Title:
Ecological Niche Modeling: What Processes do Scientists Use?

Author(s):
Valerie Ledford
Belmont Academy
valerie.ledford@gmail.com

Abstract:
Ecological niche modeling is a method of using digitized species information, perhaps from collections, and making maps of a biological species' potential distribution. Using technology allows scientists to use large data sets to get more accurate models as data is added. Ecological niche mapping is a means of showing speciation and biodiversity. Modeling thousands of species and piling distribution maps on top of each other allow scientists to get species richness data. In addition, scientists can compare current and future data models to make predictions on species distribution based on climate models or other ecological projects. It is important to note that ecological modeling looks only at the organism's fundamental niche. Interactions with other organisms and other elements may not be considered.

In this activity, students will simulate actual processes used by scientists to create ecological niche predictive models. This is a computer-based activity.

Subject, Grade, Level:
AP Biology, high school

Learning objectives:
At the end of this guided discussion, students will be able to:

- *Use tools that scientists use to complete Ecological Niche Modeling predictions*
- *Evaluate data from Ecological Niche Modeling*

Timeframe:
Teacher set up time will depend on the school technology situation. Since each computer that students will use needs software downloaded on it ahead of time, teachers will want to discuss with their school's technology coordinator to determine the best way to complete the preliminary set up.
For the activity itself, the time will depend on how technologically savvy the students are. It is estimated that the task will take 3 - 4 class periods.

List of materials:

- *PDF Instruction sheet: Ecological Modeling*
- *Computer with internet access and*
    - *Microsoft Excel*
    - *QGIS software installed (download from on-line)*

- *Maxent software installed (download from on-line)*

Procedure and general instructions (for instructor). REQUIRED.

1. Prior to class, the instructor should ensure that computers being used for this activity are able to function with the software. Teachers should ensure that computers have access to the internet and have Microsoft Excel downloaded. In addition, teachers should go to the QGIS website and download that software. In addition, teachers should ensure that Maxent is downloaded from https://biodiversityinformatics.amnh.org/open_source/maxent/

2. At the start of class, tell students that they are going to play the role of scientists doing research into ecological niche modeling. Provide students with the instructions and have them read the introduction. Ask students to share with each other what they think about what they are going to do by doing a quick "Think Pair Share". Have some of the best answers share outloud to the whole class.

3. Have students read the task so they know what they are expected to do. Answer any questions students might have.

4. Tell students that they are going to select a species to study and they need the scientific name. Tell them that there must be at least 30 viable records in idigbio.com for this species in order for the task to work so they should be prepared with alternates, just in case. The instructor might also want to have a few species selected as a back-up.

5. Allow students to work at their pace through the tutorial while the instructor circulates and assists. Be sure to note to students that there are video tutorials listed at the end of the document to help.

Procedure and general instructions (for students).

Instructions for students are provided in a PDF addendum titled "Ecological Modeling".

Reference list

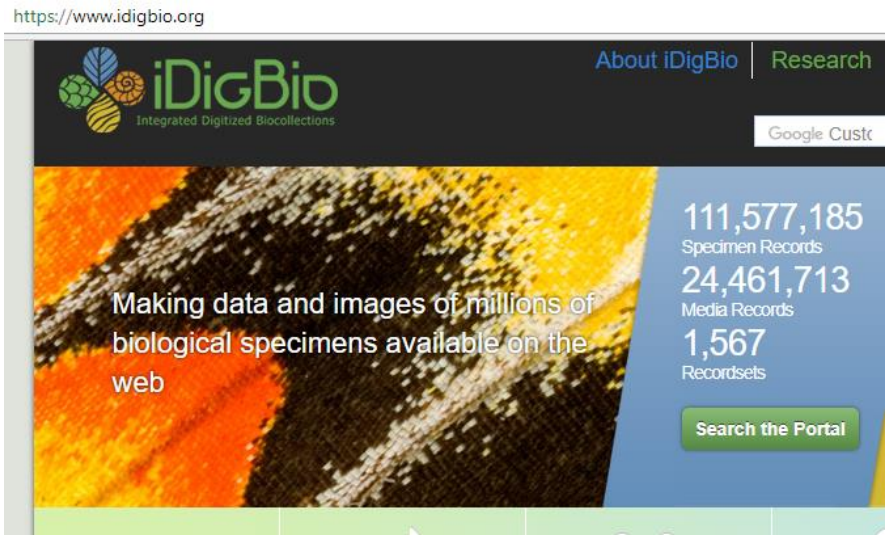
Marchant, B. (2017). Ecological Niche Modeling Tutorial. Retrieved from https://vimeo.com/idigbio

ECOLOGICAL MODELING

BACKGROUND: Ecological niche modeling is a method of using digitized species information, perhaps from collections, and making maps of a biological species' potential distribution. Using technology allows scientists to use large data sets to get more accurate models as data is added.
Ecological niche mapping is a means of showing speciation and biodiversity. Modeling thousands of species and piling distribution maps on top of each other allow scientists to get species richness data. In addition, scientists can compare current and future data models to make predications on species distribution based on climate models or other ecological projects. It is important to note that ecological modeling looks only at the organism's fundamental niche. Interactions with other organisms and other elements may not be considered.
In order to complete the process of ecological niche modeling, there are a few steps:
1. Pull **records** from georeferenced database such as idigbio.org. This is an integrated digitized biocollection of species locations and information. Because is it an on-line resource that houses data from locations from all over the world, it allows for a richness in data.
2. Pull **climate records** in layers showing various variables of an area's climate such as precipitation, temperature, etc.
3. Use the climate records and the occurrence records to run an ecological niche predictive model.
4. Read and analyze the data.

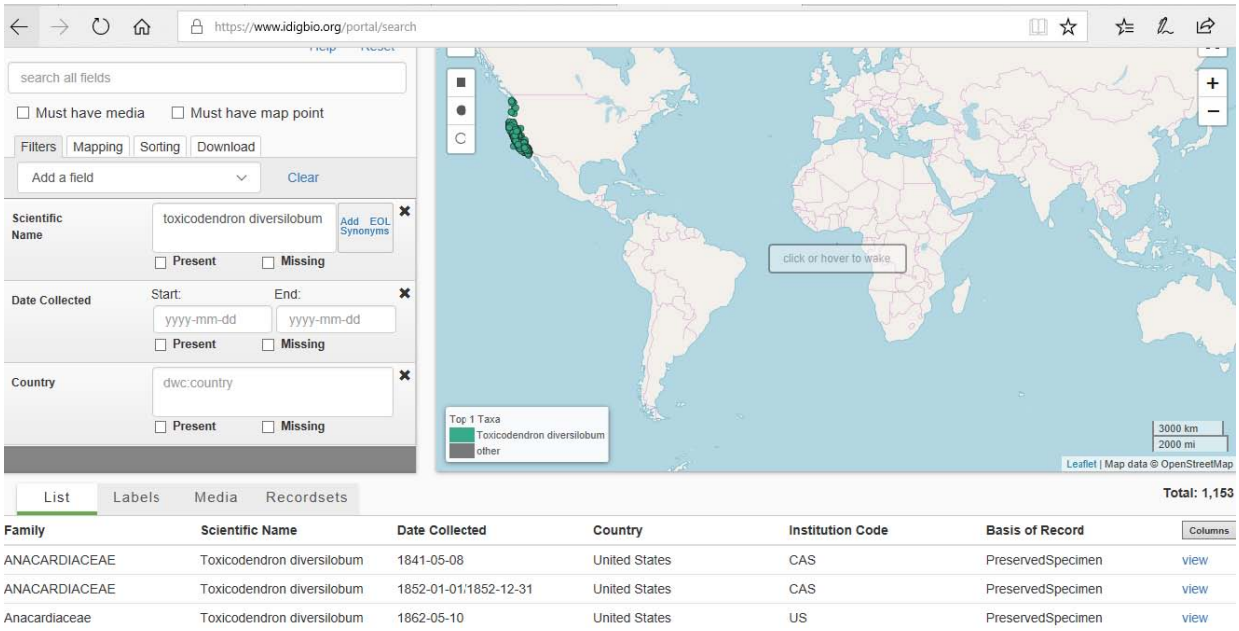
YOUR TASK: You will be acting as ecological niche modelers to collect data on a species of your choice. Please follow the directions below. At the end of the task, you will write a two-paragraph essay on:
1. An analysis of your data
2. What are potential uses and/or benefits of Ecological Niche Modeling

PULL RECORDS FROM GEOREFERENCED

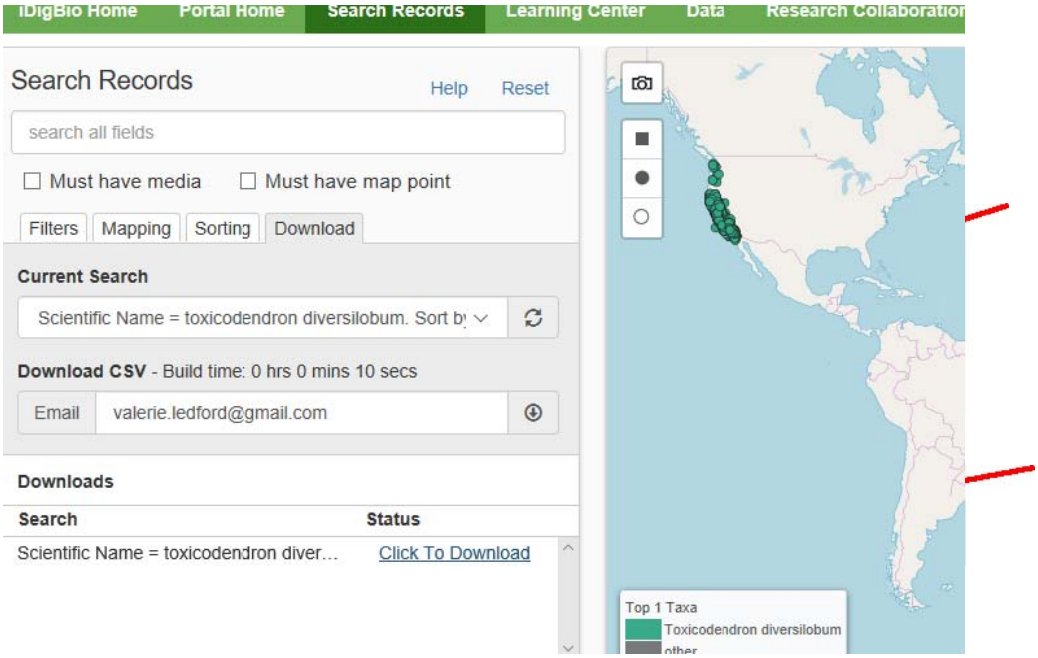


OCCURRENCE RECORDS

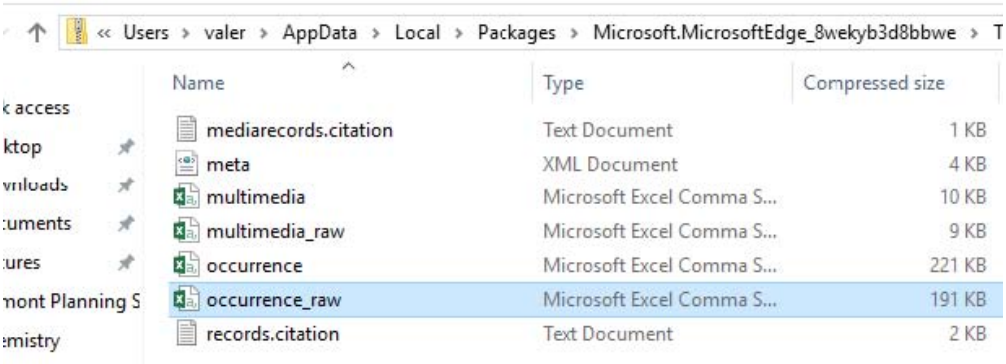
1. Go to idigbio and select “Search the Portal”



2. Go to idigbio and select a species to pull up data. Look at the bottom and ensure that you have at least 30 occurrences. More data is better. You must use scientific names.

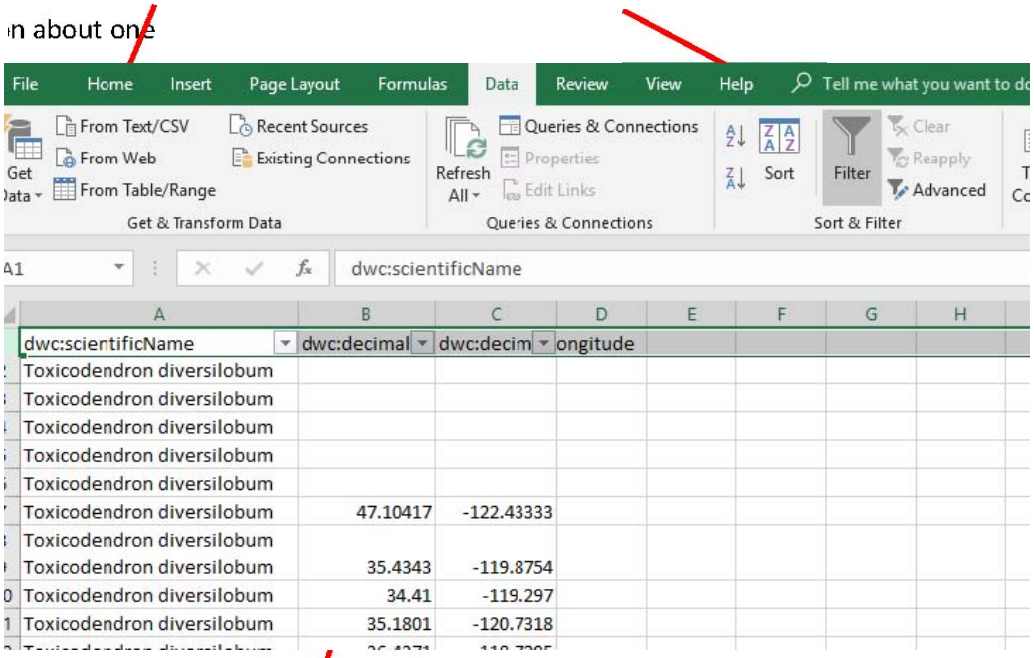


3. Go do Data and Download CSV file



4. Click “Click to Download” and save/open file

5. Select raw data file (ex. “occurrence.raw”)



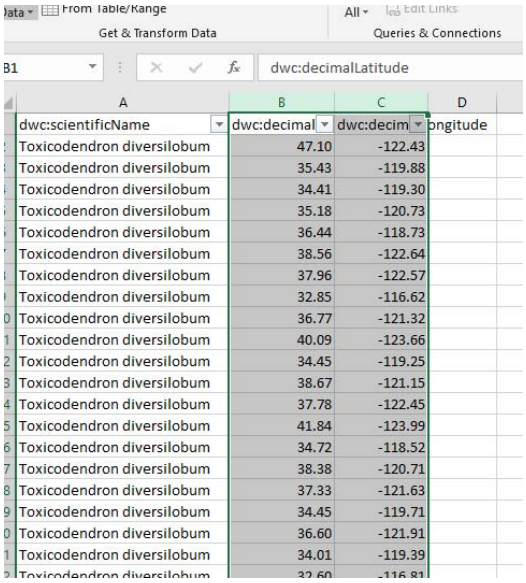
6. From the original, only select the columns for Scientific name and longitude and latitude and numerical columns have blank lat select the Data tab and click "Filter". Filter tool will give a drop down and give you information about one specific column

7. Click the drop down in the numerical columns and unclick select all, scroll down, and then select "blanks". This will show only the records that have blank latitude and longitude data.

8. Delete all the blank rows

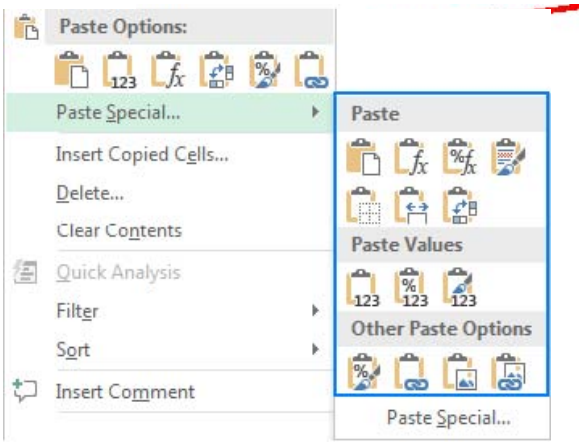
9. Go back to the drop down filter and "select all" again- we should see only data that is complete

10. Select the two number columns (column B and C) and right click to get "Format cells", go to "number" and select 2 decimal places. Your data should look something like the image to the right.



11. In a new column, we are going to round the values to 2 decimal places. The above step only gave the appearance of rounding but the original values are still "there". So, type in =round(B3,2)- this means that it will round cell B3 two decimal spaces. Label the two new columns "Latitude" and "Longitude".

12. You can copy this formatting by grabbing the little + that shows up on the bottom right corner of the cell. See image to the right that is not related to this data.



13. Copy the two rounded values and paste using "Paste Special" into a new workbook

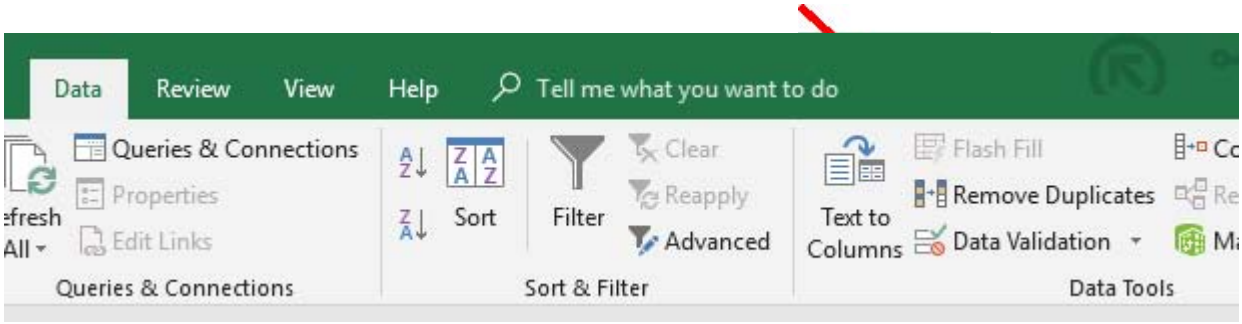
14. Make the first column has the name so that the new workbook sheet looks like this:

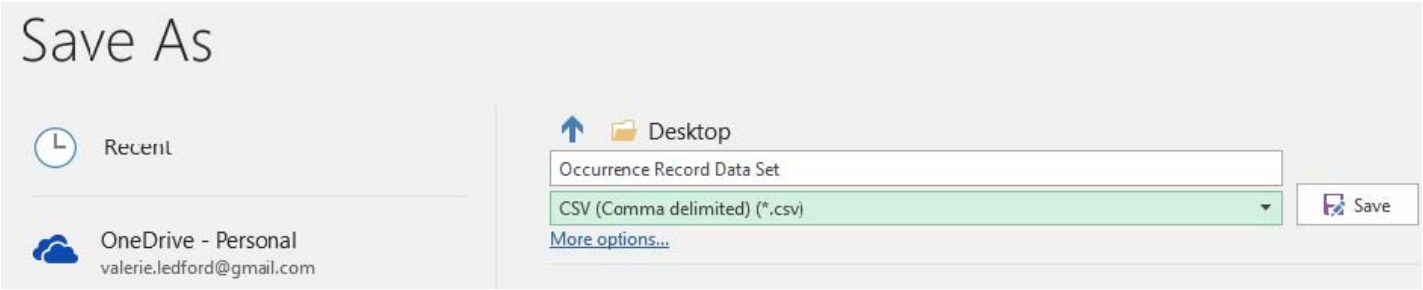
J830

	A	B	C
1	Name	Latitude	Longitude
2	Toxicoder	47.1	-122.43
3	Toxicoder	35.43	-119.88
4	Toxicoder	34.41	-119.3
5	Toxicoder	35.18	-120.73
6	Toxicoder	36.44	-118.73
7	Toxicoder	38.56	-122.64
8	Toxicoder	37.96	-122.57

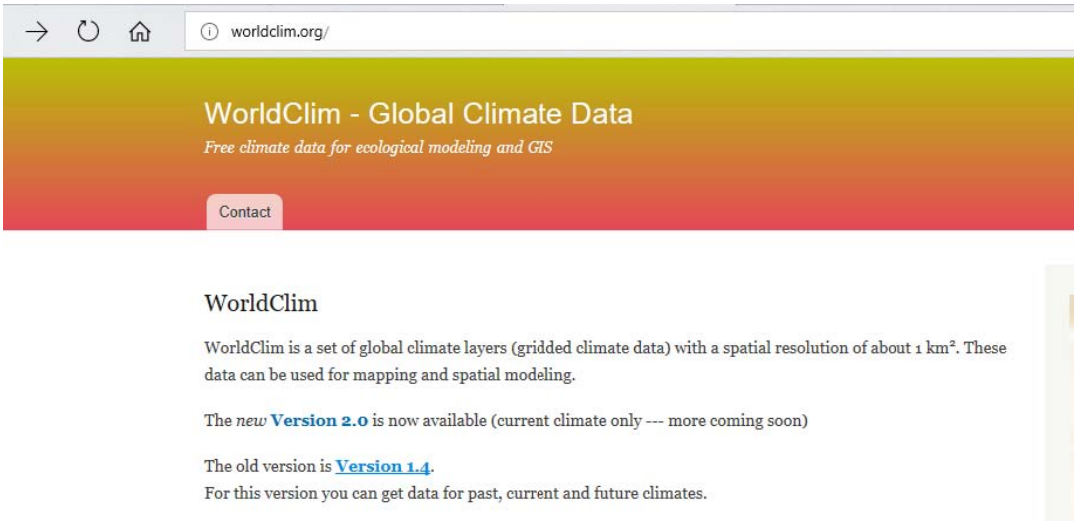
15. Select all data, click on the DATA tab and select "Remove Duplicates"

16. This is now our Occurrence Record Data Set. Save as a CSV Comma Delimited workbook.

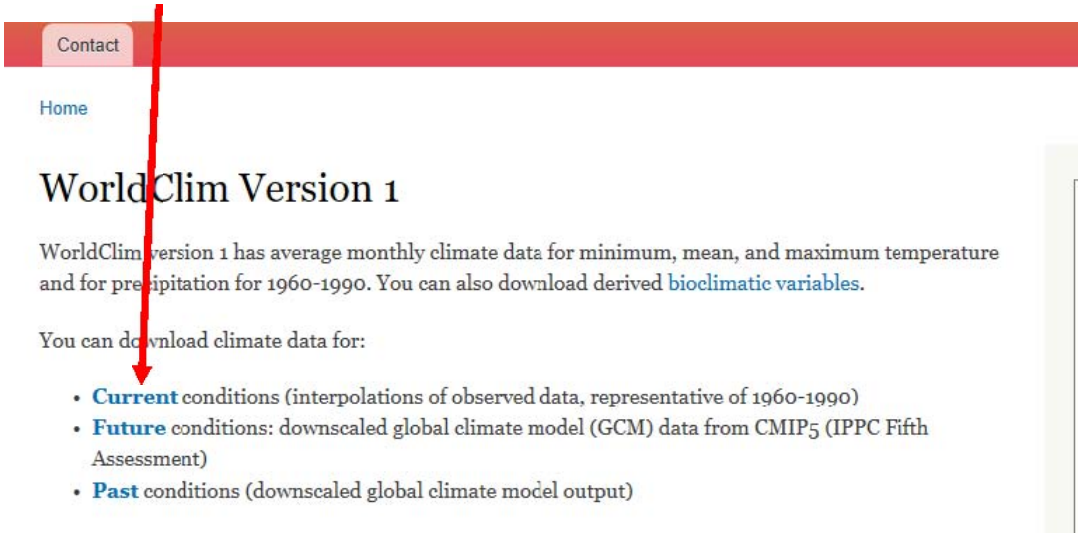




CLIMATE LAYER PROCESSING

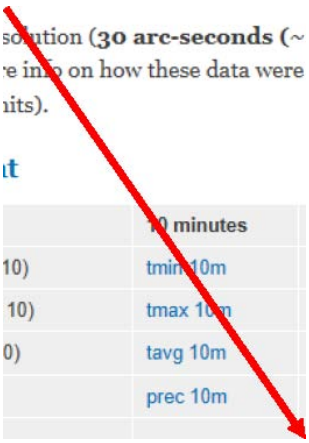


1. Go to worldclim.org/ and select Version 1.4.



2. Select the "Current" conditions.

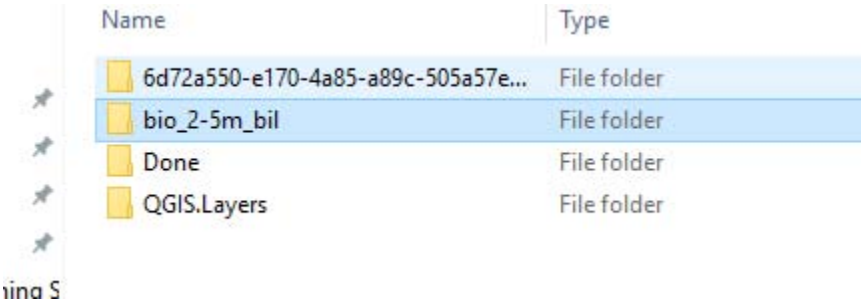
3. Select "Bioclimatic variables" based on the speed and quality of download.



If you need the highest resolution (**30 arc-seconds (~1 km)**) then you can **download by tile**. See the Methods page for more info on how these data were generated, and this page for info on details about the data (such as units).

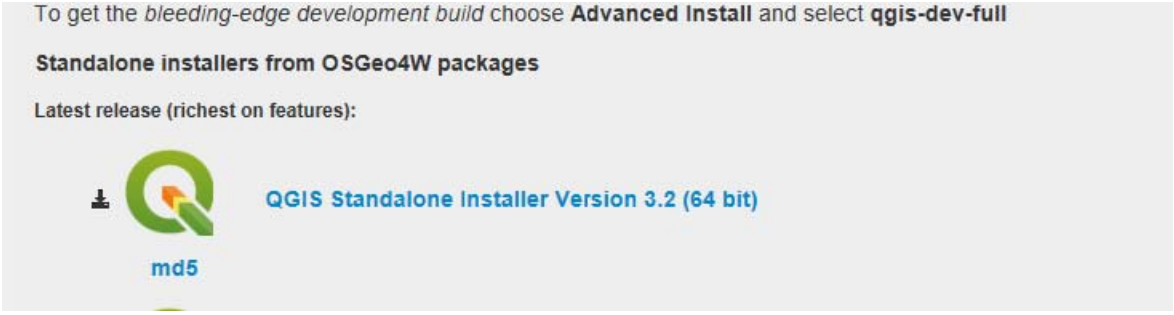
Generic grid format

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C * 10)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C * 10)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C * 10)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
bioclimatic variables	bio 10m	bio 5m	bio 2.5m	bio1-9, 10-19

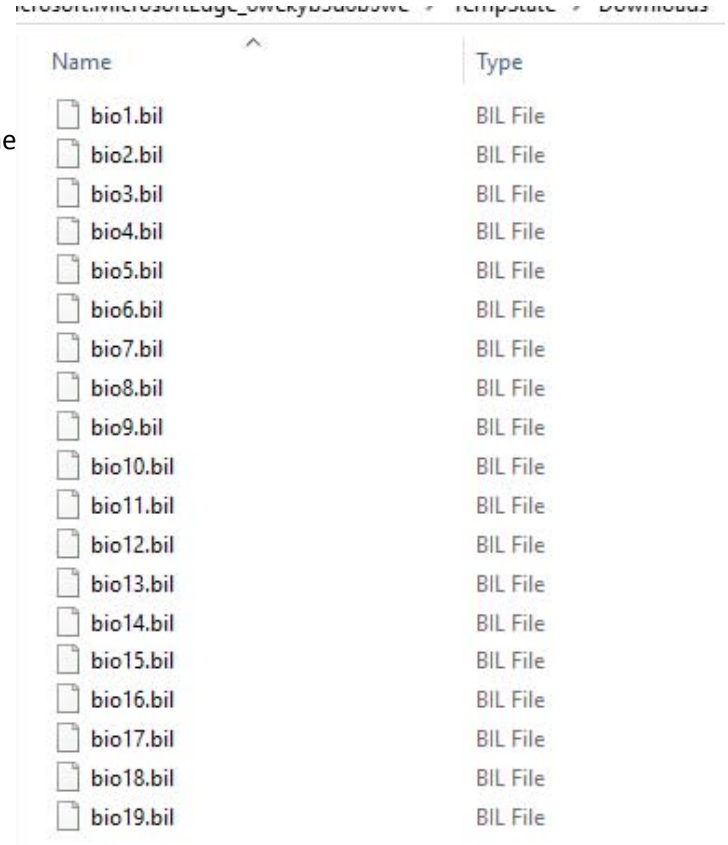


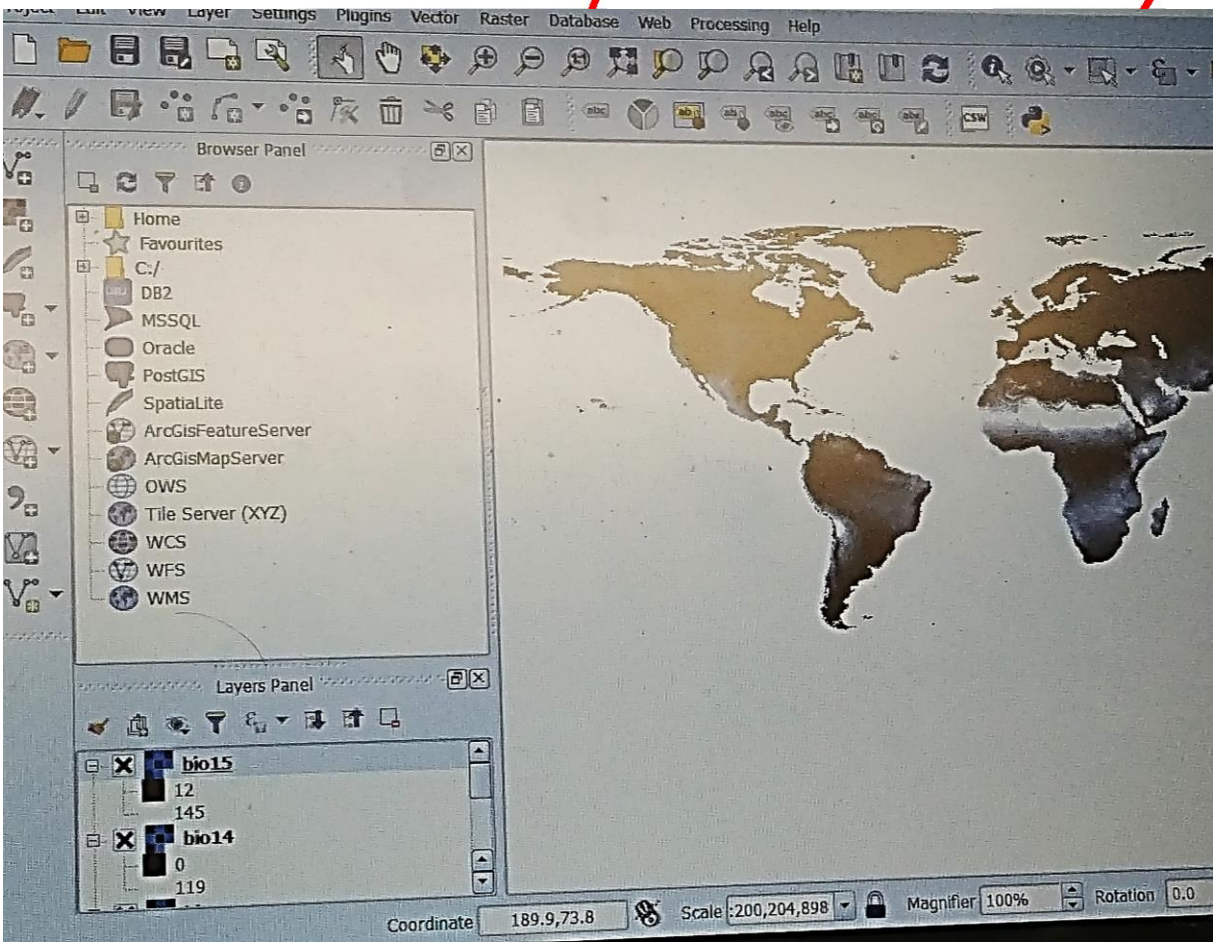
4. Be sure you know where you saved the files. The data is sorted by the variables as shown at www.worldclim.org/bioclim . You should have 19 files to show the 19 variables.

5. Go to QGIS 2.18 (downloaded from online)



6. Open and select the "Type" colum to sort by bil only



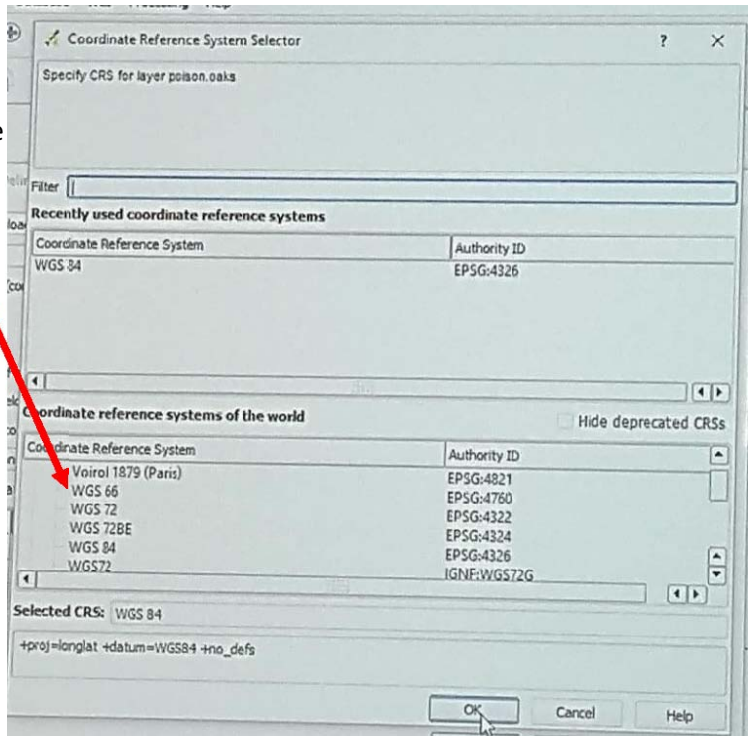


7. Drag and drop the bil files into QGIS into the “Layers Panel” on the bottom left corner to see an image

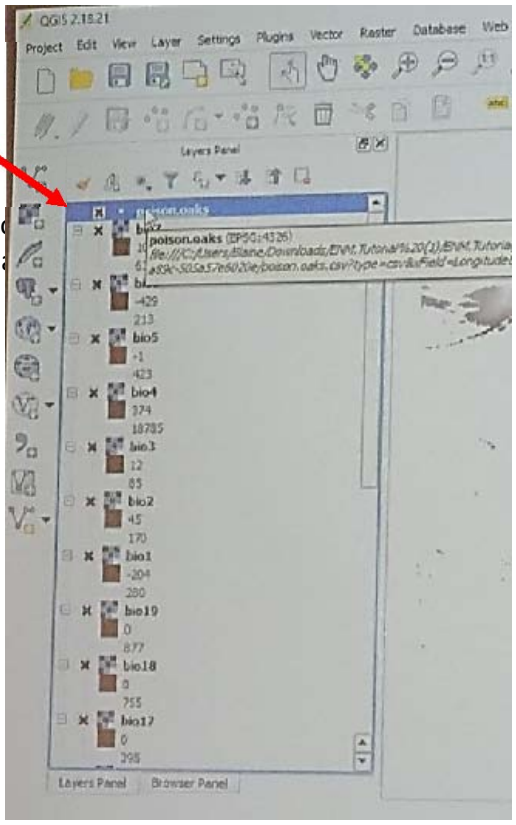
8. Now, data points must be added. On toolbar click “Layer”, “Add Layer”, and “Add delimited text layer”. Then click “Browse” in the upper right of the pop up box and find the file we saved earlier as “Occurrence Record Data” Set”. Make sure file format is in CSV

9. Click “OK”

10. The next box pops up. Select the default of WGS 84. Select “OK”.

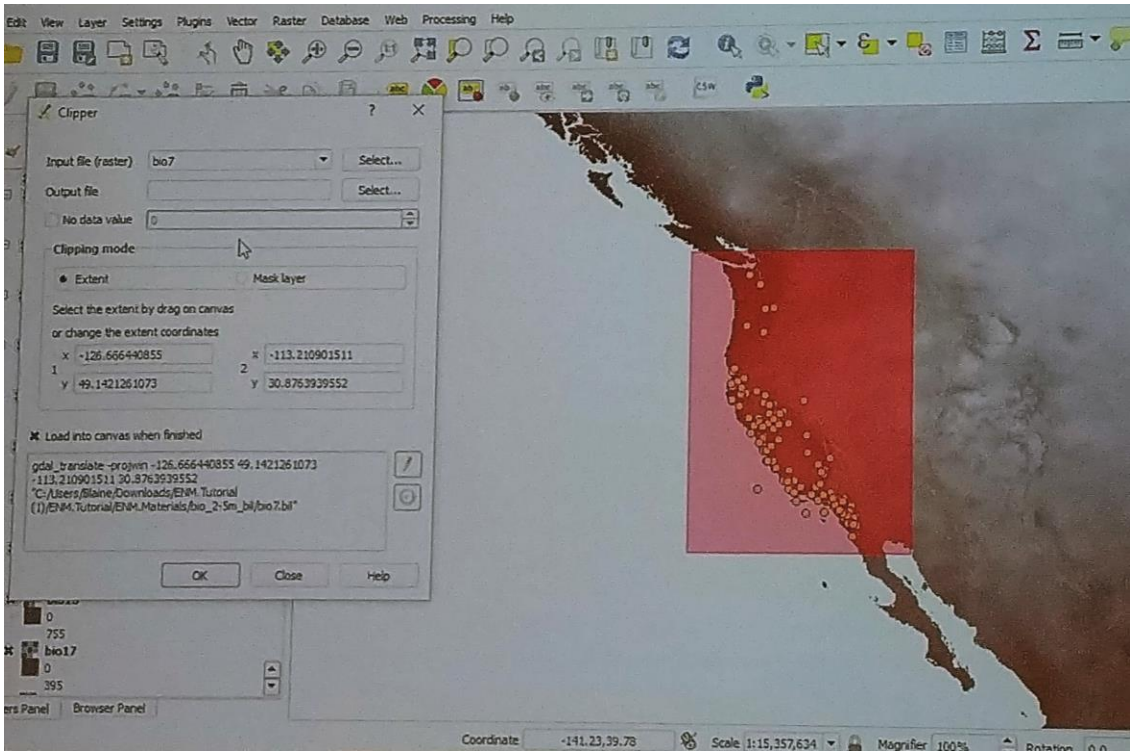


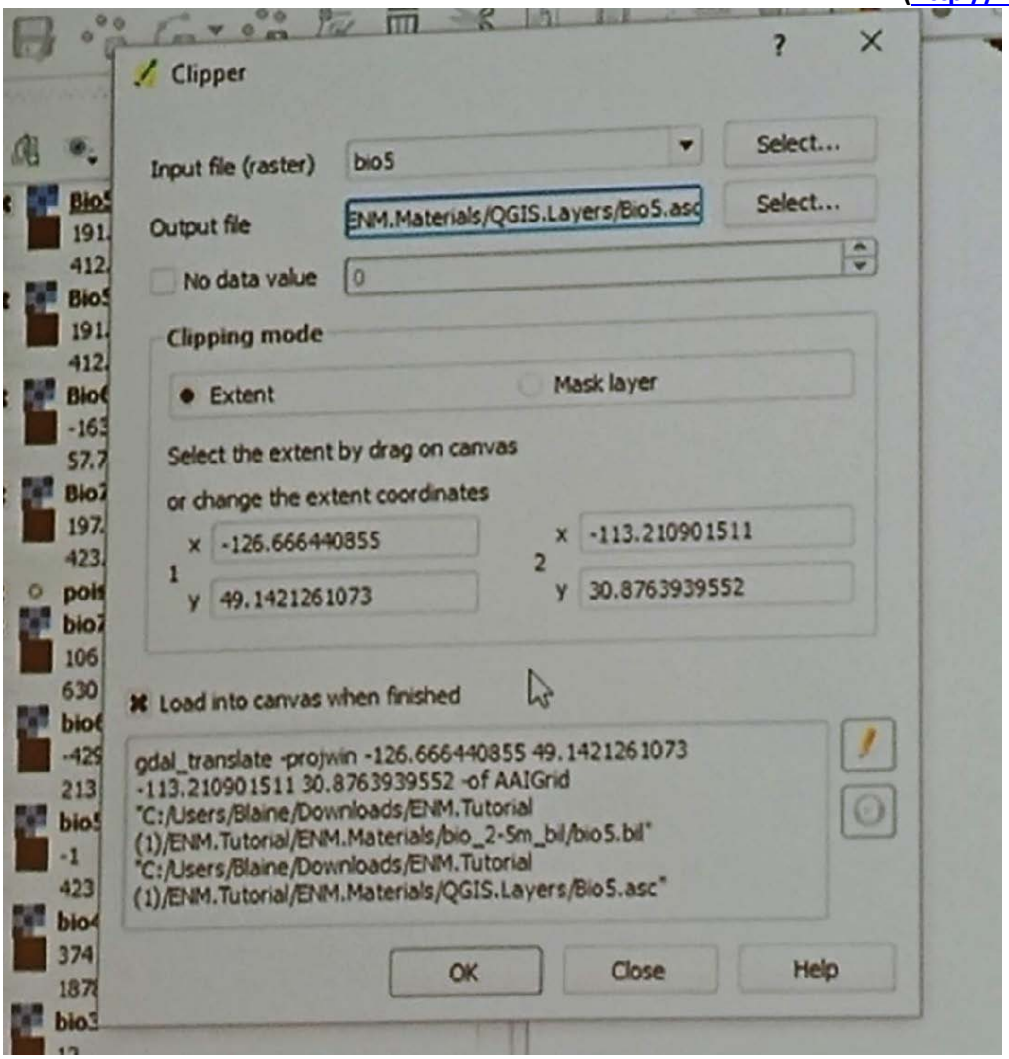
11. Be sure occurrence data is at the top of the Layer Panel on the right so that yo dots on the ma



12. Now you need to trim the layers. From the tool bar, select "**Raster"**, "Extraction", "Clipper" (from tool bar).

Move the Clipper box aside, click and select the area on the map that you want to focus on (the ones with the dots in it) by clicking in the area and dragging to form a box. This selects the area that you want to work with.





13. Data should pop up in the coordinates in the clipper box

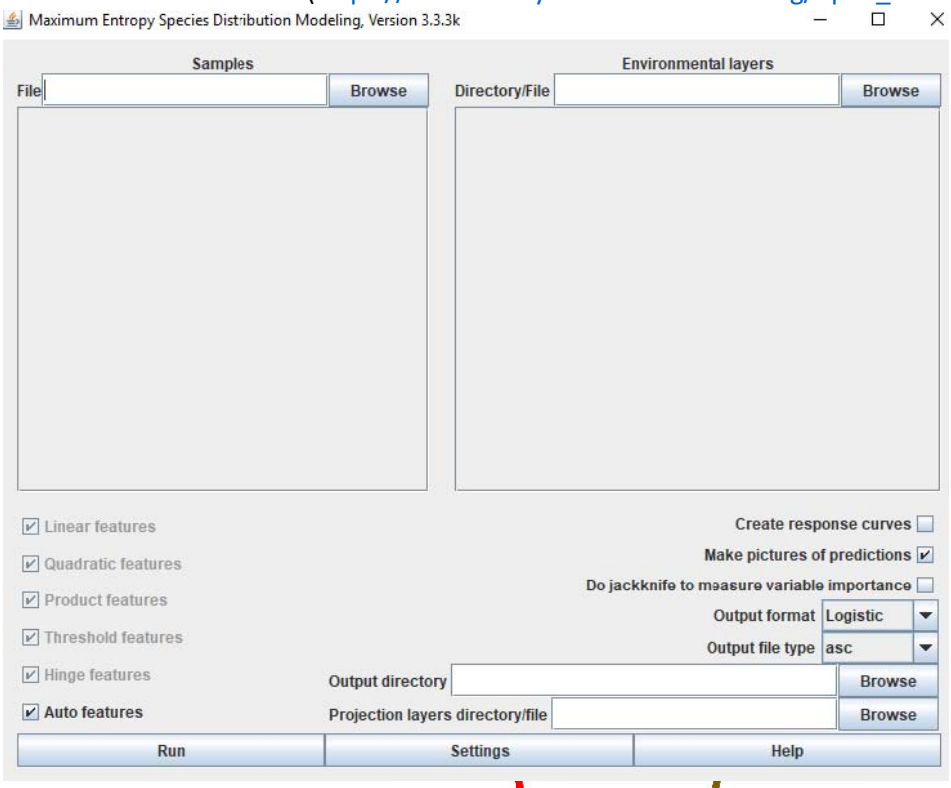
14. Select an output file by clicking “Select” in the top right area- be sure you label it and it will be Arc.Info ASCII file.

15. Repeat step 14 FOR EACH LAYER by selecting the input file (select bio7, then bio6, etc.) and then selecting an output (click Save to File, if necessary and then click ASCII)

16. Now, you should have 19 files (one for each of the variables from the Climate Data).

PREDICTION MODELING

1. Install Maxent (https://biodiversityinformatics.amnh.org/open_source/maxent/) and open the application so you see the start up menu at:

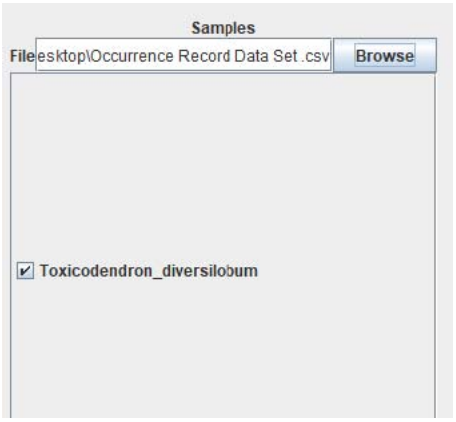


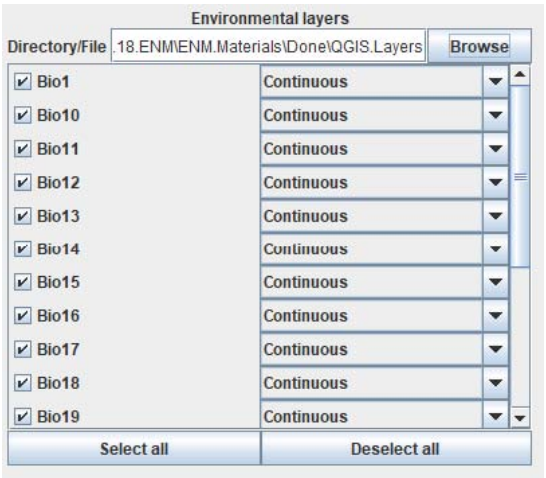
2. To upload the samples, open your occurrence data from the first section by clocking “Browse” and locating the CSV file

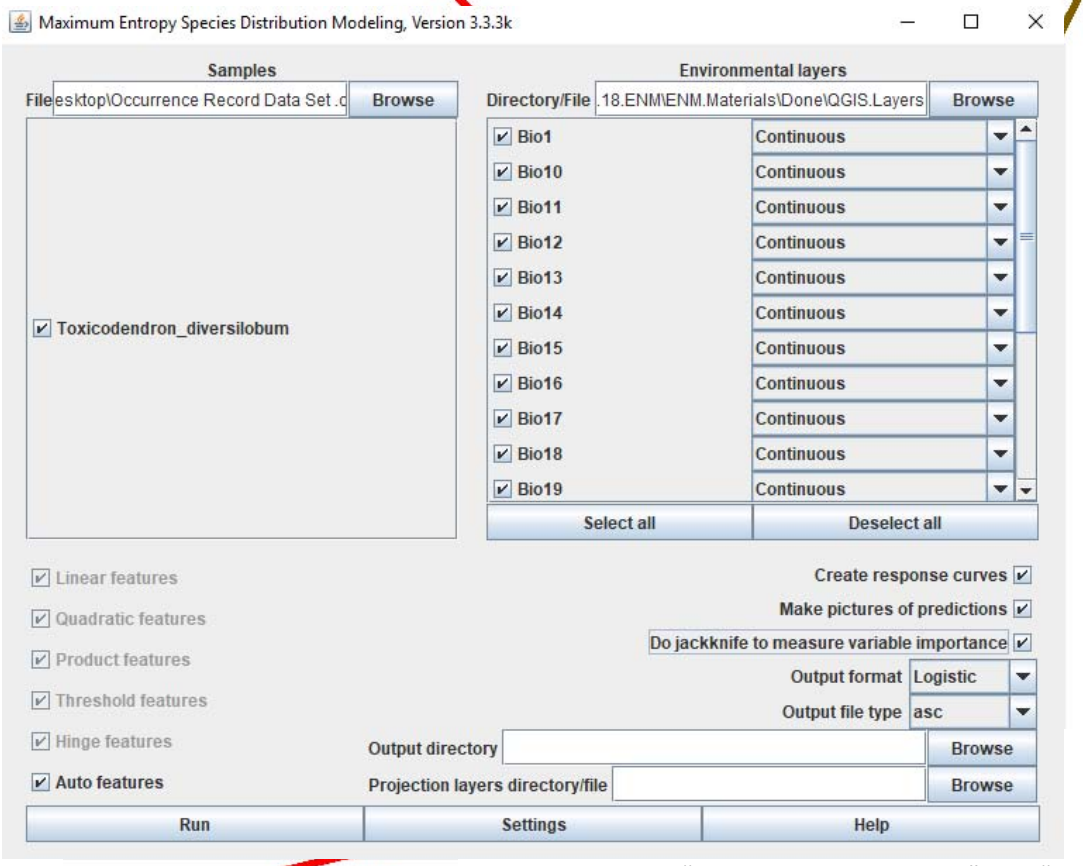
from the QGIS activity into the box for Environmental layers.

rrence data from the first
SV file

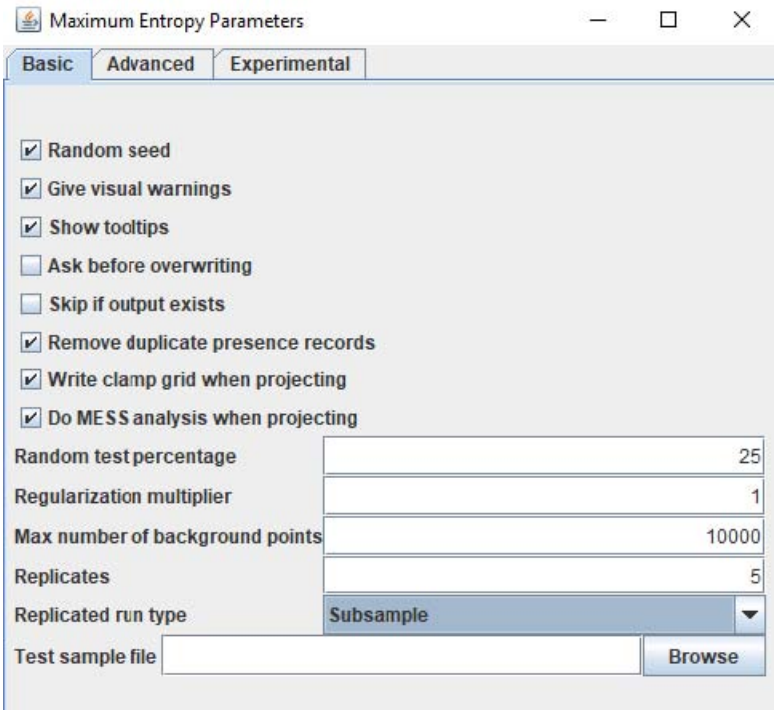
3. Upload your 19 environmental layers



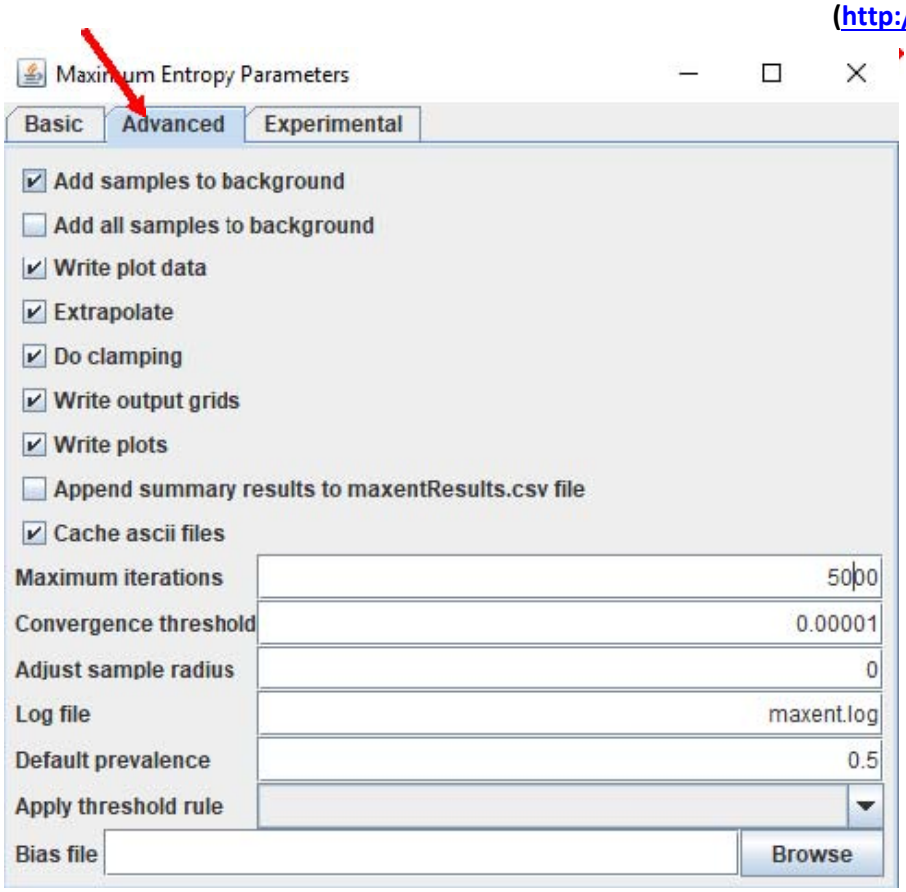




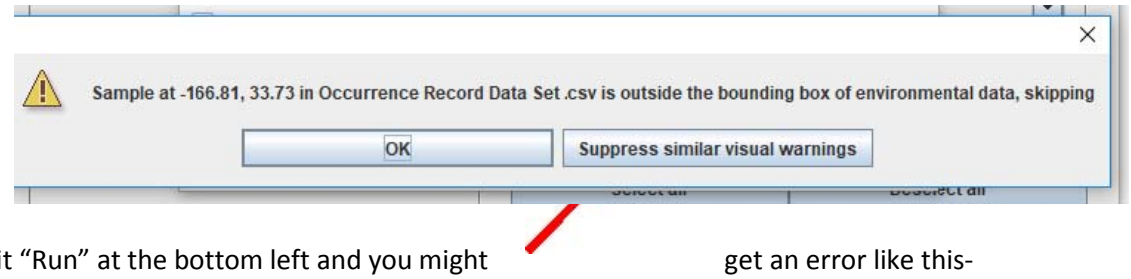
4. Check "Create Response Curve" and "Do jackknife to measure variable importance". Then, select "Browse" and identify an output directory



5. Under "Settings", set the setting to match the image below. For the random test percentage, you are selecting to use 25% of your occurrence records to test the model that you are using 75% of your data to build.



6. Under "Advanced", select the options to match the image below. Click the X when done.

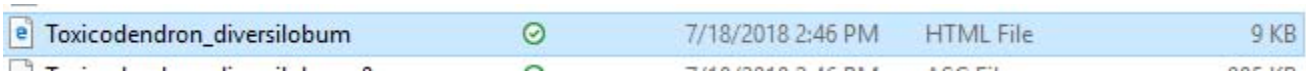


7. Hit "Run" at the bottom left and you might get an error like this-

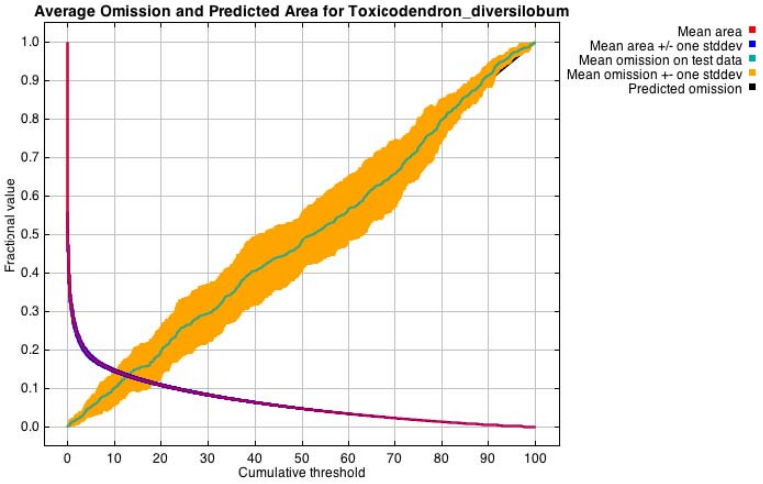
8. Hit "Suppress" because this is likely identifying that a few points are in places that make no sense- ex. You have an occurrence of a land plant in the water. This will eliminate those points from the final model and as long as there are few of these, suppressing them will not affect the outcome.

9. Ecological niche modeling is used to figure out (and map) the abiotic tolerance of a species. We will see a map of where the species is found based on the tolerance. We can project those tolerances on other time periods or locations to compare for climate change data, etc. In the section of "Projection layers director/file", you could add climate model layers that would them help you see where that species might occur based on climate change.

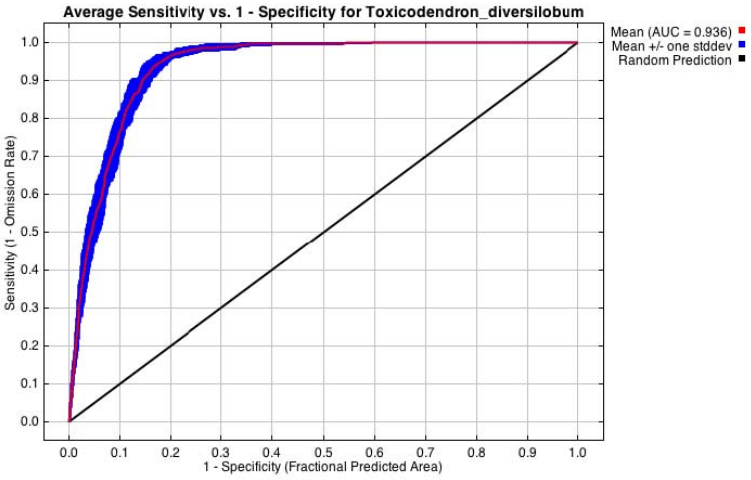
10. To run the data, it may take some time depending on sample size, layers, etc. When you are done, look for a file labeled like this in your results. Do not look at the numbered ones, as those are the replicates. The one with no number is the averages.



ANALYZE YOUR DATA



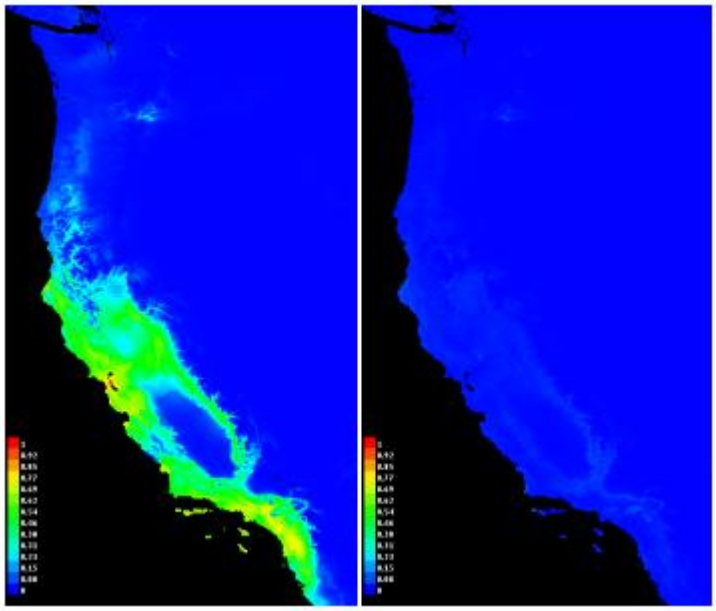
Average Omission and Predicted Area Graph: This graph is basically showing the predicted accuracy for the model you ran. The average omission should be linear to feel confident in your model. This means that when it is testing your data, it is not leaving information out and more accurately predicting what data should be included. A slope of 1 means that the "choices" being made by the program are random and are not purposely leaving out important information.



Average Sensitivity and Specificity Graph: This graph again is to help you determine the reliability/accuracy of your model. The AUC should be greater than 0.8. This identifies the Area Under the Curve and identifies how much of the graph is below the curve. This represents how well the model can predict whether a point provided from the data is actually a true point or not. Did the model select and use the best points.

Map Image: This is the graph you really want to look at. The image to the left shows the predicted suitability and habitat of the organism based on the model .

The right image is the standard deviation.



Response Curves: These are again determining the reliability of the model. These should standard deviations and relationships between the different variables from the climate data. You can research more, if you choose, but these are really for verifying the quality of the model.

Analysis of Variable Contributions: This data breaks down which of the variables were the most related to the final modeling of the 19 potential variables from the WorldClim data.

ASSIGNMENT

Write a two-paragraph essay on:

3. An analysis of your data
4. What are potential uses and/or benefits of Ecological Niche Modeling

ADAPTATION/EXTENSIONS

Confirmation Runs: To provide additional support for your hypothesis, you will likely want to run secondary, confirmation runs of the data. You can use the average raw data from the ASC file that was created in the output folder to do this and compare the outcomes.

Effects of Climate Change: If you want to see how a species might change based on climate change, go back to the climate layer processing section and select the future data from WorldClim. You have

choices between various climate change models projecting into 2050 and greenhouse gas concentrations. You can run predictive models on these and compare to the current conditions model to see differences.

Help: There are some tutorials at https://vimeo.com/idigbio created by Blaine Marchant, University of Florida