Occurrence and environmental data

Jesús N. Pinto-Ledezma and Jeannine Cavender-Bares

Set up your data and your working directory

Set up a working directory and put the two data files in that directory. Tell R that this is the directory you will be using, and read in your data:

```
setwd("path/for/your/directory")
```

Install and load the following packages

```
library(rgbif)
library(maptools)
library(CoordinateCleaner)
library(raster)
library(dismo)
library(rworldmap)
library(rgdal)
```

Occurrence data

As indicated during the lecture, there are many sources of occurrence data. Today we will learn how to obtain and process occurrence data available in the Global Biodiversity Information Facility (aka **GBIF**). The GBIF is an international network and research infrastructure with the aim of providing anyone, anywhere, open access to data about all types of life on Earth. For more information, please visit https://www.gbif.org/.

Although GBIF provide access to several occurrence data (millions of occurrence data points), some of the data contain issues, more especifically, is often unclear which spatial resolution the data represent.

Downloading GBIF occurrence data

Here we will use the R package **rgbif** to retrieve occurrence data. But first, let's explore some functions.

Check out the number of occurrences found in GBIF

```
occ_count()
```

Number of observations

```
occ_count(basisOfRecord = "OBSERVATION")
```

Number of georeferenced observations

```
occ_count(georeferenced = TRUE)
```

Number of occurrences reported for Brazil. Note that you can explore other countries.

```r
occ_count(country = isocodes[grep("Brazil", isocodes$name), "code"])
```

Number of observations reported for Brazil

```r
occ_count(country = isocodes[grep("Brazil", isocodes$name), "Brazil"], basisOfRecord = "OBSERVATION")
```

Now, let's combine information

```r
occ_count(country = "BR", georeferenced = TRUE, basisOfRecord = "OBSERVATION")
```

Until now we have explored the occurrences at country level, let's move on to species.

Check for synonyms

```r
name_suggest(q = "Furnarius rufus", rank = "species")
```

Check number of records

```r
occ_search(scientificName = "Furnarius rufus", limit = 10)
```

Now, let's download the records and plot them. Of course, the map will only help us judging the data quality if we have a rough idea where the species should occur. So, look up the species on the web first!

```r
gbif_furu <- occ_search(scientificName = "Furnarius rufus", return = "data", limit = 500)
```

```r
SA <- subset(countriesCoarse, continent == "South America")
plot(SA)
points(gbif_furu$decimalLongitude, gbif_furu$decimalLatitude, col = "red",  pch = 19)
```

That looks nice, but let's check if this data have some issues. To clean the data downloaded from GBIF, we will use the R package **CoordinateCleaner**, this package allows cleaning geographic coordinates using different cross-checks. Here, we first compare whether the coordinates for each entry match the country code provided for each entry and are no outliers.

```r
gbif_furu <- subset(gbif_furu, !is.na(decimalLatitude))
cl_gbif_furu <- clean_coordinates(gbif_furu, lon = "decimalLongitude", lat = "decimalLatitude", countri
```

```r
plot(SA)
points(gbif_furu$decimalLongitude, gbif_furu$decimalLatitude, col = "red", pch = 19)
points(gbif_furu$decimalLongitude[cl_gbif_furu$.summary], gbif_furu$decimalLatitude[cl_gbif_furu$.summa
```

Create an object with the cleaned data

```r
gbif_furu <- gbif_furu[cl_gbif_furu$.summary,]
```

Now let's try to mach the occurrence data with the geographical range of the species.

```r
FurRanges <- readOGR(dsn = "Data/Franges", layer = "Furnarii_ranges_geo")
```

Wait, those are the 652 species, we just need the Furnarius rufus geographical range. Let's filter the information.

```r
range_furu <- subset(FurRanges, SCINAME == "Furnarius rufus")
```

And now take a look if both occurrences from GBIF and the geographical range from BirdLife match in space.

```r
plot(SA)
plot(range_furu, col = "green", add = T)

# Overlay the GBIF records
points(gbif_furu$decimalLongitude, gbif_furu$decimalLatitude, col = "red",  pch = 19)
```

Hmmm, seems that there are some occurrence outside the species geographic range. *Could you explain why this is happening? Is the species geographic range underestimating the realized species distribution? Other ideas*

Bioclimatic data

During the lecture we talk about several sources of environmental data and their pros and cons, now let's try to get some data that are available online.

To do this, let's first create an empty folder where we will store the data

```
dir.create("Data/WorldClim")
```

And now using the funtion **getData()** from the package raster we will get climatic information from WorldClim **(Hijmans et al. 2005)**. These climatic variables represent 19 bioclimatic variables at a 10' resolution. Do you know what the 19 bioclimatic variables are? See here: http://www.worldclim.org/bioclim.

```
bios <- getData("worldclim", var = "bio", res = 10, download = TRUE, path = "Data/WorldClim")
```

Let's take a look at the bioclimatic data

```
bios
```

```
plot(bios)
```

Remote Sensing data

To obtain Satellite remote sensing data is not that easy as the bioclimatic data. Today insted of downloading information from the NASA servers, we will create **NEW BRAND** environmental variables. These environmental variables are known as Dynamic Habitat Index (DHI; Hobi et al. 2017, Pinto-Ledezma & Cavender-Bares 2020). In this case we will create the DHI of the Leaf Area Index (LAI) a biophysical variable that represent different ecosystem functions.

```
setwd("Data/LAI_2015_NW")
lf <- list.files(pattern = ".tif")
```

```
lai2015 <- stack(lf)
```

Explore the LAI data

```
lai2015
```

To calculate the DHIs we will use the funtion **demonExtract()** available in Jesús's GitHub. Note that it will take some time to make all calculation.

```
source("https://raw.githubusercontent.com/jesusNPL/RS-SDM_ENM/master/extractDemon.R")
```

```
dhi <- demonMETRICS(rslist = lai2015, ext = extent(SA), Nraster = length(lai2015), saveResults = TRUE)
```

Let's take a look...

```
dhi
```

```
plot(dhi)
```

```
plot(dhi$Seasonality)
# Overlay the GBIF records
points(gbif_furu$decimalLongitude, gbif_furu$decimalLatitude, col = "red",  pch = 19)
```

Joining occurrences and environmental variables

```
gbif_furu_BIOS <- gbif_furu[,
    c("key", "scientificName", "decimalLatitude", "decimalLongitude", "basisOfRecord", "speciesKey", "sp
```

Now using the coordinates from the GBIF occurrences we can extract information from the bioclimatic variables and the DHIs.

```
gbif_furu_BIOS <- cbind(gbif_furu_BIOS, raster::extract(x = bios,
    y = data.frame(gbif_furu_BIOS[, c("decimalLongitude", "decimalLatitude")])))
```

```
gbif_furu_BIOS <- cbind(gbif_furu_BIOS, raster::extract(x = dhi,
    y = data.frame(gbif_furu_BIOS[, c("decimalLongitude", "decimalLatitude")])))
```

```
gbif_furu_BIOS <- cbind(gbif_furu_BIOS, raster::extract(x = dhi,
    y = data.frame(gbif_furu_BIOS[, c("decimalLongitude", "decimalLatitude")])))
```

```
str(gbif_furu_BIOS)
```

That's all folks!

References

Hobi ML, Dubinin M, Graham CH, Coops NC, Clayton MK, Pidgeon AM, et al. A comparison of Dynamic Habitat Indices derived from different MODIS products as predictors of avian species richness. Remote Sens Environ. 2017;195: 142–152. doi:10.1016/j.rse.2017.04.018

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." International Journal of Climatology 25 (15): 1965–78. doi:/10.1002/joc.1276

Pinto-Ledezma, J.N and Cavender-Bares, J. (2020) Using remote sensing for modeling and monitoring species distributions. In Cavender-Bares, J., J. Gamon & P. Townsend (Eds.) Remote Sensing of Plant Biodiversity. Springer Nature