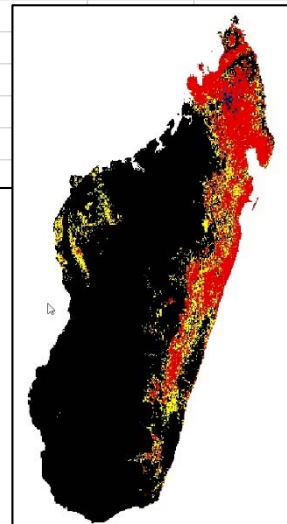
Reminder. Use the same projection for all rasters and CSVs in the study

- Check to save the raster files of results (see below)
- Output file type. Here I selected 'Tiff (.tif)' format.

Tip. I prefer TIFF files as output format because they allow for longer file names (vs. ESRI grid files that are limited to 13 characters) and don't have too many raw parts to each file. There is, however, a slight reduction in performance (vs. ESRI grid files), thus, if processing thousands of rasters this should be taken into consideration.

Results

SDM Tools: Distribution Changes of zonosaurus_madagascariensis_changes	
Output Key: -1 = range expansion; 0= no occupancy (absence in both); 1 = no change (presence in both); 2 = range contraction	
Value	Area (Km2)
-1	525.2446
0	458408.8
1	82113.24
2	47615.69



Results: CSV file of area changes and to right, a raster of results (optional).

TOOL OVERVIEW

Overprediction Correction: Clip Models by Buffered Minimum Convex Polygons

To limit over-prediction of SDMs, a problem common with modeling species distributions, 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 observed range and unsuitable habitat through the landscape.

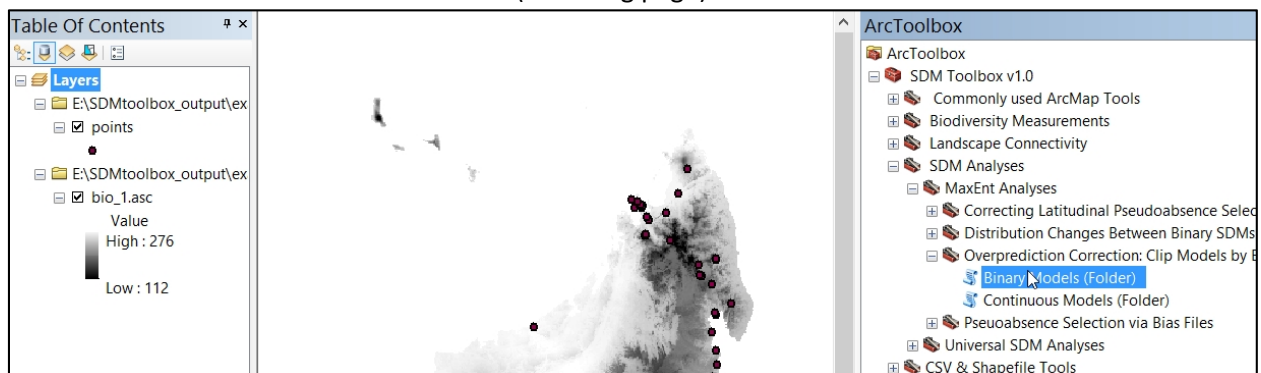
Kremen, C., A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. Louis Jr., R. A. Nussbaum, C. J. Raxworthy, A. Razafimpahanana, G. E. Schatz, M. Vences, D. R. Vieites & M. L. Zjhra (2008): Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. – *Science* **320**: 222-226.

Overprediction Correction: Clip Models by Buffered Minimum Convex Polygons

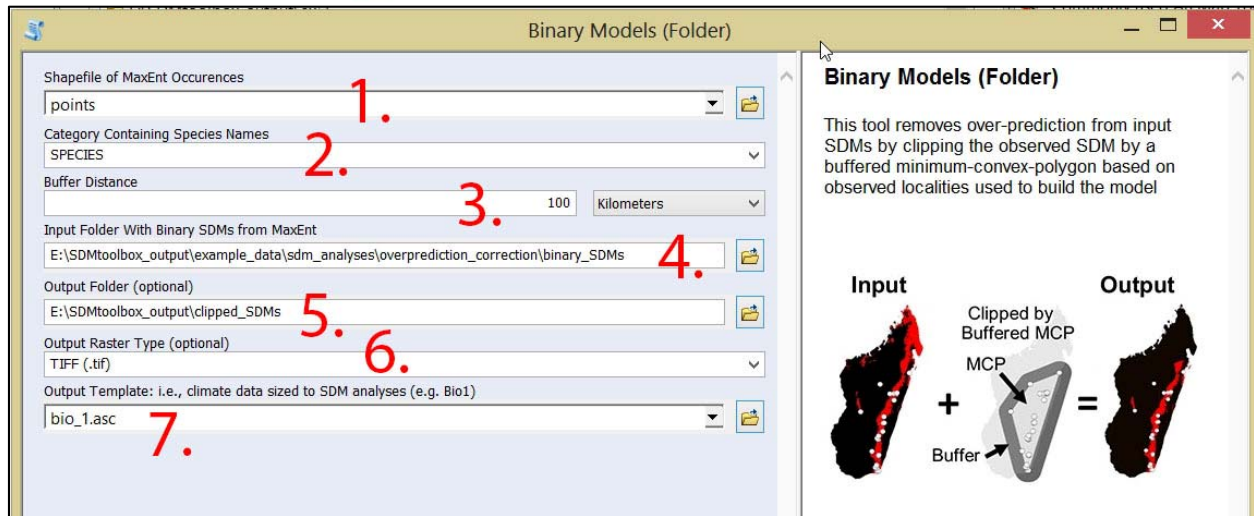
Tool: Binary Models (folder)

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document.
2. Import 'points.shp'
(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\)
3. Import 'bio_1.asc'
(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\climate_data)
4. Double-click the 'SDM Tools → 2. MaxEnt Tools → Overprediction Correction: Clip Models by Buffered MCPs → Binary Models (Folder)' tool
5. Continue to tool interface instructions (following page)



Imported Data



Binary Models (folder) tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

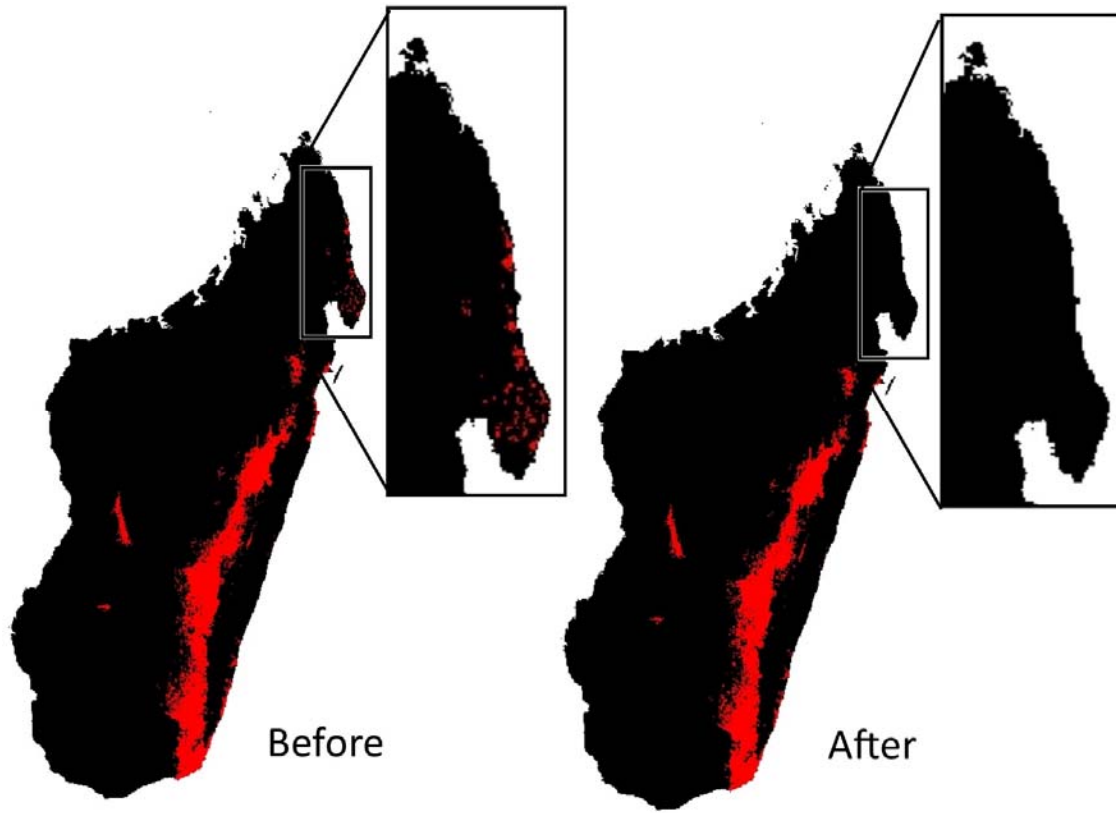
1. A point shapefile of species occurrences corresponding to input SDMs. Note: species name in shapefile must perfectly match input occurrence data species names.
2. Field indicating species names, here: 'SPECIES'
3. Distance to buffer minimum-convex polygons. Here try: 100km.
4. The folder containing binary SDMs to be clipped.

(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\binary_SDMs)

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. Output file type. Here I selected 'Tiff (.tif)' format.
7. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).



Results

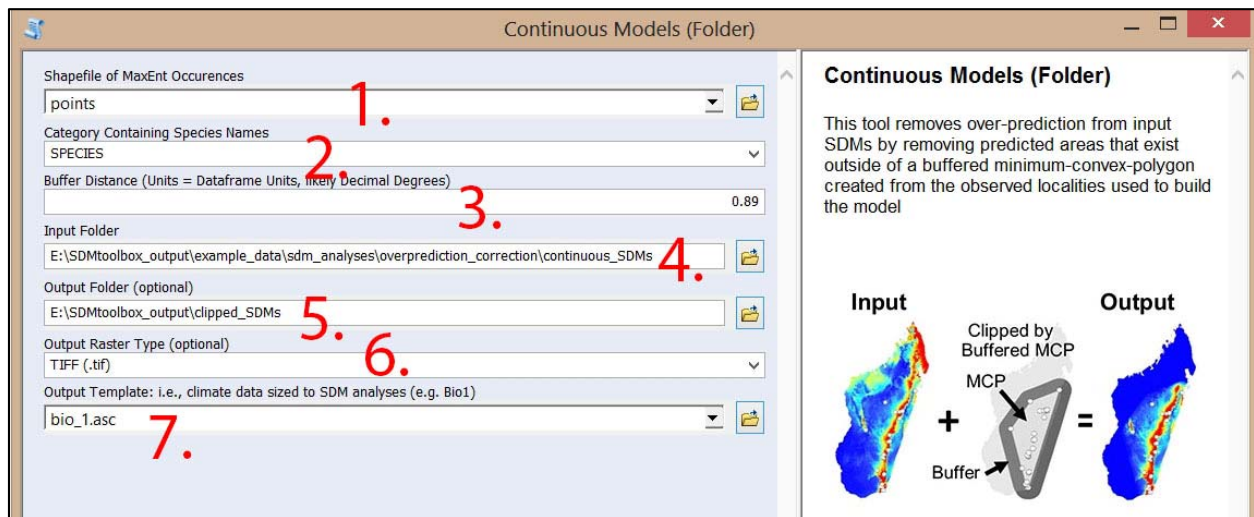


Overprediction Correction: Clip Models by Buffered Minimum Convex Polygons

Tool: Continuous Models (folder)

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'points.shp'
(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\)
3. Import 'bio_1.asc'
(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\climate_data)
4. Double-click the 'SDM Tools → 2. MaxEnt Tools → Overprediction Correction: Clip Models by Buffered MCPs → Continuous Models (Folder)' tool
5. Continue to tool interface instructions



Continuous Models (folder) tool interface

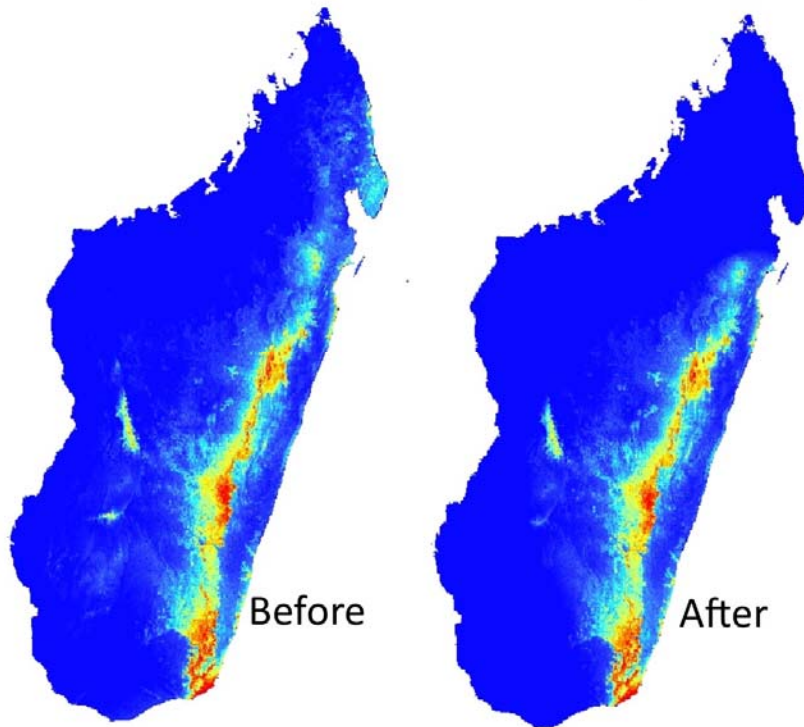
SDMTOOLBOX STEP-BY-STEP GUIDE:

1. A point shapefile of species occurrences corresponding to input SDMs. Note. Species name in shapefile must perfectly match input occurrence data species names.
2. Field indicating species names. Here 'SPECIES'
3. Distance to buffer minimum-convex polygons. Here try 100km.
4. The folder containing binary SDMs to be clipped.

(Location: ...\\example_data\\sdm_analyses\\overprediction_correction\\binary_SDMs)
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. Output file type. Here I selected 'Tiff (.tif)' format.

7. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).

Results



TOOL OVERVIEW

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 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 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 two 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).

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 neighbors 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) 1. only the occurrence points from that species, or ii) 2. 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 and likelihood of detection of your species in all surveys (e.g. sampling only near roads).

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.

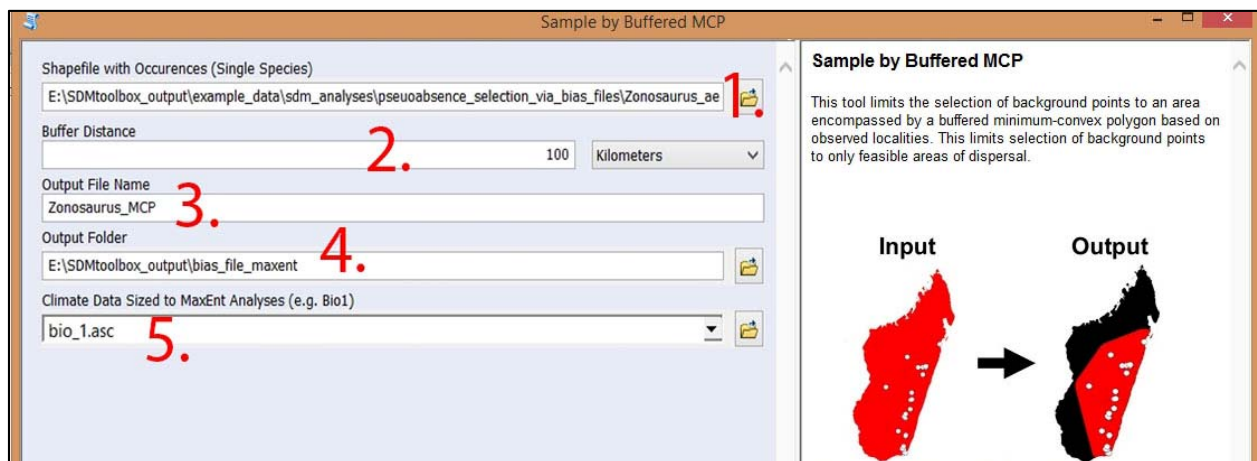
Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M. B. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373.

Background Selection via Bias Files

Tool: Background Selection: Sample by Buffered MCP

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'Uruloke_eowynae.shp' and 'bio_1.asc' (Location: ...\\example_data\\sdm_analyses\\background_selection_via_bias_files)
3. Double-click the 'SDM Tools → 2. MaxEnt Tools → Background Selection via Bias Files → Background Selection: Sample by Buffered MCP' tool
4. Continue to tool interface instructions (below)



Background Selection: Sample by Buffered MCP tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Shapefile of species occurrence.

Tip. If you need to convert a table to a shapefile for use here. You need two extra columns with latitude and longitude. Then use the tool 'Basic Tools → CSV & Shapefile Tools → 1. CSV to Shapefile'
 -Avoid non-alphanumeric characters in both names (e.g., avoid: * : \ / < > | " ? [] ; = + & £ \$, etc.).

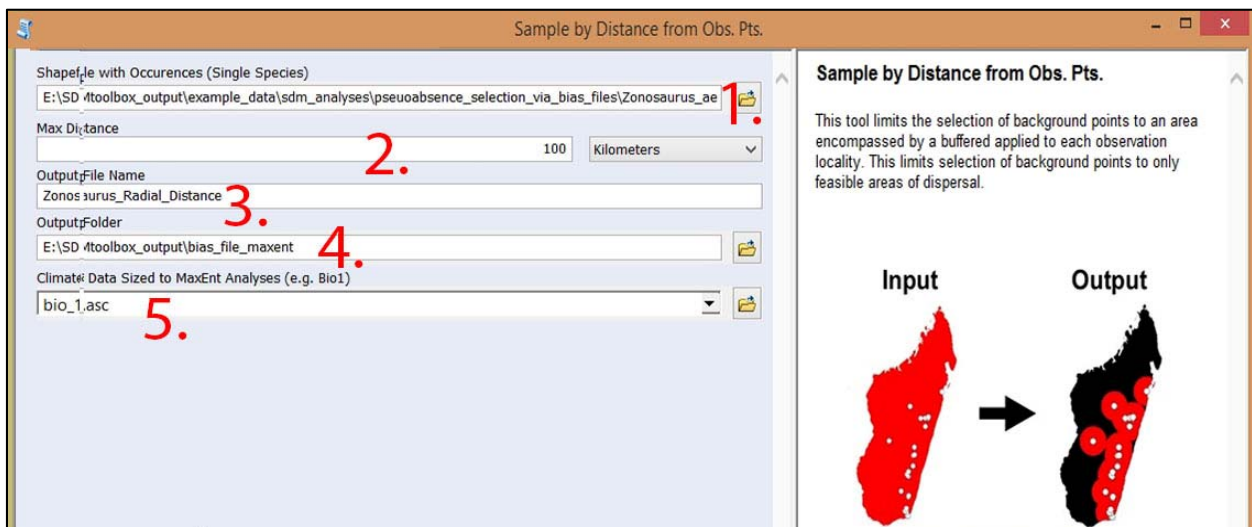
2. The distance outside of minimum-convex-polygon included in background selection.
3. Output name.
4. Select output folder location. This should be a new empty folder. If not empty this can cause the analysis to fail, particularly if temporary files from a previous analysis were not properly removed (e.g. this can happen if another SDMtoolbox analysis is terminated early).
5. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).

Background Selection via Bias Files

Tool: Background Selection: Sample by Distance from Obs. Pts.

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'Uruloke_ewynae.shp' and 'bio_1.asc' (Location: ...\example_data\sdm_analyses\background_selection_via_bias_files)
3. Double-click the 'SDM Tools → 2. MaxEnt Tools → Background Selection via Bias Files → Sample by Distance from Obs. Pts.'
4. Continue to tool interface instructions (below)



Background Selection: Sample by Distance from Obs. Pts. tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Shapefile of species occurrence.

Tip. If you need to convert a table to a shapefile for use here. You need two extra columns with latitude and longitude. Then use the tool 'Basic Tools → CSV & Shapefile Tools → 1. CSV to Shapefile'
-Avoid non-alphanumeric characters in both names (e.g., avoid: * : \ / < > | " ? [] ; = + & £ \$, etc.).

2. The maximum distance away from the observed localities to include in background selection.
3. Output name.
4. 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).
5. Climate data sized to extent of MaxEnt Modeling. Here use the imported '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).

UNIVERSAL SDM TOOLS

TOOL OVERVIEW

Create Friction Layer: Invert SDM

The use of least-cost paths and along-path distances often dramatically improve the calculation of geographic distance for testing hypotheses (such as, isolation by distance). However, few studies have access to meaningful friction landscapes. Some researchers (i.e. Broquet et al. 2006) generate friction landscapes from classified satellite images where each major habitat type represents a different value. A primary downfall to using habitat heterogeneity as a friction landscape is the weighing of each habitat class to represent relevant friction values. Doing this properly relies heavily on expert life history knowledge and when done analysis loses some objectivity. For example, Broquet et al. (2006) adjusted the friction values until they satisfied prior expectations. More recently authors used species distribution models (SDMs) as friction landscapes (Wang et al. 2008, Chan et al. 2011). This method is a more objective alternative to expert knowledge and the generation of high quality SDMs can be done with relative ease for many species.

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 LM, Brown JL, Yoder AD (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Mol Phylogenet Evol* 59(2):523-37.

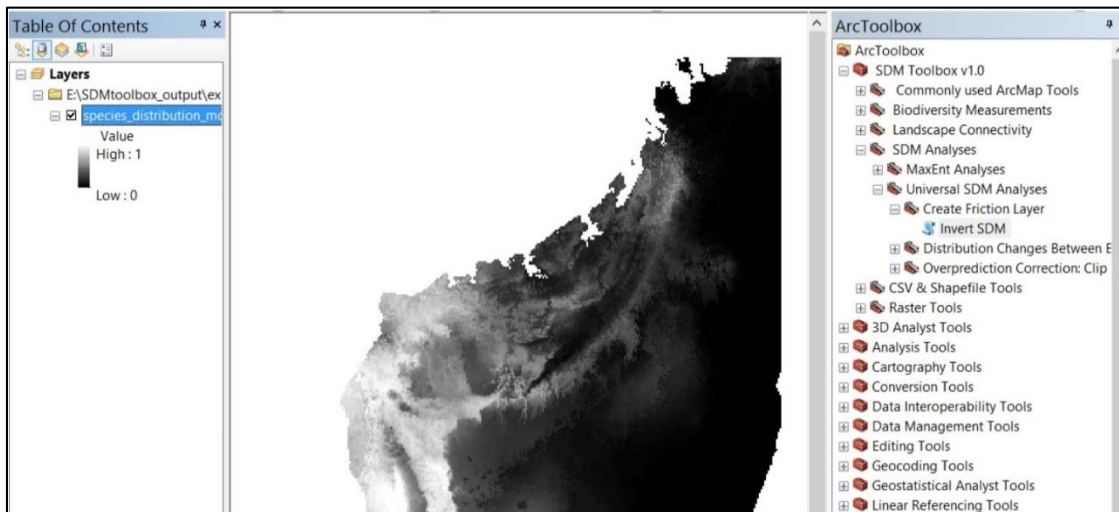
Wang, Y.H., Yang, K.C., Bridgman, C.A., Lin, L.K. (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, **23**, 989–1000

Create Friction Layer: Invert SDM

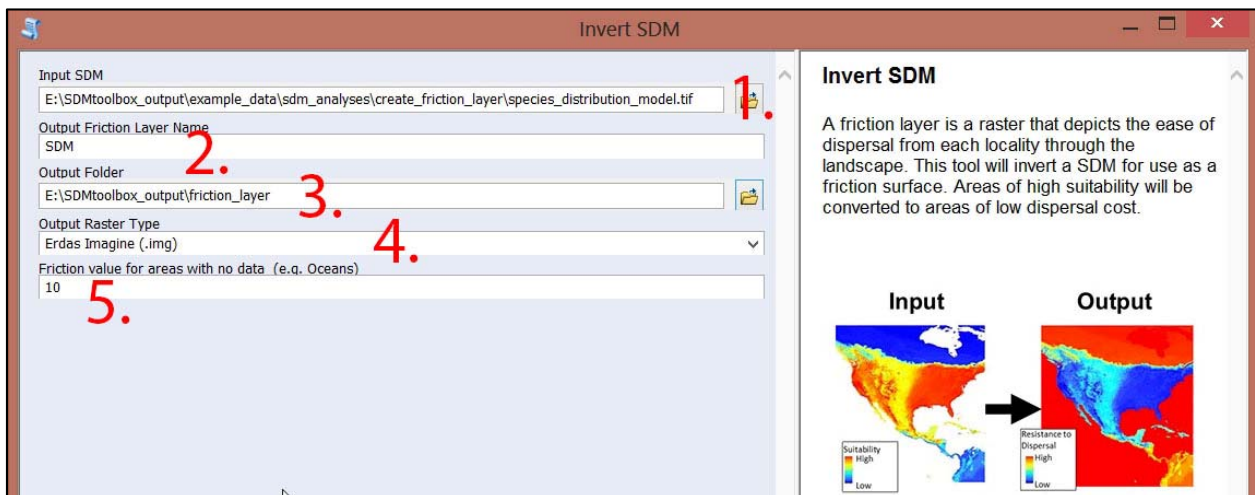
Tool: Invert SDM

ARC GIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'species_distribution_model.tif'
(Location: ...\\example_data\\sdm_analyses\\create_friction_layer\\)
3. Double-click the 'SDM Tools → 1. Universal Tools → Create Friction Layer → Invert SDM'
4. Continue to tool interface instructions (below)



Input data



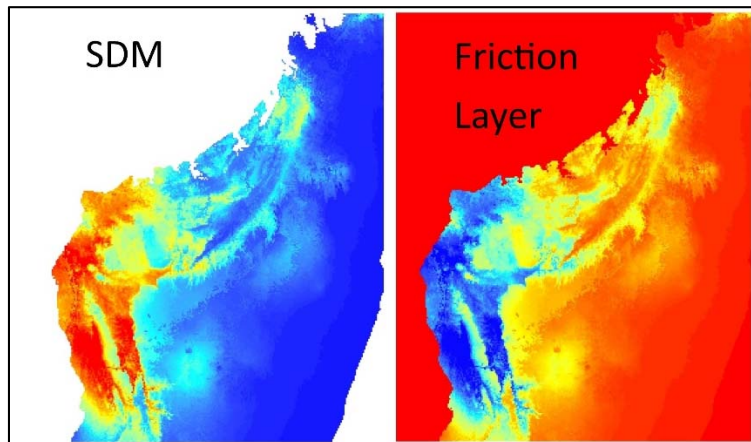
Invert SDM tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input species distribution model (SDM)

2. Name of output file. Note: '_FrictionLayer' will be appended to the end of output name.
3. 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).
4. Output raster format. Here chose 'ERDAS Imagine (.img)'
5. Areas of 'NoData' (cells that lack SDM values, i.e. oceans) can be converted to friction values. This is particularly important if using frictions layers to connect populations isolated on islands and mainland populations.

Results



Invert SDM tool. Right: Input SDM. Left: Output friction layer

UNIVERSAL SDM TOOLS

TOOL OVERVIEW

Spatially Rarefy 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 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.

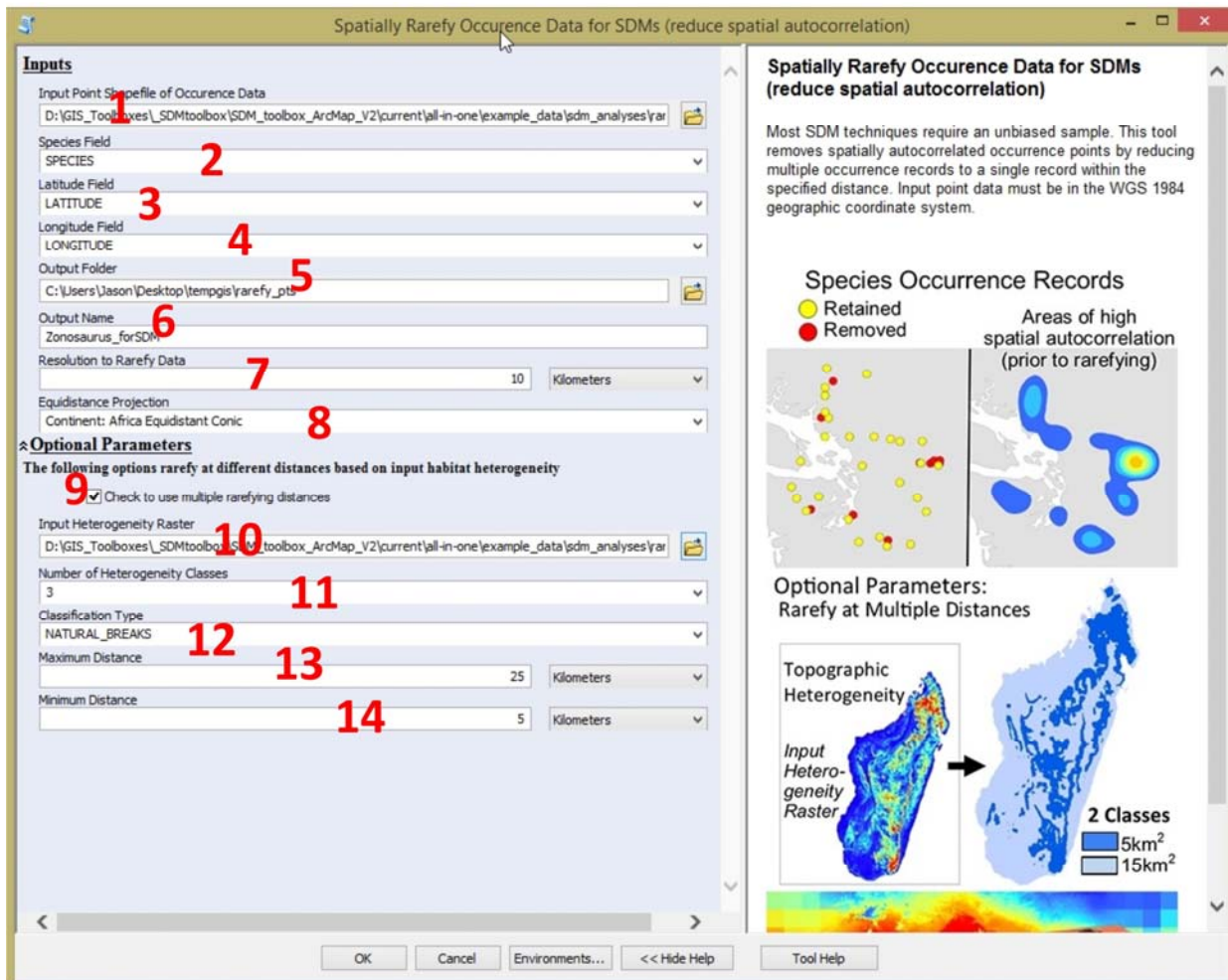
Boria R. A., Olson L.E., Goodman S.M. & Anderson R.A. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modeling*, 275, 73-77.

Spatially Rarefy Occurrence Data

Tool: Spatially Rarefy Occurrence Data

ARCgis STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'occurrences.shp' and 'climate_hetero.tif'
(Location: ...\\example_data\\sdm_analyses\\rarefy\\)
3. Double-click the 'SDM Tools → 1. Universal Tools → Spatially Rarefy Occurrence Data for SDMs (reduce spatial autocorrelation)' tool
4. Continue to tool interface instructions (below)



Spatially Rarefy Occurrence Data tool interface

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 'Continent: Africa Equidistant Conic'
9. Placing a check mark in the box to execute the multi-distance data rarefying

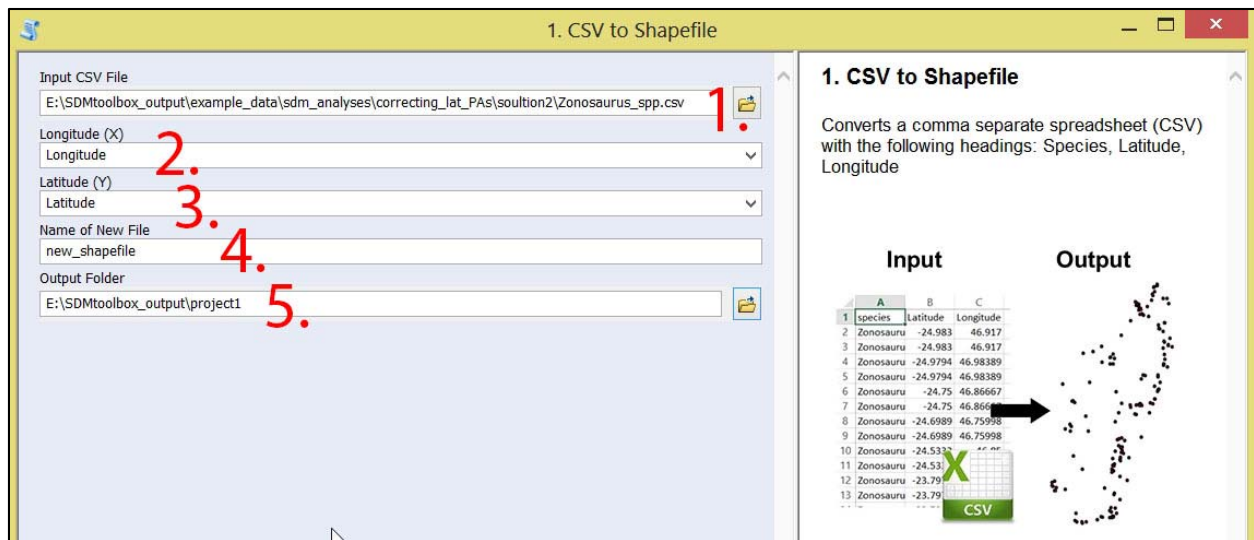
10. Input heterogeneity raster, here the "climate_hetero.tif" This is require if using the multi-distance data rarefying
11. The number of classes, here 5
12. Classification type, here 'NATURAL_BREAKS'
13. The maximum distance, here 25 Kilometers
14. The minimum distance, here 2 Kilometers

Chapter 4. Raster, CSV & Shapefile tools

Tool: CSV, TXT, XLS to shapefile

ARCGIS STEP-BY-STEP GUIDE:

1. Open a fresh ArcMap document
2. Import 'Uruloke_spp.csv'
(Location: ...\\example_data\\SDM_analyses\\correcting_lat_BPs\\soultion2\\)
3. Double-click the 'CSV & Shapefile Tools → 1. CSV to Shapefile'
4. Continue to tool interface instructions (below)



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



Chapter 5. 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

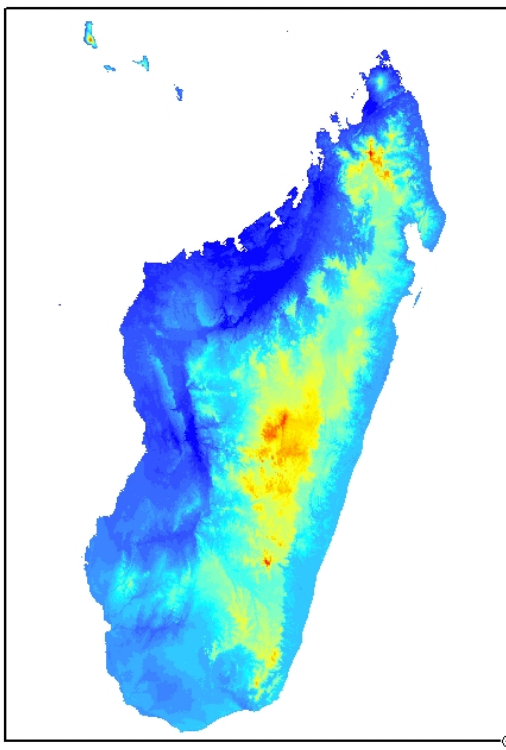
ARCGIS STEP-BY-STEP GUIDE:

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 templates (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 issues (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

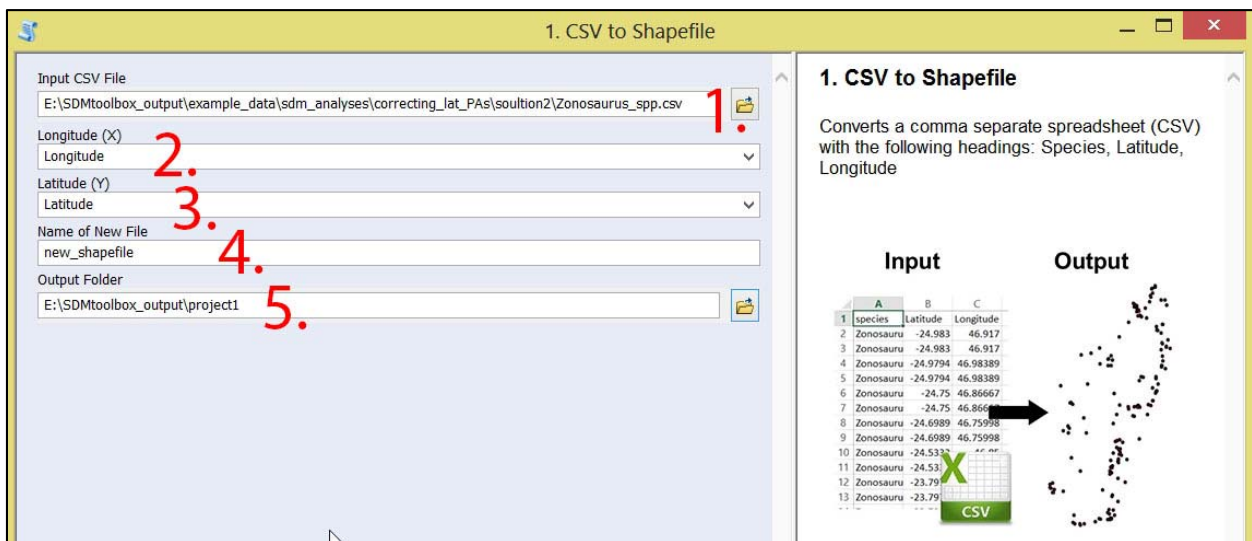
Tool: CSV, TXT, XLS to shapefile

ARC GIS 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

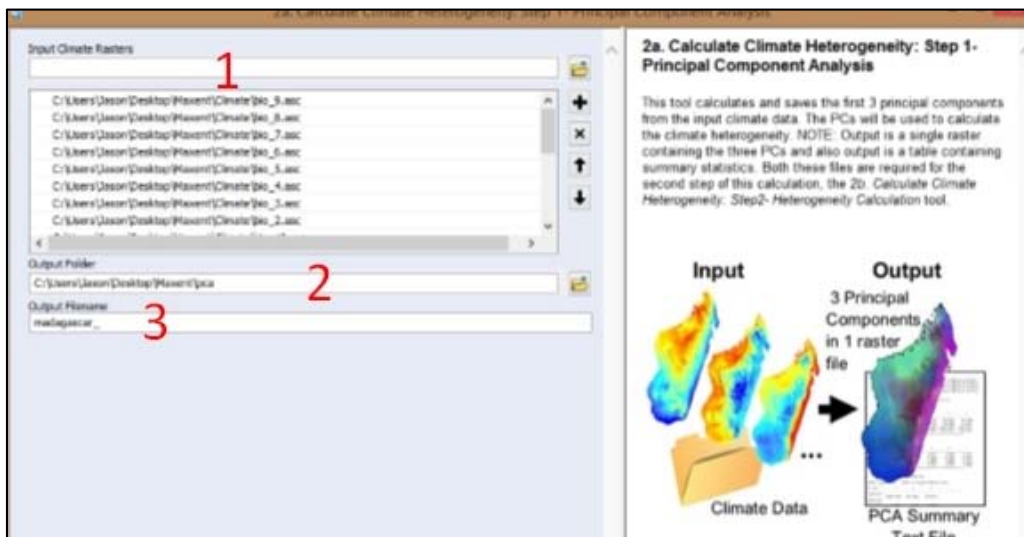
IF input data are another projection, use ArcMap's tool: "Define Projection".

3C. Preparing Occurrence Data: PCA of Climate Variables to Estimate Heterogeneity

Tool: 2a. Calculate Climate Heterogeneity: Step 1- Principal Component Analysis

ARCgis STEP-BY-STEP GUIDE:

1. Double-click the 'SDM Tools → 1. Universal Tools → 2a. Calculate Climate Heterogeneity: Step 1- Principal Component Analysis' tool
2. Continue to tool interface instructions (below)

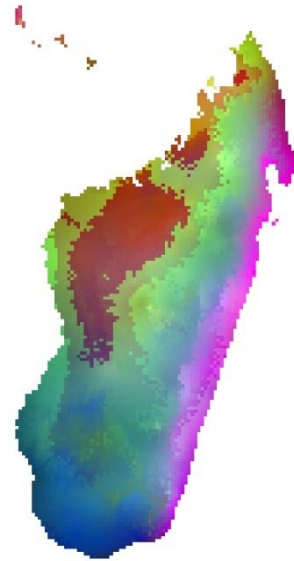
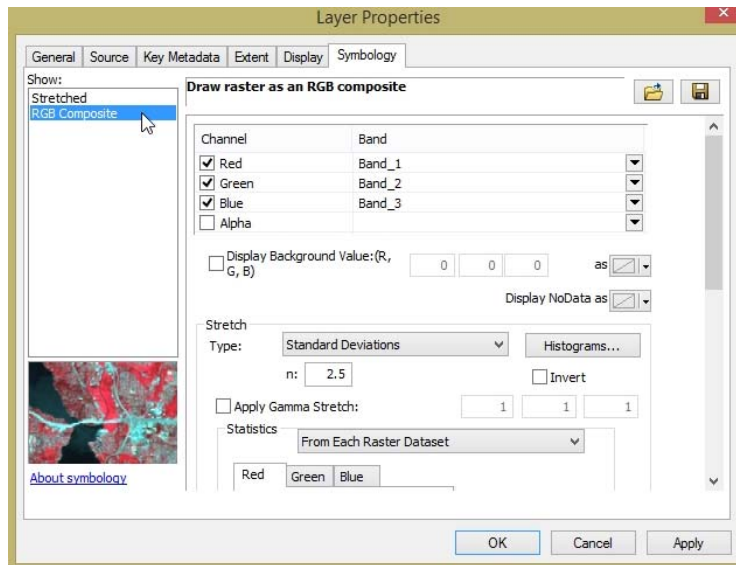


2a. Calculate Climate Heterogeneity: Step 1- Principal Component Analysis tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input all your climate rasters that depict continuous data. Note: hold 'shift' to select many rasters at once.

2. 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).
3. Name of output file. Note that "_Climate_PCs" will be appended to file name.



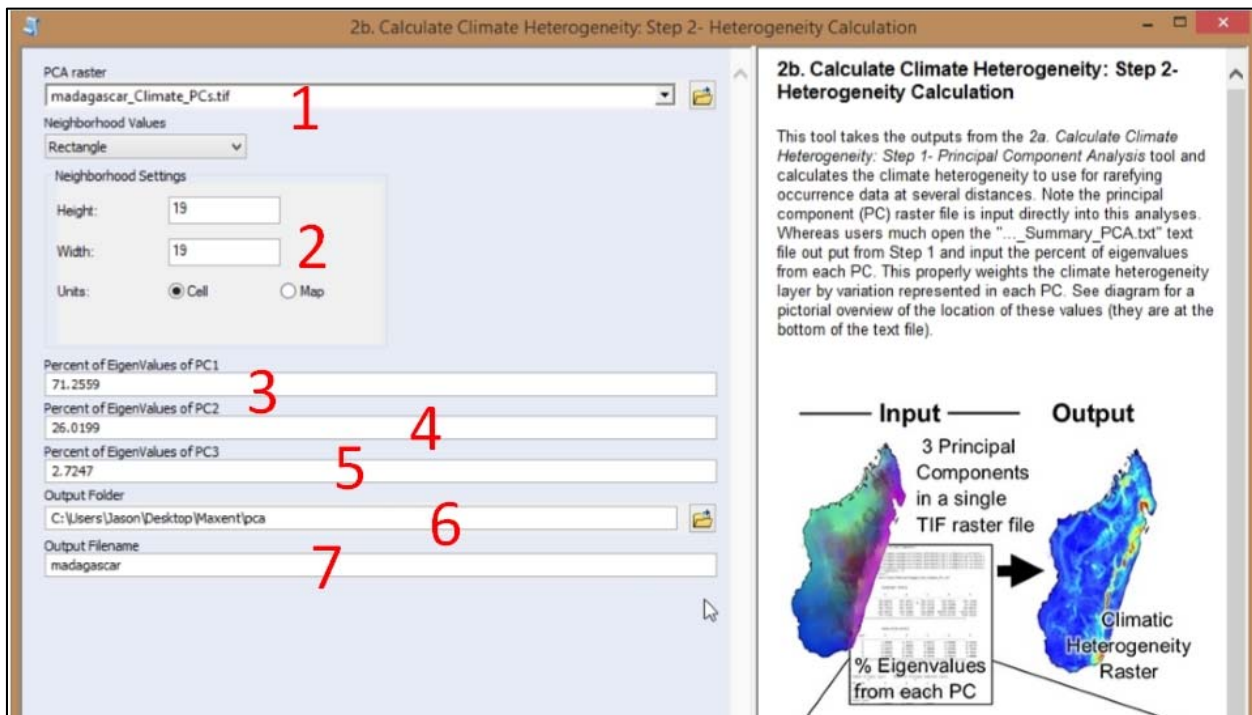
To visualize all three bands output from the PCA tool, right click layer and select 'Properties'. Then select the 'Symbology' tab and select 'RGB Composite'. The output raster depicts climate space: the more similar the colors the more similar values.

3D. Preparing Occurrence Data: Measure Spatial Heterogeneity of Climate PCs

Tool: 2b. Calculate Climate Heterogeneity: Step 2- Heterogeneity Calculation

ARCGIS STEP-BY-STEP GUIDE:

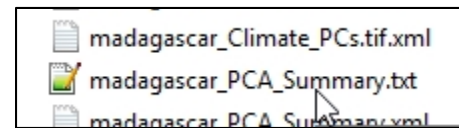
1. Double-click the 'SDM Tools → 1. Universal Tools → 2b. Calculate Climate Heterogeneity: Step 2- Heterogeneity Calculation' tool
2. Continue to tool interface instructions (below)



2b. Calculate Climate Heterogeneity: Step 2- Heterogeneity Calculation tool interface

SDMTOOLBOX STEP-BY-STEP GUIDE:

1. Input climates PCA raster output from previous step
2. This is the spatial scale used to calculate the heterogeneity of the landscape. E.g., if 3 x 3 rectangle and cell units are selected: heterogeneity values will be calculated from each raster pixel and the 8 cells neighboring the focal cell.
3. Percent of EigenValues of PC1. This value is input from the "..._PCA_summary.txt" file. Go to the bottom of the text file to the heading 'PERCENT AND ACCUMULATIVE



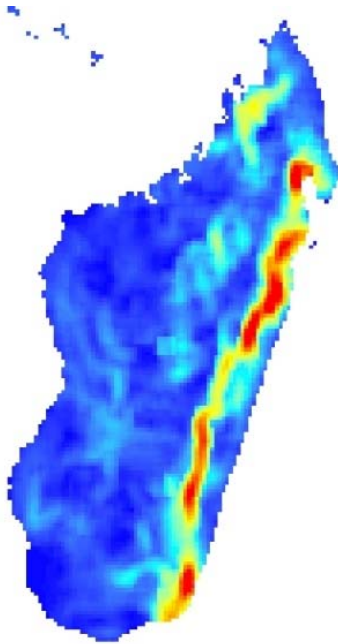
Output PCA Summary table

PERCENT AND ACCUMULATIVE EIGENVALUES			
PC Layer	EigenValue	Percent of EigenValues	Accumulative of EigenValues
1	167483.51078	71.2559	71.2559
2	61158.52418	26.0199	97.2758
3	6403.13672	2.7242	100.0000

Section of the table that is necessary for input into this tool

EIGENVALUES' and go to the row corresponding to PC1 and input the value from the 'Percent of EigenValues'.

4. Percent of EigenValues of PC2.
5. Percent of EigenValues of PC3.
6. Output Folder
7. Output filename. Note: "_clim_hetero.tif" will automatically be appended.



Above. Output climate heterogeneity raster. Here warm colors depict high areas of climatic heterogeneity.



3E. Preparing Occurrence Data: Graduated 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 particularly 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 Modelling*, 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.
9. Placing a check mark in the box to execute the multi-distance data rarefying
10. Input heterogeneity raster.
11. The number of classes, here 5
12. Classification type, here 'NATURAL_BREAKS'

- 13. The maximum distance, here 25 Kilometers
- 14. The minimum distance, here 2 Kilometers

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).

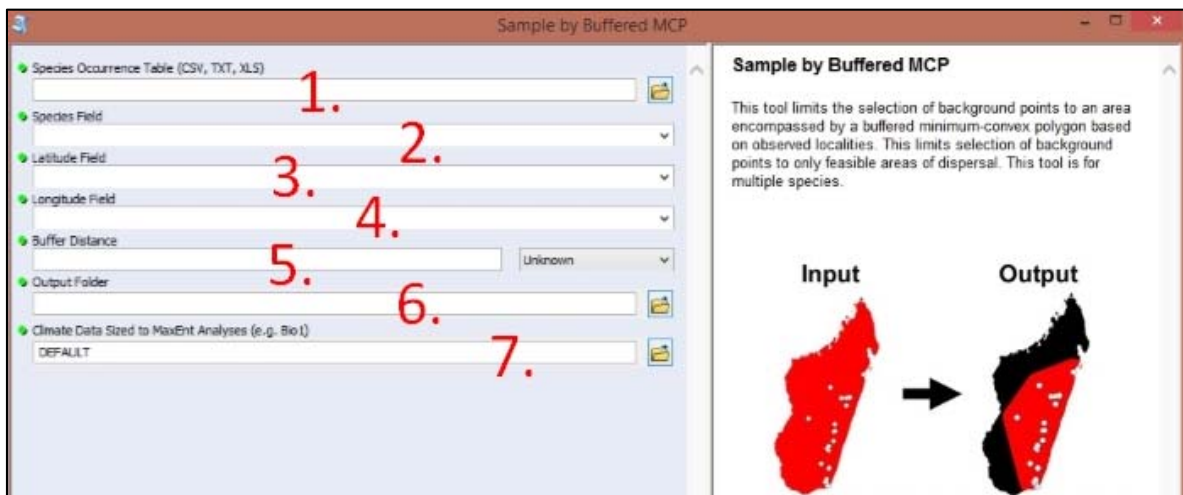
To circumvent this problem, many researchers have begun using background point and pseudo-absence selection methods that are more regional. SDMtoolbox contains three tools to facilitate more sophisticated background selection for use in Maxent. The *Sample by Distance from Obs. Pts.* tool (see: SDM Tools → 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, Dudik 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

The screenshot shows the 'Run MaxEnt: Spatial Jackknifing' application window. The interface is divided into several sections with various input fields and checkboxes. Red numbers 1 through 33 are overlaid on the interface to indicate specific settings:

- 1: MaxEnt Program
- 2: Species Occurrence Table (CSV, TXT, XLS)
- 3: Species Field
- 4: Longitude (X) Field
- 5: Latitude (Y) Field
- 6: Environmental Data Folder
- 7: Categorical Layers (optional)
- 8: Layers in environmental data folder to exclude (optional)
- 9: Bias File Folder (see notes) (optional)
- 10: Model Output Folder
- 11: Output Format
- 12: Output File Type
- 13-17: MaxEnt Settings (Basic Settings: Create Response Curves, Create Prediction Pictures, Do Jackknife to Measure Variable Importance, Skip If Output Exists, Suppress Warnings)
- 18: Main Model Parameters (Regularization Multiplier)
- 19: Model Projection Settings (Projection Climate Layers)
- 20: Clamping / Do not extrapolate
- 21-22: CPU Settings (Number of processor threads to use)
- 23: MaxEnt GUI Size
- 24: Spatial Jackknifing (Perform Spatial Jackknifing?)
- 25: Minimum number of occurrence points to spatially jackknife
- 26: Replicates of each model parameter class
- 27: Number of spatial groups
- 28: Use spatially segregated groups
- 29: Model Parameters for Species Not Spatially Jackknifed (Number of replicates)
- 30: Replicate Type (for non-spatially jackknifed species)
- 31: Percentage Random Test Points
- 32: Additional Parameters For Standard MaxEnt Runs (Auto-Feature Classes)
- 33: Manual Feature Classes (Feature Linear, Quadratic, Hinge, Product, Threshold)

On the right side of the window, there is a text area with instructions and a 'Spatial Jackknifing' section showing diagrams of spatial groups (AB, AC, BC) and training/evaluation points. Below this, there are maps illustrating the spatial jackknifing process with 6 different groupings.



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. If 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 subsampling (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 random 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 you 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
<p>Data compilation occurrence records:</p> <p>Are species presences (and absence) records representative of the actual distribution?</p>	<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
<p>Data compilation Environmental variables:</p> <p>Do environmental variables accurately capture the association between species subsistence and the environment at the relevant scale?</p>	<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)
<p>Model generation and calibration</p> <p>Is the modelling algorithm appropriate given the data available and research question?</p>	<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
<p>Model generation and calibration</p> <p>Is the model appropriately calibrated for the data available and research question?</p>	<ul style="list-style-type: none"> ▪model complexity ▪model selection procedure ▪setting of model parameters ▪variable selection strategy
<p>Model validation</p> <p>Is validation performed on truly independent data and under appropriate settings?</p>	<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
<p>Model projection</p> <p>Is the species environment relationship likely to be maintained in space and/or time?</p>	<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