**megaSDM Setup Instructions**

## Required and Optional Data:

To run this program, the following data and programs are required:

1. A ***.csv*** file listing the species to be used in the analysis with two columns:
    a. Column 1: The higher-order taxon of each species to be used: this can be any taxonomic level, but class or order is recommended
        i. If occurrence points need to be download from GBIF, the high-order taxon should follow the GBIF Backbone Taxonomy (e.g.: Soricomorpha instead of Eulipotyphla). https://www.gbif.org/species/search
    b. Column 2: The scientific name of each species to be used

| | A | B |
|---|---|---|
| 1 | Taxon | Scientific.Name |
| 2 | Mammalia | Taxidea taxus |
| 3 | Mammalia | Cervus elaphus |
| 4 | Mammalia | Martes americana |

2. Environmental data in raster form (for example ***.bil***, ***.asc***, or ***.tif***) for all desired time periods and climate scenarios
    a. These data must be projected, but any projection is valid (see steps 7 and 8).
    b. All climate variables must be named consistently across time-periods and scenarios
    c. If categorical environmental variables are used (e.g., soil type, landforms), mark those rasters with a distinguishing prefix:
        i. Ex.: *Categ_landforms.bil* has the distinguishing prefix "Categ"
3. RStudio (version 1.1 or higher)
4. The executable Java script ***maxent.jar*** (can be found on GitHub at https://github.com/mrmaxent/Maxent)
5. The configuration file accompanying this package (***config.txt***)
    a. ***Config.txt*** can be downloaded using the getConfig function or by downloading it from https://github.com/brshipley/megaSDM/blob/master/config.txt.

If available, additional data may be provided by the user:

1. Occurrence points for each species in a ***.csv*** file with three columns:
    a. The name of the species repeated
    b. The longitude of each occurrence
    c. The latitude of each occurrence
    d. Named with the scientific name of the species
        i. Ex.: *Cervus_elaphus.csv*

| | A | B | C |
|---|---|---|---|
| 1 | Species | Longitude | Latitude |
| 2 | Cervus elaphus | -83.118639 | 35.648375 |
| 3 | Cervus elaphus | -105.935073 | 32.880342 |
| 4 | Cervus elaphus | -110.846022 | 43.500302 |

2. A buffer around the occurrence points (in ***.shp*** or raster form) to preferentially generate background points close to the occurrence points
    a. Named with the scientific name of the species
        i. Ex.: *Cervus_elaphus.shp*
3. Background points (in the same format and coordinate system as the occurrence points)
    a. 3 columns: species name and coordinates
    b. Named "species*_background_X.csv*", where X is 1 → number of replicates wanted
        i. Ex.: *Cervus_elaphus_background_2.csv*
4. The dispersal rates for each species in a ***.csv*** file with two columns:

    a. List of species
    b. Dispersal Rate (km/year)
5. Protected areas shapefile polygon(s) (***.shp***) detailing protected areas within the study region and (if available) other time periods
    a. All protected areas at each time period examined must be in a single shapefile
6. Binary urban data raster(s) for the study area and (if available) other time periods
    a. 0 = non-urbanized
    b. 1 = urbanized
    c. If future/projected urbanized files are used, name all urban files with the time period the layer should be applied to
        i. Ex.: *urb_2070.bil*

## Setup, Configuration, and File Management:

To begin the setup process, download all ServalDENS files and fill out ***config.txt*** with the desired parameters:

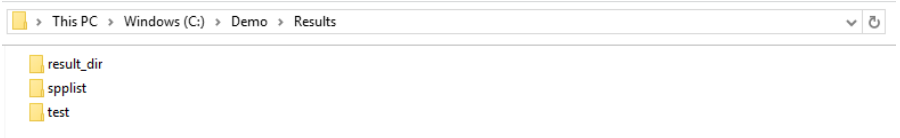
1. Fill out the 1st section of *config.txt* (Data Directory) with directory paths and create each directory and subdirectory within ***DataDirectory***
    a. Ex.: DataDirectory = "C:/Demo/Data"
    b. Ex.: scripts = "/scripts"

    After Step 1, this is what ***DataDirectory*** should look like:


2. Place all scripts and the *config.txt* into the ***scripts*** subdirectory

3. Fill out the 2nd section of *config.txt* (Trial Directory) with directory paths and create each directory and subdirectory within ***TrialDirectory***
    a. Ex.: TrialDirectory = "C:/Demo/Results"
    b. Ex.: result_dir = "/result_dir"

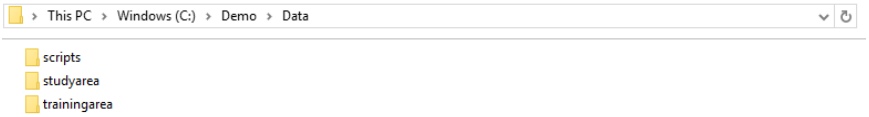
    After Step 3, this is what ***TrialDirectory*** should look like:




4. Move the list of species to be used in the analysis to the path given by ***spplist***

5. Fill out the 3rd section of *config.txt* (Environmental Data Management) and create each required subdirectory within ***DataDirectory***
    a. Ex.: trainingarea = "/trainingarea"

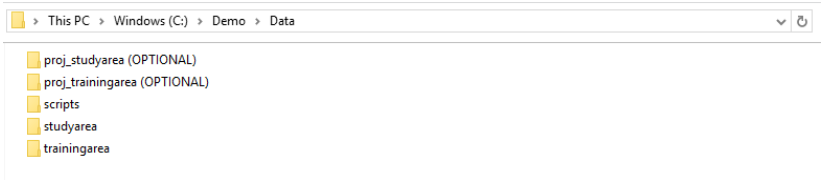
    After Step 5, this is what ***DataDirectory*** should look like:




6. Move all environmental data into the correct directories
    a. Training Area Data are environmental data for the region where the model coefficients will be generated (all occurrence and background points will be inside this region)
    b. Study Area Data are environmental data for the region of interest (i.e. where the model will be applied to)

7. Fill out the 4th section of *config.txt* (GIS Layer Projection) with the desired parameters
    a. Put CoordinateProjectionStep = “N” if environmental rasters are already in the desired coordinate reference system and units
    b. If CoordinateProjectionStep = “Y”, create the required subdirectories within ***DataDirectory***

    After Step 7, this is what ***DataDirectory*** should look like:

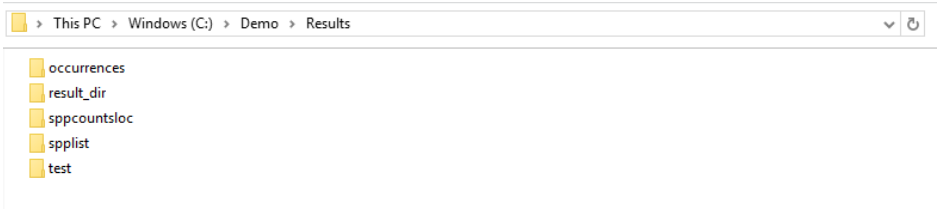



8. Fill out the 5th section of *config.txt* (Coordinate Systems) with the desired Coordinate Reference System (desiredCRS) and the Coordinate Reference System of the **occurrence data** (defaultCRS)
    a. CRS definitions should be in PROJ4 form; an overview of common coordinate systems may be found at https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/OverviewCoordinateReferenceSystems.pdf

9. Fill out the 6th section of *config.txt* (Raster Clipping and Resampling) with the desired parameters
    a. Put ClipEnvDataStep = "N" if clipping the environmental rasters to a different extent is not necessary.
    b. Put TrainingAreaClip = “Y” if provided data have a larger extent than the desired training area
        i. TrainClipLatitude = minlatitude, maxlatitude
        ii. TrainClipLongitude = minlongitude, maxlongitude
    c. If categorical data are used, set Categorical to the distinguishing prefix
        i. Ex.: make all categorical rasters begin with “Categ”, set Categorical = “Categ”

10. Fill out the 7th section of *config.txt* (Occurrence File Management) and create the required subdirectories in ***TrialDirectory***

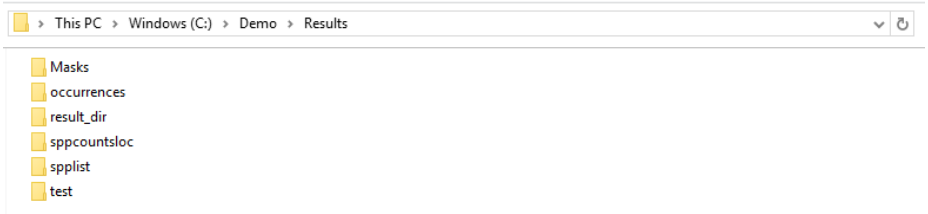
    After Step 10, this is what ***TrialDirectory*** should look like:




11. Fill out the 8th section of *config.txt* (Downloading GBIF Occurrences) with the desired parameters
    a. Put gbifstep = "N" if occurrence files have already been generated for each species
    b. Geographic extent should be provided in min, max form
        i. Ex.: decimalLatitude = 20,40
        ii. Ex.: decimalLongitude = -57, -30

12. Fill out the 9th section of *config.txt* (Generating Background Points) with the desired parameters
    a. Put backgroundPointsStep = "N" if background points have already been generated for each species
        i. Create a directory within ***test*** called "backgrounds"
        ii. Place all background points within the "backgrounds" directory
    b. If backgroundPointsStep = "Y", background points will be generated
        i. Put speciesBufferStep = "Y" if sampling background points within a buffer is desired
        ii. If speciesBufferStep = "Y", create ***buff_dir*** in ***TrialDirectory***

    After Step 12, this is what ***TrialDirectory*** should look like (mcp_dir = "/Masks"):




13. Fill out the 10th section of *config.txt* (Environmental Subsampling) with the desired parameters
    a. Recommended parameter values are provided in the configuration file.

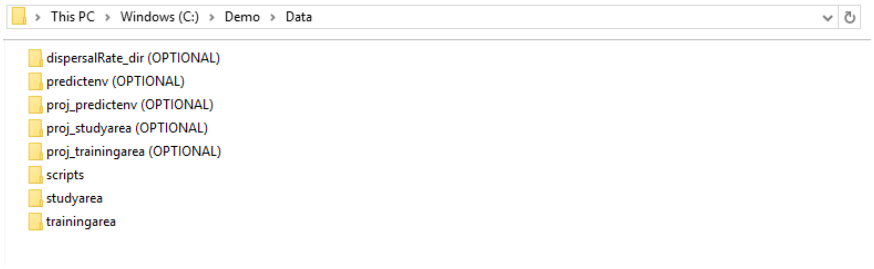
14. Fill out the 11th section of *config.txt* (MaxEnt) with the desired parameters
    a. Recommended parameter values are provided in the configuration file.

15. Fill out the 12th section of *config.txt* (Hindcasting/Forecasting) with the desired parameters
    a. If forecasting/hindcasting is desired, create subdirectories within ***DataDirectory***
        i. These directories must be formatted as:
        ii. ***DataDirectory/predictenv***-->scenario folder-->date folder-->environmental files
        iii. Ex.: *C:/Demo/Data/predictenv/scenario/year/file1.bil*

16. Fill out the 13th section of *config.txt* (Dispersal Rate) with the desired parameters

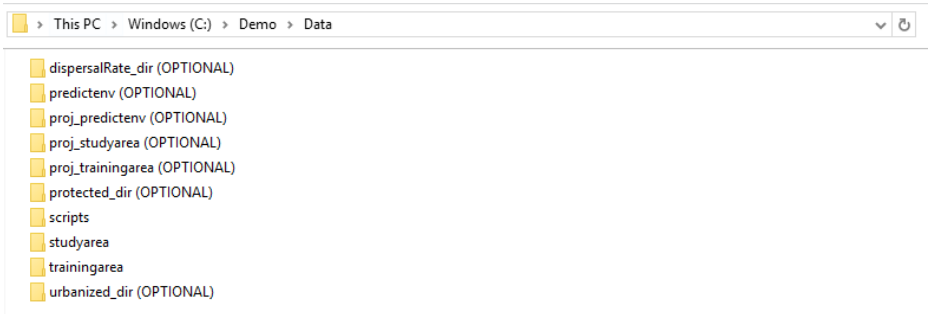
a. If dispersal rate is desired (dispersalStep = "Y"), create subdirectory ***dispersalRate_dir*** within ***DataDirectory*** and copy the dispersal rate csv file into ***dispersalRate_dir***

After Step 16, this is what ***DataDirectory*** should look like:




17. Fill out the 14th section of *config.txt* (Species Richness) with the desired parameters
    a. Put RichnessStep = "Y" if richness maps for all modelled species are desired

18. Fill out the 15th and 16th sections of *config.txt* (Urban Analysis & Protected Area Analysis)
    a. Create the subdirectories within ***DataDirectory***
    After Step 19, this is what ***DataDirectory*** should look like:




19. Copy ***maxent.jar*** into the ***occurrences*** subfolder

20. Begin running ***megaSDM_run.R***

21. When prompted ((file.choose()) in ***megaSDM_run.R***), navigate to and select ***megaSDM_run.R***

22. Once the script has entirely run, delete or move all result and test files if re-runs are necessary or desired.