A short tutorial for creating a species distribution model using QGIS, R, and MaxEnt.

Edit

#species-distribution-modelling #maxent #qgis #tutorial #nus-ape-lab Manage topics

50 commits | **1** branch | **0** releases | **1** contributor

Branch: **master** | New pull request | Create new file | Upload files | Find file | Clone or download

dondealban Added references in further reading section — Latest commit 74c7b6a 6 minutes ago

poster	Updloaded study area image; edited data prep section	2 days ago
screenshots	Edited sdm section	2 days ago
README.md	Added references in further reading section	6 minutes ago
README.pdf	Tested markdown to pdf compile using pandoc	3 days ago
Select.Species.R	Uploaded R script for selecting species	2 days ago
sdm_logistic.qml	Edited sdm section	2 days ago

README.md

A Short Species Distribution Modeling Tutorial

This repository contains a short tutorial for creating a species distribution model using QGIS, R, and MaxEnt. I prepared this documentation for the skills training sessions during the lab retreat of the Applied Plant Ecology Lab, Department of Biological Sciences, National University of Singapore held on 25-28 September 2017 in Malacca, Malaysia.

Table of Contents

Download and Installation

Software

For this tutorial, download and install QGIS, R, and MaxEnt, all of which are free and open-source software. For QGIS and R, download the versions compatible with your machine's operating system. MaxEnt is a Java-based application and runs using various operating systems. The procedures shown in this tutorial uses a Mac OSX platform but it should be applicable to other operating systems.

Data

MaxEnt will require two types of input datasets:

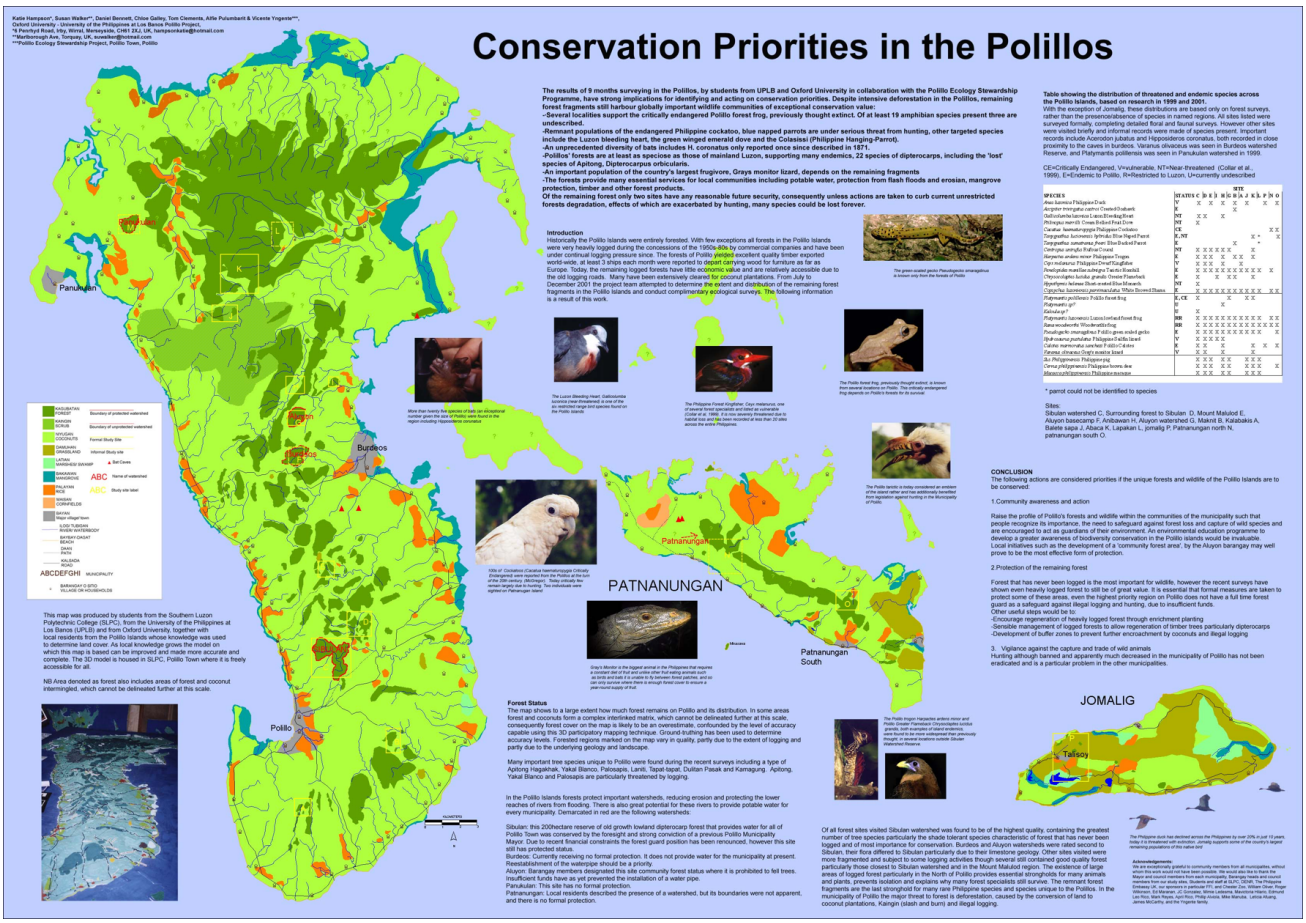
1. **Species occurrence data.** The species occurrence records are the geographic point locations or coordinates of species observations. For this exercise, we will use the georeferenced database of selected threatened forest tree species in the Philippines compiled by Ramos et al. (2011). Download the database from the World Agroforestry Centre/ICRAF Dataverse here (150 KB, CSV file).

2. **Environmental predictors.** The environmental covariates consist of raster data that contain either continuous or categorical values such as precipitation, temperature, elevation, etc. We will be using the WorldClim raster datasets. WorldClim is a set of gridded global climate data layers, which can be used for mapping and ecological modeling. For this exercise, we will use WorldClim v.1.4 Current conditions (or interpolations of observed data from 1960-1990). We will need the highest resolution data available provided at 30 arc-seconds (~1 km); hence click the `download by tile` link and choose Tile 210. After clicking the tile, download the GeoTIFF file formats of the Altitude (~2 MB, ZIP file) and Bioclim (~29 MB, ZIP file) layers. You can read Hijmans et al. (2005) for more information about the climate data layers.

To prepare the datasets, we will also need **administrative boundary** data. We can use the administrative boundary vector data from the Global Administrative Database. On GADM's Download page, select "Philippines" and "Shapefile" from the *Country* and *Format* drop-down menus, respectively, and click the download link provided (~22 MB, ZIP file).

Study Area

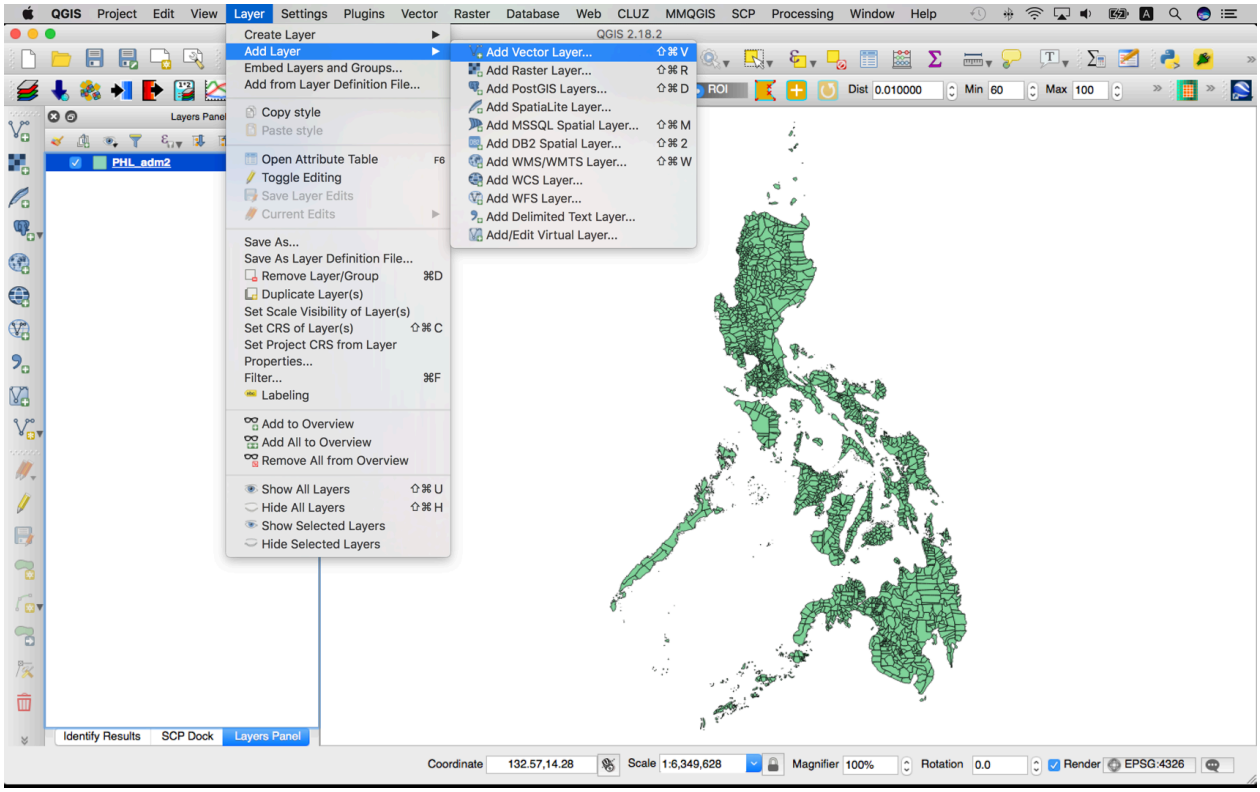
Polillo Islands, Quezon Province, Philippines. The Polillo group of islands (approx. 14.861 N, 122.038 E) is situated in the northeast part of the Philippine archipelago. The Polillos comprise 27 small islands and islets, 25km off the east coast of Luzon. They form part of the Luzon Endemic Bird Area (EBA), ranked sixth in the world listing of critical EBAs (Stattersfield et al. 2000), whilst also constituting a highly distinct sub-centre of endemicity. Amongst these Polillo-specific endemics are a frog, several reptiles and seven birds, including a goshawk, hornbill, and parrots. The islands also support important populations of globally threatened species (e.g. Philippine cockatoo, Gray's monitor lizard, Philippine jade vine), and are accorded high priority in all independent reviews of Philippine conservation priority areas. Open the image below to see more information about the conservation priorities in the Polillo Islands.



Prepare Datasets

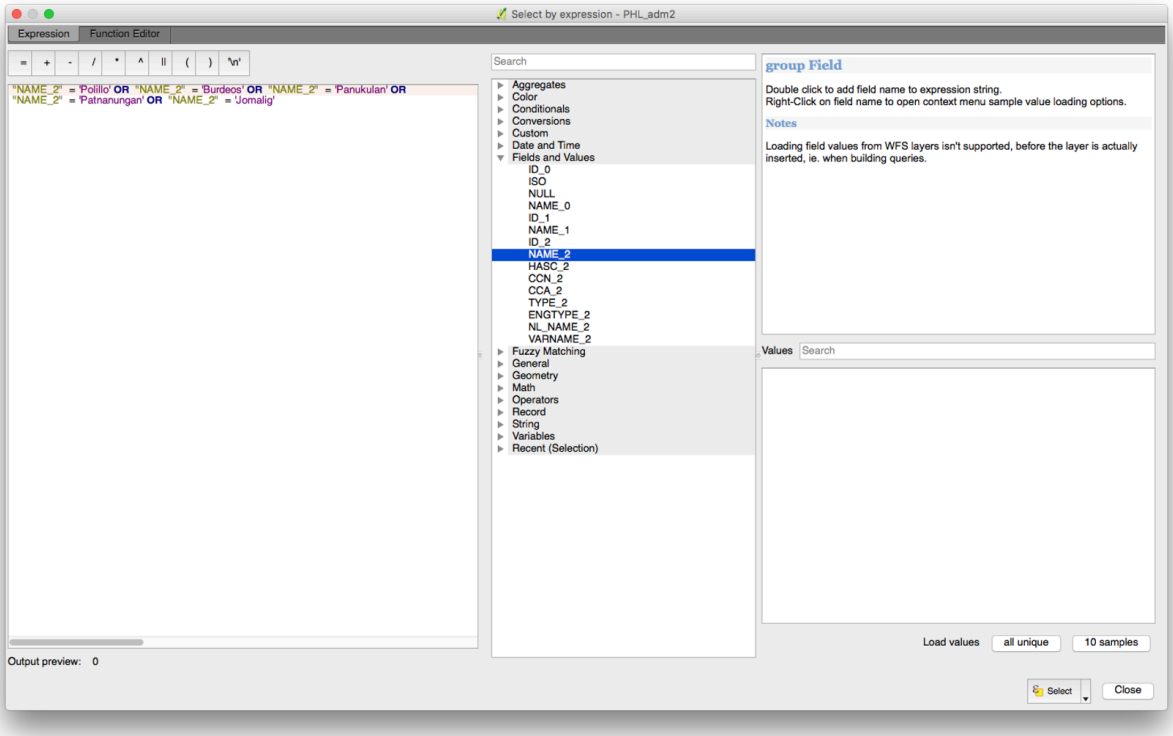
1. First, we will create subsets from the environmental rasters to focus our modeling over our study area. To do this, we will create a polygon shapefile containing the extent of the study area and use this shapefile to clip all the raster map layers. Follow these steps using QGIS:

    - Load the **PHL_adm2.shp** shapefile by adding a vector layer from the `Layer > Add Layer > Add Vector Layer...` menu. This displays the municipal-level administrative boundaries.

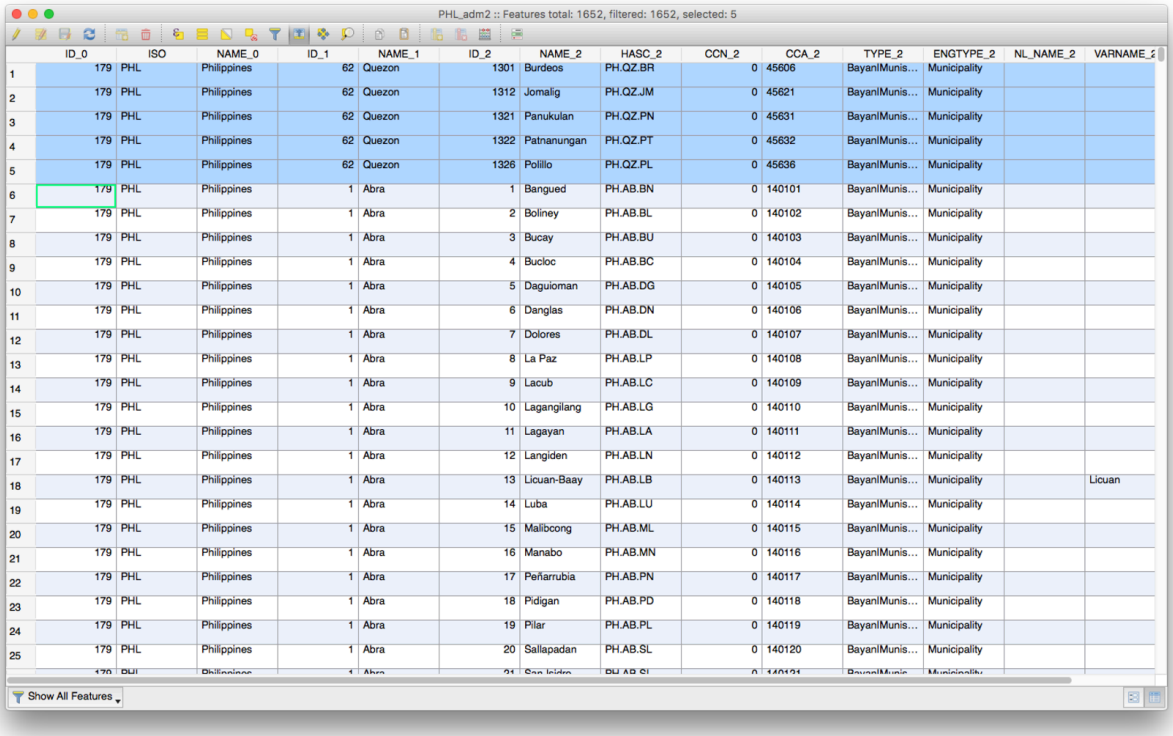


    - To select our areas of interest, we will select the municipalities from the attribute table. Open the attribute table of **PHL_adm2.shp** by right-clicking the shapefile within the `Layers Panel` and then selecting `Open Attribute Table` from the menu. Inside the attribute table window, click the `Select features using an expression` icon. Once the `Select by expression` dialog box opens, enter the following expression:

```
 "NAME_2"  = 'Polillo' OR  "NAME_2"  = 'Burdeos' OR  "NAME_2"  = 'Panukulan' OR  "NAME_2"  = 'Patnanungan'
OR  "NAME_2"  = 'Jomalig'
```



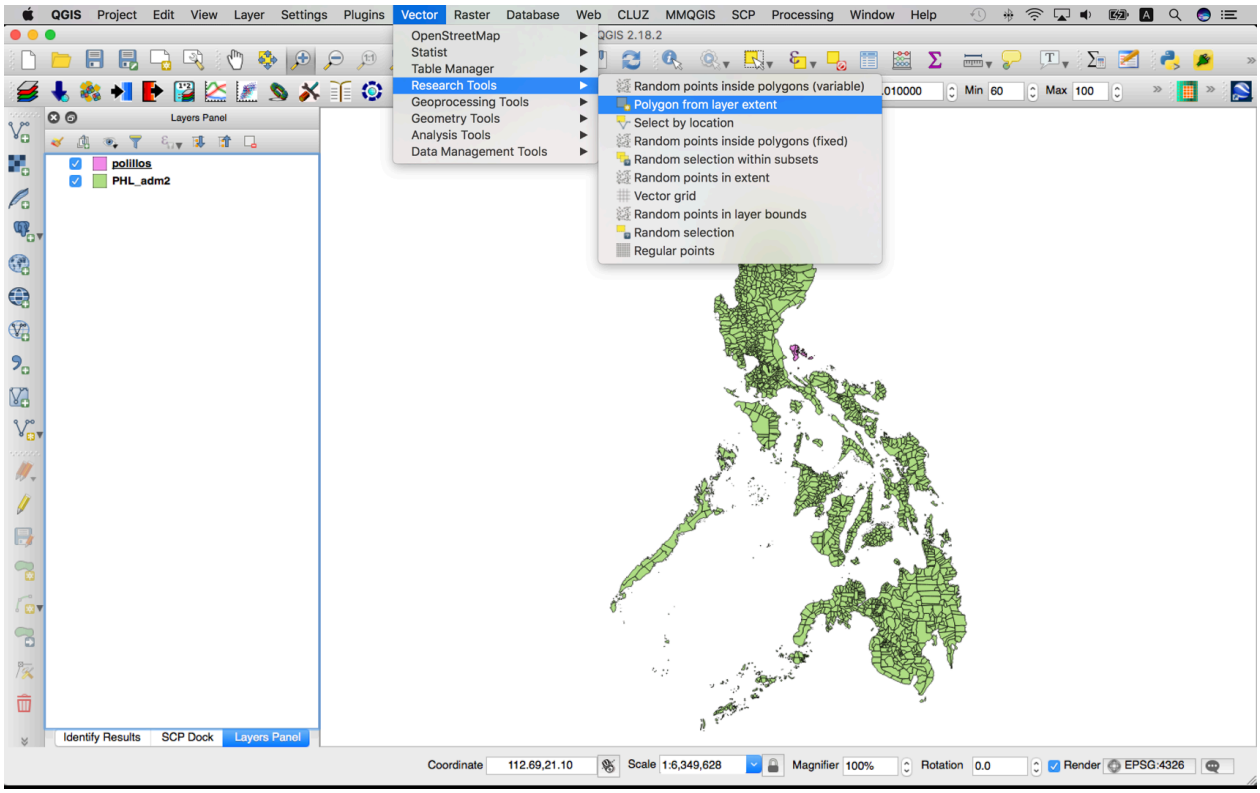
This will select the municipalities belonging to our study area. Check the attribute table if you have selected five records, which includes the following municipalities: Polillo, Burdeos, Panukulan, Patnanungan, and Jomalig towns.



	ID_0	ISO	NAME_0	ID_1	NAME_1	ID_2	NAME_2	HASC_2	CCN_2	CCA_2	TYPE_2	ENGTYPE_2	NL_NAME_2	VARNAME_2
1	179	PHL	Philippines	62	Quezon	1301	Burdeos	PH.QZ.BR	0	45606	BayanlMunis...	Municipality		
2	179	PHL	Philippines	62	Quezon	1312	Jomalig	PH.QZ.JM	0	45621	BayanlMunis...	Municipality		
3	179	PHL	Philippines	62	Quezon	1321	Panukulan	PH.QZ.PN	0	45631	BayanlMunis...	Municipality		
4	179	PHL	Philippines	62	Quezon	1322	Patnanungan	PH.QZ.PT	0	45632	BayanlMunis...	Municipality		
5	179	PHL	Philippines	62	Quezon	1326	Polillo	PH.QZ.PL	0	45636	BayanlMunis...	Municipality		
6	179	PHL	Philippines	1	Abra	1	Bangued	PH.AB.BN	0	140101	BayanlMunis...	Municipality		
7	179	PHL	Philippines	1	Abra	2	Boliney	PH.AB.BL	0	140102	BayanlMunis...	Municipality		
8	179	PHL	Philippines	1	Abra	3	Bucay	PH.AB.BU	0	140103	BayanlMunis...	Municipality		
9	179	PHL	Philippines	1	Abra	4	Bucloc	PH.AB.BC	0	140104	BayanlMunis...	Municipality		
10	179	PHL	Philippines	1	Abra	5	Daguioman	PH.AB.DG	0	140105	BayanlMunis...	Municipality		
11	179	PHL	Philippines	1	Abra	6	Danglas	PH.AB.DN	0	140106	BayanlMunis...	Municipality		
12	179	PHL	Philippines	1	Abra	7	Dolores	PH.AB.DL	0	140107	BayanlMunis...	Municipality		
13	179	PHL	Philippines	1	Abra	8	La Paz	PH.AB.LP	0	140108	BayanlMunis...	Municipality		
14	179	PHL	Philippines	1	Abra	9	Lacub	PH.AB.LC	0	140109	BayanlMunis...	Municipality		
15	179	PHL	Philippines	1	Abra	10	Lagangilang	PH.AB.LG	0	140110	BayanlMunis...	Municipality		
16	179	PHL	Philippines	1	Abra	11	Lagayan	PH.AB.LA	0	140111	BayanlMunis...	Municipality		
17	179	PHL	Philippines	1	Abra	12	Langiden	PH.AB.LN	0	140112	BayanlMunis...	Municipality		
18	179	PHL	Philippines	1	Abra	13	Licuan-Baay	PH.AB.LB	0	140113	BayanlMunis...	Municipality		Licuan
19	179	PHL	Philippines	1	Abra	14	Luba	PH.AB.LU	0	140114	BayanlMunis...	Municipality		
20	179	PHL	Philippines	1	Abra	15	Malibcong	PH.AB.ML	0	140115	BayanlMunis...	Municipality		
21	179	PHL	Philippines	1	Abra	16	Manabo	PH.AB.MN	0	140116	BayanlMunis...	Municipality		
22	179	PHL	Philippines	1	Abra	17	Peñarrubia	PH.AB.PN	0	140117	BayanlMunis...	Municipality		
23	179	PHL	Philippines	1	Abra	18	Pidigan	PH.AB.PD	0	140118	BayanlMunis...	Municipality		
24	179	PHL	Philippines	1	Abra	19	Pilar	PH.AB.PL	0	140119	BayanlMunis...	Municipality		
25	179	PHL	Philippines	1	Abra	20	Sallapadan	PH.AB.SL	0	140120	BayanlMunis...	Municipality		

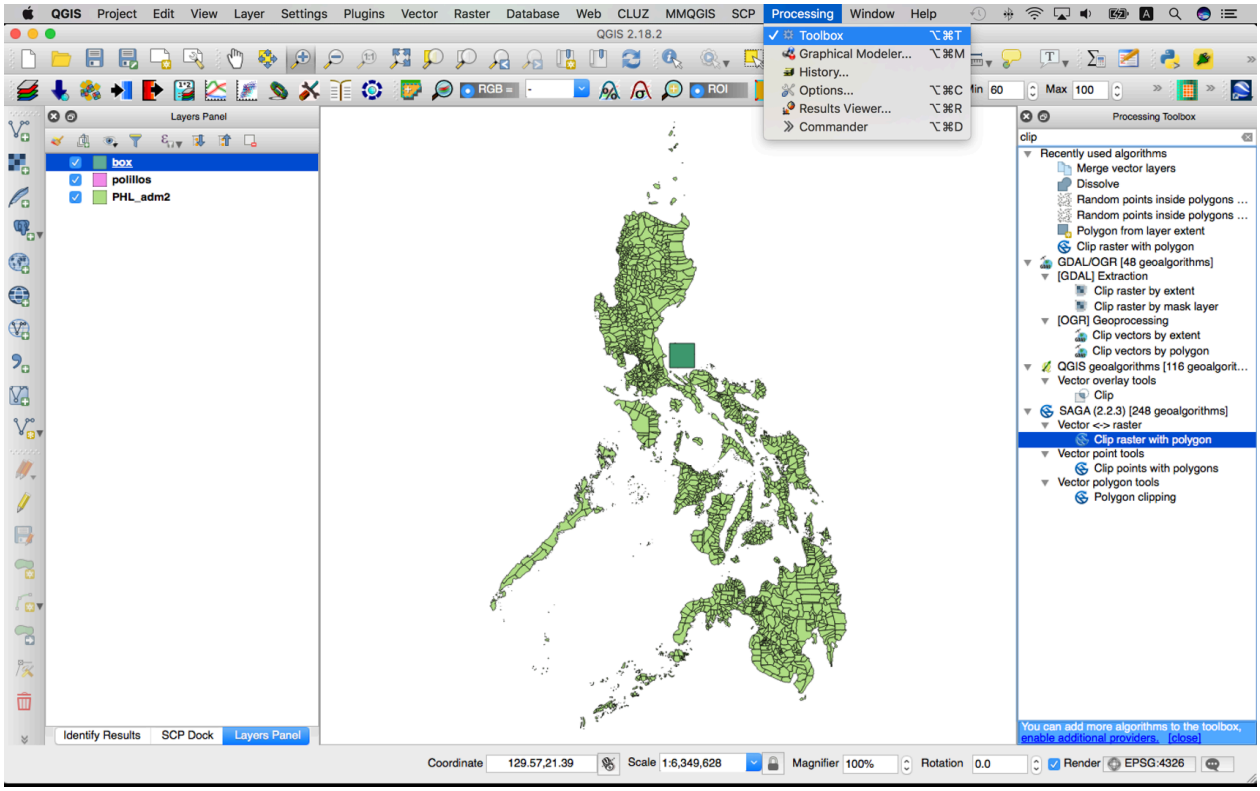
- In the main QGIS window, right-click on **PHL_adm2.shp** and select `Save As...` from the menu. Once the `Save vector layer as...` dialog box opens, tick the `Save only selected features` to ensure that we save a new shapefile containing only the selected municipalities. Then, enter the file name of the output shapefile as **polillos.shp** to your working directory, and click `OK`. The new shapefile should appear in the QGIS `Layers Panel`.
- Next, we will create a polygon from the extent of the municipalities shapefile that we have just saved. Go to `Vector > Research Tools > Polygon from Layer Extent` menu.
  - Under the `Input Layer` drop-down menu, select the newly created **polillos.shp** shapefile.
  - Under the `Extent` input line, select `Save to File` from the menu to save the file in your working directory.

Then, click `Run` to create another shapefile called **box.shp**, which consists of a polygon covering the extent of the study area.



- Next, go to **`Processing > Toolbox`** menu, which opens the `Processing Toolbox` panel. Search for the `Clip raster with polygon` function under the SAGA geoalgorithms and select this function. This will open the **`Clip Raster with Polygon`** dialog box.

    - Under the `Input` drop-down menu, click `...` and navigate through your working directory and select one of the raster layers, say **biol1_210.tif**.
    - Under the `Polygons` input line, select the **box.shp** shapefile from the drop-down menu. The rasters will be clipped using the extent of this polygon.
    - Under the `Clipped` input line, click `...` and select **`Save to File`** from the menu to save the file in your working directory using the same file name, but this time, change the output file type to **ASC** as this is the file type requirement used by MaxEnt. Then, click `Run` to generate the clipped raster file, **biol1_210.asc**.



Repeat this for all other raster layers by following the same process. You may also opt to run this through batch processing by clicking on the `Run As a Batch Process...` button.

2. Next, we will extract the occurrence points of the species we are interested in modeling.

    - Inspect the threatened tree species database using tools like R or Excel. For this exercise, let us model the distributions of `Cinnamomum mercadoi` that were observed in the Polillo Islands.

    - To select the species from the CSV file, we will use a few lines of code in R as follows:

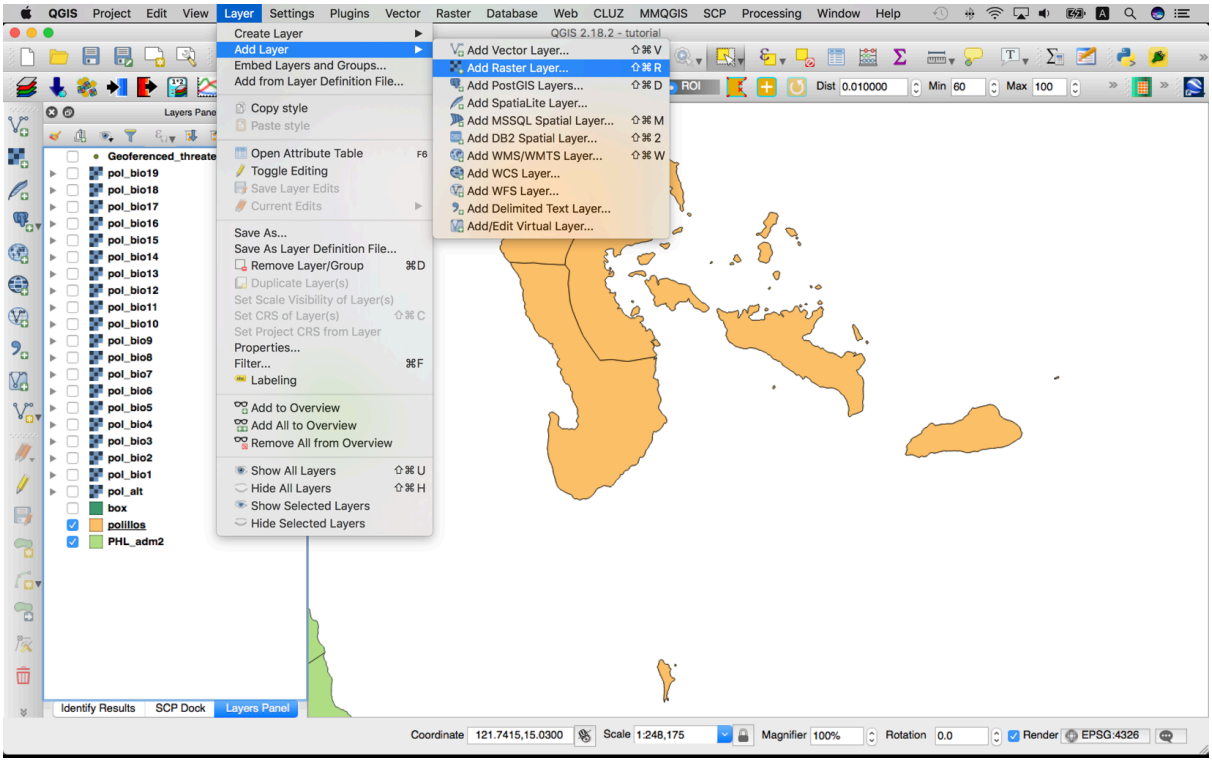
```
# This line reads the CSV file and stores it in a variable
# Note: change file path to your working directory
data <- read.csv(file="Geoferenced_threatenedforesttreespecies.csv", header=TRUE, sep=",")

# This line selects the species in our study area and stores it in a variable
polillo_cm <- subset(data, Species=="Cinnamomum mercadoi" & Source=="Clements, 2001", select=c(2,10:11))

# This line saves the selected species in a CSV file
write.csv(polillo_cm, file="polillo_cm.csv", row.names=FALSE)
```

    Here, note that the search terms used include the species name and the source of the data based on the database. Also, only columns 2, 10, and 11 were selected and saved in the final CSV file, which corresponds to the columns 'Species', 'Lat', and 'Long'.

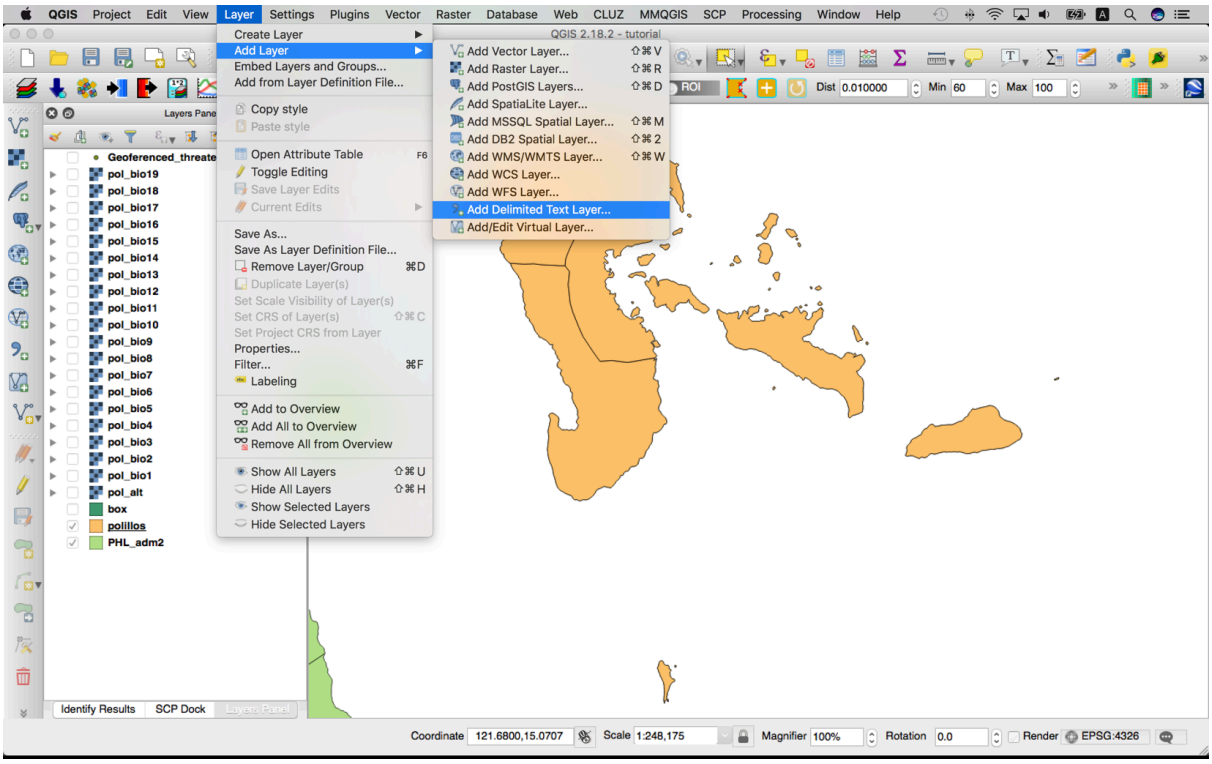
    Alternatively, you can download the R script and run this in R or RStudio.

3. We are almost ready to create our first species distribution model. But before we do that, load all of the clipped environmental rasters and the species occurrence file in QGIS:

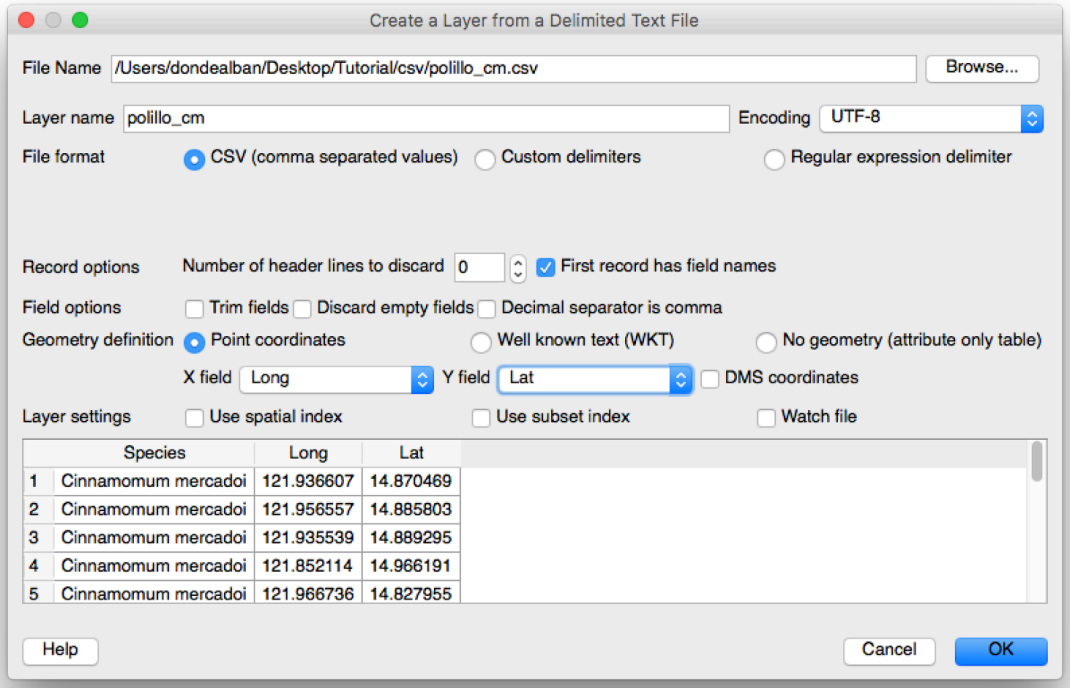
- Load the clipped environmental raster layers by adding them from the `Layer > Add Layer > Add Raster Layer...` menu. Remember that these are the **.ASC** files.



- Load species occurrence data CSV file by adding it from the `Layer > Add Layer > Add Delimited Text Layer...` menu.



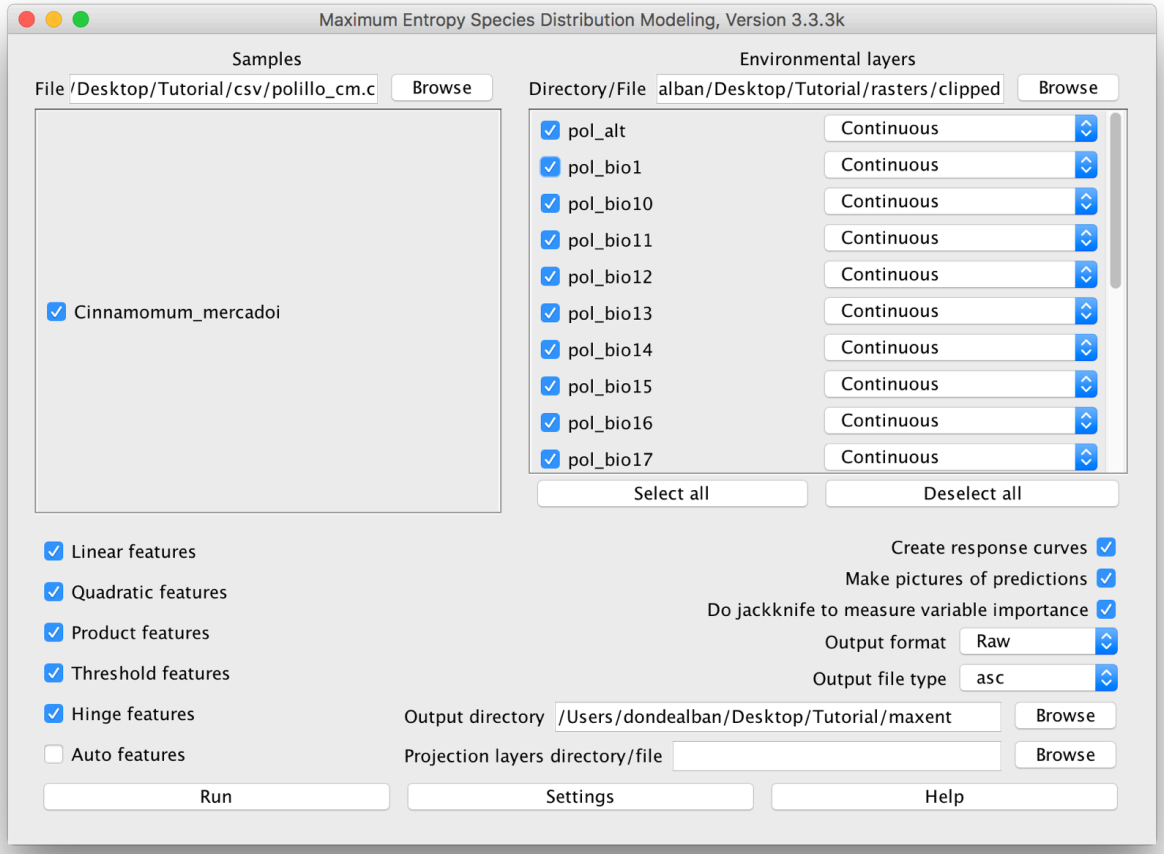
- In the `Create a Layer from a Delimited Text File` dialog box, select the CSV file by navigating to the file in your working directory. Once it is opened, the species records and their coordinates will be shown in the lower part of the dialog box. In the `X field` and `Y field` drop-down menus, select 'Long' and 'Lat' columns, respectively.



Model Species Distributions

We are now ready to create our first species distribution model using MaxEnt.

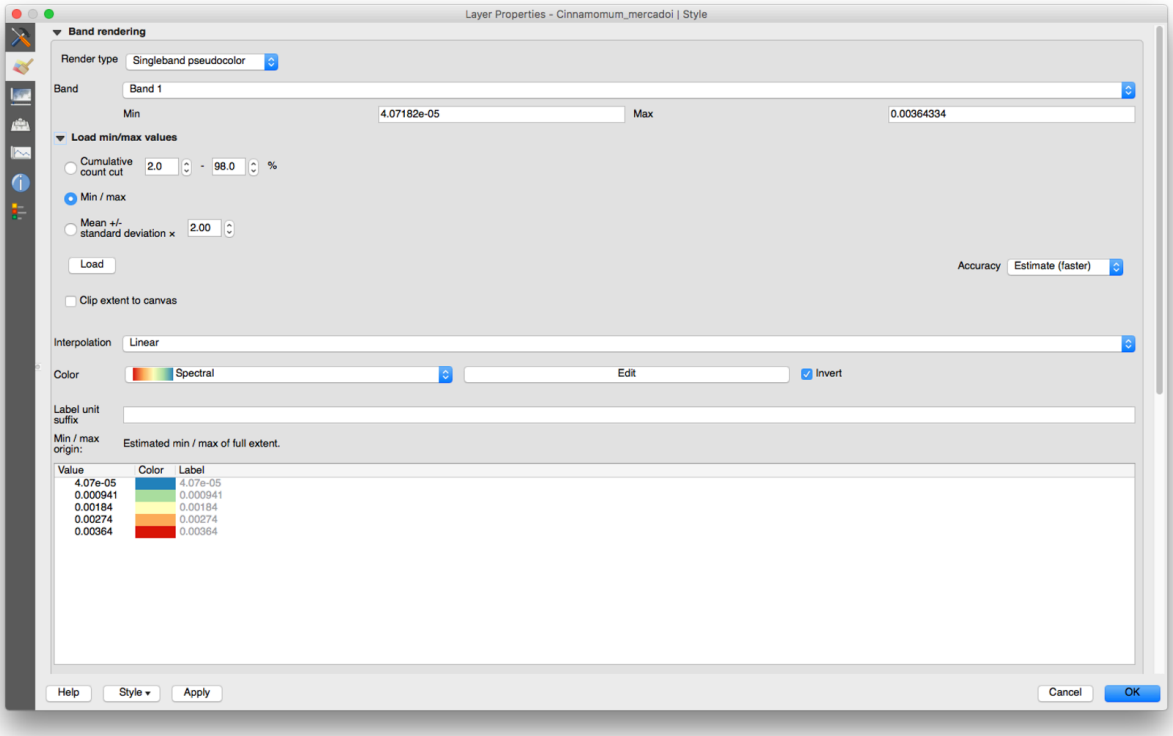
1. Open MaxEnt and load the `Samples` and `Environmental Layers` by navigating to the respective directories of those files. Ensure that the tick boxes of all files are checked, and that the `Environmental Layers` files are all 'Continuous' types.



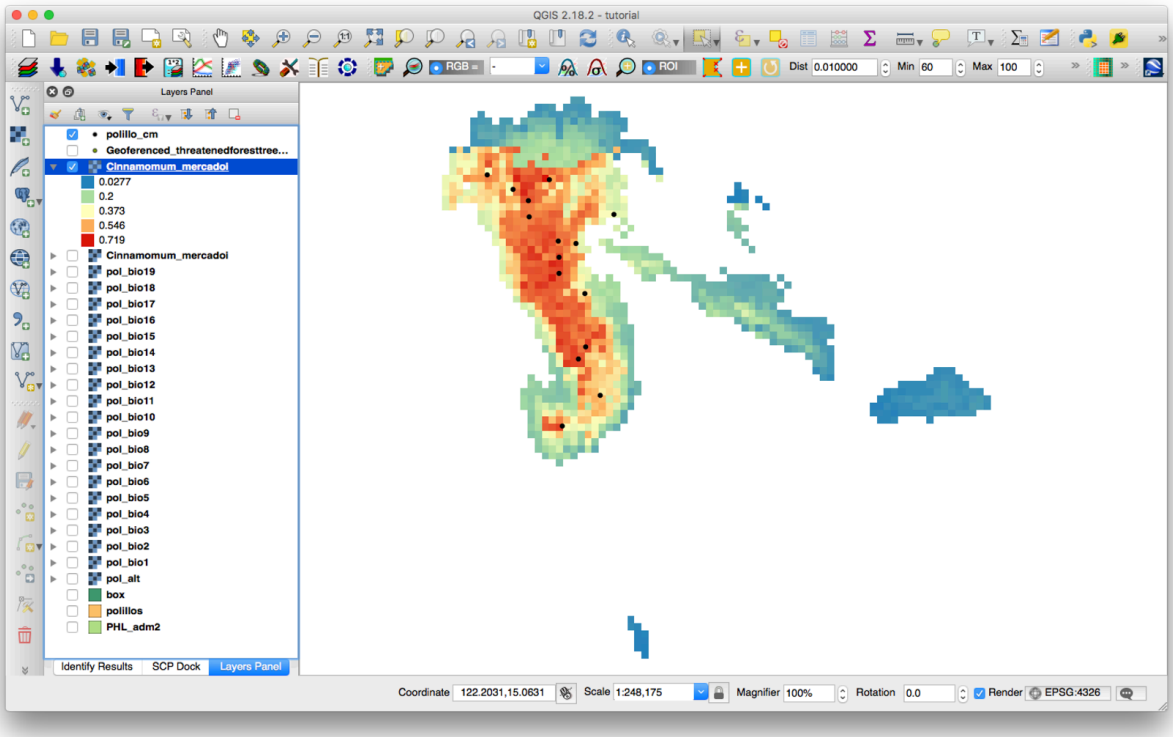
2. Also, in the main MaxEnt window, check tick boxes or select the following options:
    - Linear/Quadratic/Product/Threshold/Hinge features
    - Create response curves
    - Make pictures of predictions
    - Do jackknife to measure variable importance
    - `Output format` : Logistic
    - `Output file type` : asc

Leave the other advanced settings in their default for now. Then, click `Run` and wait for the processing to finish.

3. Once the MaxEnt software completes its data processing:
    - Load the resulting **ASC** file in QGIS from the `Layer > Add Layer > Add Raster Layer...` menu. Then, change the styling of the raster layer by going to `Layer > Properties...` menu, or double-clicking on the layer under the `Layers Panel`. Change the styling of the raster layer as shown on the image below. Alternatively, you can also load the layer styling using this QML file (only applicable to the logistic model output).



The styling of the raster layer has been changed similar to the image below, which shows the logistic output of the MaxEnt's species distribution model.



Congratulations! You have now made your first species distribution model using QGIS, R, and MaxEnt.

Further Reading

To learn more about MaxEnt such as analysing and interpreting MaxEnt's outputs, adjusting model settings, etc. the following materials are suggested for further reading:

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29, 129–151.

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

Morales, N.S., Fernandez, I.C. & Baca-Gonzalez, V. (2017) MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. PeerJ, 5, e3093.

Phillips, S.J., Anderson, R.P., Dudik, M., Schapire, R.E. & Blair, M.E. (2017) Opening the black box: an open-source release of Maxent. Ecography, 40, 887–893.

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

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution, 4, 236–243.

References

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. (DOI)

Ramos, L.T., Torres, A.M., Pulhin, F.B. & Lasco, R.D. (2011) Developing a georeferenced database of selected threatened forest tree species in the Philippines. *Philippine Journal of Science*, 141, 165–177. (PDF)

Stattersfield, A.J., Crosby, M., Long, A.J. & Wege, D.C. (1998) *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. The Burlington Press, Ltd., Cambridge, United Kingdom.

License

Want to Contribute?

In case you wish to contribute or suggest changes, please feel free to fork this repository. Commit your changes and submit a pull request. Thanks.