EEB 5301; Problem Set 2
Due February 28

There are two tutorials. I posted two R scripts, one for point pattern analysis and a more complete one for species distribution models. Recreate the examples presented in the tutorials. Then, you will apply the same methods to a new data set.

Part 1. *Analysis of point patterns.* See the document titled "Spatial Analyses in R 2017", which is available from the HuskyCT site for the course.

Go through the tutorial following the Japanese pines and longleaf pines examples. The more involved code (used to make a figure resembling Fig. 5.6 in the text) is in the ProblemSet2SpatialPattern.r file.

Then carry out the same steps to analyze the positions of white oaks in Lansing woods, Michigan. The "lansing" data set is included in the spatstat package. This is an example of a "marked" point pattern – each point is marked with the species identity of the tree. To work with just the white oaks, they can be extracted from the entire data set as follows:

```
whiteoaks <- lansing[lansing$marks == "whiteoak"]
```

This "whiteoaks" object can now be used in the same way as "japanesepines" in the tutorial.

Answer the following questions. You can copy graphs from R and paste them into a Word document to show your results.

(a) What is the Clark and Evans index for the white oaks at Lansing woods? (Use Donnelly's edge correction).
(b) Is the Clark and Evans index significantly different from 1?
(c) Produce a plot of Ripley's *L* for the white oaks, including confidence envelopes for spatial randomness. Include a copy of this plot in your answer. What does this plot indicate about the spatial pattern?
(d) Divide the plot into a 10 by 10 grid of quadrats. (Notice that the longleaf pine example used a 15 by 15 grid, so you need to change 15 to 10 in several places.) Plot the observed and predicted frequencies of quadrat counts and test whether the observed values are consistent with the null hypothesis of complete spatial randomness.
(e) Interpret these results. What do they tell you about the spatial distributions of the white oaks at various spatial scales?

Part 2: Species distribution models. See the document titled "Species Distribution Models in ", which is available from the HuskyCT site for the course.

Go through the tutorial following the bradypus (*Bradypus variegatus*) example. You'll need to download MaxEnt and move a file to a particular folder (see the tutorial). The R

code is in the "Species_Distribution_Models.R" file. Executing the commands in the R script should be easy if you have the packages installed and the other software (MaxEnt and Java) installed. (So it's a good test to see if you have everything set up correctly.)

Then, apply the same methods to a new data set on fishers, *Martes* (or *Pekania*) *pennanti*, a carnivore in the weasel family that occurs here in Connecticut. The **fisher.csv** file is in the RStuff folder in the HuskyCT site. You'll need to copy it into your R working folder. (I made the file by downloading data from gbif, cleaning it up, and thinning records that are close together. I can't vouch for the quality of the data, but it's suitable for this exercise.) To fit MaxEnt models and boosted regression trees to the data on fishers, you will need to change several of the commands used for the bradypus example. Details on the needed changes are given below.

Answer the following questions. You can copy graphs from R and paste them into a Word document to show your results.

(a) Show the following plots for your species distribution models of fishers (like those produced in the tutorial for bradypus).

A plot of the occurrences of fishers on a map of one of your predictors.

A plot of your MaxEnt model showing the contributions of the variables.

Response plots for the predictors (or just for the most important predictors).

The pair of maps showing the MaxEnt raw values and of the region divided by some threshold into suitable habitat (presence) and unsuitable habitat (absence).

The pair of maps showing the predicted probabilities from the random forest model and of the region divided by some threshold into suitable habitat (presence) and unsuitable habitat (absence).

The plot showing the relative influence of each predictor in your gbm model.

The plots showing the fitted functions for each predictor of your gbm model.

(b) Compare the predictive abilities of the MaxEnt and boosted regression trees models.

(c) Compare the influences of different predictors in your MaxEnt and regression tree models. Which variables are most important? How do they affect the predicted occurrence of fishers?

These are the changes needed to fit MaxEnt models and boosted regression trees to the data set on fishers. Most of the lines in Species_Distribution_Models.R, hereafter referred to as the bradypus R script, do not have to be changed.

(i) Put the fisher.csv file in your working folder. The command to set the working folder looks like this: setwd("C:/Users/Adams/Documents/RWorking")
.. but you must change it to fit the path for your own working folder.

(ii) These two commands make a data frame called "fishers," read in the data, and drop the first column (which isn't needed). The fishers.csv file must be in your working folder or you will get an error message. (These lines substitute for lines 20-23 in the bradypus R script.)

```
fishers <- read.csv(file = "fishers.csv")
fishers <- fishers[,-1]
```

(iii) Anywhere "**bradypus**" is used in the tutorial, substitute "**fishers**".

(iv) Read in climate data and clip it to the extent of the fisher data (or slightly larger). (These lines substitute for lines 30-37 in the bradypus R script.) The first command goes online to download the bioclim layers – this may take a few minutes. The next line declares an extent that is appropriate for the data set. The third line crops the climate layers to that region and stores the results in climN.

```
climate <- getData('worldclim', var='bio', res=2.5)
extf <- extent(-142, -65, 32, 69 )
climN <- crop(climate, extf)
```

Choose a subset of these layers as the predictors. There are 19 in total (listed at the end of this document), so you can pick any or all numbers from 1 to 19. The next line selects bio1, bio5, bio6, bio12, bio15, and bio19, which works fairly well, but you're welcome to add others.

```
predictors <- subset(climN, c(2, 5, 6, 12, 15, 19))
```

(v) The extent (latitudinal and longitudinal limits) of the fishers data set differ from those of the bradypus data set. Replace line 35 in the bradypus R script with the following:

```
extb = extent(-138, -65, 33, 62)
```

(vi) When you make the plot of the training and test sets for the presences and background points, be patient. It may take a minute or a few minutes. (The commands are lines 76 to 87 in the bradypus R script.)

(vii) You may want to select different predictors for inclusion in the boosted regression tree models. Modify line 153 in the bradypus R script as desired; e.g.,

```
model <- pa ~ bio2 + bio5 + bio6 + bio12+ bio15 + bio19
```

(viii) Depending on how many predictors you are using, you will probably need to change gbm.x = 2:9 in lines 172-174 of the bradypus R script. For example, if you use five predictors, they will be in columns 2 through 6 of the "envtrain" data frame.

```
rf2 <- gbm.step(data=envtrain, gbm.x = 2:6, gbm.y = 1,
```

```
family = "bernoulli", tree.complexity = 5,
learning.rate = 0.005, bag.fraction = 0.5)
```

These are the 19 bioclim layers:

- bio1 : annual mean temperature
- bio2 : mean diurnal range [mean of monthly (max temp - min temp)]
- bio3 : isothrmality (bio2/bio7)*100
- bio4 : temperature seasonality (st dev *100)
- bio5 : max temperature of the warmest month
- bio6 : min temperature of the coldest month
- bio7 : temperature annual range (bio5 - bio6)
- bio8 : mean temperature of wettest quarter
- bio9 : mean temperature of driest quarter
- bio10 : mean temperature of warmest quarter
- bio11 : mean temperature of coldest quarter
- bio12 : annual precipitation
- bio13 : precipitation of wettest month
- bio14 : precipitation of driest month
- bio15 : precipitation seasonality (coeff. var.)
- bio16 : precipitation of wettest quarter
- bio17 : precipitation of driest quarter
- bio18 : precipitation of warmest quarter
- bio19 : precipitation of coldest quarter