**GIS and remote sensing for conservation and evolutionary biology**

**Session 5: computer lab exercise**

**Aim:** To run niche-based distribution models in R.

The basis for this week’s lab is a tutorial written by Robert J. Hijmans and Jane Elith: *Species distribution modeling with R* (2017). Please open the pdf of this document, and follow the tutorial. The pdf is located in the data folder for this week, or go to: https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf

The tutorial provides an introduction to many powerful methods that are available for modeling species distributions and niches in R. You are encouraged to work through as much of the tutorial as possible – there is a huge amount of very useful information that will set you up to do some fairly sophisticated analyses if you work through it. However, the notes below pick out the essential things that you must cover as part of the course.

The species and environmental data you have prepared in previous weeks can be used in this tutorial, so you should use your own case study species instead of the *Bradypus* example. Because you will not be using the same data, we suggest you do not just copy the code from the tutorial; by typing each command you will familiarize yourself with the R language and will make you input the right file paths (rather than the file paths written in the tutorial).

**Mac Users:** The most recent versions of Mac OSX may have problems running Maxent directly from R. Running the script below before loading ‘dismo’ should fix these problems.

Sys.setenv(NOAWT=TRUE) #only for certain macs before loading rJava

library(rJava) # required to run Maxent within dismo

options(java.parameters = "-Xmx1g" ) #optional, to set the memory allocated to java, hence maxent (has to be done before loading dismo)

library(dismo) # should also load automatically the required packages sp and raster

**Note to Mac OS X El Capitan users:** If you are receiving errors regarding rJava, see this link: [http://charlotte-ngs.github.io/2016/01/MacOsXrJavaProblem.html](https://na01.safelinks.protection.outlook.com/?url=http%3A%2F%2Fcharlotte-ngs.github.io%2F2016%2F01%2FMacOsXrJavaProblem.html&data=01%7C01%7Cpgalante%40amnh.org%7C5aa8bc4a5c9f419f74a308d4b4e84cfd%7Cbe0003e8c6b9496883aeb34586974b76%7C0&sdata=Viv0hVmDVPCh8aXs8SwekIPankL8RhUp2oT6Vguf6LE%3D&reserved=0)

**Part 1:**

If you do not already have it, first you’ll need to download the R software package from [*http://cran.r-project.org/*](http://cran.r-project.org/). If you do not have any experience with R we recommend that you also download the R Manuals “An Introduction to R” and “R Data Import/Export” so that you can explore the basics of the R language. You may also want to download a workspace viewer for R such as R-studio *(*[*http://www.rstudio.com/ide/*](http://www.rstudio.com/ide/)*)*, which allows you to view your code, your data, your file workspace, and graphics/figures at the same time. R-studio also includes some user-friendly tabs that make it much easier to install packages, navigate within your workspace, and import data.

Having installed the necessary packages (*Following the instructions provided in the file “GIS\_Setup\_in\_R.R”*), you should learn to import your occurrence data. Specifically, load the dismo library, and create an object with your occurrence data by reading in your csv species file, and keeping only columns 2 and 3 (see p.4-6 section 2.1). It is important that you create an object instead of just reading the file. Remember to set your working directory before importing files using *setwd* – and, if you are using R Studio, you can import the csv file using the “Import Dataset” tool under Workspace). Note that your latitude and longitude data should be in decimal degrees. After you import them you can clean your data directly in R (see p.8-9 section 2.2); alternatively, you can clean your locality points before creating the CSV file using QGIS.

Next, you should load your environmental data and create a stack of all your raster layers (*rasterStack*, p.23-24). Note that when using your GTiff (or ASCII) grids in each of the above steps, you will need to specify that you are looking for .asc instead of .grd files. We have used ASCII grids because the format is fairly easy to understand, but the tutorial rightly notes that other GIS raster file formats can be much more efficient. A significant advantage of using R, as opposed to the Maxent software, is that it can incorporate a huge variety of different file types. You should then create background data (e.g., for use with Maxent) using the *randomPoints* function (p.18-19). Instead of the 500 used in the tutorial, you should use a minimum of 10,000. Note that you’ll need to exclude areas with no data using *mask* but you may not need to restrict the *extent.*

If you are having trouble at any point, utilize R’s help function using “*?”* or *help( )*; for example, type *?read.table* or *help(read.table)* for help with the *read.table* function. R-Seek is also a great web resource to search for R code *http://www.rseek.org/*.

That is all you need for the essentials to get your data into R. The next task is to actually fit some models.

**Part 2:**

First, divide your occurrence records into separate datasets for training and testing (use ‘*kfold*’, p. 39). Do the same for your background points. Now you have the data to fit a suite of many different methods (BIOCLIM, DOMAIN, Maxent, GLMs, GAMs, Boosted Regression Trees, and many more!). For this lab, you should fit a bioclim model (p.46-48) and a Maxent model (p. 56-58). Can you recreate in R what you previously did using the Maxent software?

Finally, you should save the data in R such that you can then use it in a GIS. Once back in the GIS, then you can use all the power of that tool to undertake spatial analyses on the model results and draw clear and accurate maps.

To get your model predictions out of R and into a GIS, use the *writeRaster* function. For example,

*> writeRaster(pb, filename=”bioclim.asc”)*

Use a filename of your choice – you can export in a lot of raster file formats including asc and grd. The file will be saved to whatever folder you have defined as your workspace.

To conclude the formal, required parts of the exercise, you should reach the point where you have maps of a bioclim and Maxent model for your species, drawn neatly in a GIS. Questions to think about, for discussion at the start of next session: Are the Bioclim and Maxent models similar? If not, why might that be? Are there any interesting biological findings suggested by the models?

Finally, we emphasize that we have just scratched the surface of the possibilities that using R opens up. Our goal in this exercise is to give you a small introduction to a potentially important set of tools. Those of you who are interested in pursuing this line of research further are encouraged to go through the tutorial in much more detail.

For example, the lambdas file (covered here: <https://docs.google.com/viewer?a=v&pid=forums&srcid=MDk5ODk4MTMxNjU4MDYzMDM0NzcBMTcyNTk1MTUxODc1NDE4MDU3ODYBOVlQQlJMNEgyellKATAuMQEBdjI&authuser=0>) may shed some light on variables that contribute to your model.

For more on using R as a GIS and on how to import and export vector data including shapefiles see R help, or also this web resource: <https://sites.google.com/site/rodriguezsanchezf/news/usingrasagis>

**R web resources**:

R software download and manuals: <http://cran.r-project.org/>

R-studio workspace: http://www.rstudio.com/ide/

R-seek (for help or to find R code): http://www.rseek.org/

Quick-R (more on statistics in R): http://www.statmethods.net/index.html

Using R for GIS: <https://sites.google.com/site/rodriguezsanchezf/news/usingrasagis>