

Multiscale MaxEnt ArcGIS toolbox guide

Information

Background

This toolkit has been made to help users create and prepare multiscale environmental variables for building Habitat Suitability Models with MaxEnt (Phillips, Anderson & Schapire 2006).

Citations

If you use this toolkit, please cite these papers in any published material:

Bellamy, C., Scott C., & Altringham, J.D. (2013) Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species. *Journal of Applied Ecology* 50 (892-901)

Further details

Dr Chloe Bellamy (c.bellamy@leeds.ac.uk)

Version control

Version 2: June 2014

Updates from version 1 (uploaded January 2014):

- Corrections & updates to the user guide made.
- The mask and bias file models ("F" models) have been updated to prevent the use of a default cell size. A model has been added to allow bias files to be created with polygon type data on sampling effort.
- Errors have been fixed to prevent incorrect inline variable substitution

Model descriptions

This toolkit allows users to create multiscale environmental variables in ArcGIS using the methods described in Bellamy, Scott & Altringham (2013). It organises and names all of the gridded output, ensuring they are all of the same extent and cell size, as required by the Habitat Suitability Modelling software, MaxEnt. Users must provide a polygon study site boundary and the environmental GIS data they wish to create environmental variables with (e.g. landcover or terrain data). To prevent boundary effects and erroneous results, these **environmental data must be provided for the entire study site plus a buffer area extending the same distance as the maximum scale at which data will be analysed**. Separate models are provided for measuring different variable types at multiple scales - see the list below to choose the most appropriate models for your analysis. Models are also provided to help with the creation of MaxEnt bias files and masks.

List of models

A_PrepareSnapRaster (required)	This model is a <u>prerequisite to all other models</u> . It creates the snap rasters to which outputs will be matched
B_MultiscaleLandcover	Measures the proportional cover of different habitat, landcover, or other categories at a specified spatial scale
B_MultiscaleLine Density	Measures the density of linear features at a specified spatial scale
B_MultiscalePatch Metrics	Measures a statistic on a user provided patch metric at a specified spatial scale
B_MultiscalePoint Density	Measures the density of point features at a specified spatial scale
C_Distance	Measures Euclidean distance to features
C_Terrain	Creates & prepares DTM, slope and aspect variables
D_MultiScaleRaster Input	Measures a statistic on a user specified raster at a specified spatial scale (e.g. can be used to measure mean altitude or distance with the outputs from the C_ models)
E_NonScalarRasters	Prepares any other user-provided rasters for MaxEnt
F_MakeBiasFile	Enables users to make a bias file with data on sampling effort across the study site
F_MakeMask	Enables users to make masks that can be provided to constrain the extent of the MaxEnt model training area
XSubModel_Don'tRun	This is a component of the C_Terrain model that shouldn't be run separately

Software requirements

This tool requires the following:

- ArcGIS version 10 with and ArcInfo licence
- ArcGIS Spatial Analyst extension enabled.

User requirements

The toolbox has been designed for people who understand the basics of both ArcGIS and MaxEnt. The models are not password protected and users are encouraged to open them, review the modelling process and check for any errors or issues. The outputs are provided in both ASCII and file geodatabase raster format and should be checked before attempting to use them in MaxEnt. The authors accept no responsibility for any errors or misuse of the toolkit.

Installation

1) Download the folder and unzip the entire contents to your computer

Copy the “Data” folder onto a location on your computer where you will store the inputs and outputs of the models. Do not alter the folder structure inside this folder. Making changes to the folders or geodatabases (apart from adding input data) will alter the pathways to the inputs and outputs directories, preventing the tools from working. The names of the data layers that are produced can be changed, apart from those outputs stored in the SnapRasters.gdb, as all other models rely on these data and locate them by name.

2) Enable the spatial analyst extension

This can be done via the Customize menu bar in ArcMap or ArcCatalog. Go to “Extensions” and tick the Spatial Analyst option.

3) Allow outputs to be overwritten

Go to the Geoprocessing menu bar in ArcMap or ArcCatalog, select “Geoprocessing Options” and tick “Overwrite the outputs of geoprocessing operations”. The models are designed to name outputs with information on how they were derived and include information on the scale at which they were measured. If you re-run models using the same inputs and parameters your outputs will be replaced, therefore, change the names of previous outputs if you would like these to be kept.

You may want to change this setting back for other ArcGIS projects.

4) Add the toolbox to ArcMap or navigate to it in ArcCatalog

If you are using ArcMap for your analysis, make sure the ArcToolbox window is turned on. A toolbox can then be added by right clicking in the space of this window

and selecting “Add Toolbox”. Alternatively, toolboxes can be dragged and dropped into the ArcToolbox window from the ArcCatalog window. Toolboxes can also be run from ArcCatalog, by navigating to the tools within the EcoServGIS folder in the Catalog Tree.

Folders & geodatabases

Data folder components

Inputs

NonScalarRasters.gdb

Add any pre-prepared raster environmental variables here and use the “E_NonScalarRasters” model to prepare these for MaxEnt (these will not be analysed at multiple scales)

RasterInputs.gdb

Add any raster environmental variables here that you want to analyse at multiple scales and use the “D_MultiScale_RasterInput” model to analyse and prepare these for MaxEnt

StoreYourOtherDataHere.gdb

You can add any other project data you wish to keep in the project folder e.g. your study site boundary or terrain data inputs

VectorInputs.gdb

Add any vector data here that you want to analyse at multiple scales and choose from the “B_” or “C_Distance” models according to the environmental variable you would like to produce

Outputs

Bias_Masks

Any bias files or masks produced with the “F_” models will be saved here in ASCII format

MaxEntEGVS

The environmental variables produced will be stored here in ASCII format, ready to be used in MaxEnt

Intermediates.gdb

Some intermediate data are stored here that you may want to review

Rasters.gdb

The environmental variables produced will be stored here. These have the same names as the equivalent ASCII versions stored in the “MaxEntEGVS” folder, allowing you to view and check the outputs easily in GIS

Scratch.gdb

Intermediate data are stored here, but will only be kept if models are run from the model edit window

SnapRasters.gdb

Snap rasters produced from the “A_PrepareSnapRasters” model are stored here. Do not move, rename or delete these as they are required for all models to run.

Model instructions

A PrepareSnapRasters (required)

This model is a prerequisite to all other models in the toolkit. It prepares the snap rasters that will be used to ensure that all environmental variables match in their cell size, geographic extent and coordinate system. Two snap rasters and a buffer outline will be added to SnapRasters.gdb.

Inputs & parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Study site boundary**
Select your study site boundary data. A buffer will be added to this by the model, therefore, environmental data provided in later models must cover the study area plus this buffer (see below)
- 3) **Maximum analysis scale (m)**
The maximum scale size at which variables will be measured. This defines the buffer distance which is added to the study site boundary, preventing boundary effects (erroneous results around the site boundary)
- 4) **Cell size**
The output cell size for all environmental variables created. This must be informed by the locational accuracy of the species presence data and the resolution of the raw environmental data inputs. All outputs will be produced at this resolution as MaxEnt does not allow for variation. Increasing the resolution will increase environmental variable production time and MaxEnt runtime
- 5) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

B MultiScaleLandcover (optional)

Before running this model, export individual landcover/habitat/other type polygon feature classes to VectorInputs.gdb. This model converts these to rasters at the resolution specified. It then uses focal statistics to calculate the percentage of the analysis scale window that each landcover type covers around every raster cell. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The model can only be used for measuring environmental variables at a single scale at a time.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select feature classes for analysis by their name (e.g. *Wood). If left blank, all polygons stored in VectorInputs.gdb will be used for analysis
- 3) **Cell size**
Provide the cell size you're working with – this must be the same as the StudySiteMask selected in the "A_PrepareSnapRasters" model
- 4) **Scale**
The distance (m) around each cell within which landcover type cover will be measured. This is used to label the outputs
- 5) **Analysis window (scale)**
As above - the distance (m) around each cell within which landcover type cover will be measured. **This must be the same as the scale value and remain as a circle window in map units**
- 6) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

B MultiScaleLineDensity (optional)

Before running this model, export individual line type feature classes to VectorInputs.gdb. This model calculates line density (line length/analysis scale area) within the circular analysis scale window around every raster cell. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The model can only be run at a single scale at a time.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select feature classes for analysis by their name (e.g. *WoodEdge). If left blank, all line features stored in VectorInputs.gdb will be used for analysis.
- 3) **Cell size**
Provide the cell size you're working with – this must be the same as the StudySiteMask selected in the "A_PrepareSnapRasters" model
- 4) **Scale**
The distance (m) around each cell within which line density will be measured. This is also used to label the outputs
- 5) **Area units**
The area units of line density measurements
- 6) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

B MultiScalePatchMetrics (optional)

Before running this model, export input landcover/habitat type feature classes to VectorInputs.gdb before ensuring that the required patch metrics have been already calculated and added as attribute fields (e.g. shape area or shape index). This model uses focal statistics to calculate a statistic (e.g. maximum), on a patch metric within an analysis scale window, around every raster cell (_raw output) and standardises these by dividing the raw output by the analysis window area used (_standard output). This model could also be used to measure a statistic on other data e.g. the variety of habitat types may give an indication of habitat richness. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The model can only be run at a single scale at a time.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select feature classes for analysis by their name (e.g. *Wood). If left blank, all features stored in VectorInputs.gdb will be selected according to any feature type specifications (option below)
- 3) **Feature type (optional)**
Use this to select feature classes for analysis by their type. If left blank, all features stored in VectorInputs.gdb will be selected according to any wildcard specifications (option above)
- 4) **Cell size**
Provide the cell size you're working with – this must be the same as the StudySiteMask selected in the "A_PrepareSnapRasters" model
- 5) **Value field**
The patch metric value field to be used
- 6) **Statistics type**
The statistic that will be calculated on the patch metric within the analysis scale window. It is important to first consider whether a statistic provides useful and valid information (e.g. the mean of a categorical metric may not be valid)
- 7) **Scale**
The distance (m) around each cell within which the statistic will be measured. This is also used to label the outputs

8) Analysis window (scale)

As above - the distance (m) around each cell within which the statistic will be measured. **This must be the same as the scale value and remain as a circle window in map units**

9) Output coordinate system

This must be a Projected Coordinate System that matches the species data coordinates

B MultiScalePointDensity (optional)

Before running this model, export individual point type feature classes to VectorInputs.gdb. This model calculates point density (number of points/analysis scale area) within the circular analysis scale window around every raster cell. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The model can only be run at a single scale at a time.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select feature classes for analysis by their name (e.g. *WoodCentroid). If left blank, all point features stored in VectorInputs.gdb will be used for analysis
- 3) **Cell size**
Provide the cell size you're working with – this must be the same as the StudySiteMask selected in the "A_PrepareSnapRasters" model
- 4) **Scale**
The distance (m) around each cell within which point density will be measured. This is also used to label the outputs
- 5) **Analysis window (scale)**
As above - the distance (m) around each cell within which point density will be measured. **This must be the same as the scale value and remain as a circle window in map units**
- 6) **Area units**
The area units that point density measures are made with
- 7) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

C Distance (optional)

Before running this model, export the required feature classes to VectorInputs.gdb. This model uses the Euclidean distance tool to measure distance to each feature across the study site. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The raster outputs are also added to RastersInputs.gdb (with a "Distance" prefix) in case users want to measure a statistic (e.g. mean) on these distance variables at multiple scales, using the MultiScaleRasterInput model.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select feature classes for analysis by their name (e.g. *WoodEdge). If left blank, all features stored in VectorInputs.gdb will be used for analysis
- 3) **Feature type (optional)**
Use this to select the type of feature classes for analysis (e.g. LINE). If left blank, all features types will be used
- 4) **Maximum distance (optional)**
The maximum distance at which distance will be measured (m)
- 5) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

C Terrain (optional)

This model uses a user-provided digital terrain model to map altitude, slope and aspect across the study site. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The raster outputs are also added to RastersInputs.gdb (with a “DTM” prefix) in case users want to measure a statistic (e.g. mean) on these terrain variables at multiple scales, using the MultiScaleRasterInput model.

Parameters & inputs:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Digital terrain model**
Select your DTM. It should have a cell size that is equal or smaller to the cell size chosen for the snap rasters and data must be provided for an area equal to or larger than the study site buffer
- 3) **Slope measurement**
Chose how you would like to measure slope
- 4) **Method used to reclass aspect into categories**
The continuous aspect raster is classed in to categories, to avoid the issues associated with using circular variables. The default method creates 9 classes that indicate flat areas and N, NE, E, SE, S, SW, W, NW aspects
- 5) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

D MultiScale RasterInput (optional)

Before running this model, export rasters you wish to measure at multiple scales to the RastersInput.gdb (terrain and distance rasters will be added here automatically if you have run the C_ models already). This model uses focal statistics to calculate a statistic (e.g. maximum) on these rasters within the analysis scale window around every raster cell (_raw output) and standardises these by dividing the raw output by the analysis window area used (_standard output). Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling. The model can only be run at a single scale at a time.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select rasters for analysis by their name (e.g. Distance*). If left blank, all rasters stored in RasterInputs.gdb will be selected for analysis
- 3) **Statistics type**
The statistic that will be calculated within the analysis scale window. It is important to first consider whether a statistic provides useful and valid information (e.g. the mean of a categorical metric may not be valid)
- 4) **Scale**
The distance (m) around each cell within which the statistic will be measured. This is also used to label the outputs
- 5) **Analysis window (scale)**
As above - the distance (m) around each cell within which the statistic will be measured. **This must be the same as the scale value and remain as a circle window in map units**
- 6) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

D NonScalarRasters (optional)

If you have other raster environmental variables you would like to prepare for MaxEnt and ensure they match in their extent etc., save them to NonScalarRasters.gdb. This model will match these to the extent and resolution of the snap rasters and replace any NoData values. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling.

Parameters:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Wildcard (optional)**
Use this to select rasters for analysis by their name (e.g. Landcover). If left blank, all rasters stored in NonScalarRasters.gdb will be selected for analysis.
- 3) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

F Make BiasFile points (optional)

Use this model if you have point data (e.g. GPS trail) detailing sampling effort across the study site with which you would like create a bias file. Information on the sampling effort at each point must be provided e.g. number of visits or amount of time spent surveying. This process will sum sampling effort at each raster cell to provide a total measure of sampling effort. MaxEnt uses this information to factor out any bias introduced by sampling certain environments more or less often than they were available within the study site. See Dudik, Schapire & Phillips (2005) for more information. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling.

Parameters & data inputs:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Cell size**
Provide the cell size you'd like the outputs to be in – this should be the same as the StudySiteMask
- 3) **Sampling points**
The point location of sample sites from which species data were recorded. The radius over which species were recorded from these points should be smaller than or equal to the cell size
- 4) **Sample effort field**
Provide a field in the sampling points attribute table describing the sampling effort at each point location (e.g. time spent recording). If this effort was uniform, populate this field with a constant value of 1
- 5) **Cell assignment type**
Choose a method which will determine how raster cell values are assigned if sample point locations fall within the same raster cell, or separate sampling effort data are provided for multiple visits to the same site. Summing these values is the default option
- 6) **Background sample effort value**
Non-sampled areas cannot be assigned a value of zero, therefore, a small decimal value must be assigned to these cells. The default value is 10^{-6}
- 7) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

F Make BiasFile polygon (optional)

Use this model if you have polygon data showing sampling effort across the study site (e.g. buffers around survey platforms within which species locations were recorded) with which you would like create a bias file. Information on the sampling effort must be provided e.g. number of visits or amount of time spent surveying at a location. Please note that this model does not account for spatial overlapping survey polygons and so sampling effort should already be summed in the sampling effort field where overlap occurs. MaxEnt uses this information to factor out any bias introduced by sampling certain environments more or less often than they were available within the study site. See Dudik, Schapire & Phillips (2005) for more information. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling.

Parameters & data inputs:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Cell size**
Provide the cell size you'd like the outputs to be in – this should be the same as the StudySiteMask
- 3) **Sampling areas**
The polygon sample sites from which species data were recorded
- 4) **Sample effort field**
Provide a field in the sampling points attribute table describing the sampling effort at each point location (e.g. time spent recording). If this effort was uniform, populate this field with a constant value of 1.
- 5) **Background sample effort value**
Non-sampled areas cannot be assigned a value of zero, therefore, a small decimal value must be assigned to these cells. The default value is 10^{-6}
- 6) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

F Make Masks (optional)

Use this model if you want to produce a mask that constrains the area which MaxEnt uses to train the model. A separate mask is produced that then allows the models to be projected to the entire study site. Raster outputs are added to Rasters.gdb for viewing and checking. Equivalent ASCII files are added to the MaxEntEGVs folder, ready for MaxEnt modelling.

NB. Once you have created these masks you will have to follow MaxEnt guidance on their use. They need to be copied to separate training (mask_constrain) and projecting (mask_project) environmental variable folders and renamed with the same title before MaxEnt modelling.

Parameters & data inputs:

- 1) **Data folder**
Navigate to where the data folder is stored on your computer
- 2) **Model training area**
Select a feature that delineates the area which you would like to constrain model training to
- 3) **Cell size**
Provide the cell size you'd like the outputs to be in – this should be the same as the StudySiteMask
- 4) **Output coordinate system**
This must be a Projected Coordinate System that matches the species data coordinates

References

Bellamy, C., Scott C., & Altringham, J.D. (2013) Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species. *Journal of Applied Ecology* 50 (892-901)

Dudik, M., Schapire, R.E. & Phillips, S.J. (2005) Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems*, 18 (eds Y. Weiss, B. Scholkopf & J. Platt), pp. 323–330. MIT Press, Cambridge, Massachusetts, USA.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.