Homework 4

Climate Change Ecology, Spring 2015

Due Monday, April 6, 2015 before class

Your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Please answer the following questions in a different font (**like this**) to help me find your answers. Turn in a print out of your homework. You may work in a group, but your answers and your words must be your own.

Species distribution modeling. You will download, run, and analyze a species distribution model called “MaxEnt” (maximum entropy). For this homework, you will get the software working and run it with supplied input data.

***NOTE: run MaxEnt on UI computers to avoid issues with permissions, etc. I am happy to help you if you have problems, but please run on UI computers.***

In this computer exercise, you will run MaxEnt to study the distribution of pikas (*Ochotona princeps*), a small alpine mammal that lives in the western United States. You will:

* study the potential explanatory variables;
* ascertain model quality;
* identify which explanatory variables are important in describing pika distribution and how those variables influence the distribution;
* create a map of current pika potential habitat;
* assess the sensitivity of model results to choice of pika presence data;
* quantify how climate change will affect the distribution of pikas.

A. Download and install MaxEnt and pika files.

1. Download and unzip MaxEnt version 3.3.3.k from http://www.cs.princeton.edu/~schapire/maxent (supply your information as requested). The model is coded in Java and should be run on UI computer lab PCs. There may be various ways of unzipping a file; one is to double-click on the zip file, then “extract” to a folder.
2. Download and unzip the tutorial zip file from the above web site. IMPORTANT: unzip or move these files to the same folder as where MaxEnt is (from step above).
3. Download the zip file for Homework 4 from the course web site (homework page) and unzip the file. Here, I have provided you with a CSV file that contains both the locations of the presence and background points as well as the values of the potential explanatory variables at those locations. IMPORTANT: unzip or move these files to the same folder as where MaxEnt is.

B. To begin, species distribution modelers collect (perhaps themselves) and/or assemble (perhaps from other studies, museum, etc.) species locations (presences or observations). Here, we will use observations provided by the tutorial. Species names (“FID”; in this case, only “pika”), observed locations of that species (longitude, latitude), and potential explanatory variables are taken from pika\_presences.csv. Open this file with Excel.

1. How many potential explanatory variables are there (latitude and longitude are not considered explanatory variables here)? (1 point)

2. Find the pika observation point that is -113.786º E longitude, 48.6135ºN latitude. What is the mean annual temperature (MAT) at this point (units are ºC)? (2 points)

I also provided you with the background points needed by MaxEnt in the file background\_pts\_random.csv. Background points are points distributed throughout a region to test model results (both presence and absence). These points were randomly located within the western US.

B. MaxEnt builds a model based on presence points (more specifically on the values of explanatory variables at these points). You can think of this model as analogous to an ordinary least squares regression model that finds a line that best fits a set of x, y data, and whose equation is y = mx + b. In the regression model, y is the dependent variable, x is the independent (explanatory) variable used to predict y, and m and b are parameters of the regression model that are determined by the least squares method. In MaxEnt,

* + the probability of presence of pikas is the dependent variable;
	+ the environmental layers (like precipitation) are the independent (explanatory) variables;
	+ there can be more than one explanatory variable (unlike the regression model above that has only one explanatory variable);
	+ the relationship between the explanatory variable and the response variable can be nonlinear (unlike the regression model above);
	+ and the model parameters and the method to compute them are complex.

To run MaxEnt, from Windows Explorer, double-click on maxent.bat. This should open up a MaxEnt window and allow you to specify inputs and run the model. Specify pika\_presences.csv as the “samples” file and background\_pts\_random.csv as the “environmental layers”; specify the output\_current\_climate output directory. Make sure map (mean annual precipitation), mat (mean annual temperature), mmax (maximum monthly temperature) are checked. Check the “Create response curves” and “Do jackknife…importance” boxes. For the output directory, select “output\_current\_climate”, and for the “projection layers directory”, specify layers\_current\_climate. Leave all other settings as defaults. Select the “Run” button. When it completes, open the html file in the output directory.

C. Assess the quality of the model (how well do the model results fit the observations (data used to develop the model)) by scrolling to the “Analysis of omission/commission” section and look at the table whose first column is “Cumulative threshold”. Let’s look at the “training omission rate” column. The rows in the table list various ways of converting predictions in continuous format to 0 or 1 for matching the response (independent variable).

3. What is the greatest omission rate of all the rows? Omissions are the locations where observations indicate the presence of pika but the model estimates an absence. Multiply by 100 to get omission percentage; is this percentage high (a poor model) or low (a good model)? (2 points)

D. We can study the model to learn which variables are important for explaining this species distribution and how each variable affects the distribution. Regarding variable importance: often the controls on a species’ distribution are not known, and so more variables are included in a model than actually control the distribution. Two methods are available to assess which variables are important and which variables are not needed. Within a set of variables, MaxEnt reports which variables contribute the most to explaining a distribution; this is in the html output file under “Analysis of variable contributions”. Inspect this table, and based on the “percent contribution” column, answer the following questions:

4. Which variable is most important? (2 points)

5. Which other variables contribute more than 10% to the distribution of pika (these could be identified as also important in explaining the distribution of the species)? (2 points)

A second method that is useful assesses variable important by rerunning the model with only one explanatory variable at a time, then compares models. This “jackknife” procedure builds one model per variable, and the model takes some time to run. Under the “Analysis of variable contributions” section of the resulting output html file, there is a color bar chart. Focus on the dark blue bars, which indicate the amount of variance explained by each variable when that variable is the only one in a model.

6. Which variable has the highest contribution (is most important) (“regularized training gain”)? (1 point)

7. Which has the lowest contribution? (1 point)

E. A second use of the model to increase the scientific understanding of the controls on a species’ distribution is to examine the modeled relationship between each explanatory variable and the probability of observing that species. In the output html file from the step above, go to the section called “Response curves”. There are two sets of three plots each; focus on the top set. Each of these three plots is associated with one of the explanatory variables (along the x axis); the y-axis is the probability of observing pikas. Ignore the small wiggles in the curves and focus on the general patterns.

8. What is the relationship between the most important variable and the probability of observation? Does the probability go up as the variable increases, go down, or remain flat? Is there any evidence of broad nonlinear patterns? (4 points)

9. What is the general relationship between another variable, mean annual precipitation, and the probability of observation? (4 points)

F. A person developing species distribution models should at this point compare the understanding gained from the model with any available information about the species of interest.

10. Read the Wikipedia page on this species (http://en.wikipedia.org/wiki/American\_pika) and consider the discussion of climate influences under “Conservation and decline”. Is there general consistency among climate, ecological understanding of this species, and the model results? How so? (4 points)

G. Once a model is developed, a user can then apply (“project” in the MaxEnt tutorial document) this model using other values of the explanatory (environmental) variables to predict the probability of a pika being observed given the values of these explanatory variables. In Maxent, usually this step involves specifying grids (rasters, or maps) of explanatory variables (the same as used to build the model), which results in a map of the probability. The output html file has provided a map that predicts the probability of observation of pikas for the full range of the layer grids, which is the western US.

Inspect the first map in the HTML file under Pictures of the model” that shows the model training and test points and the model prediction. Click on the map to open the larger version.

11. Where are the training points (white squares) located? Compare with a map of the western US that shows states and other geographical features. (2 points)

12. Where does the model predict the highest probability of pika habitat (generally; multiple locations are possible)? (2 points)

H. Now assess the sensitivity of the model predictions to availability of pika observations. Pretend you only have pika observations east of -109º E longitude (in pika\_presences\_east.csv). Rerun MaxEnt with that as your samples file (change the output directory and make sure all explanatory variables are checked; also, uncheck “Do jackknife...” to speed up the calculation) and compare the new map with the old map.

13. Where on the map (about what state or major geographic feature) is the westernmost pika observation? (2 points)

14. What are the largest differences between the two maps? (2 points)

15. Why do these differences occur? Open the Excel file called temperature\_frequency\_distributions.xlsx and view the two charts for MAT (first worksheet) and MMAX (second worksheet). Each chart is a frequency distribution: the y-axis indicates how many presence points have values at the x-axis value. Each chart shows the frequency distribution for all points (blue) and for just the eastern points (red). What is the main general difference between the frequency distributions for all versus east (same general patterns occurs for MAT and MMAX) (hint: what is the differences in means and why?)? (2 points)

16. What are the implications of this sensitivity test for species distribution modelers? (2 points)

I. Now estimate how climate change will affect the modeled distribution of pikas. Rerun MaxEnt with all the presence points (change samples files back), specify a new output directory, and change the projection layers directory to layers\_climch\_projection\_GFDLCM21\_A2. Make sure all explanatory variables are checked. No need for jackknifing or creating response curves. After running, click on the projection map within the html file to open the larger version of the map.

17. What are the general differences between the current predicted distribution of pikas and the modeled future distribution given warming? (2 points)

18. If you were employed by the US Fish and Wildlife Service, which was charged to protect the species, how would these maps inform your decision making about where to invest in protection actions (such as creating reserves)? (3 points)

Email me this Word document as well as the three HTML files that resulted from your three MaxEnt runs (in the three output folders).