

## Environmental Niche Modeling Exercise

Climate change is expected to have significant impacts on species' ranges by altering the distribution of key climate variables that determine where species can live. Environmental niche modeling (ENM) is a statistical tool for estimating which environmental variables are important determinates of species' ranges and projecting species' geographic distributions based on those environmental variables. ENM can then be used to estimate how species' geographic ranges will respond to climate change by projecting species' distributions onto maps of predicted future climate variables.

The goal of this module is construct and compare “niche models” for interacting species under current and future climate scenarios.

By the end of this module you should be able to:

- Construct an environmental niche model of a species and use it to project a species distribution in the future.
- To articulate and test a hypothesis about a biological responses to climate change for a particular species
- To synthesize information across species and consider how interactions *between* species may mediate their responses to climate change
- To evaluate the strengths and limitations of the environmental niche modeling approach while interpreting your results.

Over the last two days, your group identified an assemblage of four species (five if you are in the group of five) that have a well-documented biological interaction. As you continue to research your focal species and your assemblage you will articulate three sub-hypotheses that relate to that species assemblage.

- a. A sub-hypothesis concerning what bioclimatic variables may be most important for driving the niche for each of your species—define for all four/five species.
- b. A sub-hypothesis concerning how species range sizes and/or locations will change in the future—articulate hypothesis for all species.
- c. A sub-hypothesis concerning how species ranges will change in relation to each other and in terms of niche overlap—articulate for at least 2 species in your assemblage. No need to calculate for all combinations of species.

Once you have come up with your biologically motivated hypothesis, each group member will be responsible for producing MaxEnt analysis for one of these species. Then the group will pool output to calculate using R the changes in range area and niche overlap between species and over time. This document takes you through this process step by step.

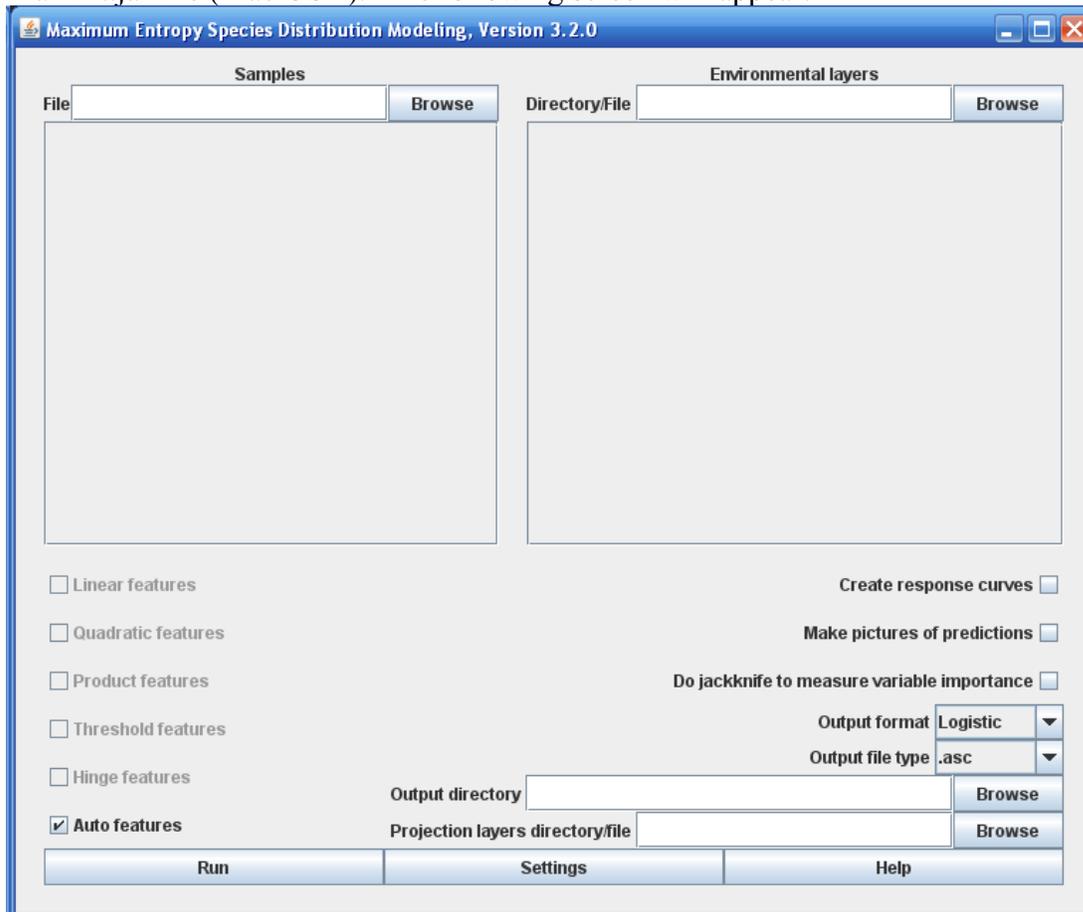
**Ultimately your group will produce two products from these results:**

- 1) a 10-minute group presentation on your findings
- 2) a group report on the potential impacts climate change may have on the group of species you investigated.

One presentation and one report will be prepared for each group. To make sure everyone is contributing, all group members will be asked to assess the contribution of other group members to the project (rubric is at the end of this document). I will summarize these responses and provide each of you individual feedback. This process serves two purposes. First it gives us an idea of whether any group members should not get full credit for the assignment, and secondly it provides you feedback on how successful you are at group work.

## RUNNING MAXENT FOR YOUR SPECIES

Open MaxEnt by either clicking on the MaxEnt.bat file (Windows) or clicking on the MaxEnt.jar file (Mac OSX). The following screen will appear:



First, import your species geospatial data. Click the 'Browse' button under samples and select your species .csv file. After you select it, your species name should appear in the upper left box. Select your species by checking the box beside it.

Second, import the climate data. Click the 'Browse' button under 'Environmental layers' and select the folder where you have saved your current\_bioclim folder (this should contain both .bil and .hdr files). A list of all the variables will appear in the right panel. Make sure all of the variables are selected so that they are included in the analysis.

To determine the environmental factors that predict the current geographic distribution of your species, do the following. Select the boxes ‘Make picture of predictions’, ‘Do jackknife to determine variable importance’, and ‘Create response curves.’ The first option will create a heat map of each species’ predicted distribution based on their associated climate variables; the colors indicate the probability of the species occurring in a given location. The second option will determine which of the climatic variables best explain the distribution of a given species. The third option quantifies how the probability of a species’ occurrence changes as a function of each environmental variable.

Designating a percentage of random observation points as ‘test points’ helps the model ‘train’ its predictions. What this means is that it first uses observed climate data and presence/absence data to characterize the combinations of climatic factors in the locations in which a species is present, and it then estimates parameters that describe the strength of the relationship between the factor and the probability that the species is present. It then uses the rest of the data to test how well the parameters estimated from the “training” data set actually predicted the probability that a species was present in the locations not used on the training set. We will designate 25% of your observations to be used as test points in the training data set. Click the ‘settings’ button and enter 25 next to “Random test percentage.” Close the settings dialogue box.

To project the geographic distribution of your species given future climatic conditions, do the following: In the bottom right corner define the “projection layers directory/file” by selecting the folder that contains the data that will be used for projecting your species into the future: `future_bioclim`. This folder contains the .asc files.

Finally, select the folder where you want the output from the model run to be stored. The box for this is labeled “Output Directory”.

To run your model, press the “Run” button. A progress monitor describes the steps being taken. This may take awhile depending on your computers capabilities and how much occurrence data you have for your species.

## **INTERPRETING RESULTS FROM MAXENT**

MaxEnt outputs a series of files encapsulating numerous results. This output is summarized in one .html file. Open the .html file. I will outline some of the main output below, but there is also a MaxEnt tutorial posted in the resources/assignments/niche modeling folder that describes the output in more detail.

*Pictures of the Model:* This image, like all distribution maps produced by MaxEnt, uses colors to indicate the predicted probability that conditions are suitable for the species, with red indicating a high probability of suitable conditions, green indicating conditions typical of those where the species is found, and lighter shades of blue indicating low probability of suitable conditions.

Below this picture, is a projection for your species in the future. Click on the picture to see a bigger, interactive version.

If you want to copy these images, or want to open them with other software, you will find the .png files in the directory called “plots” that have been created as an output during the run (These will be an important component of your results).

The maps below these first two show areas of the globe where the future predictions may not perform well because they are “no analogue” climates. Because these future climate combinations do not occur currently, it is hard to infer how species will respond from their current distribution. This is related to the classic problem of extrapolating beyond the range of your data. Read the descriptions of the various measures in the html file.

#### *Analysis of variable contributions*

MaxEnt provides three main tests that an environmental factor predicts a species’ presence. The first is the “**percent contribution**,” which indicates the percentage of the variation in the presence/absence of the species that can be explained by variation in the environmental factor.

Second, the “**permutation importance**” is based on a method called “permutation.” MaxEnt first estimates the strength of the relationship between the presence of a species and an environmental factor using the training test data points. Next, it scrambles the values of the environmental factors in that training data set, and re-calculates the strength of the relationship. If the environmental factor is actually important in predicting a species presence, then the strength of the relationship between the species presence and that factor would be higher in the real data set than in the scrambled data set. The “permutation importance” is a measure of how much the strength of the relationship changes between the real and the scrambled data sets. The larger the change, the more “important” that factor is in predicting the species’ presence.

Take note of the three bioclim factors that were identified as important predictors of species presence/absence through each of these two methods. Look up what those bioclim variables actually stand for <http://www.worldclim.org/bioclim>

	Percent contribution	Permutation importance
1.	_____	_____
2.	_____	_____
3.	_____	_____

Are these variables the same for the different species? Are they in the same order for the different species?

Third, while not required for your write-up, consider the results from the **jackknife test**. These tests identify the trait that provides the “most useful information by itself” to the model and the trait that provides the “most information that isn't present in the other variables.” These can be found listed in the paragraph above the graph labeled “Jackknife of regularized training gain for Your Species”. [A jackknife analysis is a way to measure the consistency of a relationship (in this case, between the probability of species' occurrence and an environmental factor). It randomly deletes observations from the data set and re-calculates the parameters, and it does this hundreds of times. This gives a range of estimates of the parameters: one for each data set. This is used as an estimate of the consistency (or lack thereof) of the relationship. ]

How do these traits compare to the traits identified above as being the most important variables?

You see here that MaxEnt gives three ways of measuring the strength of the relationship between an environmental factor and the probability of species' occurrence. These are complementary statistical approaches that give slightly different information about the importance of the environmental factor and the degree of consistency with which it predicts species' presence, and researchers often present all three.

#### *Response curves*

Finally, examine the of the three traits you identified by their ‘permutation importance.’ Look at the second set of graphs, not the first. Describe how the logistic prediction (y-axis, which is a measure of the probability of the species occurring) changes with the variable in question (x-axis). For example, ‘bio1’ is mean annual temperature and ranges from -256 (-25.6 C) to 302 (30.2 C). Under what values of an environmental factor (e.g. mean annual temperature) is your species most likely to occur? Which environmental factor(s) have a large effect on the probability of the species occurring?

In your report, present these statistics and describe how environmental factors identified by these methods as significant might be important to your species. Do they make sense based on your understanding of the biology of your species?

### **ANALYSIS OF DATA IN R**

First, combine the data files from all the members of the group into one folder on one person's computer that has successfully installed R. Each person will have two files in the folder that represent their species. The first is the niche predictions under current climate conditions and the second expected niche in the future. The first will be an .asc file with just the species name. The second will have the species name with the name of the future climate folder appended. Here is an example.

```
Aedes_aegypti.asc  
Aedes_aegypti_future_bioclim.asc
```

Detailed explanations for the R code below can be found in the “NicheModelingDemoRcode.r” file in the Sakai Niche Modeling folder. I recommend using this file as a template for your R analysis.

Open up R and change the ‘working directory’ to be the folder that holds all the species data.

- If you are using the program R directly this is under the heading ‘Misc’. Click “change working directory” and navigate to the folder where you saved your MaxEnt outputs.
- If you are using Rstudio, go to the heading “Session”, click “set working directory”, then “choose directory”. Navigate to the folder where you saved your MaxEnt outputs and click “choose”.

If you are starting a new R session, load in the four R packages you installed earlier.

```
library(ape)
library(phyloclim)
library(maptools)
library(sp)
```

### CALCULATING SPECIES RANGE AREA IN R

Now load the output file for one of your species into R. Note that you will need to replace the file name in quotes in the R code below to reflect the name used for your file.

```
Species1.current<-readAsciiGrid("Species1.asc")
```

We now want to estimate the area of the geographic range of your species (in the present and the future). We can very roughly approximate the range area for each species using the knowledge that each of our grid cells reflects approximately 10 km<sup>2</sup>. To put your ranges in perspective, the whole of earth’s surface is about 510 million km<sup>2</sup> and the land area of the earth is approximately 150 million km<sup>2</sup>. Copy the following code into R.

```
Species.raster<-raster(Species1.current)
plot(Species.raster)
Species.raster<-getValues(Species.raster)
Species.raster<-Species.raster[!is.na(Species.raster)]
range.size.km2<-
length(Species.raster[Species.raster>.33])*10
range.size.km2

prop.landcover<-range.size.km2/150000000
prop.landcover
prop.landcover*100
```

Repeat this process for your future projection for that species and for all the other species in your group. Record the current and future range areas for all of your species and

present this as a table or graph in your report. How are ranges expected to change in the future? Are all of your species responding to changes in climate in similar ways?

### **CALCULATING NICHE OVERLAP IN R**

Niche overlap is a measure of the similarity of niches and is highly correlated with geographic range overlap. You can use it to compare the similarity of the niches two species. As before, load the output files into R that you wish to compare. Note that you will need to replace the file names in quotes in the R code below to reflect the names used for your files.

First, you can calculate species area overlap in the present. Again, make sure your file names match those used in R. Use the following code.

#### *Current niche area overlap of Species one and Species two*

```
Species1 <-readAsciiGrid("Species1.asc")
Species2 <-readAsciiGrid("Species2.asc")

grid.list<-"NA"
grid.list<-as.list(grid.list)
grid.list[[1]]<-Species1
grid.list[[2]]<-Species2
NO.Sp1vsSp2<-niche.overlap(grid.list)
```

When you type in NO.Sp1vsSp2 a 2x2 matrix will appear containing two measures of niche overlap: the D statistic (upper right value) and the I statistic (bottom left corner). These two measures of niche overlap were proposed by Warren et al., 2008; read if you want more information on these measures. Both statistics scale from 0 (no overlap) to 1 (complete overlap). For the most part, the two statistics will give similar results. Report both in your final write up. These values should be calculated and reported for all species.

In a similar fashion, you can calculate overlap for species pairs in the future.

To load the prediction files for your species based on future climate models and calculate their predicted niche overlap, use the following code.

#### *Species 1 niche overlap with Species 2 in the future*

This is the same as for above but read in your future files from maxent instead of the current files.

```
Species1<-readAsciiGrid("Species1_future_bioclim.asc")
Species2<-readAsciiGrid("Species2_future_bioclim.asc")

grid.list<-"NA"
grid.list<-as.list(grid.list)
grid.list[[1]]<-Species1
grid.list[[2]]<-Species2
NO.Sp1vsSp2<-niche.overlap(grid.list)
```

How does the degree of niche overlap between your species change over time? What could be the biotic consequences of these changes?

**DUE DATES:**

By class time Oct 2<sup>nd</sup> - Prove you have generated your Maxent output

- 1) Email me- Save your html file as a pdf or print it to a pdf and email it to me
- 2) Meet with me before then and show me your results and we can discuss your interpretation.

By class time Oct 7<sup>th</sup> - Presentation and Group Report

**CLASS PRESENTATION:**

You will present a 10-minute presentation to the class. In this presentation, first introduce the species you chose and their relationship to each other, and present a hypothesis for how climate change might alter their interactions. Then present your results. Finally, interpret your results in terms of how climate change might alter interactions among your species. Make sure to indicate how pervasive non-analogue climates were in your simulations. See grading rubric on next page.

**GROUP REPORT:**

Format the report in the style of a scientific article with an Introduction, Methods, Results, and Discussion.

- In the introduction provide a summary of the species you chose and how they interact. Introduce your first sub-hypothesis about which climate variables will be most vital to determining the range of each species (these may vary by species). Next motivate and introduce your second hypothesis regarding how species range area and distributions may change and third hypotheses regarding how climate change may alter the interactions (and niche overlap) between these species.
- In the Methods section describe how you characterized the niche of both species. Include details like the number of observations you have for each species. You do not need to present in detail how you did the analysis, as that is already written in the instructions for the exercise. What you should include is: a description of what taxa you used, and why; what specifically you evaluated (eg distribution? environmental variables?) and what method you used to evaluate them (indicate which procedure in MaxEnt, but no need to describe in more detail); what specific comparisons you made, and why. You should indicate which statistics/tests you used to address each of your questions.
- In the Results section, provide a systematic and organized presentation of the results, including the statistical information. Keep in mind the purpose of each analysis and the question it addresses, and present your results in terms of answers to those questions. Highlight your major findings. What climate variables matter

most for each species? How does range size change with climate change? What is the degree of niche overlap for under current versus future climate conditions?

- Finally, in the Discussion, interpret your results in terms of how you expect climate change to affect your species. Did you find what you expect? Did any of your results surprise you? *Use this exercise to generate hypotheses that can be tested with other methods. What future questions and hypotheses emerged from your results?*

Notes: There is no official page limit for this, we expect it to be as concise as possible to fully report, explore, and evaluate your hypothesis. A rough guideline is: 1-1.5 page intro, 1 page methods, 1-2 pages of results and 2 pages of discussion. Consider 8 single-spaced pages without the figures as an absolute maximum. This assignment will require outside research and citations of relevant literature to support your hypothesis of how species respond to environmental factors and how your species assemblage interacts.

Rubric for presentation (20% of project grade)

Criteria	
1. Equal contribution of group members. All group members speak clearly and are easy to understand and the time frame of the presentation is respected	/20
Notes:	
2. Information is presented in an organized way and hypotheses and biological motivations are clear	/20
Notes:	
3. Results are interpreted thoroughly and accurately.	/20
Notes	
4. Presentation includes some details but clearly emphasizes a few take home messages	/20
Notes:	
5. Group uses visual aids such as the white board or power point effectively	/20
Notes:	
Total score: /100	

## Rubric for group report (80% of project grade)

Criteria	
1. Quality and motivation of the hypotheses (25%)	/25
<ul style="list-style-type: none"> <li>a. Clear motivation-</li> <li>b. Hypotheses indicates directionality-</li> <li>c. Sub hypothesis concerning what bioclimatic variables may be most important for each species</li> <li>d. Sub-hypothesis concerning how species range sizes and locations will change in the future for each species</li> <li>e. Sub-hypothesis concerning how species ranges will change in relation to each other in terms of niche overlap for at least one species pair</li> </ul>	
2. Correct identification of necessary comparisons (25%)	/25
<ul style="list-style-type: none"> <li>a. included which bioclim variable are most imp. for each species-</li> <li>b. included maps of all species now and in the future-</li> <li>c. included table or graph of range size estimates for each species-</li> <li>d. included appropriate niche overlap comparisons to test the stated hypothesis-</li> <li>e. clearly articulated major findings in results section</li> <li>f. other-</li> </ul>	
3. Appropriateness and creativity of interpretation (25%)	/25
<ul style="list-style-type: none"> <li>a. demonstrate an understanding of and discuss cogently how to interpret with bioclimatic variable are most important for each species-</li> <li>b. discussion of changes in range/niche overlap within a species-</li> <li>c. interpretation of changes in ranges and niche overlap between species-</li> <li>d. consideration of the limitations of niche models in regards to results-</li> <li>e. other</li> </ul>	
4. Clarity of writing and presentation of results (25%)	/25
<ul style="list-style-type: none"> <li>a. quality of organization</li> <li>b. clear take home messages</li> <li>c. other</li> </ul>	
Total score:	

Rubric for grading other group member's participation

Criteria	
1. Contributed the projection of assigned species and did so in a timely fashion	/20
Notes:	
2. Offered help to others, or sought help when needed	/20
Notes:	
3. Asked questions that moved the discussion along	/20
Notes	
4. Contributed ideas, opinions	/20
Notes:	
5. Provided constructive feedback to other group members	/20
Notes:	
Total score: /100	