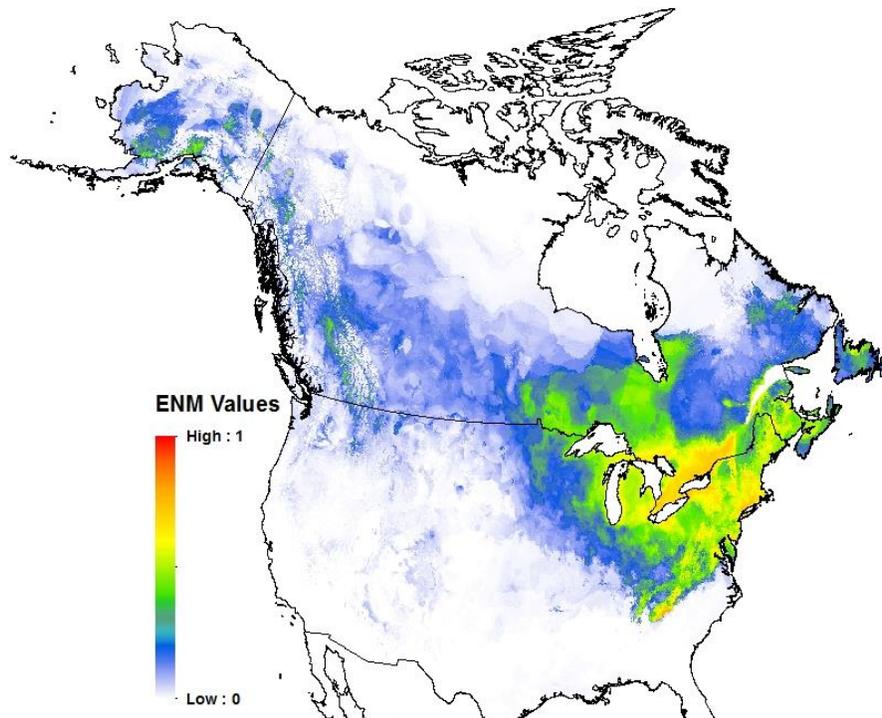


Lab 5: Top Down Ecological Niche Modeling Part 1

****This lab requires at least 4 GB, preferably more so you can use the higher resolution files.**

Introduction

Today we are going to use ecological niche modeling to determine the species distributions of amphibians and reptiles. As introduced in last week's lecture, top down models use known presence locations and correlations with landscape data to create a species distribution model. The logic here is that if we know the climate where a species is, we can estimate other locations the species may also be. We will be using Maxent (Phillips, 2006) to create our models. Output files (after doing some cleaning) look like a heat map of relative probabilities of presence. Warmer colors represent areas of higher relative probability of presence. See below for an example using the wood frog (*Lithobates sylvaticus*):

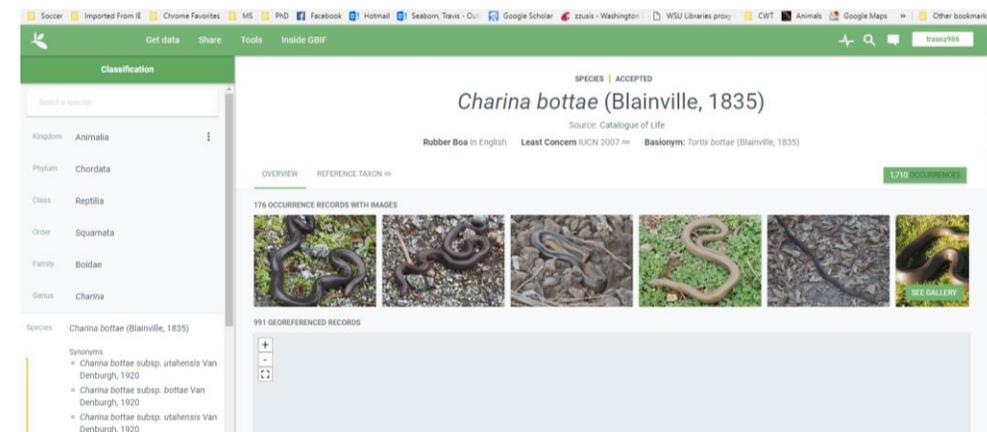


There are two primary steps to working through the correlative models. The first, acquiring the presence locations of a species, and second, acquiring landscape layers. In our case, we are going to be using climate data for the landscape layers. To begin with, please create a new folder on your computer called "Ecological Niche Lab." Within that folder, create several more folders, "Points," "Layers," "Maxent," "QGIS," "Clipped," "Albers," and "Outputs." For general GIS data manipulation and map preparation throughout the lab we will be using QGIS, available at <http://www.qgis.org/en/site/forusers/download.html>

Species distribution data

For species distribution data, we will use online databases that are linked to museum and university collections. This will serve as our locations of known presence. There are several options available, but we are going to focus on the Global Biodiversity Information Facility (GBIF). To download data follow the following steps:

- 1) Go to www.gbif.org. You will be setting up a free account with them.
- 2) Click species above the search bar. You can search by scientific or common names. In addition, you can also browse by taxonomic group.
- 4) To begin with, type the common name of your favorite North American species in the search bar, and click the corresponding name in the search results.
- 5) You should now see a map of occurrence points (see below as an example with the rubber boa). Make sure that the species you want to work with has at least 50 occurrences. To download them, click "Occurrences" and then "Download" on the subsequent page, then "CSV". You then will click the download box once it is done preparing. If you have not logged in or created an account, you will be prompted to do so now. Save the file to your "Points" folder.



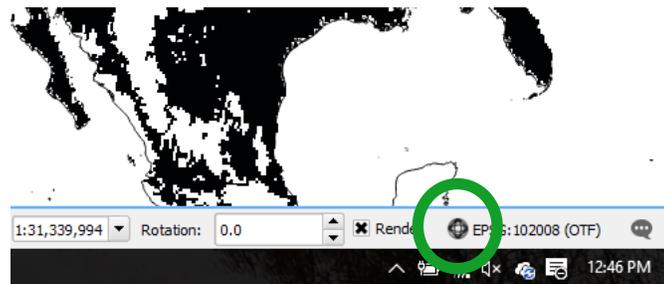
- 6) Open Microsoft Excel, do not double click the file, instead open a blank workbook.
- 7) Under the data tab, click "from text/csv," and open the file as a tab delineated file for you. Click "Next" and then "Finish." This will separate out the data into appropriate columns. Check the preview pane in the import wizard to make sure things are going to be done correctly, and then push okay.
- 9) You will now see your data laid out. Note that there are way more columns than we need. Delete everything except the columns "species," "decimallatitude," and "decimallongitude." Make sure that any occurrence data without GPS locations are deleted. (Note: if modelling for a publication, other data "clean-up" would need to be performed, e.g. removing records that were older than our landscape layers or removing records without a designated data frame.). The easiest way to do this is sort by one of lat/long columns largest to smallest, click yes for expand selection. Now all of the blanks should be at the bottom.
- 10) Save this file as a .csv with the species name in the Points folder. If Excel asks if you are sure you want to keep using this format select yes.

Landscape Layer Data Download, Clip, Reproject, and Convert

Next, we need to download the climate layers that we are going to correlate with the model. The caveat with the layers we download is that they are not in an equal area projection and they cover the entire world. We don't want to build a model using areas that the bullfrog is not currently located. The reason for this, is that we don't have true absence points (it is really hard to show that a species is not in a certain location) so Maxent picks random background points. If you background/extent of your climate data is too large, the model will be of poor quality.

- 1) Go to www.worldclim.org so that you can download landscape climate data
- 2) Click the "Version 2.0" link
- 3) To start, click the link for "Bioclimatic variables"
- 4) Download the 10 arc-minutes if your computer is short on hard disk space; if not select 5 arc-minutes. Select the "Bioclim" rasters. The finer resolution grids are necessary for publication, but to save on computer space and computational power required, we are going to stick with 5 or 10 arc-minutes. 10 will make uglier maps and be less precise on the regressions, but will run faster and use less space. www.worldclim.org/bioclim has information on what each variable is. Bioclimatic variables are derived from temperature and rainfall values to create variables that are more biologically relevant. All information you need for your write up, including the citation, is on this page.
- 5) Save the zip file to your "Layers" folder, and extract the files.
- 6) Move all files to a single folder new folder called "Raw". The expectation of this first lab is that you will build a rough model for your favorite species by picking 3 or 4 layers you think are biologically relevant.
- 7) You will now want to trim the layers just to North America. Open "QGIS Desktop with GRASS" and make a new project for today's work. Save this project in the QGIS folder. If you have used ArcGIS, you may find yourself familiar with the overall layout of the program. The great thing about QGIS is that it is completely free.

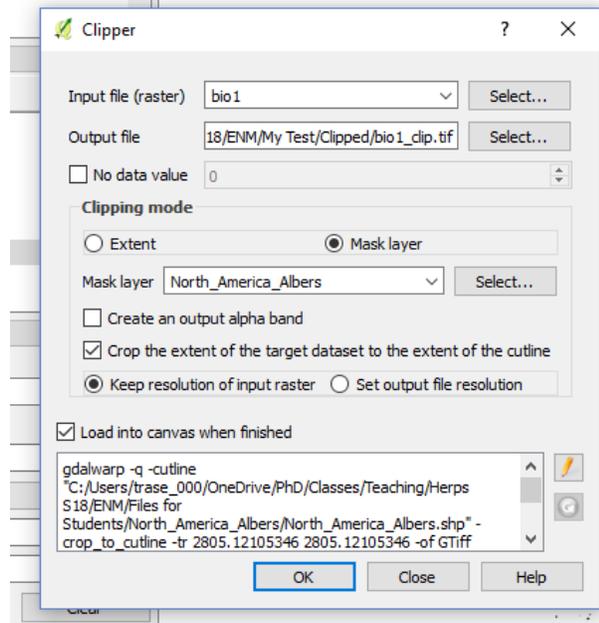
8) Select Layer, Add Vector, and add the North American Albers Shape (.shp) file (you will need to unzip it first). It is the largest file that will appear after unzipping. Now click the coordinate system selection button on the bottom right of the screen (starts often with EPSG and has a circle logo next to it, see picture with green circle). Set the project coordinates to 102008, the North American Albers Equal Area projection. You might need to check the box at the top for on the fly projection. Then, double click the file under the layers panel, click style, then the box next to simple fill. From there, change the fill to transparent, so that we just have the outline left.



9) Next, we need to reproject the layer. Select Raster, Projection, Warp. For the input file, click Select, and select your clipped and converted files from step 10. For the output file, save them to a new folder

named "Albers." Save output as "bio_x_albers" where x is the number of the variable. Set Target SRS to EPSG 102008. Source should be EPSG 4326. You should now see the layers stacked on top of the North America Shape File. This process is sometimes slightly glitch using QGIS.

10) Now we will clip the raster so the extent is just North America. Select Raster, Extraction, Clipper. Now, select one of the data layers you extracted in step 5. Set the output file to a new folder (preferably called "Clipped"). Name the file "bio_x_albers_clip" where x is the appropriate number and keep the tif format because it is relatively small. Under clipping mode, select Mask layer, and select the North American Shape file from above. Make sure to select the "Crop" checkbox.



If you want to save space, you can also create a shapefile mask that only covers your point and not all of North America.

11) Next, we need to convert the files from WorldClim. Click Raster, then Translate. Input file is where the clipped and reprojected files are (in .tif format), and then output them into the folder Converted as an .asc files. **This will take a while. Tip: you can drag the menu to the side and work on the next part. If you look in the folder you should see the file getting larger.**

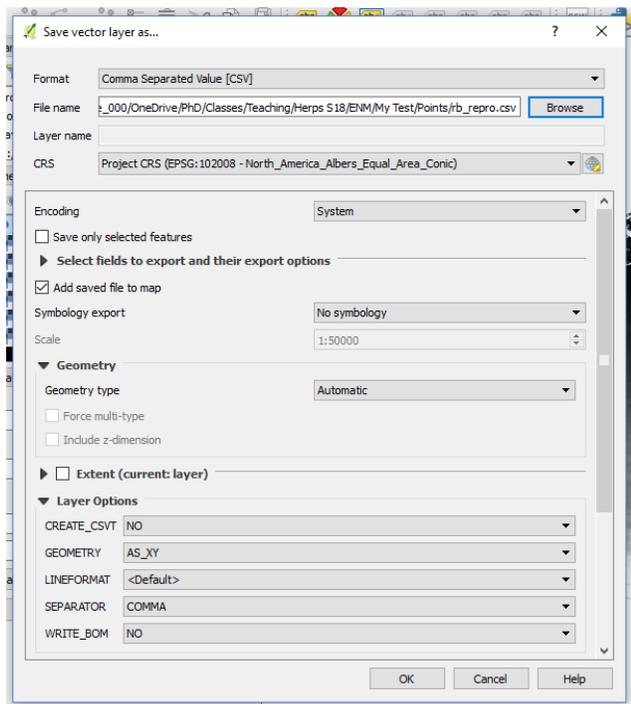
Reprojecting Species Occurrence Data

1) Now that we have fixed our landscape layers, we need to fix our presence points

2) To add presence points, select Layer, Add Layer, Add Delimited Text Layer. Click Browse and select your data. Next, make sure CSV is selected for file format. The table below should look like an Excel sheet.

3) When you click okay, you will be prompted to select the coordinate system. Select WGS 84 (EPSG 4326). You should now see the points on your map.

4) Right click in the Layers Panel, and select Save As. Change format to CSV. Click browse to select a new location and save these new points with a new name. Change CRS to our



EPSG 102008. Under Layer Options, change geometry to XY. This will now save a .csv file with the added columns of latitude and longitude in meters (which is what the North American Albers equal area projection uses).

5) Remove the CSV file from QGIS and open in Excel (using the same import method from before, but make sure you select comma delimited on the second page of the prompt after you click "Next").

6) Delete decimallatitude and decimallongitude. Put the species column first. Do this by cutting and then right clicking and going to "Insert Cut Cells."

Downloading Maxent

1) Go to <http://www.cs.princeton.edu/~schapire/maxent/>

2) You will need to register your download, so fill in the Name, Institution, and Email fields.

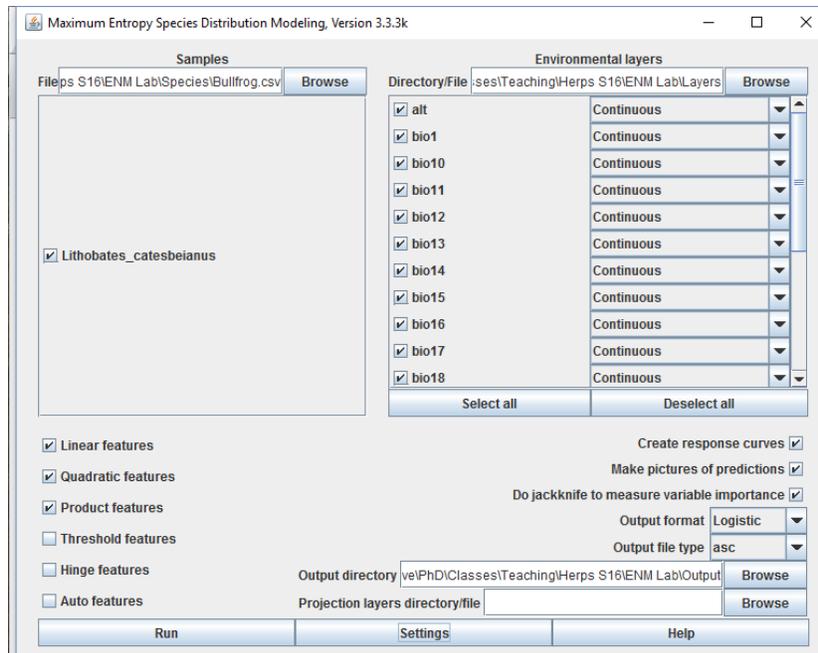
3) Download the current version by clicking "Accept terms and download"

4) You will be taken to a site where three files can be downloaded "readme.txt," "Maxent.bat," and "Maxent.jar." Download all of these to your Maxent folder. The .jar file is the one you will double click to open Maxent. You need to have Java installed.

Running Maxent

1) Click browse under Samples, and point to your .csv file for your species that is in meters. You should now see a list of species in the box.

2) Click browse under environmental layers and go to the folder where all of your environmental layers are located that have been reprojected (should be under the folder "Converted"). You should now see a list of your environmental layers in the box. Note that all of the climate variables are continuous, so do not change that drop down menu. Also, make sure all of the layers have their boxes checked. **If you have not finished converting all of the files, that is fine. Try running Maxent with just Bio1.**



3) Under the features section (on the bottom left), uncheck Auto features. Have only linear, quadratic, and product features highlighted. This reduces the model complexity and response curves to be more biologically relevant (and mirror our thermal performance curves). Make sure all the check boxes on the bottom right (there are three of them) are checked.

4) Click browse to the right of Output directory, and specify your Output folder.

5) Click Settings. We will use default settings, but realize this where you can run replicates. Start with running Maxent once, but if you run more than one later you will be presented with a mean model.

6) Click Run. If there are any warnings, click "Suppress similar visual warnings." This is usually just because some of your points are outside the range of your environmental layers. Note that depending on the size of extent, the number of species, the number of replicates, and the resolution of your layers, this can be computationally intensive for your computer and will take *around* 15 to 30 minutes.

While you wait, to make things easier for the next steps, you will want to remove the individual climate layers from your QGIS file. To do this, right click the layer, and select remove. If you are waiting for a Maxent model and need to leave, you can close your laptop, and as long as the power does not die completely, it will pause and then resume the model when you open your laptop.

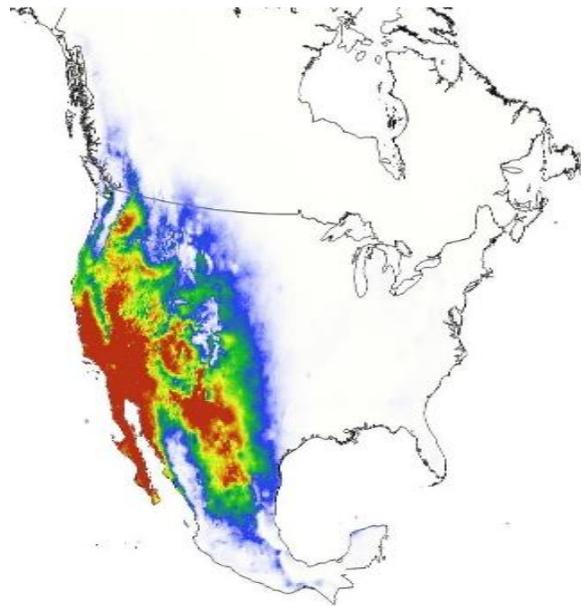
Maxent Results

7) **Open the folder to view and click the .html file to read the output.** You will see a few plots of interest. The first is model fit. You will see an AUC value with the second plot. An AUC value of 0.5 would be expected from chance, so higher values here the better. You will next see a map showing the mean and standard deviation of the models (if you did more than one run). Response curves are next. This

part is really important. This is the response to different climate variables. Note that the units for bioclim are a bit funky, so you should look them up on the website if you are confused. These should look like the performance curves from previous labs. There are two sets of plots, one for the entire model, and then the response if you only look at those variables in a Maxent model by themselves. Next you will see how much each variable contributes to the final model, both in table and jackknife format. Let your instructor know if you are confused by these.

8) To edit the raster (and make it much more visually pleasing), open QGIS and your project. Click Layer, Add Layer, Add Raster Layer. For the datum, select the North American Albers Equal Area (EPSG 102008). Make sure you are opening the average model if you ran multiple trials. The file extension should be .asc

9) Open Layer Properties by double clicking or right clicking the layer. Click Style. Change render type to singleband pseudocolor. Make sure the minimum value is 0. Start with the Spectral default, and then click Classify. This now creates the color scheme. Note that you can now tweak the colors as you see fit by double clicking values and colors. Generally, you want colors to be warmer for higher relative probability of presence. Here is an example using only climate variables and species occurrence for the *Uta stansburiana* (note that it is missing a figure caption).



10) To export maps, simply select Project, Save as Image. Note that you can change the size of the image through this menu.

Assignment

Write out a figure caption and turn in your ENM figure for your favorite North American herptile. This can all be on single page. It will be worth 10 points, with five points for a nice looking figure and 5 points for a complete figure caption.

Lab 6: Top Down Ecological Niche Modeling Part 2

This week we will begin building the model for your lab report. Use the species results under genus "*Lithobates*," which has more points than what is catalogued under the old name (*Rana*). Make sure to bump the number of replicates under settings, because we want to work with the average species distribution data. Also, you will want to use all BioClim layers and make sure to use all of North America. I will provide you with those files.

While you wait for the Maxent model to run, you should start to write your methods section for this lab and check/finish your threshold calculations with your classmates from the physiology lab. These values will be used next week

There is no assignment to turn in next week but you are expected to finish your model and make it presentable today. Next week we will use our physiology experiments to create a bottom-up model based on your threshold calculations. **These two labs will be combined into a single ecological niche model report for the American Bull Frog. Please contact me immediately if you have trouble. We will discuss your results from this lab extensively next week.**

I can't emphasize this enough: Before leaving lab you should finish the figure, caption, methods, and results so that I can take a look at it before you go.