A MaxEnt Model v3.3.3e Tutorial (ArcGIS v10)

Last Modified on September 1, 2011

This tutorial was created for internal and educational purposes at the Natural Resource Ecology Laboratory at Colorado State University and the National Institute of Invasive Species Science.

Nick Young*, Lane Carter and Paul Evangelista

*Corresponding author: neyoung@nrel.colostate.edu

Introduction

In this tutorial, we demonstrate the application of the maximum entropy modeling or MaxEnt model (Phillips et al. 2006) for predicting the distribution of Alligator Weed (*Alternanthera philoxeroides*), a southeastern US invasive plant. An important component of MaxEnt modeling is the data preparation, which requires the understanding of several other software and file formats including Microsoft Excel©, ESRI ArcGIS©, and Notepad©. This tutorial will help guide users on how to format data using different software prior to running the MaxEnt model. Additionally, we provide sample data on Alligator Weed distribution that was downloaded from the archive at the Natural Resource Ecology Lab at Colorado State University. We encourage users to send us comments and suggestions on how to improve this tutorial. For more detailed instructions on how MaxEnt operates, interpreting results, and advanced modeling options, users should refer to www.cs.princeton.edu/~schapire/maxent/ .

The data used in this tutorial are for educational purposes only, and in some cases may lack the quality required to produce accurate and precise results.
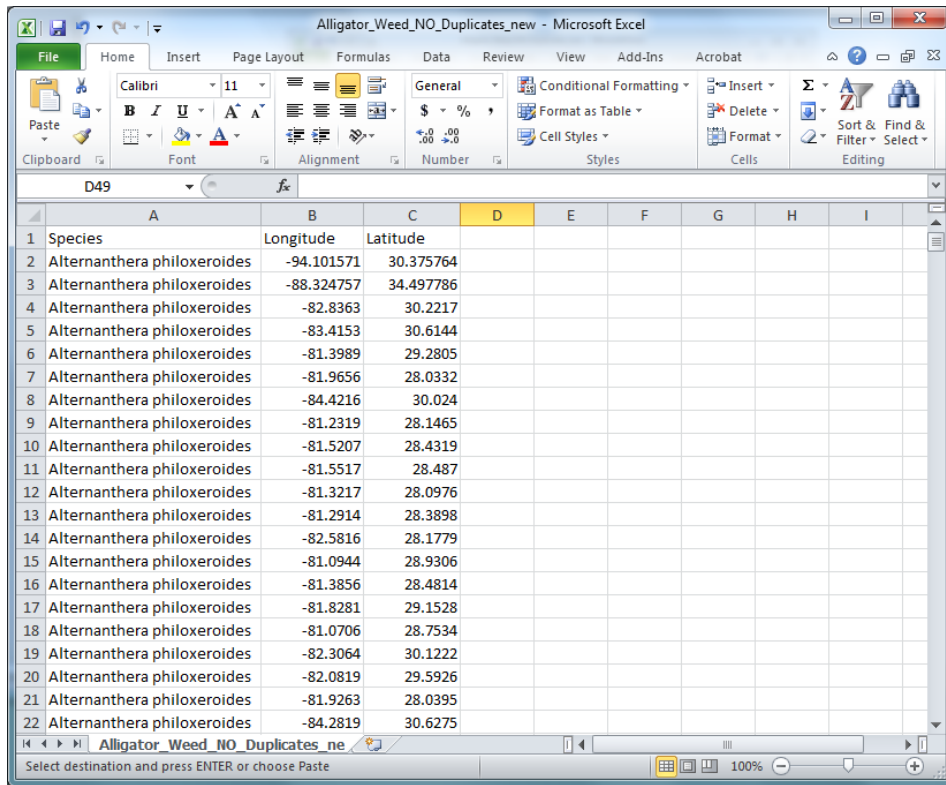
Part 1: Producing a Comma-Separated Value (.csv) File from Species Occurrence Coordinates in an Excel Spreadsheet

In this tutorial, you will create a comma-separated value (.csv) file from longitude and latitude coordinates of Alligator Weed (*Alternanthera philoxeroides*) from an Excel spreadsheet. A .csv file is the required format for the input samples in MaxEnt.

Step 1: Setting up an Excel spreadsheet

In Excel, you will first need to add your data into three columns. This may also be done in Notepad or Wordpad, but Excel has some operational features that may prove beneficial when working with large datasets. The first row of your datasheet is a header line, and should have three column headings titled: **Species, Longitude,** and **Latitude**. Longitude and latitude can be

substituted with Easting/Northing, X/Y, etc. Under these headings, your data should be entered or copied. If your data is in a shapefile format, Excel is able to open .dbf formatted data.


IMPORTANT! Before you continue, make sure the coordinates you are using are in longitude and latitude for this example! MaxEnt can handle most coordinate systems provided that the .csv file coordinates match the coordinate system of the spatial data layers.

When you have completed the data compilation, save the file in your working folder (e.g. all of the steps following will save documents under the Alligator_Weed working folder).

Step 2: Converting your Excel file to a csv file

Once your data in the Excel file are ready, you will need to convert and save it in your working folder as a .csv file. In our case, the original file is labeled Alligator_Weed.xls

To convert this file to a csv file click **File** then **Save As.** A window will appear. Save the file in the Alligator_Weed folder and under **Save as Type**, choose **CSV**
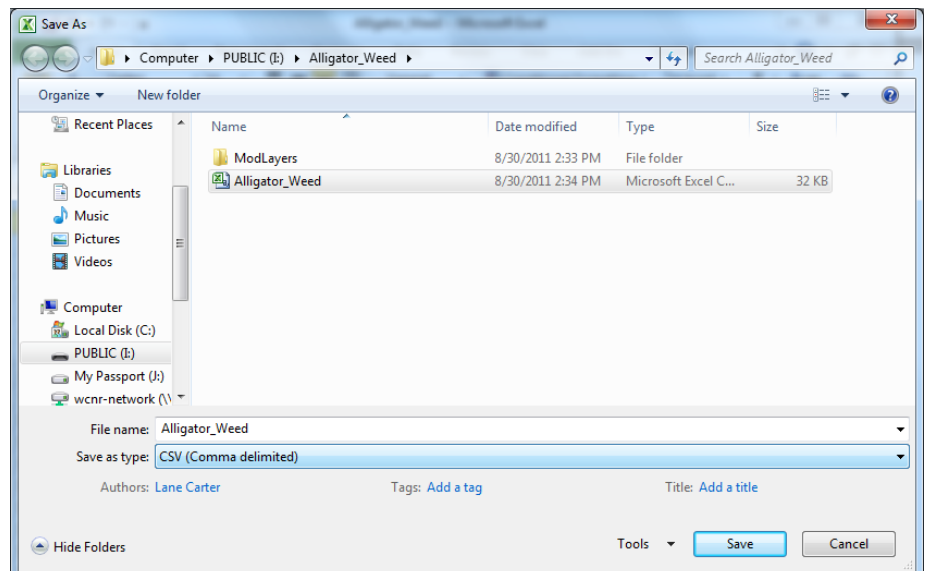
(Comma delimitated) (*.csv). Then click **Save**.

Now you have created a .csv file that is ready to be used with MaxEnt. The new file should be in your **Alligator_Weed** folder. You can check to see if the conversion worked by making sure that there is a .csv at the end of the file name.

Well done! You have created a comma separated value (.csv) file from coordinates in an excel spread sheet!

Part 2: Modifying Environmental Layers to be the Same Extent (geographic bounds and cell size) Using ArcGIS

Background:
This part of the tutorial walks you through the steps required to modify environmental layers in ArcMap so that all your spatial data (i.e. independent or predictor variables) have the same extent (same geographic bounds and cell size). MaxEnt requires all the environmental layers be in raster format and have the exact same cell size, extent and projection system (e.g., geographic or UTM) in order to execute a model.

**Step 1: Loading layers into ArcMap and opening the Extract by Mask tool**

Before you start modifying layers, make sure you have adequate space to work with large files. You will need at least 6 giga bytes (GB) of free space to complete these modifications. Also, you may want to create a new folder to store and separate the modified layers. Create this folder in the **Alligator_Weed** folder and call it **ModLayers**.

Open a new map in ArcMap. Click on the add data icon [icon] and navigate to the location of your environmental variables. In this example they are within a folder called **BioRasters**. Add all the environmental layers to the map. In our example, this will be the bio_1, bio_2, bio_3 etc. (The environmental layers are coded to correspond with the list of variables at the end of this tutorial) ArcMap may pop up a window that says if you would like to create pyramids, click **Yes**. This will help display the raster layers quickly, but it might take a while for ArcMap to load the raster layers.

First, you will need to turn on the spatial analyst toolset. Click on **Customize** at the top of your ArcMap screen, and scroll to **Extensions.** An extensions window will appear. Check the **Spatial Analyst** box and click **Close**.
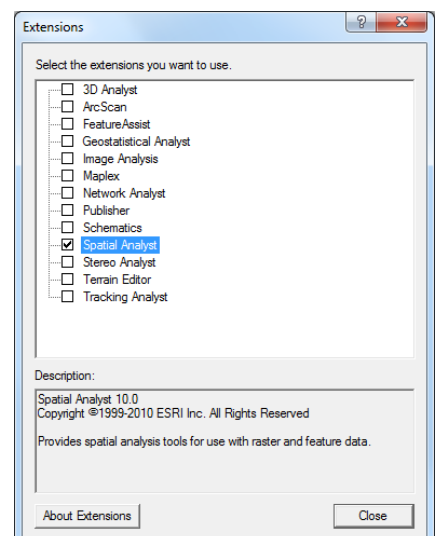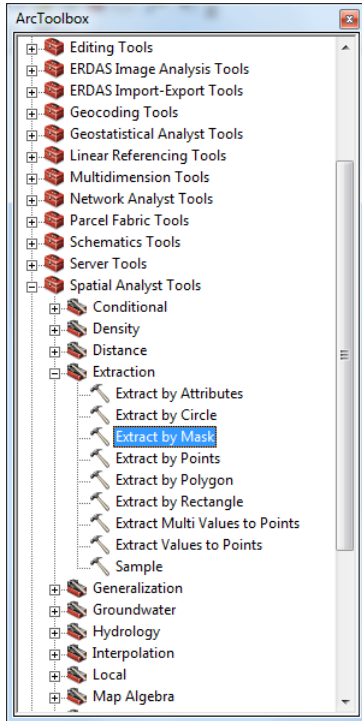
Click on the **Toolbox** icon [icon] and from the list of tools extend the **Spatial Analyst** tool. Under **Extraction** select **Extract by Mask.**

Step 2: Setting up Extract by Mask and the extent and cell size in spatial analyst

If your environmental layers happen to be larger than the area you are interested in modeling, this step allows you to clip them down to your area of interest and set all layers to have the same extent, cell size, and coordinate system (a requirement of MaxEnt). This step can also be used to set your extent, cell size and coordinate system if your environmental layers are already clipped.

A window should appear that looks like the one below for the **Extract by Mask** tool (location described above).

For the **Input raster** click on the folder icon [icon] and browse to the **BioRasters** folder and select the bio_1 variable.

For the **Input raster or feature mask data** field, use the [icon] icon to browse to a raster or polygon that represents the spatial boundary you intend to model.

Note: For this Alligator Weed example, the species is known to be most prominent in the southeastern states, therefore the mask is a shapefile named SE_states.shp that was already created.
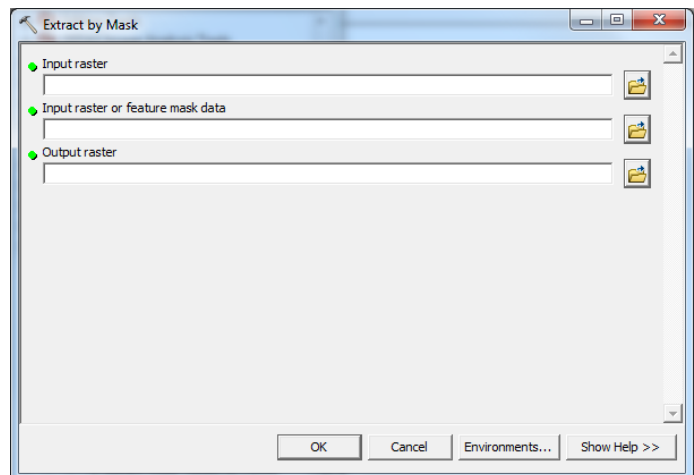
Save the **Output raster** as bio_1 under the folder **ModLayers** you created previously. This is where all the modified layers you create from this step will be saved.

In order to set the coordinate system, cell size and extent, click on the **Environments…** button in the bottom right of the window*. Extend the variables **Output Coordinates**, **Processing Extent**, and **Raster Analysis**.

*Note: If this is the first environmental layer that you want to clip, you do not need to set the **Environments…**The first clipped layer you create will become the layer you use to define your environments for all other environmental layers in the model (in this example we first used the boundary layer to mask bio_1 which will then be used in the environments for the remaining environmental layers [bio_2-bio_19]). It may be a good idea to use the Environments for the first environmental layer if it is not already in your intended coordinate system.

The window to the right is displayed:

Use the icon to set the **Output Coordinates** variable to bio_1 in the **BioRasters** folder. The name of the coordinate system will appear in the window below the input field.
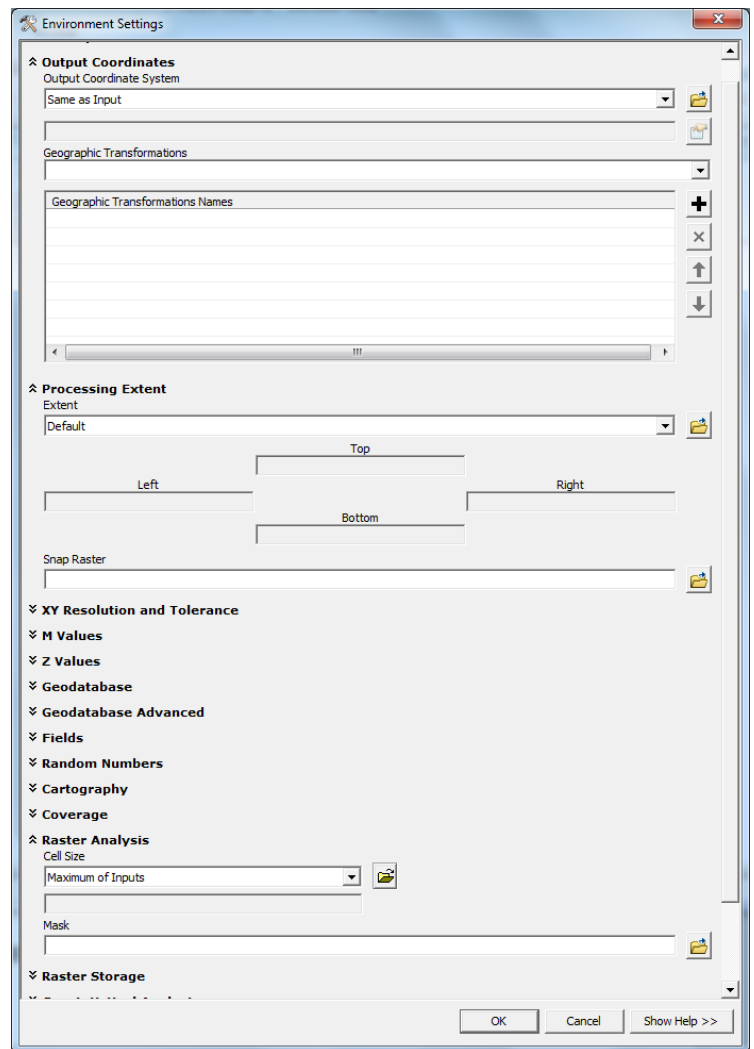
For the **Processing Extent** and **Snap Raster** variables, browse to a raster you know that already has the extent you wish to match. In this case it is bio_1 as well.

For the field of **Cell Size** under the **Raster Analysis** variable, you will want to enter your environmental layer that has the smallest cell. In this case we will choose the same layer the other variables: bio_1. The true cell size value will appear in the window below the input field. It will reflect the units from which your coordinate system is based.

Click **OK**.

By the end the windows should look something like this:
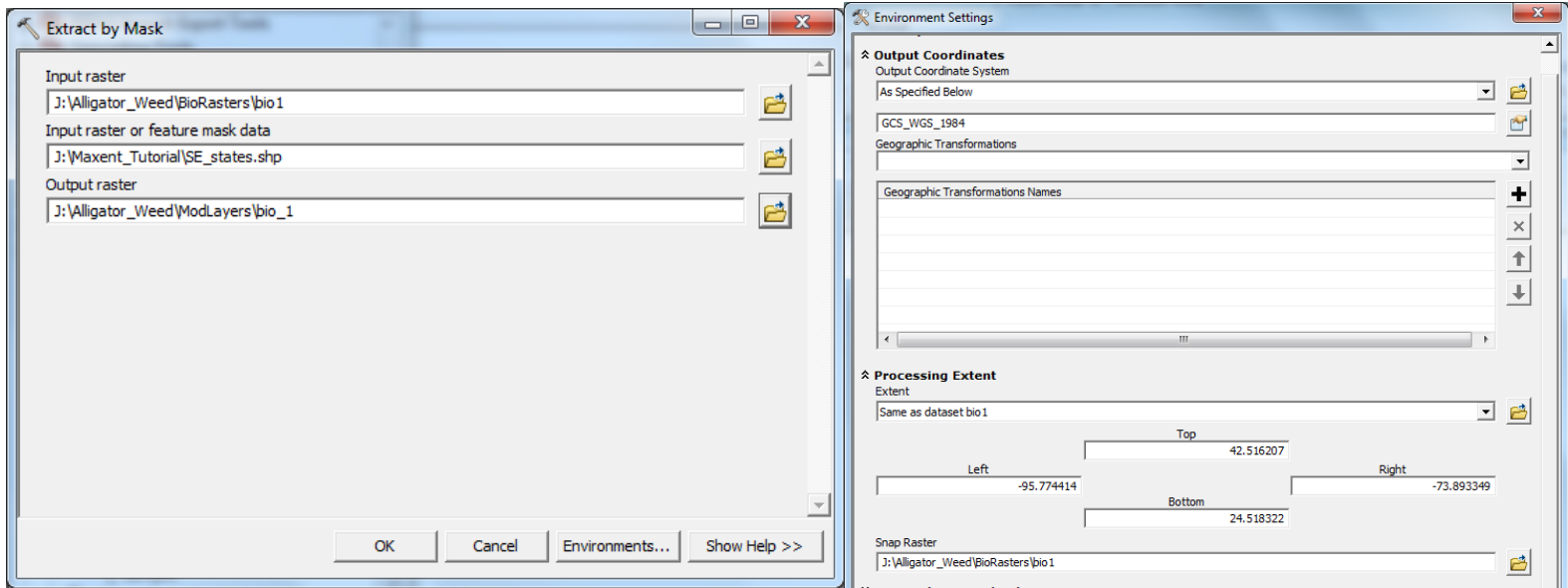
Now Click **OK.**

Repeat this process until all of your environmental rasters are extracted to the extent and contain the cell size and coordinate system you intend.

IMPORTANT! Within the **Environments**, you must use one consistent layer to set the coordinate system, extent, and cell size to ensure all are exactly the same. Here the layer bio_1 is used for all environmental layers to match.

Congratulations! You have modified all your environmental layers to be the same cell size and extent. This can be most time consuming part of the data preparation.

Part 3: Converting Environmental Rasters to ASCII Format

This section demonstrates how to convert environmental raster layers to ASCII format. All environmental layers must be in ASCII format to run MaxEnt.

Step 1: Loading raster layers into ArcMap and opening the Raster to ASCII tool.

Before you start, create a new folder in the **Alligator_Weed** folder labeled **ASCII_Environmental_layers**.
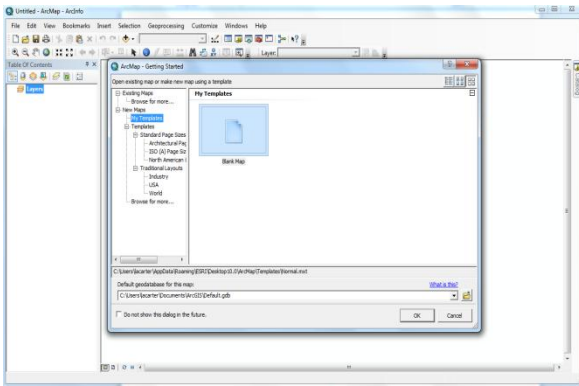
Open ArcMap and select a **New Empty Map**.

Click **OK.**

Click on the add data icon and navigate to the **Alligator_Weed** folder, then to the **ModLayers** folder and select one of the environmental rasters that has already been modified to be the same extent and cell size as the other layers (i.e. bio_1).

Click **Add.**

Open the **Toolbox** window by clicking on the toolbox icon .

In the Toolbox window double click **Conversion Tools** then **From Raster,** and then double click on the **Raster to ASCII** tool.

Step 2: Converting raster layers to ASCII files.

For the **Input Raster** click on the folder icon and navigate to the bio_1 file in the **BioRasters** folder.

For the output ASCII raster file, click on the folder icon and navigate to the new folder you created in the **Alligator_Weed** folder labeled **ASCII_Environmental_layers**. Save the ASCII layer as the same name as the original layer name bio_1.

IMPORTANT! Make sure the **File (*.ASC)** is selected in **Save as type.**

Click **Save.**

7
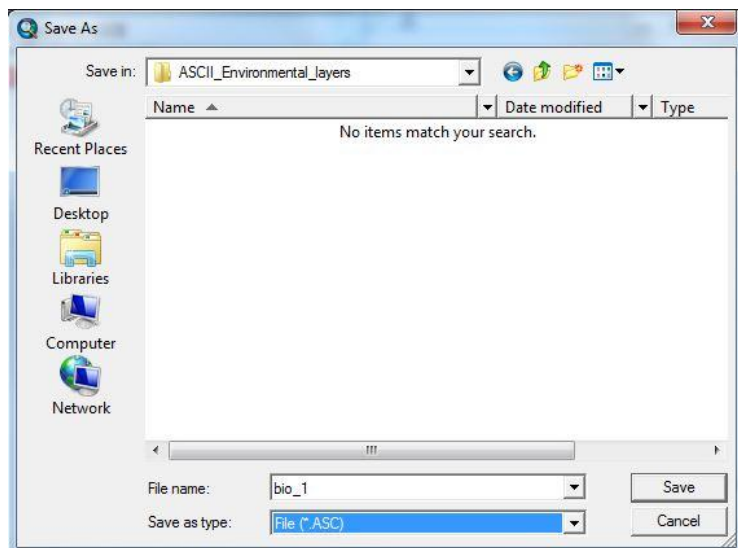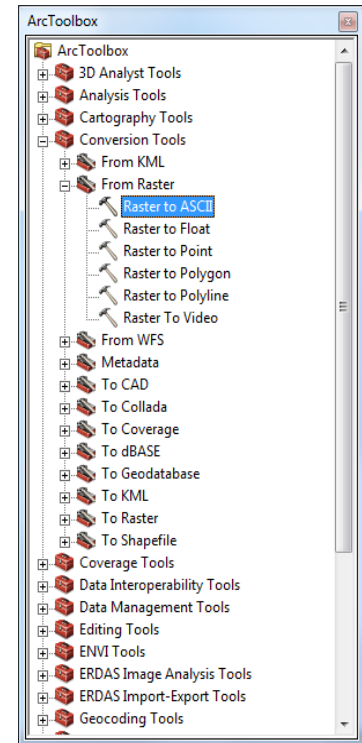
The **Raster to ASCII** window should look like this:
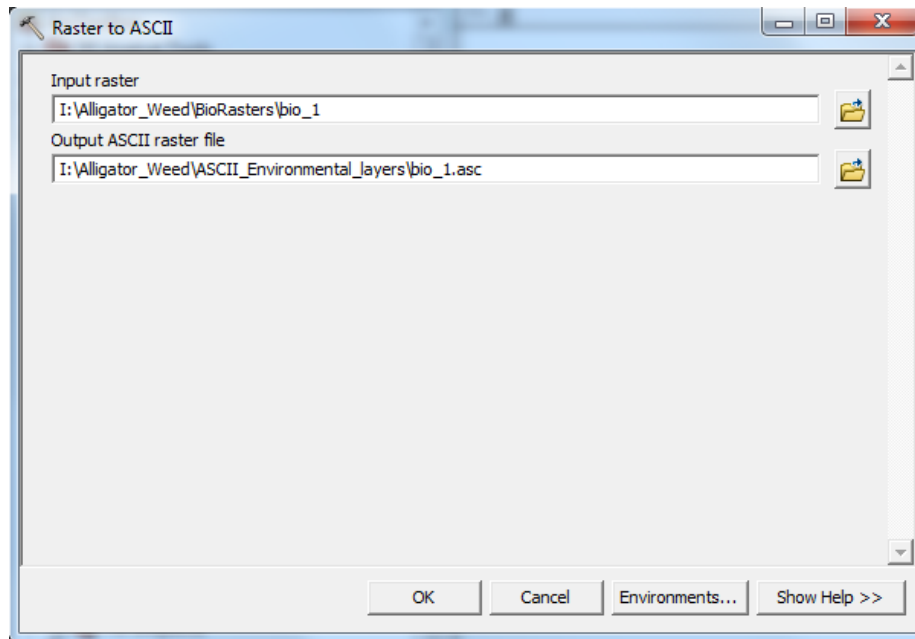
You may need to manually change the file name to .asc instead of .txt or .asc.txt. If the file is not is .asc format, MaxEnt will not be able to read it.

Click **OK**.

A progress scroll will appear as it runs the conversion. This can take a while.

Now you should have a file called bio_1.asc in the **ASCII_Environmental_layers** folder. Repeat the above steps for the remaining raster layers that need to be converted to ASCII for the MaxEnt model.

Congratulations! Now you have created ASCII files from your environmental layers that you can enter into MaxEnt.

Part 4: Running the MaxEnt Model

In this part of the tutorial, you will learn how to download and run a MaxEnt model. This tutorial will also explain where to find MaxEnt on the web and how to download the module. MaxEnt, a machine learning program (Phillips et al. 2006), is a powerful tool used to predict species' spatial distribution (current and potential) using presence point locations and environmental layers. This tutorial covers the minimum information needed to run the MaxEnt model. Additional model capabilities and information can be accessed at MaxEnt website (www.cs.princeton.edu/~schapire/maxent/).

Step 1: Downloading MaxEnt

To begin, go to the website http://www.cs.princeton.edu/~schapire/maxent/ to download the MaxEnt software. On this page, you will find links to two papers that may help inform you more on mathematical and theoretical aspects of how the MaxEnt algorithms work. Further down the page under **Terms of Use**, you will see where to enter your *Name*, *Institution*, and *Email* address. Fill in the requested information, and select **version 3.3.3e** (or latest available version).

When you are finished, click the **Accept terms and download** button.
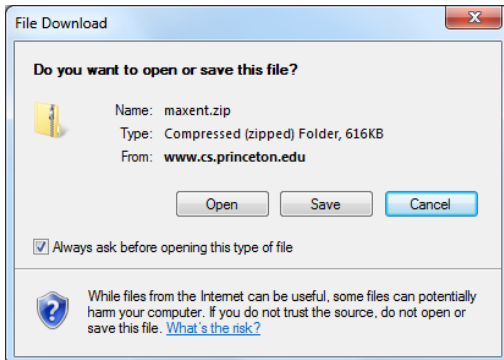
A new page should come up that explains how to download the software. You can download each of the MaxEnt's three files separately (i.e. maxent.jar, maxent.bat, readme.txt) or you can download all of them in .zip file.

Save the .zip folder to whatever workspace you are working off of.

Once your download is completed, you should see the three MaxEnt files in your folder.

To open the MaxEnt model, go to this folder and click on the maxent.bat file.

Step 2: Defining MaxEnt Background Selection

This section describes how to define how MaxEnt selects background samples (also called pseudo absences) or how to provide MaxEnt with absences. These samples will be used to estimate the environmental layers across the entire extent or landscape used to model the species distribution. The background samples used when developing a distribution model can have significant impacts on the model results (for more information on this topic see Elith et al. 2011).

Often we are modeling species within the United States. One method used to limit where the background points are selected from is to only let MaxEnt select from counties where you have sample locations. This limits the background point to areas that we assume were surveyed for the species which provides MaxEnt with a background file with the same bias as the presence locations.

Often, we may be modeling a species distribution in another country or at small extent where the use of counties to select background points from does not make sense. In these cases, it may be better to use the minimum convex polygon to define the area where MaxEnt will draw its background samples from. For the purpose of this tutorial, we will not describe in detail how to do this, but it important to be aware of this technique.

Another method to select background samples is to provide your own background points to MaxEnt. This serves two purposes: (1) you select the specific background locations MaxEnt uses to develop the model and (2) this will speed up the MaxEnt model process. Deciding what background samples to include in the model can be difficult. The best suggestion is to use presence points for another species that was part of the same survey as the focus species.

Which method you use for background samples for MaxEnt depends on the data available, the geographic characteristics of the area, and the size of the extent you are modeling.

For this example, we will select background points from the counties where we have presence data for alligator weed. To accomplish this, you will have to clip all the environmental layers you use for the model (see Part 2, step 2 for details) to the counties that have alligator weed presence locations. You will then use the clipped environmental layers to develop the model and project the model (more on this in step 3) to the extent you wish to model the distribution of Alligator Weed. The steps to create the clipped environmental layers are outlined below:

First, open ArcMap10 and load your Alligator_Weed.csv file of the locations points and make it a shapefile by right-clicking on the layer selecting **Data…** then select **Export Data**. Save the shapefile in the Alligator_Weed folder and name it Alligator_Weed.shp. Make sure to add the new layer to the map. Now add the US_counties_GCS_wgs84.shp to the map located in Alligator_Weed.
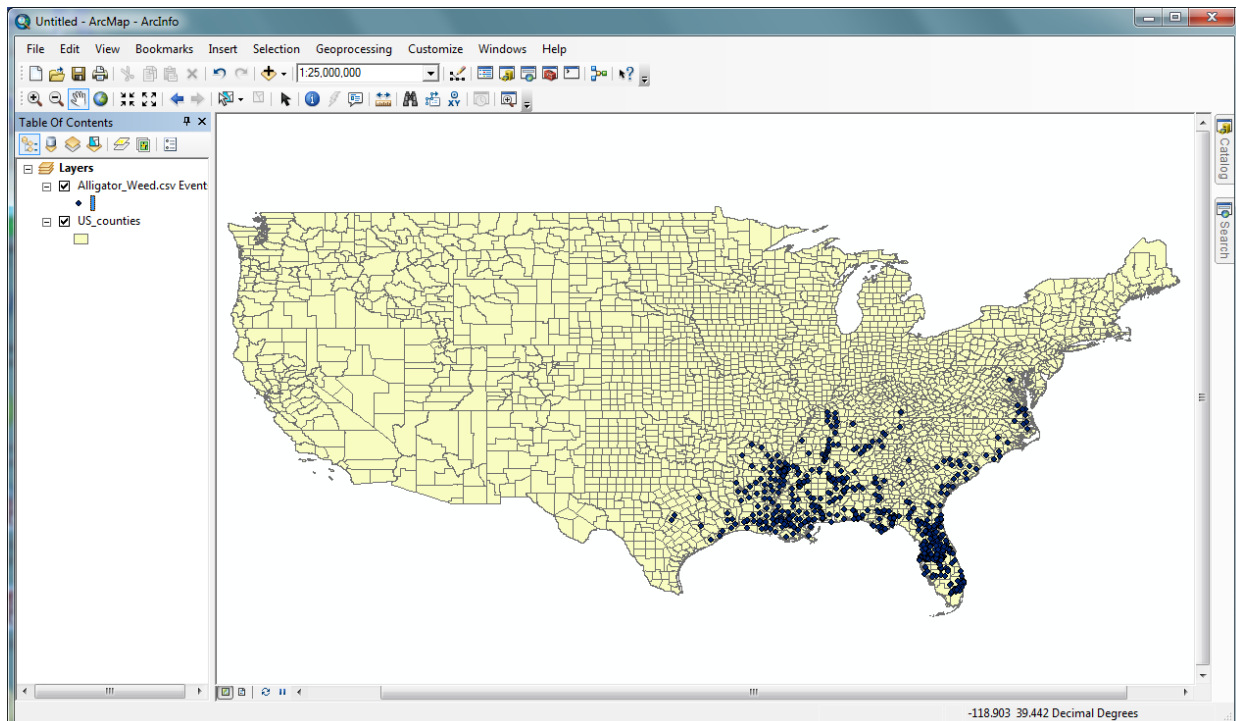
We are now going to select only the counties that have a recorded presence within them. To do this, go to **Selection** in the top menu bar then **Selection By Location…**
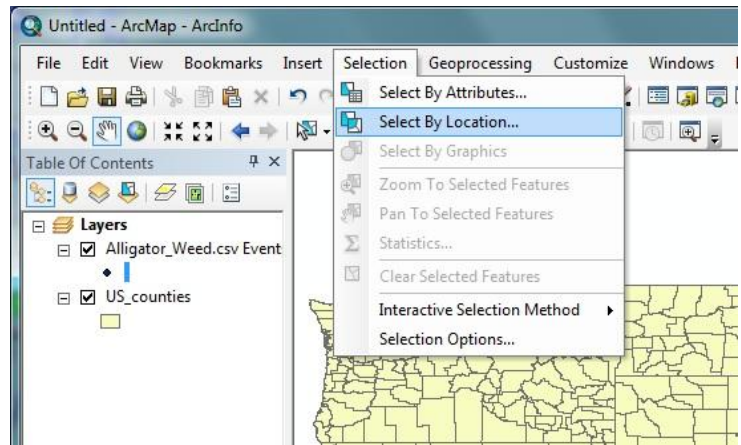
In the **Select By Location** window, select US_counties_GCS_wgs84 as the **Target layer**, make Alligator_Weed the **Source layer,** and choose **Target layer(s) features intersect the Source layer feature** as the **Spatial selection method**. Leave the apply search distance unchecked.

Click **Apply**. Then click **OK**.
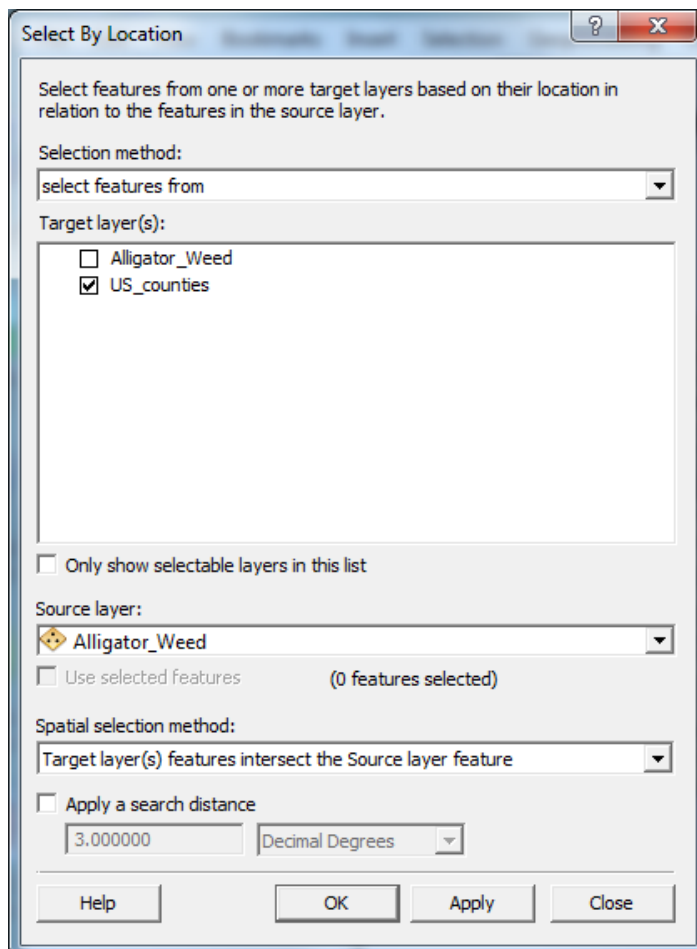
The counties with presence points should be highlighted. To create a new layer with these counties right click the US_counties_GCS_wgs84 layer in the table of contents and select **Selection** and then **Create Layer From Selected features**.

A new layer should appear in the table of contents called US_counties_GCS_wgs84 selection.

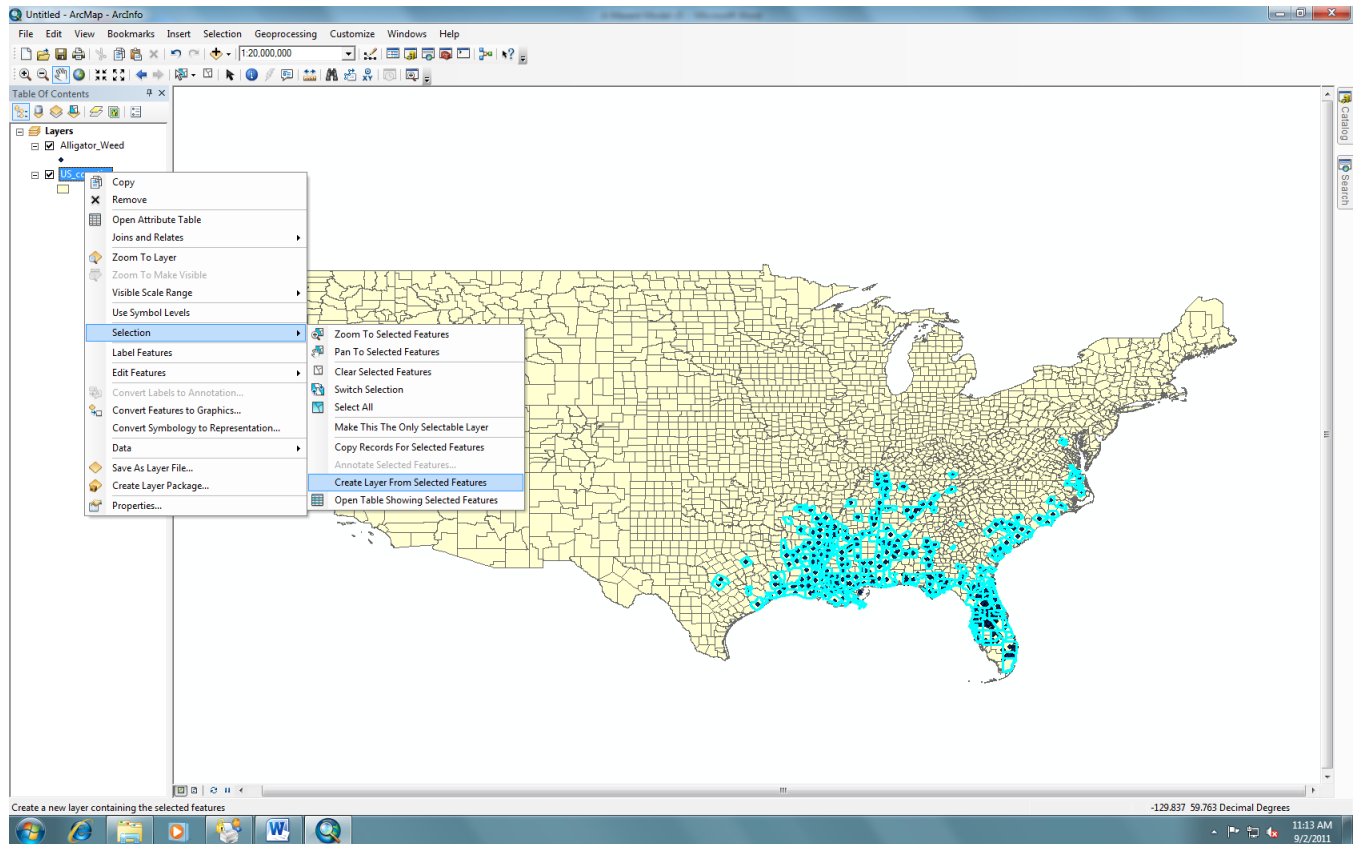Now we will convert the new layer to be a raster. To do this go to **ArcToolbox** then **Conversion Tools** then **To Raster** and select the tool **Polygon to Raster**. Select US_counties_GCS_wgs84 selection for the **Input features**, save the file in **Alligator_Weed** as bias_file and set the **cell size** to be the same as the bio_1 grid file in the **BioRasters** folder. Leave the other fields blank and click on the **Environments…** button.

In the **Environments** window, set the **Extent**, **Snap Raster**, **Cell Size**, and **Mask** all to be the bio_1 grid in the **BioRasters** folder. Click **OK**.
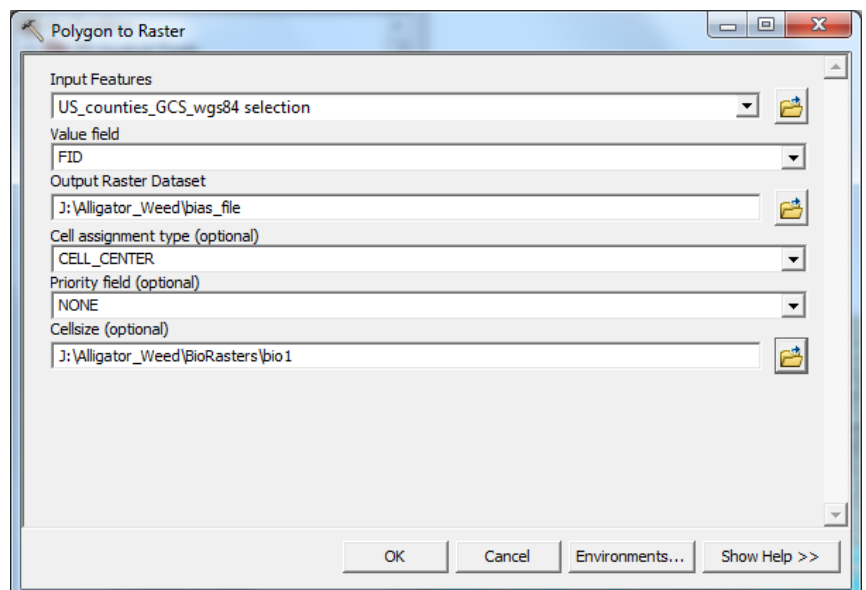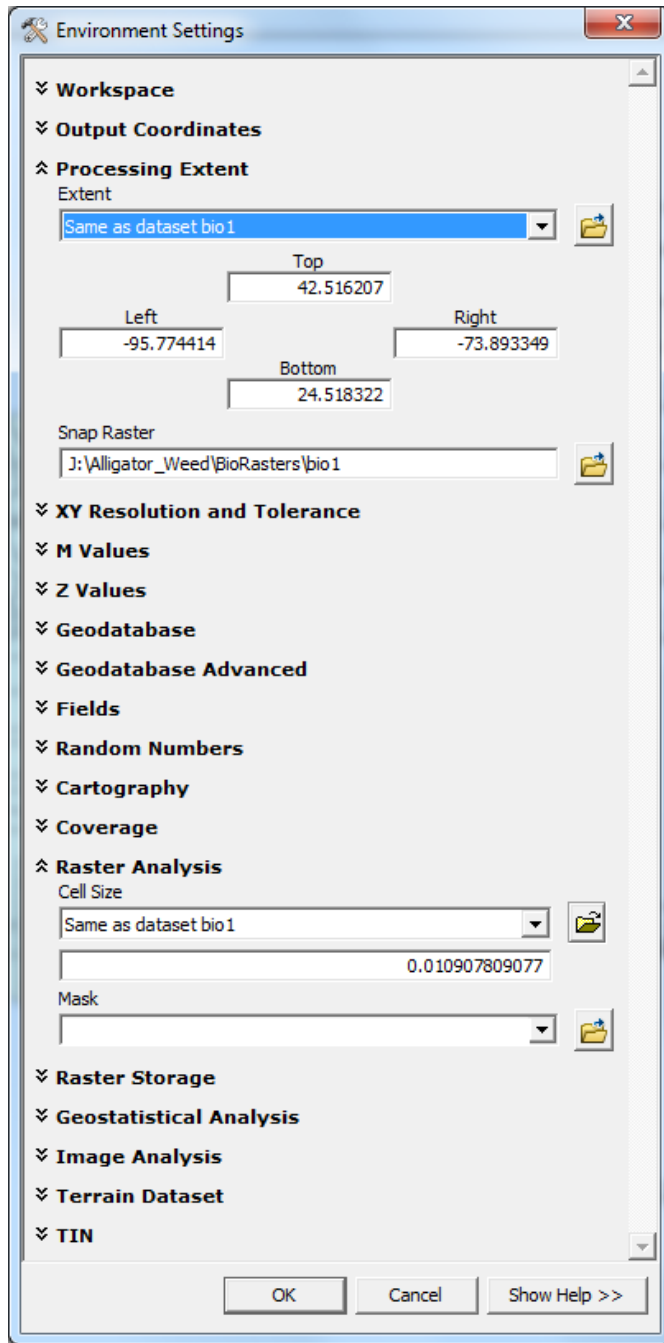
A new raster should appear that only covers the counties selected. We will want to convert the new raster to have a value of 1 in all selected counties and a value of "NoData" everywhere else.

To do thism, go to **ArcToolbox** then to **Spatial Analyst Tools** then **Map Algebra** and select **Raster Calculator**. Enter the following Con statement into the window and save the file under **Alligator_Weed** as bias_filecon:

Con Statement: *Con ("bias_file" >= 0, 1, "bias_file")*
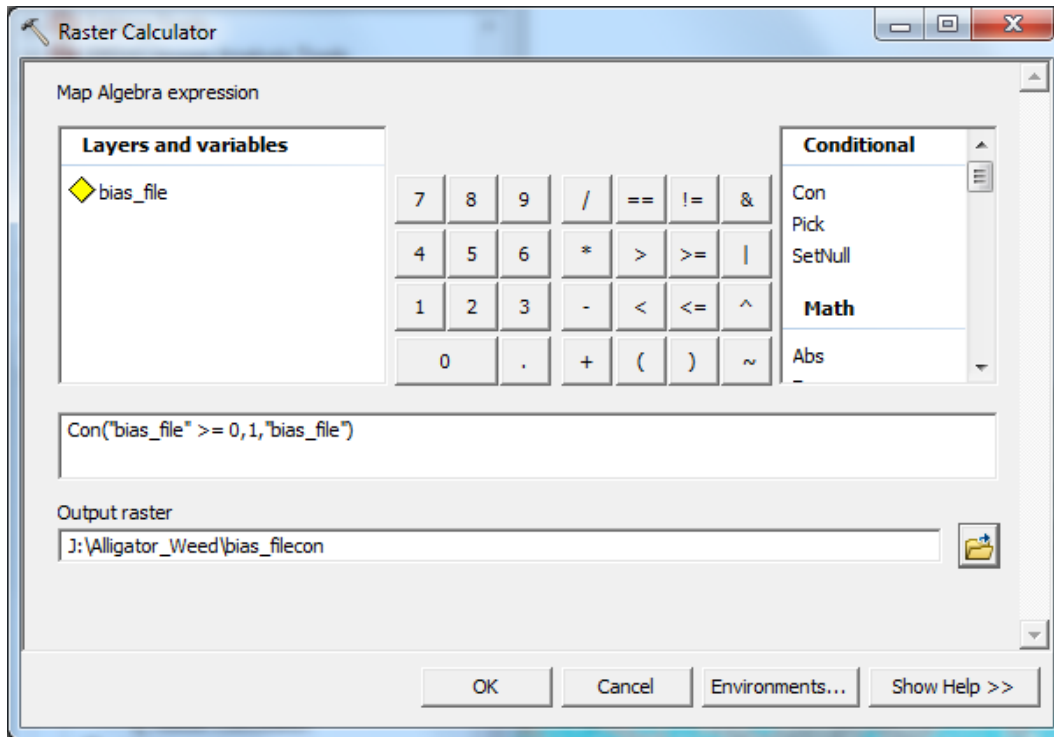
This will create a new layer in the table of contents. We will need to convert this layer to be an ASCII file. To do this go to **ArcToolbox** then **Conversion Tools** then to **From Raster** then select **Raster to ASCII**. Make sure to set the **Extent**, **Snap Raster**, **Cell Size**, and **Mask** all to be the bio_1 ASCII file in **ModLayers** that you created earlier. Save the output file as bias_file_asc. You will include this file in the **Bias** window in MaxEnt settings (see **Bias** in step 4).

Step 3: Opening MaxEnt and Setting Up a Run

Before starting MaxEnt, navigate to the **Alligator_Weed** folder and create a new folder labeled **Outputs**. This will be the folder where the outputs created by MaxEnt will be stored.

IMPORTANT! When environmental layers are very large, you may get an "out of memory" error when you run the program. The best way to fix this problem is to give MaxEnt access to more memory. To do this, you will edit the maxent.bat file and increase the memory from 512 to 1024 (or multiples of 512 such as 512x2, 512x3, 512x4 depending on your computer's RAM). Memory size of 512 indicates the 512mb of your computer's total RAM will be allocated for running MaxEnt model.

To do this, first right click on the maxent.bat file in your working folder and select **Edit**.

A Notepad window should appear like the one shown on the right.
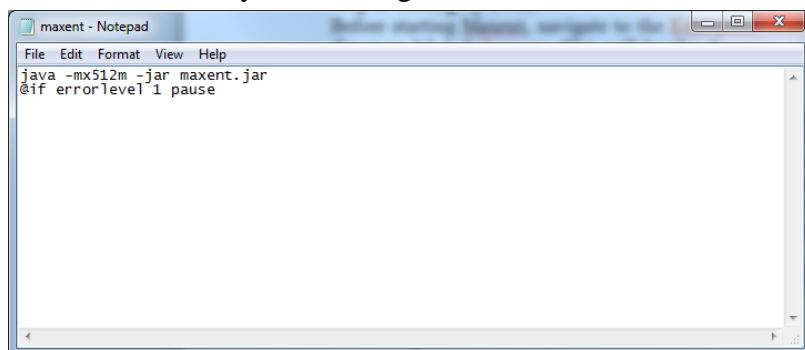
You can change the amount of memory that MaxEnt uses by simply changing 512 to 1024. This is how the window should look.
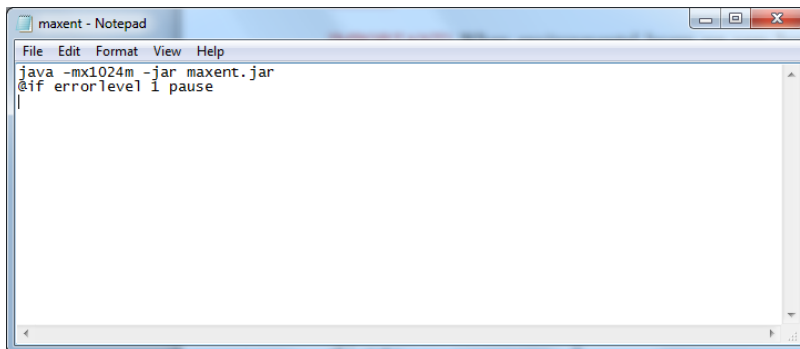
Once you have changed the memory, click **File** and **Save** and close the window.

To open MaxEnt, click on the maxent.bat file. A window should open that looks like this:

To begin, you must provide a **Samples** file. This file is the presence localities in .csv format (see Part 1). For this demonstration, we will use Alligator_Weed.csv within the **Alligator_Weed** folder. Navigate to this file by clicking the **Browse** button under **Samples**, or you can type in the file path.
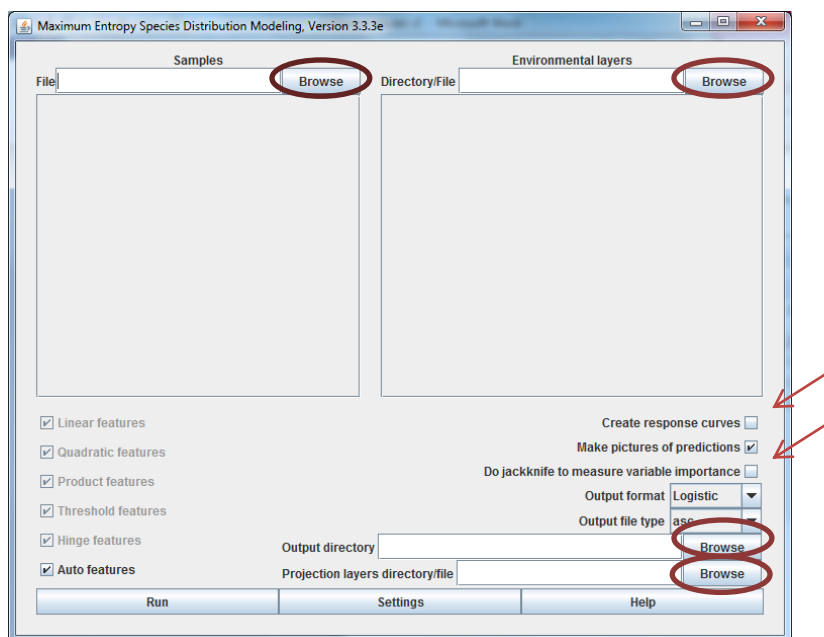
Next you have to provide the **Environmental Layers** used for the model. This will be the folder that contains all your environmental layers in ASCII format (they must have an .asc file extension) with the same geographic bounds, cell size, and projection system. In our case, the folder containing our layers is found in the **ASCII_Environmental_layers** folder. Navigate to this folder by clicking the **Browse** button under **Environmental Layers**, or type in the file path. Notice how you can change the environmental layers to either continuous or categorical. If any of the layers you include in your environmental layers are categorical (e.g. vegetation type), make sure you change them by clicking on the down arrow and choosing categorical.

An **Output** folder also needs to be selected. This will be the folder where all the MaxEnt outputs will be stored. For this exercise, we will use the folder created earlier named **Outputs**. Navigate to this folder by clicking the **Browse** located next to the **Output Directory,** or type in the file path.

You can leave the **Projection Layers Folder/File** window blank if you do not intend on producing future scenarios. For the Alligator_Weed we will input one projected model for 2020.

Assuming these projected climate variables are already available, navigate to them by clicking on the **Browse** located next to **Projection Layers Folder/File.**

Depending on where your projected climate variables are located, the browse window will look similar to this:
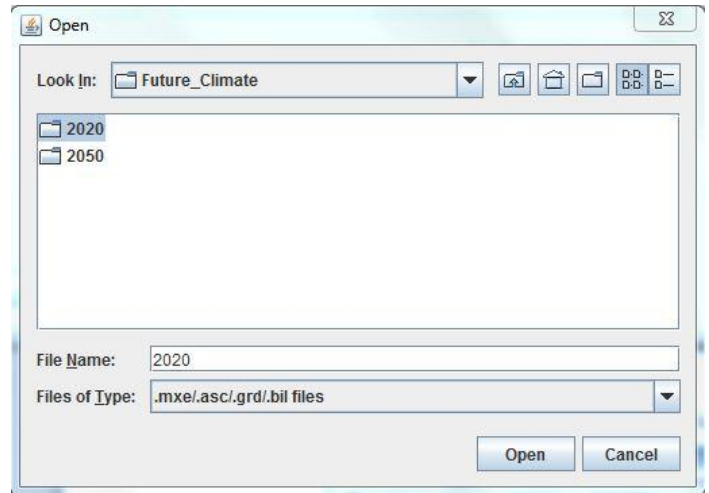
To add a second or third projection, one cannot navigate using **Browse.** Instead, place a comma at the end of the first file path, enter a space, then type in the second path name (if your projected variables are within the same folder simply copy and paste the first path name but be sure to change the last file name from **2020** to your new projection file name).

Make sure that the **Create Response Curves**, **Make Pictures of Predictions**, and **Do jackknife to Measure Variable Importance** boxes are all checked.

Keep the **Auto Features** box checked and leave the **Output Format** as **Logistic** and the **Output file type** as **.asc**

Now, the MaxEnt GUI (graphic user interface) should look like this:
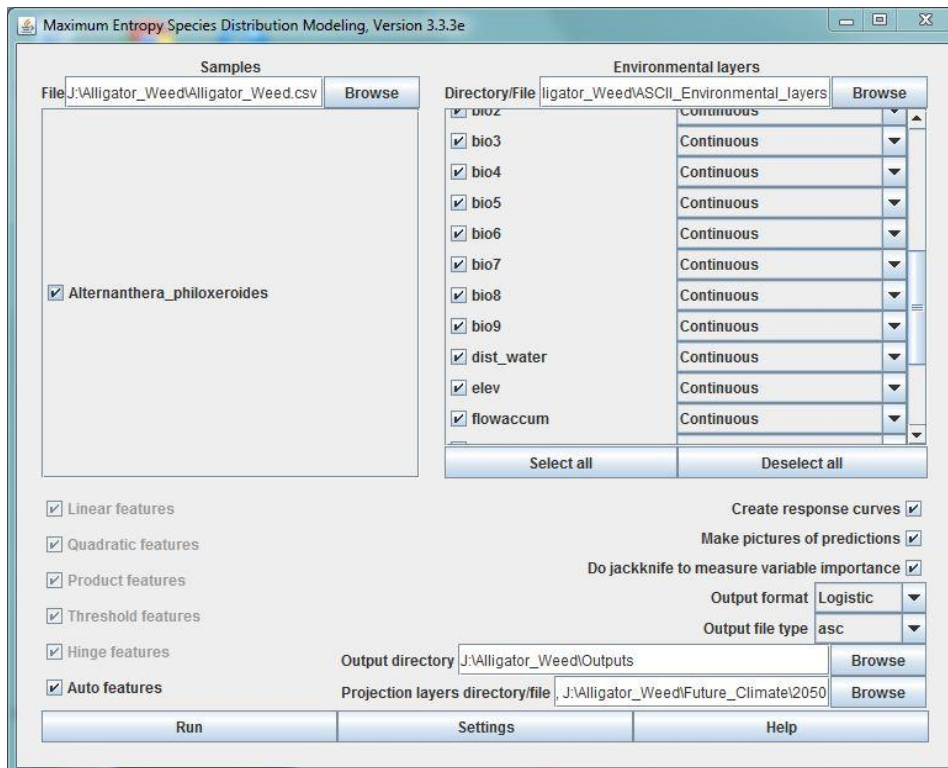
Step 4: MaxEnt Settings

Replicates (Number of Runs)
MaxEnt allows the ability to run a model multiple times and then conveniently averaging the results from all models created. Using this feature in combination with withholding a certain portion of the data for testing (see Random Test Percentage below) ena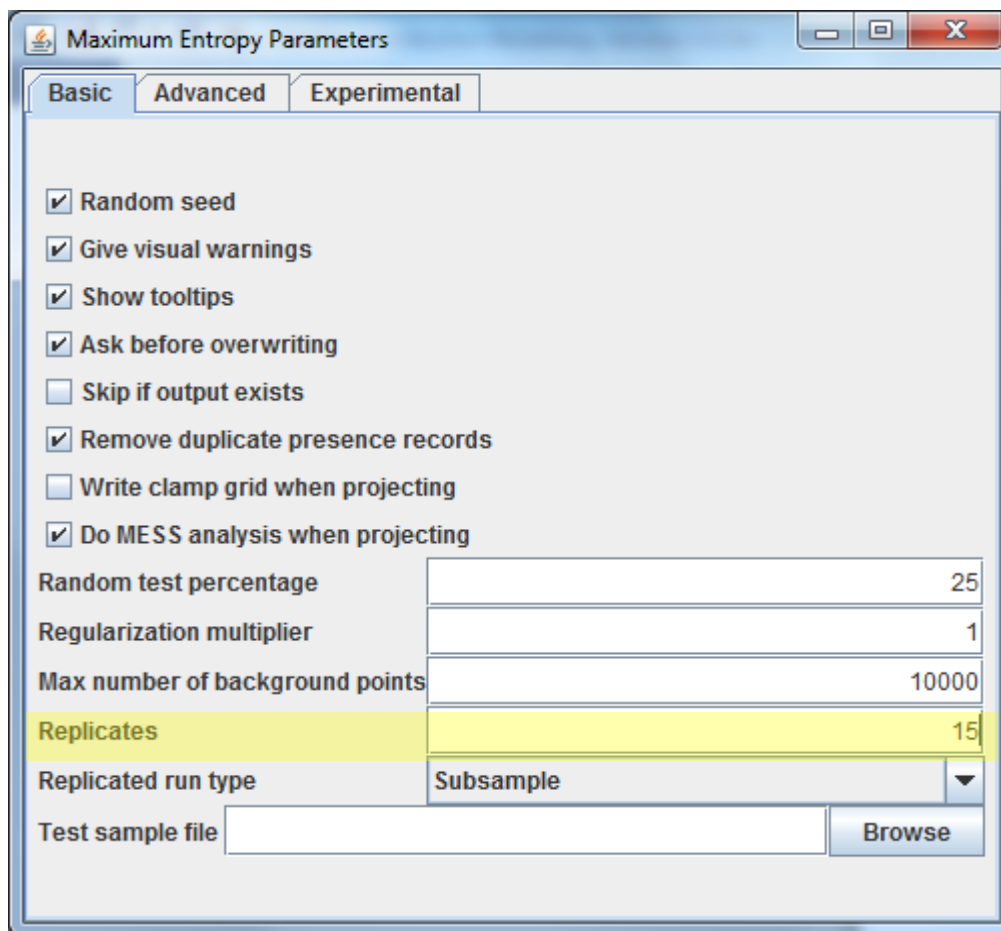bles the ability to test the model performance while taking advantage of all available data without having an independent dataset. Executing multiple runs also provides a way to measure the amount of variability in the model. To set the replicates in MaxEnt, go to **Settings** and then enter the number of run in the **Replicates** field. In this example we will use 15.

Reducing disk space and increasing speed (optional setting)
When the only output needed from a MaxEnt run is the averaged results from multiple runs (replications), you can change the model setting to turn off the "**write output grids.**" This will prevent MaxEnt from writing output grids from individual runs and only output the summary statistic grids (e.g. Average, Minimum, Maximum, etc.) from all the runs. This will speed up the

total run time and decrease disk space. You can turn off the "**write output grids**" option by going to **Settings,** selecting the **Advanced** tab and then de-check **write output grids.**
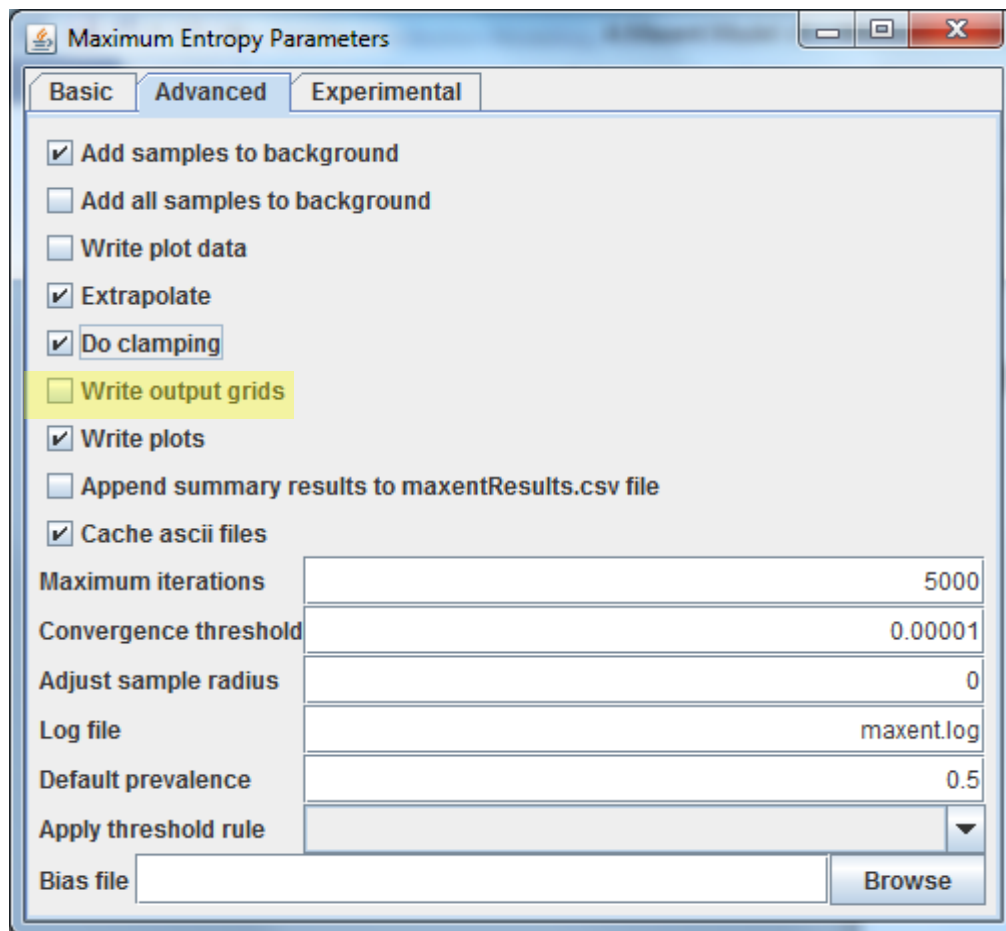
Random Test Percentage (Test data)
One way to evaluate model performance is to use the Random test percentage setting in MaxEnt. This setting allows you to withhold a certain percentage of your presence data to be used to evaluate the model's performance. This is important because without these test data, the model will employ data used to develop the model (also called training data) to evaluate the model. This is a bias method and will provide an inflated measure of model performance.

There are three different sampling techniques (replicate run types) that are available in MaxEnt; Crossvalidation, Subsampling and bootstrap. We use sub-sample for the majority of the models we run. To select the Subsample replicate run type go to **Settings** and chose **Subsample** in the drop down for the field **Replicated run type.**
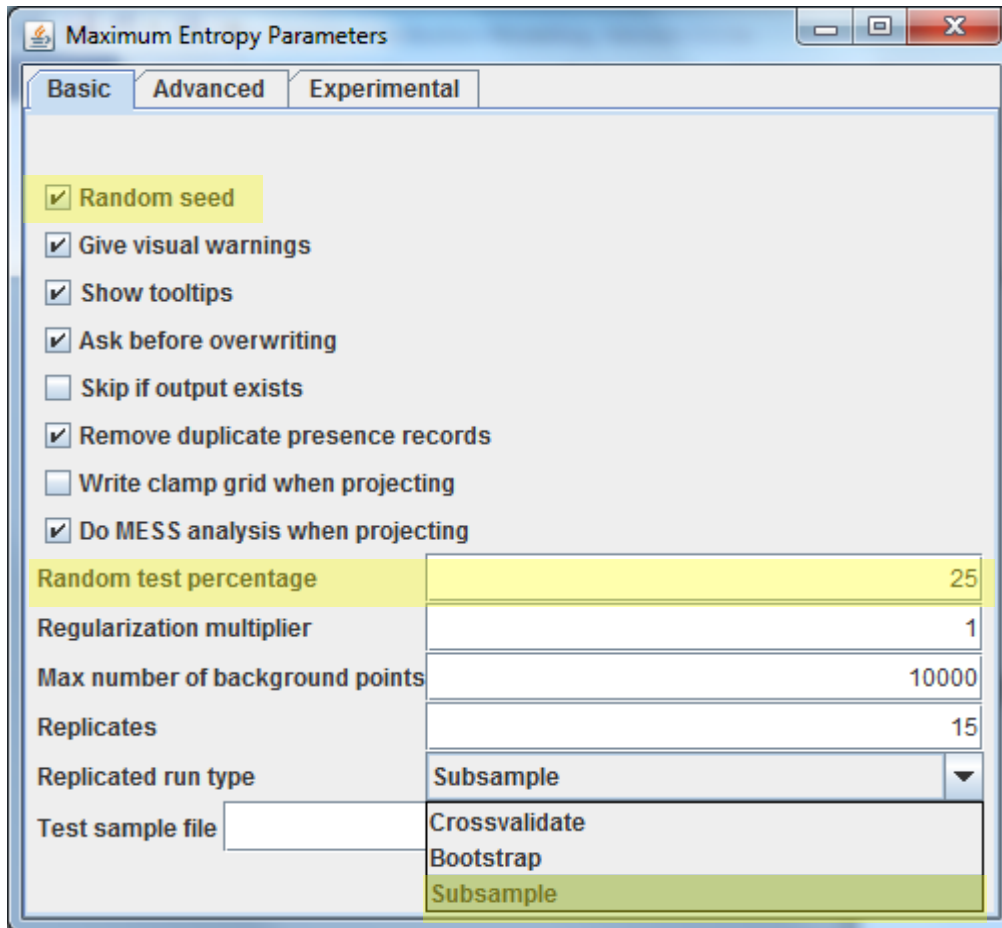
NOTE: You will need to check the "**Random seed**" box when using test data. If you forget to check this, MaxEnt will pop up an error and force you to check this box.

Number of Iterations (Convergence)
Normally set to 500, increase this amount to 5000. This allows the model to have adequate time for convergence. If the model doesn't have enough time to converge, (in the form of number of iterations) the model may over-predict or under-predict the relationships. To increase the amount of iterations go to the MaxEnt **Settings** select the **Advanced** tab and then enter **5000** in the field **Maximum iterations**.
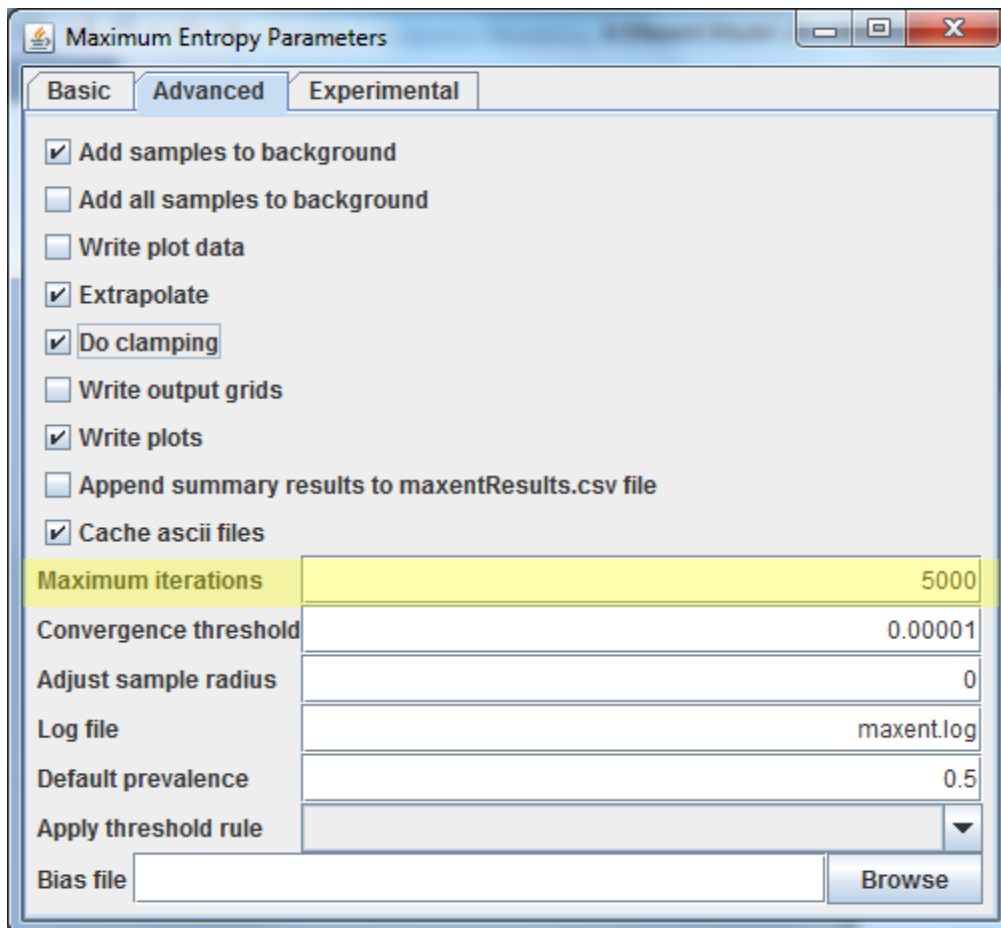
Regularization

Regularization provides a method to reduce model over-fitting and, when used with ENMTools, can help find the most parsimonious model (for more on ENMTools, see Warren and Seifert, 2011). Regularization can be thought of as a smoothing parameter, where larger values increase the amount of smoothing. For the purpose of this tutorial, we will leave the default setting at 1 for this value.

Bias

A bias file can be included in the run to represent sampling effort to reduce the sampling bias (see Philips et al., 2009 for more information on this feature). For this example, we will assume that only the counties with presence locations were sampled. We can represent this bias using the grid you created in Part 4, step 2. To add the bias layer, go to **Settings** and browse to the bias_file_asc.asc file for the **Bias file** field. This defines where MaxEnt selects the background points from.
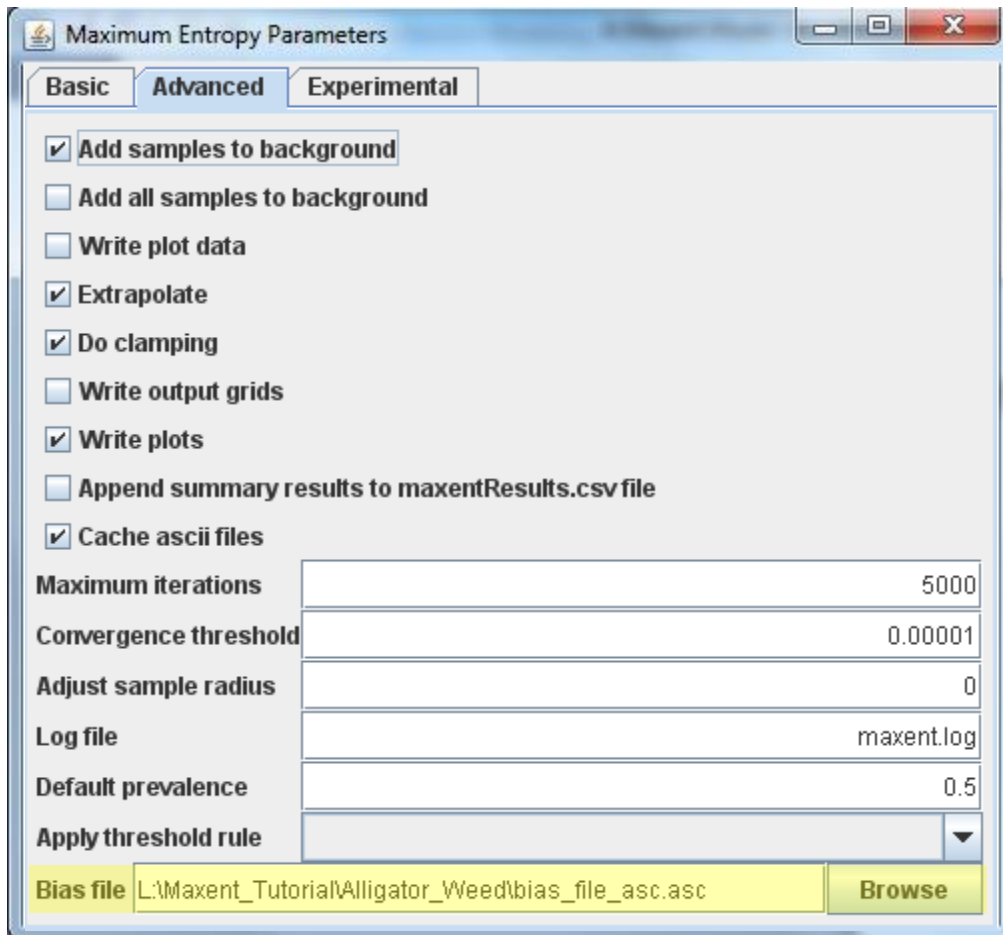
Step 5: Running the MaxEnt model

Now that everything has been entered into the MaxEnt program, simply press the **Run** button to begin modeling. A progress window will appear describing the modeling process.
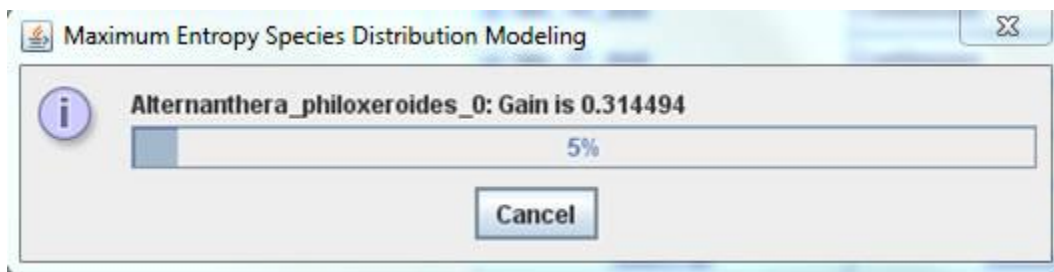
You will also be able to see the gain for each environmental variable while the model is running. The gain is similar to a measure of goodness of fit. Specifically, the gain is a measure of the closeness of the model concentration around the presence samples. So a gain value of 2, would translate to the average likelihood of the presence samples is exp(2) or about 7.5 times higher than that of a random background pixel.

Once the MaxEnt model has completed its run, the progress window will disappear. You will be able to find all the outputs created by MaxEnt in the **Outputs** folder you created earlier.

Congratulations! You have now completed a MaxEnt model and can start interpreting the outputs.

Part 5: Interpreting MaxEnt Outputs

This part of the tutorial explores the different outputs of MaxEnt. Many, but not all, of MaxEnt's outputs will be discussed and explained. The layers used for this tutorial were bio_1 - bio_19. These have corresponding environmental characteristics which include Annual Mean Temperature, Max Temperature of the Warmest Month, Precipitation of the Driest Month, etc. The full list of codes and their bioclimatic variables is attached at the end of this tutorial. For further explanation and interpretation of MaxEnt's outputs, users should refer to www.cs.princeton.edu/~schapire/maxent/ .

Step 1: Exploring MaxEnt Outputs

First, open the folder that contains the MaxEnt outputs (i.e. **Outputs**). When you open the folder, you should see another folder labeled plots, and a list of other files.
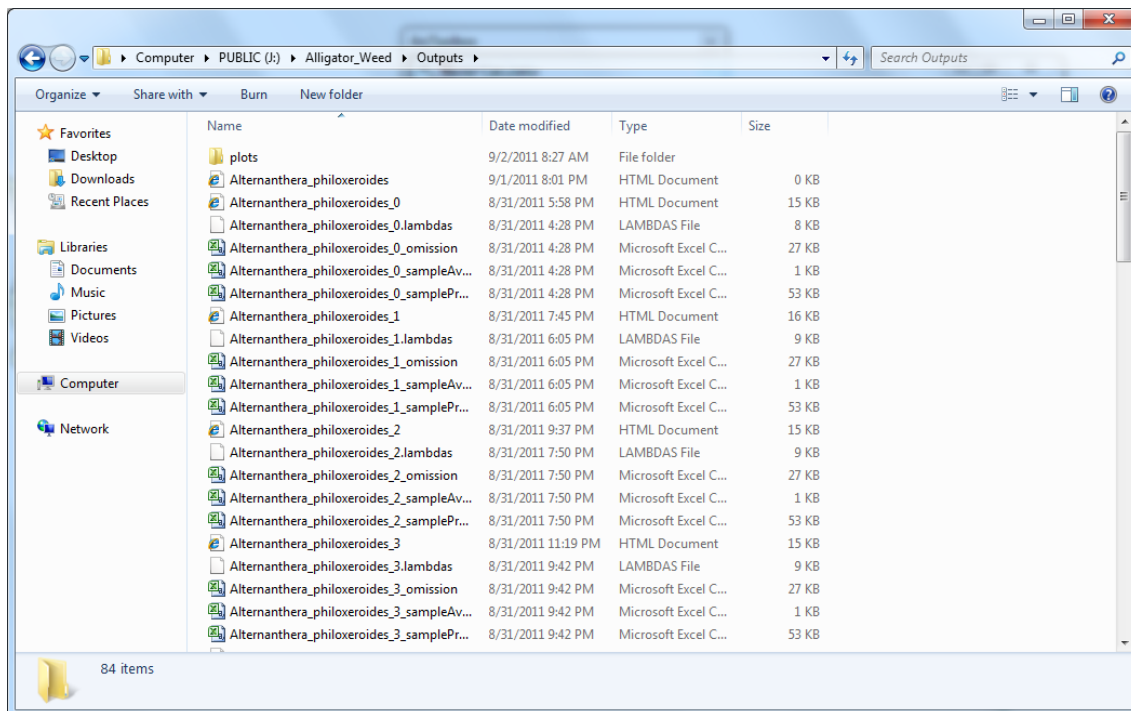
Open the file called Alternanthera_philoxeroides.html file. In this example, it has an Internet Explorer icon which is based on the default browser. This is the main output file for the MaxEnt model. This file contains information on the overall averaging of all model runs that were specified with statistical analyses, plots, model images, and links to the other files and runs.

Also contained in this file are the control settings and parameters that were used to run the model, and the code to run the MaxEnt model from the command line.

The first graph you see in this file is the **Analysis of Omission/ Commission.** This graph displays the omission rate and predicted area at different thresholds. The orange and blue shading surrounding the lines on the graph represent variability.
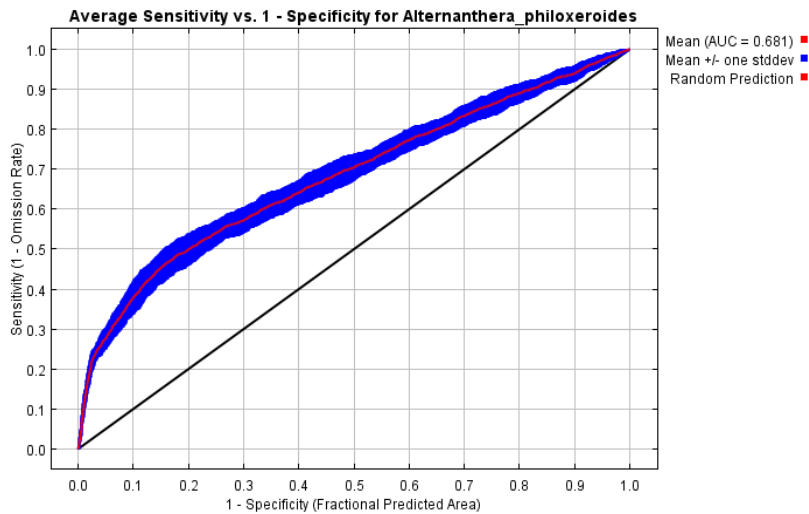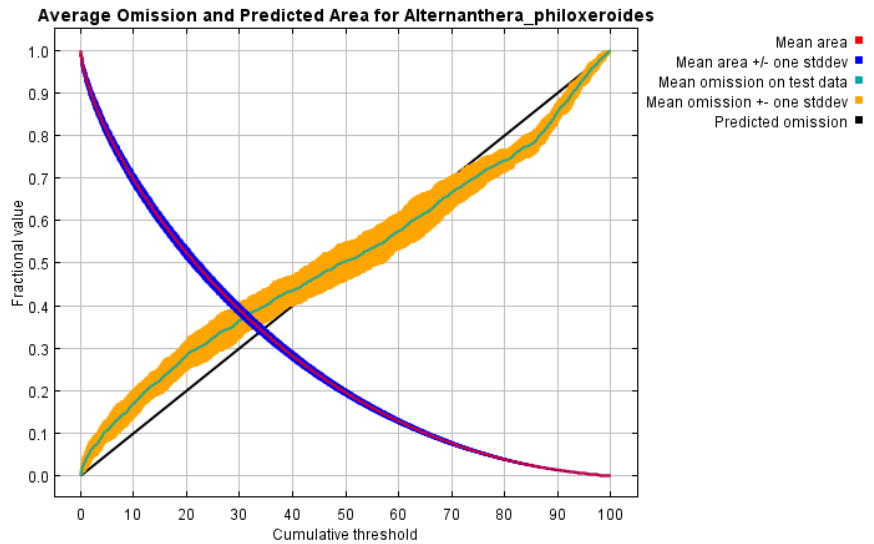
The next graph you see when you scroll down is the **Sensitivity vs 1 –Specificity for Alternanthera_philoxeroides**. This is a graph of the Area Under the Receiver Operating Characteristic (ROC) Curve or AUC. The AUC values allow you to easily compare performance of one model with another, and are useful in evaluating multiple MaxEnt models. An AUC value of 0.5 indicates that the performance of the model is no better than random, while values closer to 1.0 indicate better model performance.

Further down the page, you will see a picture of the model. You can click on the picture to see an enlarged version. You can also find this image in the **Plots** folder in the outputs as a Portable Network Graphic (.png) file.
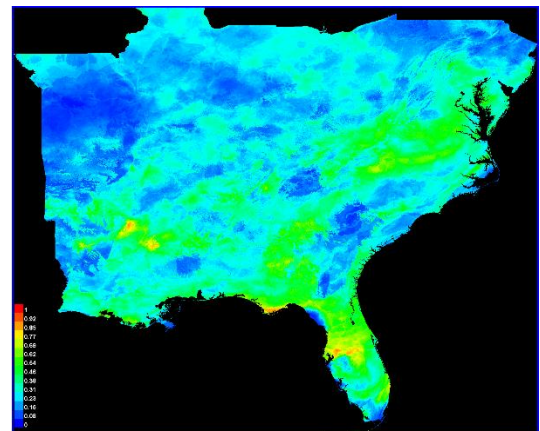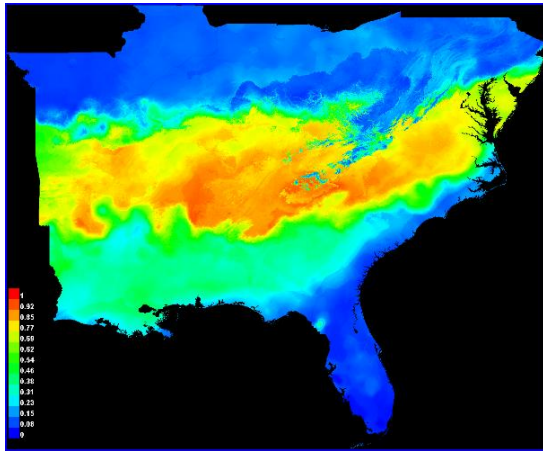
Another image that is displayed looks something like the one to the left. This is an image that represents the model of species distribution projected to the future climate data (**2020**). Both models are also accompanied by mapped standard deviation, but these are not pictured here.

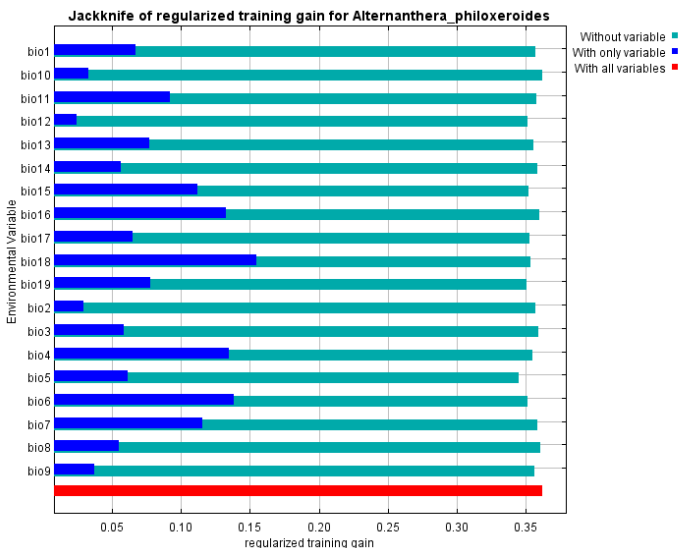| Variable	Percent contribution	Permutation importance
bio15	22.8	20.5
bio18	20.3	2
bio5	10.6	2.7
bio6	8.1	8.4
bio16	6.4	2.2
bio4	6.1	4.4
bio9	4.3	1.3
bio17	3.4	7.8
bio2	2.9	4.9
bio14	2.4	1.1
bio13	2.2	2.3
bio7	2.2	2.4
bio19	2	7.6
bio1	1.8	7.3
bio12	1.4	6.4
bio3	1.3	5.5
bio11	0.9	12.1
bio8	0.4	1.2
bio10	0.4	0.1

Charts are provided for each individual bioclimatic variable that was used in the model. In the Alligator Weed example you will see charts for all 19 variables in two sections. The first set of charts represents the change in each variable with all others held constant, and the second set shows the scenario if a model were run using only the charted variable.

Farther down the page, you will see a table that shows the **Analysis of Variable Contributions**. This table shows the environmental variables used in the model and their percent predictive contribution of each variable. The higher the contribution, the more impact that particular variable has on predicting the occurrence of that species. In this example Precipitation Seasonality (i.e. bio_15) had the highest predictive contribution of 22.8%.

Below the variable contributions is a graph of the **Jackknife of Regularized Training Gain**. The Jackknifing shows the training gain of each variable if the model was run in isolation, and compares it to the training gain with all the variables. This is useful to identify which variables contribute the most individually. The Alligator Weed model also provides a jackknife for test gain of the species and AUC.

At the top of the page, you will find links to other run outputs created from the model. You can click on these here or go back to the MaxEnt **Outputs** folder and open them there. Further

links may also appear at the bottom of each run's .html file.

Raw data outputs and control parameters

The data used in the above analysis is contained in the next links. Please see the Help button for more information on these.
The coefficients of the model
The omission and predicted area for varying cumulative and raw thresholds
The prediction strength at the training and (optionally) test presence sites
Results for all species modeled in the same Maxent run, with summary statistics and (optionally) jackknife results

Regularized training gain is 0.376, training AUC is 0.766, unregularized training gain is 0.594.
Unregularized test gain is 0.268.
Test AUC is 0.668, standard deviation is 0.024 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2).
Algorithm converged after 3060 iterations (276 seconds).

Step 2: Thresholds

MaxEnt (unless you change the default settings) produces a continuous raster with values from 0-1 representing habitat suitability. Often, we are interested in displaying the information with discrete classifications. For example, we may only want to display two value; suitable habitat and unsuitable habitat. In these cases, a decision must be made as to what threshold value constitutes as suitable habitat (i.e. what probability value is the minimum value for suitable habitat). There is no set rule to set these thresholds and how you decided a threshold can depend on the data used, or the objective of the map and will vary from species to species. The final decision needs to be made taking all these considerations into account and with a good understanding of the species of interest.

MaxEnt provides threshold values based on a variety of statistical measures in the maxentResults.csv included in the results. Some of the most common thresholds used are: minimum training presence logistic threshold, 10 percentile training presence logistic threshold, and equal training sensitivity and specificity logistic threshold

For this example, we will use the 10% minimum threshold to define the minimum probability of suitable habitat. We are using this threshold primarily because of the data we used which is likely to have some errors. By using this threshold, we are defining suitable habitat to include 90% of the data we used to develop the model. If we were more certain of the quality of the data we may have used the minimum presence threshold or perhaps a 5% threshold.

To set the 10% minimum threshold, first navigate to the maxentResults.csv file in the results section. Open this file and find the column titled "**10 percentile training presence logistic threshold**." Make sure that it is the logistic threshold not cumulative. Use the value in the last row of this column which represents the average of all runs performed. Use this value along with the averaged model results (for this example Alternanthera_philoxeroides_avg.asc) to reclassify the averaged model results to match the selected threshold in ArcGIS. Your final map should

have two classifications: a value representing unsuitable habitat and another value representing suitable habitat.

Step 3: MaxEnt's other outputs

In addition to the .html file, there are the several other outputs by MaxEnt. We provide a brief description as follows. For further explanation and interpretation of MaxEnt's outputs, users should refer to www.cs.princeton.edu/~schapire/maxent/ .

The **Plots** folder contains all the pictures of graphs, maps, and charts that MaxEnt created from the run.

maxent.log is a file that records the settings and options that you selected for the run. This file is especially important for troubleshooting problems you may encounter with MaxEnt.

maxentResults.csv is a file containing the number of training samples used for learning, values of training gain and AUC. In addition, if the jackknife option is selected, the regularized training gain and AUC for each part of the jackknife are included in this file.

Alligator_Weed.asc is the ESRI ASCII output that contains the probabilities of the species distribution.

Alligator_Weed.lambdas contains the computed values of the constants

Alligator_Weed_omission.csv describes the predicted area and training for various raw and cumulative thresholds.

Those are your MaxEnt results. There is a lot of information contained in the outputs and it may take a while to get an understanding of everything included. This tutorial, however, should allow you to start modeling using the basic features of the MaxEnt model.

Part 6: Converting MaxEnt's ASCII output to a Raster

This is the final step in the series of tutorials. In this tutorial, you will learn how to convert the ASCII layer (produced from the MaxEnt output) to a raster layer. The conversion needs to be performed in ArcMap.

Step 1: Open ArcMap and ArcToolbox

Before you begin, you will want to create a new folder to store the final raster. In the **Outputs** folder, create a new folder called **Raster_Output**.
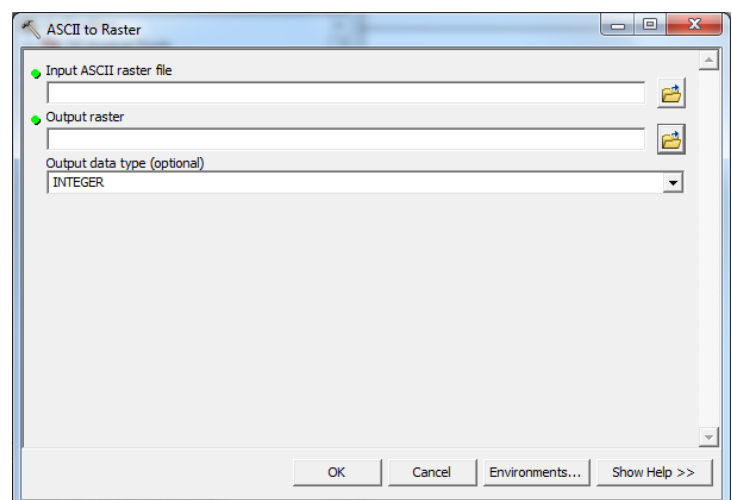
Open ArcMap and select **A blank map** and click **OK.** Next, Open the **Toolbox** window by clicking on the toolbox icon [icon]. In the Toolbox window, double click **Conversion Tools** then double click **To Raster** then double click the tool **ASCII to Raster**.

For the **Input ASCII Raster File** window to browse to the average ASCII file created by MaxEnt. Be sure to select **File (*ASC)** in **File of Type** when navigating to the ASCII file; otherwise, you won't be able to see the file. In our case, the file is located in the **Alligator_Weed** folder under the **Outputs** folder, and should be labeled Alternanthera_philoxeroides_avg.asc. There are also minimum and maximum rasters if you are more interested in those values.Select this file then click **Open.**

For the **Output Raster,** you will want to navigate to the folder you created earlier called **Raster_Output**. Save the raster in the folder and call it AW_Rast. Once you have named your raster, click **Save**.
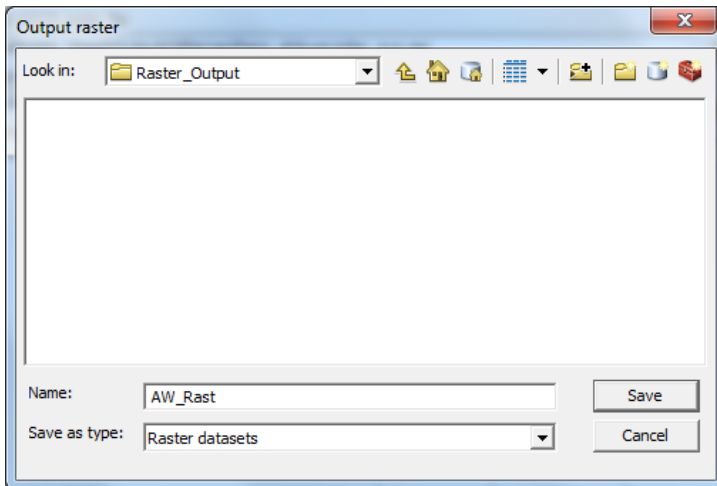

Now you will want to change the **Output Data Type** from INTEGER to **FLOAT**. We want to do this because the values in the ASCII output are much smaller than 1.0 and will be converted to 0 unless we select **FLOAT** data type.

Now your ASCII to Raster window should look like this:

Now click **OK.**

This process may take a long time (possibly up to ten minutes or more), so be patient. In the lower right-hand corner of ArcMap a scrolling text will begin to move indicating the current task.
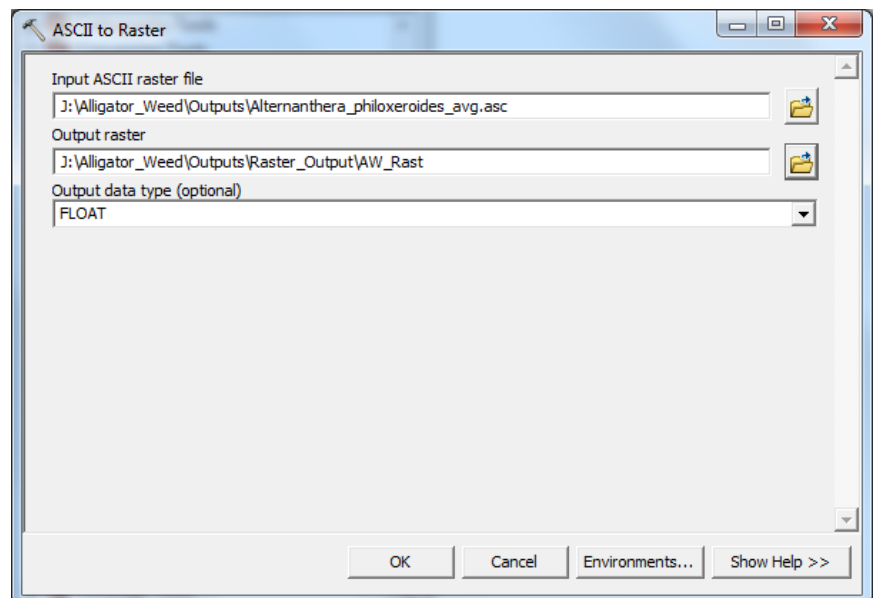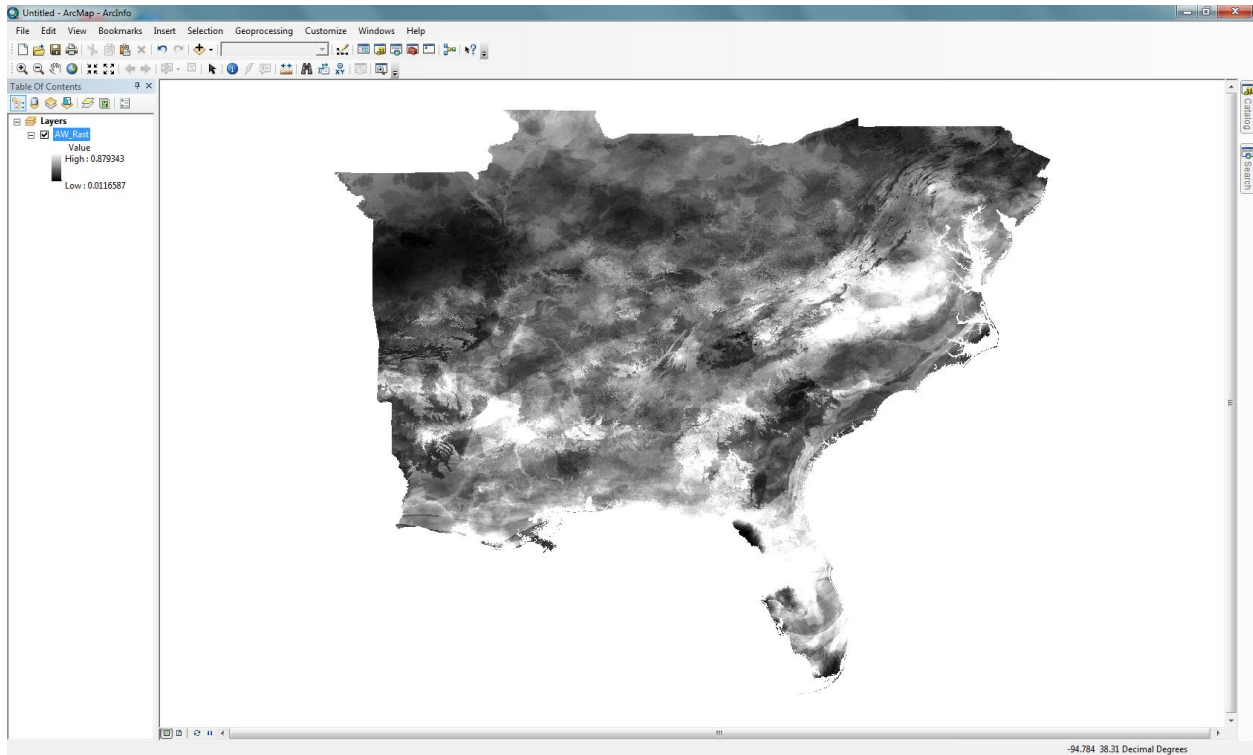
Once it has completed, a new raster layer should appear in ArcMap with the predicted probability that conditions are suitable for the species you have selected for the run.

We recommend trying different color schemes in the **Symbology** option and changing classification thresholds to visually check your model output and find potential patterns of interest.
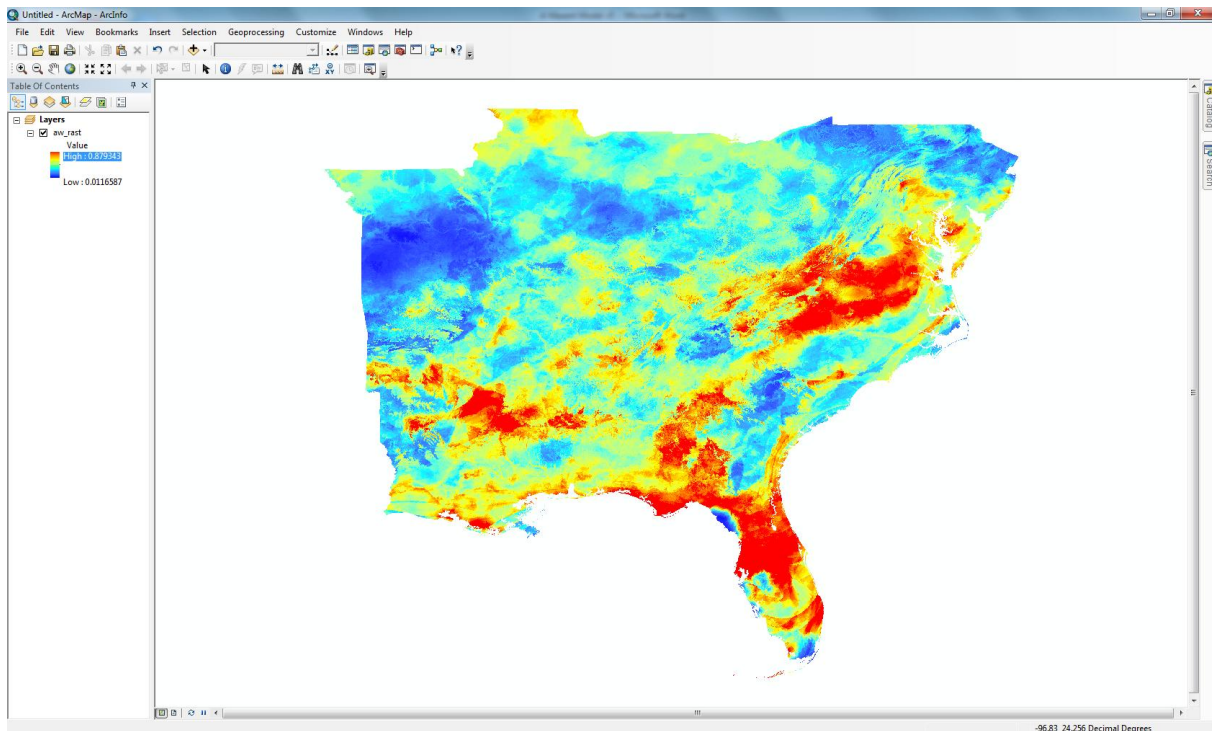
Here is the final map with a 10 % threshold and a different color scheme with Red indicating suitable habitat and and green representing unsuitable habitat.
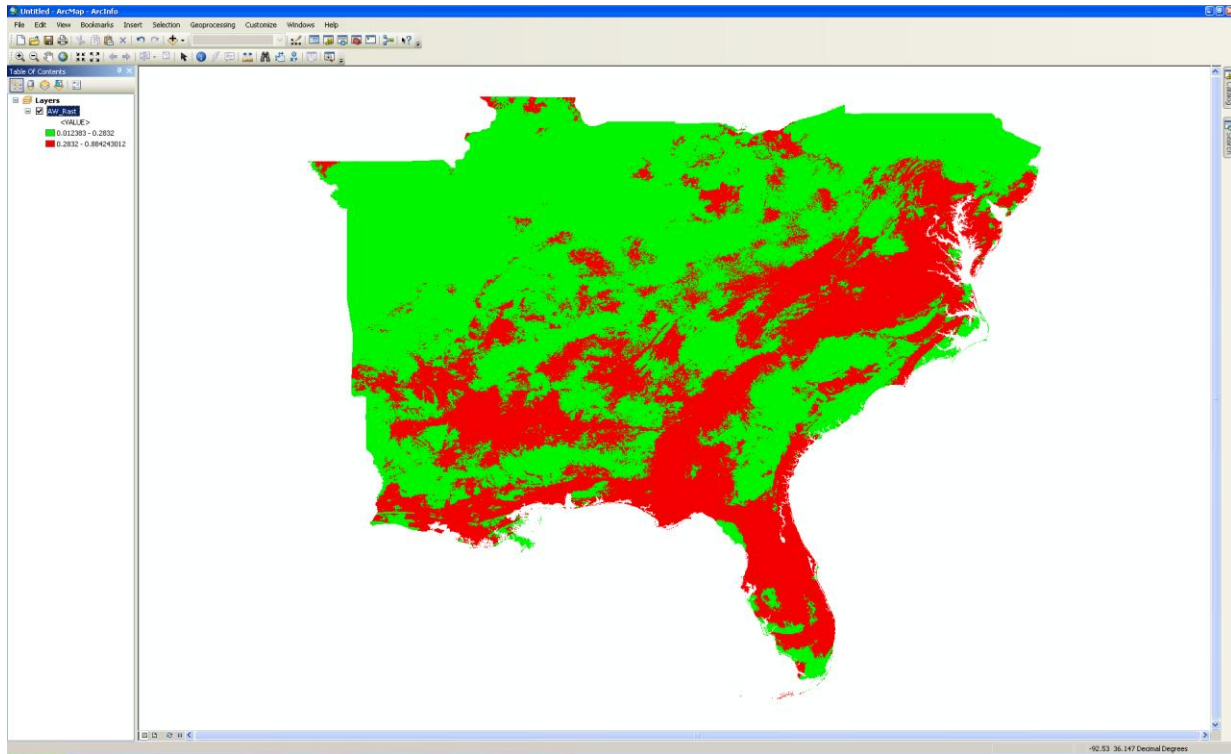
Below are the coded bioclimatic variables and what each code represents:

Bio_1 = Annual Mean Temperature
Bio_2 = Mean Diurnal Range (Mean of monthly (max temp – min temp))
Bio_3 = Isothermality (P2/P7)*(100)
Bio_4 =Temperature Seasonality (standard deviation*100)
Bio_5 = Max Temperature of Warmest Month
Bio_6 =Min Temperature of Coldest Month
Bio_7 =temperature Annual Range (P5-P6)
Bio_8 =Mean Temperature of Wettest Quarter
Bio_9 =Mean Temperature of Driest Quarter
Bio_10 =Mean Temperature of Warmest Quarter
Bio_11 =Mean Temperature of Coldest Quarter
Bio_12 =Annual Precipitation
Bio_13 =Precipitation of Wettest Month
Bio_14 =Precipitation of Driest Month
Bio_15 =Precipitation of Seasonality (Coefficient of Variation)
Bio_16 =Precipitation of Wettest Quarter
Bio_17 =Precipitation of Driest Quarter
Bio_18 =Precipitation of Warmest Quarter
Bio_19 =Precipitation of Coldest Quarter

Note: If any area calculations are to be made, the maps will need to be in a projection and datum that defines each pixel (cell) to be an equal area.

Congratulations! You now have successfully converted MaxEnt's ASCII output into a raster that you can use in your report or analysis of species distribution.

One last time…we strongly recommend looking at the MaxEnt website or MaxEnt Google Group discussion (http://groups.google.com/group/Maxent?pli=1) for further explanation and interpretation of MaxEnt's features, outputs, and interpretation.
The site is found at: www.cs.princeton.edu/~schapire/maxent/

Additional resources can be found at: http://homepages.inf.ed.ac.uk/lzhang10/maxent.html

Good luck with your modeling adventures!!

References:
Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.