**Systematic Conservation Planning Workshop Exercise:Maxent**

**Introduction**

In this series of activities, we demonstrate the application of the maximum entropy modeling or Maxent model (Phillips et al. 2004, 2006) for predicting the distribution of Sandhill crane (*Grus canadensis*) during spring in California. An important component of Maxent modeling is the data preparation, which requires the understanding of several other software and file formats including Microsoft Excel© and ESRI ArcGIS©. This tutorial will help guide users on how to format data using different software prior to running the Maxent model. For more detailed instructions on how Maxent operates, interpreting results, and advanced modeling options, users should refer to [www.cs.princeton.edu/~schapire/Maxent/](http://www.cs.princeton.edu/~schapire/maxent/) .

The data used in this activity are for educational purposes only, and in some cases may lack the quality required to produce accurate and precise results.

Note, the number parts are a continuation from the understanding and assembling model input exercise

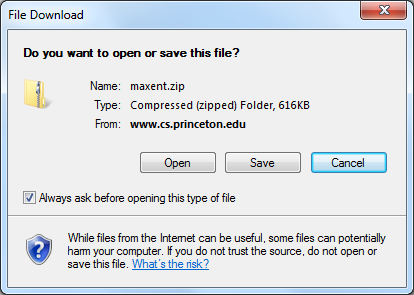
**Part 6: Running the Maxent Model**

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 data layers. Additional model capabilities and information can be accessed at Maxent website ([www.cs.princeton.edu/~schapire/Maxent/](http://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 3.3.3k.**

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.

**Step 2: Default Run**

First, we are going to create model using all the defaults. This is not recommended but it will introduce you to the Maxent GUI interface and how to define the required input. We will build off of this initial run in later section.

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

**Side note.** 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.

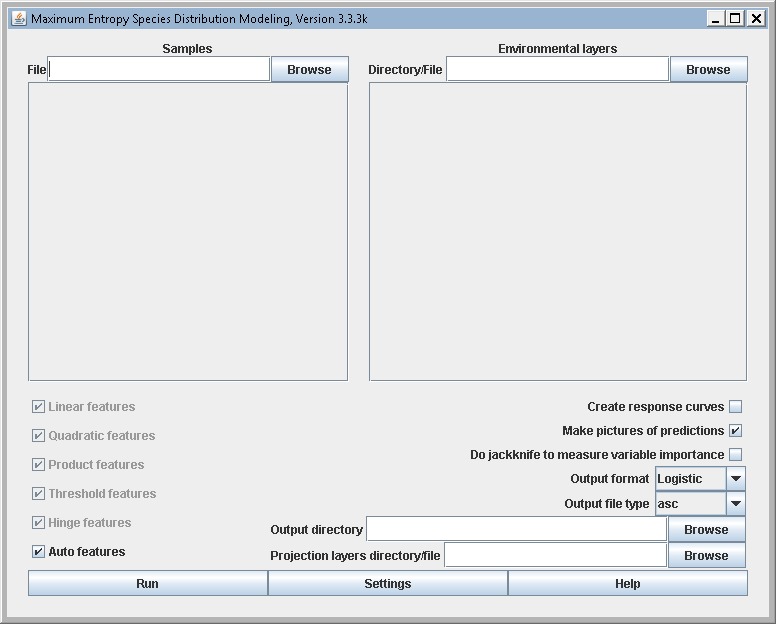
To do this, first right click on the Maxent.bat file in your working folder and select **Edit**.

A Notepad window should appear.

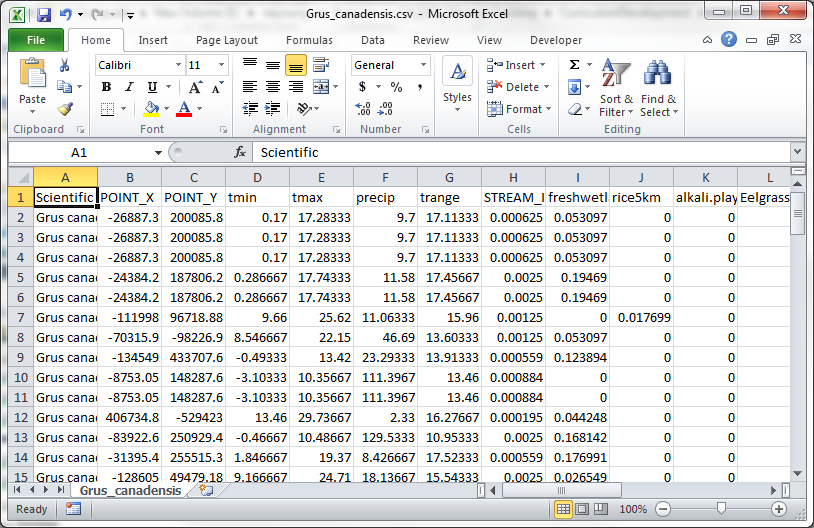
You can change the amount of memory that Maxent uses by simply changing 512 to 1024. This is how the window should look.

Once you have changed the memory, click **File** and **Save** and close the window.

1. Open the Maxent GUI by double clicking the **Maxent.bat** file. A window should open that looks like this:



To begin, you must provide a **Samples** file. This file contains the presence localities in comma separated value (.csv) format. For this demonstration, we will use *Grus\_canadensis.csv* within the **Workshop\_StartData** folder. We will be using the ‘Samples With Data’ (SWD) format to run this model. What this means is that we have already attributed each record the environmental variable values at that location like that shown in the figure below. This is important for this application because we want to match each occurrence record with PRISM climate data from the season/year that the observation was made. As you can imagine, there can be a lot of year to year variability in climate. This will allow us to have multiple records for a specific locations.



1. Navigate to the *Grus\_canadensis.csv* within the **Workshop\_StartData** folder 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 **EnvironmentalData** folder.

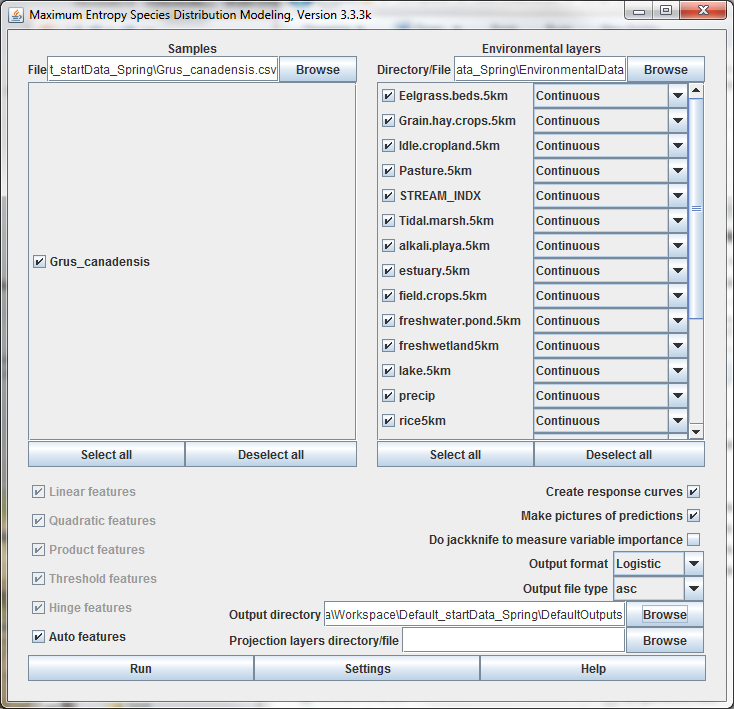
1. Navigate to the **EnvironmentalData** folder that contains all your environmental data in ascii format by clicking the **Browse** button under **Environmental Layers**, or type in the file path.

Notice how you can change the environmental layers to either continuous or categorical. All the variables for this data set are continuous. If any of the layers were categorical (e.g., land cover type), you would need to change the type for that variable 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.

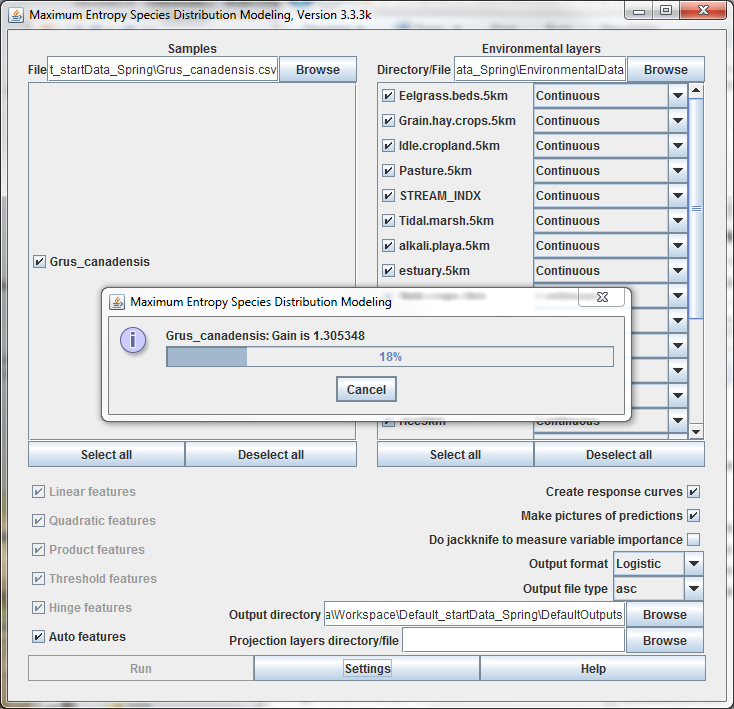
1. Create a folder in the **Workshop\_StartData** folder called **DefaultOutputs** to store the Maxent output
2. Once created, set the **Output Directory** in Maxent by navigating to the **DefaultOutputs** folder by clicking the **Browse** located next to the **Output Directory,** or type in the file path.
3. Although not part of the default settings, make sure that the **Create Response Curves** is checked which will allows us to make comparisons to this model in future. Leave all other settings as their default value

Now your Maxent GUI should similar to the one below



1. 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 like that shown below



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 sample is exp(2) or about 7.5 times higher than that of a random background pixel.

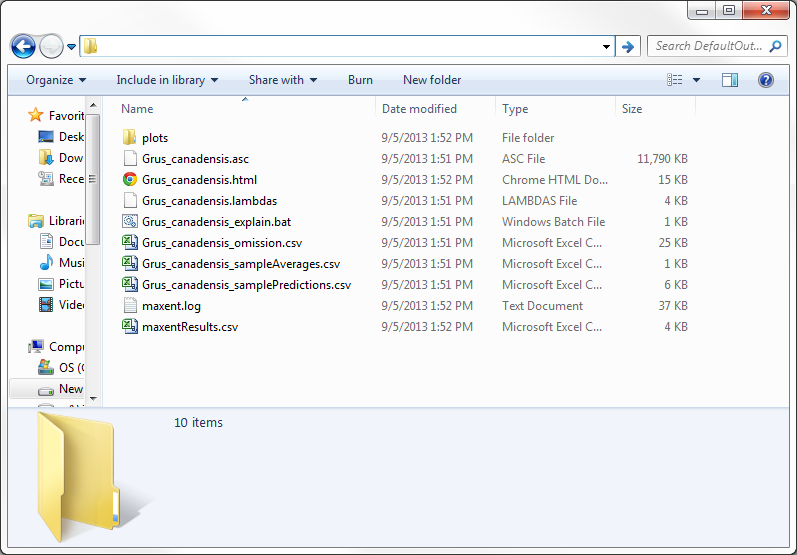
Once the Maxent model has completed its run, the progress window will disappear. You will be able to find all the outputs created by Maxent in the **DefaultOutputs** folder you created earlier.

**Part 7: Interpreting Maxent Outputs**

This part explores the different outputs of Maxent. Many, but not all of Maxent’s outputs will be discussed and explained. For further explanation and interpretation of Maxent’s outputs, users should refer to [www.cs.princeton.edu/~schapire/Maxent/](http://www.cs.princeton.edu/~schapire/maxent/).

**Step 1: Exploring Maxent Outputs**

1. First, open the folder that contains the Maxent outputs (i.e., **DefaultOutputs**). When you open the folder, you should see another folder labeled plots, and a list of other files like those shown below.

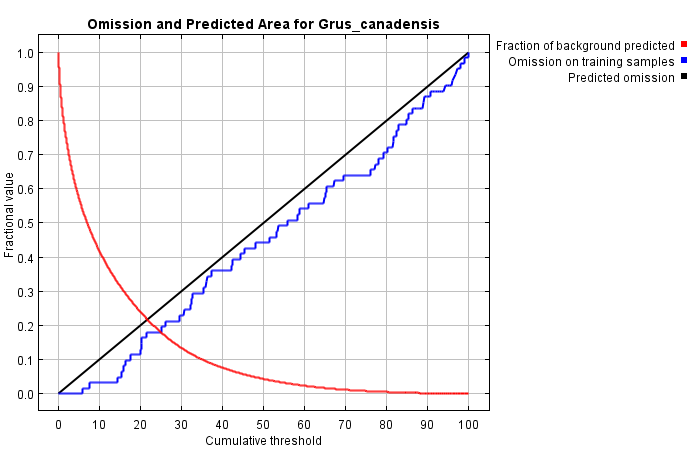


1. Open the htlm file. This will be named called *Grus\_canadensis.html*.

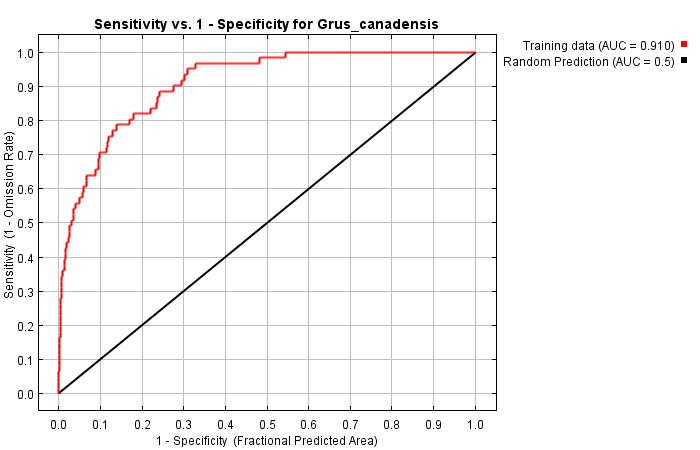
This is the main output file for the Maxent model. This file contains information on the model results including statistical analyses, plots, model images, and links to the other files. Also contained in this file are the control settings and parameters that were used to run the model, and the code to run the Maxent model from the command line. This provides the exact settings used to create the model.

1. Review the various outputs on your own. We will go over some of these together once everyone has completed their model.

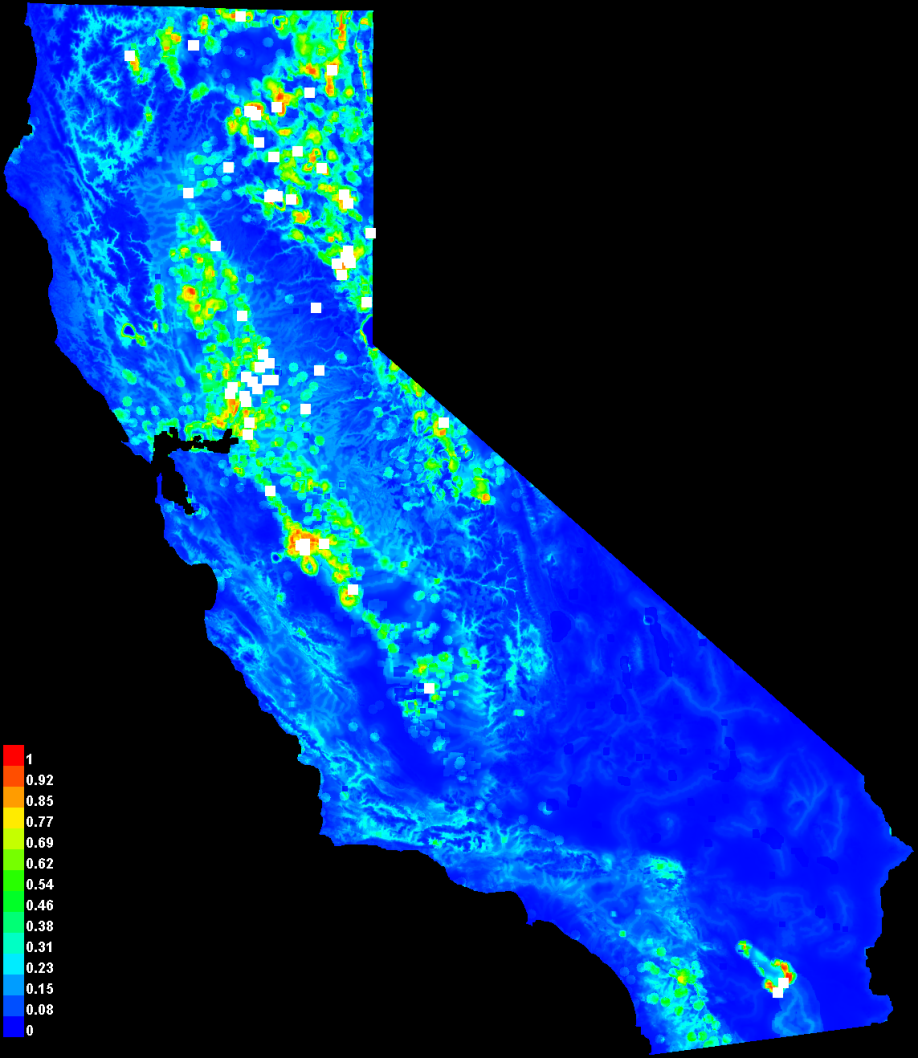
The first graph you see in this file is the **Analysis of Omission/ Commission.** This graph displays the omission rate and predicted area at different thresholds. Omission is the false negative rate (e.g., where the model predicted unsuitable conditions but where we had a occurrence location)



The next graph you see when you scroll down is the **Sensitivity vs 1 –Specificity**. This is a graph of the Area Under the Receiver Operating Characteristic (ROC) Curve or AUC. The AUC values allow you to 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.



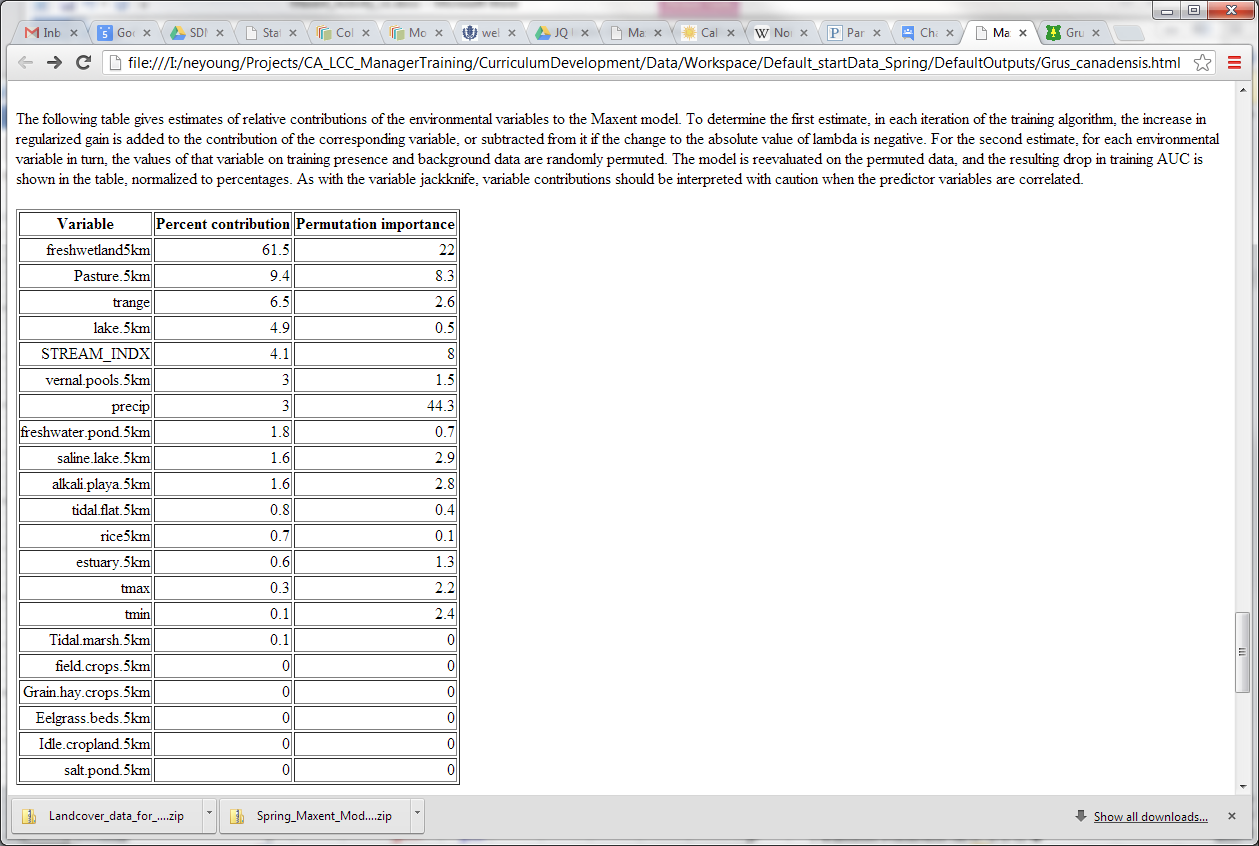
Further down the page, you will see a picture of the model. You can click on the picture to see an enlarged version. You can also find this image in the **Plots** folder in the outputs as a Portable Network Graphic (.png) file.



Response plots are provided for each individual environmental variable that was used in the model. You will see plots for all variables in two sections. The first set of plots 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. Take some time to review how these are described in the html file and how to interpret the results.



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 with the percent predictive contribution and permutation importance for each variable. The percent contribution values are calculated during model development from changes in the gain while the permutation importance is calculated by having each variable’s values changed at the training and background locations and then re-evaluating the model.



* **According to the AUC, how well did this model perform?**
* **What were the top three important variables (Use permutation importance)**
* **How would describe the pattern of the predicted habitat suitability?**

**Part 8: Customizing the model**

Now that we have created a default model, we want to go back and start changing the settings so that we can develop a model that is more appropriate for the species, system, and data we have available.

**8.1: Background sample**

First, let’s change how we define the background sample. How the background sample is defined can have drastic impacts to model results. When we select the defaults, the program randomly samples across the extent of the environmental data we set as the input. This would only be appropriate if the occurrences points were also randomly sampled across the landscape, which is rarely the case and if so, you should be considering other modeling methods that will take advantage of this data property. The goal with the background data is to represent the available environment and account for any bias in the occurrence data – both of which can be identify and represent.

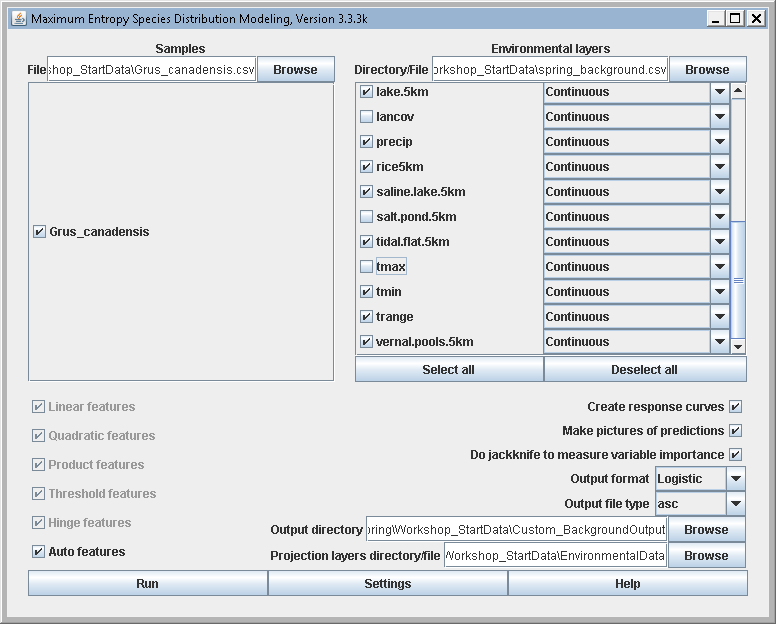
* **What might be some of the biases associated with this data set?**
* **How would you define available environment?**

1. Using the Maxent model you ran with default settings, change the **Environmental Layers file** to point to the *spring\_background.csv* file in the **Workshop\_StartData** folder.
2. For the **Projection layers directory**, change this to point to the **EnvironmentalData** folder.

Since we are providing environmental information with both the occurrence sample and the background sample we need to provide a projection layers to get a mapped output of the distribution.

1. Uncheck the landcov variable in the list of Environmental variables (this layer was generated for different species) and uncheck the environmental variables you identified yesterday to remove.

The GUI should now look similar to the one shown below:

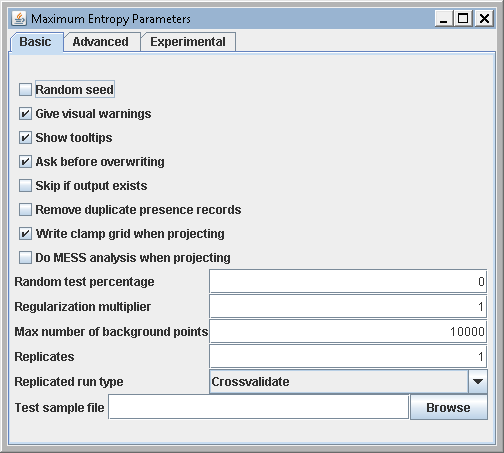


**Settings**

Now we will start changing various settings to help refine our model

*BASIC settings*

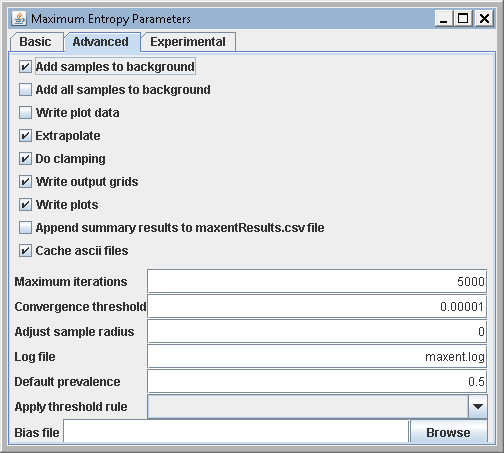
1. Uncheck the **Remove duplicate records** option. Because the occurrence data we are using are from multiple time periods, we could have two records with the same location but they will have different environmental variables associated with them.
2. Uncheck the **Do MESS analyses when projecting.** We will want to use this, but we will run a separate model to accomplish this.



*ADVANCED settings*

1. Change the **Maximum iterations** to 5000

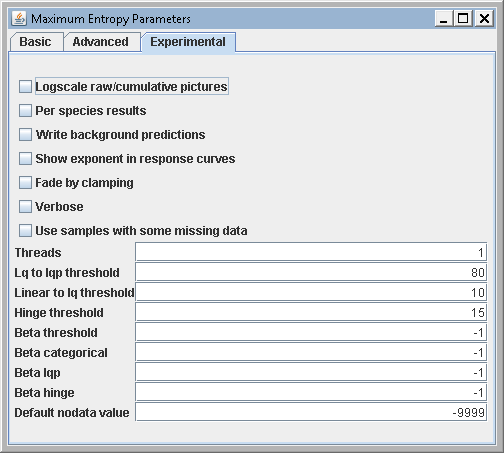
This allows for the model to reach convergence. Often, you will see that the Maximum number of iterations was reached before the model finished if you leave this at 500.



* **How many iterations were required for the default model? Did the model finished before it reached this threshold?**

*EXPERIMENTAL settings*

We will not be changing any setting in this tab. One setting that can be of interest is the **Write background predictions**. This will write out a file with the coordinates of the background locations.



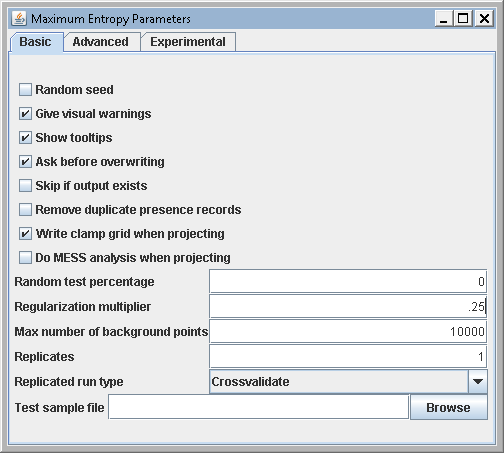
Run the model under these settings and save the model output in a folder called **Custom\_BackgroundOutput.**

* **Compare the output of this model compared to the default. What are some differences?**
  + **Predicted map?**
  + **AUC?**
  + **Variable Importance?**

**8.2: Regularization**

Now we are going to look at the effect of the regularization parameter. 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). We will not be using this specific tool in this workshop but it is worth exploring if you continue modeling with Maxent. Regularization can be thought of as a smoothing parameter, where larger values increase the amount of smoothing.

Keep the same settings as above, but change the **Regularization** parameter in the basic settings to be 0.25



Run the model with this change and put the output in a new folder called **Custom\_Reg.25Output**

Once the model has finished, Run another model with a Regularization of 3 and put the output in a new folder called **Custom\_Reg3.0Output**

* **How does changing the regularization impact:**
  + **Model mapped predictions?**
  + **Response curve shapes?**
  + **AUC evaluation?**

**8.3: Model testing**

Once we’ve identified a regularization value to use, we can run model with the intention of testing its performance.

1. Uncheck **Write Output Grids** in basic settings

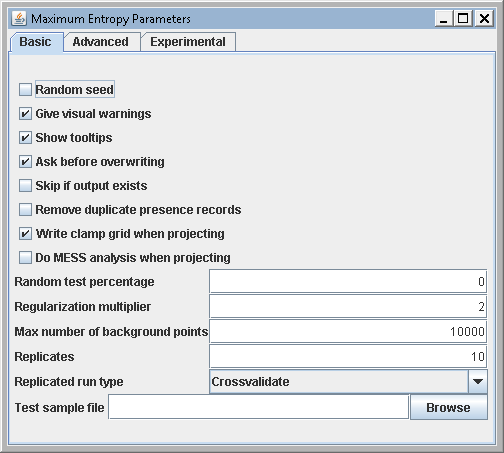
This helps reduce the size of the outputs and speeds up the model. It will still create an overall model result, just not each replicate model result.

1. Change the number of **Replicates** to 10

Maxent allows the ability to run a model multiple times and then conveniently averages the results from all models created. Using this feature in combination with withholding a certain portion of the data for testing enables 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.

There are three different sampling techniques (replicate run types) that are available in Maxent; Crossvalidation, Subsampling and Bootstrap. 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.

Under the setting we set, we are running a 10-fold cross validation to evaluate model performance. This will partition the sample data into 10 parts that will each be held out of the model development in turn for 10 different model replicates while the remaining 90% of the data will be used for training the model.



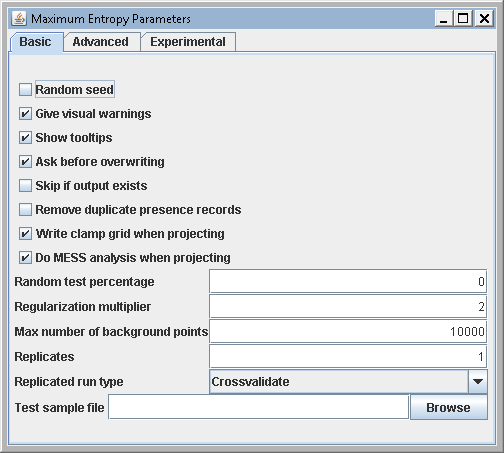
Run the model using these setting and save the output in a folder named **Custom\_TestingOutput**. This may take some time to complete. This would be a good time to go and investigate the differences between the default setting and how regularization impacts model results.

* **What as the test and train AUC?**
* **Where did you find this information?**
* **How main sample did the model use for testing and training for each model replicate?**

**8.4: MESS analysis**

Now that we have a model that is getting close to a final product, let’s create a MESS map to evaluate areas we are extrapolating to.

1. Use the same settings that you used in your last model,
   1. Change the number of **Replicates** to 1 in the Basic settings tab
   2. Re-Check the **Write Output Grids** option in the Advanced settings tab
   3. In the Basic settings tab, check the **Do MESS analyses when projecting** option



Run a model using these settings and save the output in a folder called **Custom\_MESSOutput**

* **Where are the novel environments? Is there a lot of the environment that was not captured in the training data?**
* **What environmental variable appears to be the least sampled?**
* **If you wanted to be sure to capture as much of the environmental range as possible, how could you use this output?**

**Part 9: Converting Maxent’s ASCII output to a Raster**

In this section, you will learn how to convert the mapped prediction ASCII layer (produced from the Maxnt output) to a raster layer. The conversion needs to be performed in ArcMap.

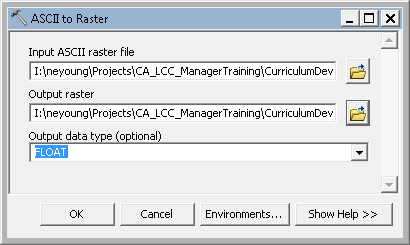
Before you begin, you will want to create a new folder to store the final raster. In the **Workshop\_StartData** folder, create a new folder called **Raster\_Output**.

Open ArcMap and select **A blank map** and click **OK.** Next, Open the **Toolbox** window by clicking on the toolbox icon . 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 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. Navigate to the **Custom\_TestingOutput** folder. The mapped prediction should be called *grus\_canadensis\_avg.asc*. Select the average 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 *Grus\_c\_avg*. 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.



Click **OK.**

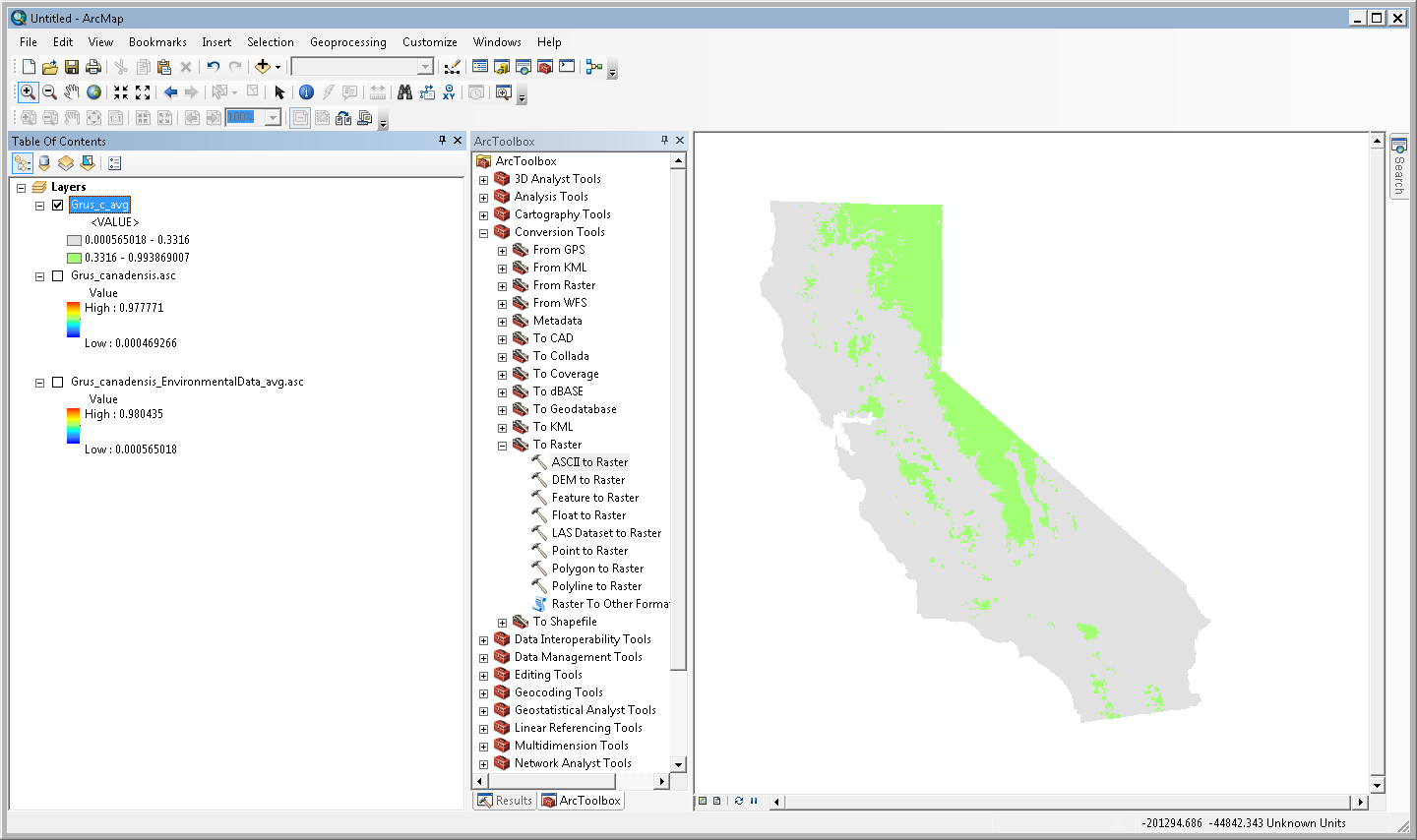
Once it has completed, a new raster layer should appear in ArcMap.

**Part 10: Thresholds**

Maxent (unless you change the default settings) produces a continuous raster with values from 0-1 representing an index of habitat suitability. Often, we are interested in displaying the information with discrete classifications. For example, we may only want to display two values; 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 value is the minimum value for suitable habitat). There is no clear 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.

Let’s use the Maximum of sensitivity plus specificity logistic threshold. This threshold “minimizes the mean of the error rate for positive observations and the error rate for negative observations. This is equivalent to finding the point on the ROC curve whose tangent has a slope of one (Cantor et al., 1999)” Freeman et al., 2008.



As a comparison, reclassify the output to look at how using the 10% Minimum training presence logistic threshold value changes the binary mapped prediction.

**Part 11: 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/](http://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.

If you finished early, spend some time evaluating how manually choosing which features the model will use and how this impacts model results.

* **How does changing which features are included change the model results?**

You could also look at how turning on or off clamping changes predictions and response curves.