Exercise 0: Getting Started in R

Adam B. Smith, Danielle Christianson, & Camilo Sanín

August 15, 2017

This workshop will walk you through all of the fundamental steps in modeling species' niches and distributions. This initial exercise will simply get you started in R.

# Workshop materials

All materials for the workshop (including a copy of this very first tutorial) can be found at [Earth::Sky::Sea (www.EarthSkySea.org)](http://www.earthskysea.org/workshops-classes/). Please download the zip file, uncompress it, and save the contents into a single folder.

# R

If you haven't already installed R, please [download](https://www.r-project.org/) the installer and do so.

Note that if you have a 64-bit computer (made, say, in the last 5 yr), then the 64-bit version of R has some advantages over the 32-bit version.

# Working folder

**The folder into which you uncompressed the files will become your working folder in which all files we make will be saved. All of the scripts in this workshop assume that files are saved and available in this directory, so don't save them in folders outside this folder.** We'll be making subfolders within this directory called "Models", "Species Records", etc. Modeling requires drawing data from many different sources, so we highly recommend using a single folder in which to store all scripts, models, and pre-processed results for a single project. Otherwise, it can be very difficult to recreate a modeling project.

It's also a lot easier to transfer a project to a different computer because you can use just change the working folder of a project and run your script:

# set working directory (Windows)  
setwd('C:/Ecology/SDM Workshop')  
  
# set working directory (Mac)  
setwd('/Users/<username>/Ecology/SDM Workshop')  
  
# get current working directory  
getwd()

# R libraries

On any computer you can install packages using the code below. However, if you are on a PC you can install packages from the Install packages(s) command in the Packages menu. If you are on a Mac, you can install packages from the Package Installer command in the Packages & Data menu. If you're running Linux you will need to install the GDAL spatial libraries outside of R. We are sorry, we probably can't help you beyond that!

We'll be using a number of R libraries to do the modeling.

install.packages(  
 c('dismo',  
 'raster',  
 'maxnet',  
 'rgbif',  
 'rgdal',  
 'rgeos',  
 'geosphere',  
 'scales'  
 )  
)

dismo and raster are the main "workhorse" packages we'll be using. dismo (="distribution modeling") is led by Robert Hijmans (UC Davis) and has an ever-evolving set of functions that enable modeling in geographic contexts. raster, also by Hijmans, handles raster data. The sp package (which is installed automatically when you install dismo or raster) handles polygon (shapefile) data. In turn, these libraries often use the GDAL (or rgdal) code to handle spatial data. The maxnet library runs "MaxNet"", an open-source version of the "MaxEnt" modeling algorithm. rgbif connects to the largest online repository of specimen records. The other libraries have very useful utility functions.

We've also created two packages, **ombinus** and **enmSdm** that you will need to install in R. The packages are located in the folder .\R Packages, and you will have to manually install them.

Using RStudio: Packages tab --> 'Install-->Install from-->Package archive`. Select the two package files.

Using R in Windows: Install each of the packages using the menu Packages --> Install package(s) from local files....

Using R in iOs:

install.packages('/Users/<username>/<path to uncompressed zip file>/SDM Workshop/R Packages/omnibus\_0.1.0.tar.gz', repros=NULL, type="source")  
install.packages('/Users/<username>/<path to uncompressed zip file>/SDM Workshop/R Packages/enmSdm\_0.2.0.0.tar.gz', repros=NULL, type="source")

Load the libraries into R's memory:

# modeling/biodiversity packages  
library(dismo)  
library(maxnet)  
library(rgbif)  
  
# geographic tools packages  
library(raster)  
library(rgeos)  
library(geosphere)  
library(rgdal)

# extra  
library(scales)  
  
# custom  
library(omnibus)  
library(enmSdm)

# Memory in R

Distribution modeling and geographic data are memory-intensive, so having ample computing capacity is essential to producing timely results.

If you are running a 32-bit version of Windows, then the amount of memory available to R is limited. Even increasing the RAM installed on the computer won't help because Windows limits memory given to programs to 2 GB.

If you're running a 64-bit version of Windows then your memory is still limited, but you can increase it to 3 GB using this command:

memory.limit(memory.limit() \* 2^30)

Note that if this gives you an error try reducing the exponent (30) to a lower number.

If you're running Linux it should give R all the available memory the computer has!

If you're running iOS, you're less limited--you should get 3 GB on a 32-bit system and essentially no limit on a 64-bit system.

# Tips

**Any operating system**: We suggest you use a different script editor from the one provided in R. Other editors have very useful features such as syntax highlighting (e.g., comments, numbers, function names, etc. are each shown in different colors) which can be very useful for debugging. I use [Notepad++](https://notepad-plus-plus.org/), but [RStudio](https://www.rstudio.com/) is even fancier and provides some very user-friendly functionality. Mac users might enjoy of [TextWrangler](http://www.barebones.com/products/textwrangler/).

**Windows**: If you find yourself changing Window's directory names' backslashes to forward slashes too often for your liking, the program [Path Copy Copy](https://pathcopycopy.codeplex.com/) offers a very convenient right-click menu that will copy a file's/folder's name with forward slashes. You have to install it then enable an optional copy command in the "Settings" dialog.

**Copying commands**: The tutorial files are made in rmarkdown which doesn't automatically nicely indent long lines of code. As a result, long lines of code would look like a jumbled mess:

plot(cbind(randomBg$WC01, randomBg$WC12), pch=16, col=alpha('mediumseagreen', 0.2), xlim=c(min(randomBg$WC01), max(randomBg$WC01)), ylim=c(min(randomBg$WC12), max(randomBg$WC12)), main='Random Background', xlab='MAT', ylab='MAP', cex=1.2)

As a result, in some cases we've added line breaks manually so you can read them more easily:

plot(  
 cbind(randomBg$WC01, randomBg$WC12),  
 pch=16,  
 col=alpha('mediumseagreen', 0.2),  
 xlim=c(min(randomBg$WC01), max(randomBg$WC01)),  
 ylim=c(min(randomBg$WC12), max(randomBg$WC12)),  
 main='Random Background',  
 xlab='MAT',  
 ylab='MAP',  
 cex=1.2  
)

Some people have found (especially those using RStudio on a Mac) that copying these lines with manual breaks and pasting them into R introduces an error. For some reason code on multiple lines has an invisible character embedded in it that throws an error in R. To fix this copy the line into your script editor then put it all on one line, being sure to delete (then re-enter if you want) spaces that appear between lines.

Finally, you will probably find that files paths and names that are split won't work unless you put them on the same line:

# may not work if copied and pasted as-is  
myFile <- read.csv('C:/My Folder is Awesome/  
 So Is My File.csv')  
# will work  
myFile <- read.csv('C:/My Folder is Awesome/So Is My File.csv')

# Crash but don't burn

The exercises in this workshop are meant to be completed in a series. Later exercises often depend on products make by earlier exercises. Owing to the intensity of some of the processes, your R session may crash or hang (seems to be especially a problem when running RStudio). Please save your workspace periodically!!! Use either:

save.image('SDM Workshop', compress=TRUE)

or in R, use the File menu then select Save Workspace....

However, at some point you may need to restart everything because you forgot to save your workspace. Even if you did, some of the kinds of data objects we're using probably won't actually load. **If you need to start all over, restart R and copy and paste the script below into R. Do not run the script below right now!!!**

Note that:  
1. You will need to change the working folder to your working folder.  
2. Unless you completed all of the tutorials and thus created all the data files, at some point you will get errors because R won't be able to find the given data files. Don't worry--just wait until it's fully pasted then continue.  
3. If you're lucky you can continue from the point in a tutorial where you left off. But if not you will have to restart or at least backtrack in a tutorial.

# CRASH BUT DON'T BURN!  
  
# set working directory to YOUR working directory  
# PC:  
setwd('C:/ecology/SDM Workshop')  
  
# Mac:  
setwd('/Users/<username>/SDM Workshop/')  
  
# load packages  
library(dismo)  
library(maxnet)  
library(raster)  
library(rgbif)  
  
library(rgeos)  
library(geosphere)  
library(rgdal)  
library(scales)  
  
library(omnibus)  
library(enmSdm)  
  
# load countries map  
countries <- rgdal::readOGR('./World Borders Dataset', 'TM\_WORLD\_BORDERS-0.3', verbose=FALSE)  
  
# load species data... will load the latest version you created  
records <- readRDS('./Species Records/00 Species Records - Merged All Raw Data Sets.rds')  
records <- readRDS('./Species Records/01 Species Records - Removed Observations.rds')  
records <- readRDS('./Species Records/02 Species Records - Retained Records between 1970-2000.rds')  
records <- readRDS('./Species Records/03 Species Records - Removed Observations with Bad Coordinates.rds')  
records <- readRDS('./Species Records/04 Species Records - Matched with WORLCLIM Data.rds')  
records <- readRDS('./Species Records/05 Species Records - Eliminated All But One Record Per Cell.rds')  
records <- readRDS('./Species Records/06 Species Records - Randomly Thinned Presences.rds')  
  
# create spatial version of records  
recordsSpatial <- SpatialPointsDataFrame(  
 coords=cbind(records$longitude, records$latitude),   
 data=records,  
 proj4string=CRS('+proj=longlat +datum=WGS84 +no\_defs +ellps=WGS84 +towgs84=0,0,0')  
)  
  
# load range map  
rangeMap <- readOGR('./Urocitellus columbianus', 'urocitellusColumbianus\_iucnRangeMap', verbose=FALSE)  
  
# study extent  
load('./Study Region/Study Region Extent.Rdata')  
  
# load background sites  
load('./Background Sites/Random Background Sites across Study Region.Rdata')  
load('./Background Sites/Target Background Sites Drawn from All Spermophilus Downloaded from GBIF.Rdata')  
load('./Background Sites/Bias Backgound Sites Drawn from Road Density.Rdata')  
  
# define selected predictors  
predictors <- c('WC03', 'WC10', 'WC15', 'WC19')  
  
# load clipped climate stack  
climate <- stack(list.files('.//WORLDCLIM/1970-2000/Study Region', full.names=TRUE, pattern='.tif'))  
climate <- subset(climate, predictors)  
  
# load every model up to latest trained  
load('./Models/Model 01 Predictors - Automated Selection/Model.Rdata')  
load('./Models/Model 02 Predictors - Manual Selection/Model.Rdata')  
load('./Models/Model 03 Bias Correction - Target Background/Model.Rdata')  
load('./Models/Model 04 Bias Correction - Bias Background/Model.Rdata')  
load('./Models/Model 05 Bias Correction - Spatially Thinned Presences/Model.Rdata')  
load('./Models/Model 06 Bias Correction - Env Thinned Presences/Model.Rdata')  
load('./Models/Model 07 Model Tuning - Feature Selection/Model - Hinge Features.Rdata')  
load('./Models/Model 08 Model Tuning - Beta Parameter/Model.Rdata')  
load('./Models/Model 11 Niche Model/Model.Rdata')  
  
# load model rasters up to latest written  
autoSelectMap <- raster('./Models/Model 01 Predictors - Automated Selection/maxentPrediction1970to2000.tif')  
manualSelectMap <- raster('./Models/Model 02 Predictors - Manual Selection/maxentPrediction1970to2000.tif')  
targetBgMap <- raster('./Models/Model 03 Bias Correction - Target Background/maxentPrediction1970to2000.tif')  
biasBgMap <- raster('./Models/Model 04 Bias Correction - Bias Background/maxentPrediction1970to2000.tif')  
spatialThinMap <- raster('./Models/Model 05 Bias Correction - Spatially Thinned Presences/maxentPrediction1970to2000.tif')  
envThinMap <- raster('./Models/Model 06 Bias Correction - Env Thinned Presences/maxentPrediction1970to2000.tif')  
hingeMap <- raster('./Models/Model 07 Model Tuning - Feature Selection/maxentPrediction1970to2000.tif')  
tunedMap <- raster('./Models/Model 08 Model Tuning - Beta Parameter/maxentPrediction1970to2000.tif')  
futureMap <- raster('./Models/Model 08 Model Tuning - Beta Parameter/maxentPrediction2070sRcp85.tif')  
nicheMap <- raster('./Models/Model 11 Niche Model/maxentPrediction1970to2000.tif')  
  
bioclimMap <- raster('./Models/Model 12 Other Algorithms/bioclimMap.tif')  
domainMap <- raster('./Models/Model 12 Other Algorithms/domainMap.tif')  
glmMap <- raster('./Models/Model 12 Other Algorithms/glmMap.tif')  
bioclimMap <- raster('./Models/Model 12 Other Algorithms/bioclimMap.tif')  
gamMap <- raster('./Models/Model 12 Other Algorithms/gamMap.tif')  
brtMap <- raster('./Models/Model 12 Other Algorithms/brtMap.tif')  
rfMap <- raster('./Models/Model 12 Other Algorithms/rfMap.tif')