Exercise: Model your chosen species' habitat suitability under present and future climate conditions

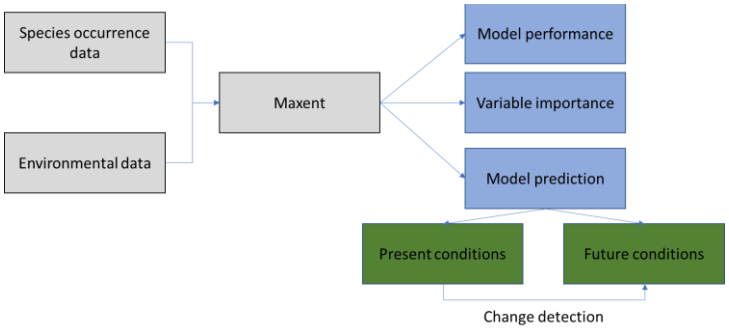
Introductory notes:

This manual serves as an example of how to prepare data and use it in MAXENT. This example can be different from what perhaps is needed for your case study species. Take note of that, but it should be enough to give you an example of all the options and alternatives available to you for setting up your own experiment.

In the first part of the tutorial we will "review" the occurrence selection, as this should have been done by you beforehand. On the second part we will show you an approach to select the environmental variables you can use, how to prepare them for your case study and finally how to do some statistical tests to help you infer if the group of variables selected is autocorrelated. In the last part we set up a MAXENT trial run so you can explore the diagnostics of this modelling tool. The last two parts will explore these diagnostics further and also the actual outputs of the model.

The following tutorial explores the SDM of Rhinolophys Euryale, the Mediterranean horseshoe bat. As the name implies, it is a type of bat characteristic of the Mediterranean region, implying already a climatic niche of dry summers and wet winters. In this case, we want to know if this species will potentially survive climate change.

This is a general overview of the main steps but do notice that each step has multiple steps within:



Step 0: Setting up your working environment:

Take some time to organize your working folders as this will facilitate your work on the report later. I chose the general following structure: One folder for all the original downloaded data, one folder for maxent with different subfolders for the results of each run. And independent folders for each of the environmental variables.

- Main folder: c:/Practical/
    - Downloads
        - All the original downloads go in here, as is, from GBIF and worldclim (zip format)
    - Occurrences
        - Store all the different occurrence data here, CSV or shapefile formats
    - Future
        - The original .tif files of each scenario are stored here, in **independent subfolders**
            - E.g. he26bi50; he45bi50 (..)
            - In each folder, the .tif files are stored
    - Future_AOI and Future WLD
        - Same structure as above, meaning **a subfolder for each future scenario with _WLD or _AOI**:
            - AOI – Refers to the environmental variables cropped to the Area Of Interest in .asc format
            - WLD – Refers to the uncropped environmental variables – WORLD in .asc format
    - Maxent
        - The maxent java files are stored here
        - Results folder
            - All the results should be stored here
            - OPTIONAL: have a subfolder here for each of the different experiments
    - Present
        - The original unzipped .tif files are stored here
    - Present_AOI and Present_WLD
        - On each of these folders you store:
            - AOI – Environmental variables cropped to the area of interest in .asc format
            - WLD – Uncropped environmental variables in .asc format
    - R_scripts
        - A folder where we can store different R scripts that we might have used

Notice:

- MAXENT requires that the occurrences data is in CSV format, tabular or comma separated (semi-colon versions will not be read correctly
- MAXENT requires that the environmental variables are in .asc format, hence why we need to create different folders to store this data.

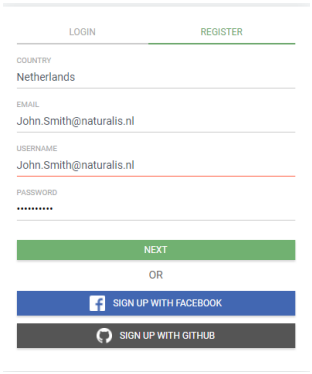
The script: 00_SettingUpWorkEnvironment.R creates the same set of directories as I used on this tutorial if you opt to do so.

Download maxent from: https://biodiversityinformatics.amnh.org/open_source/maxent/

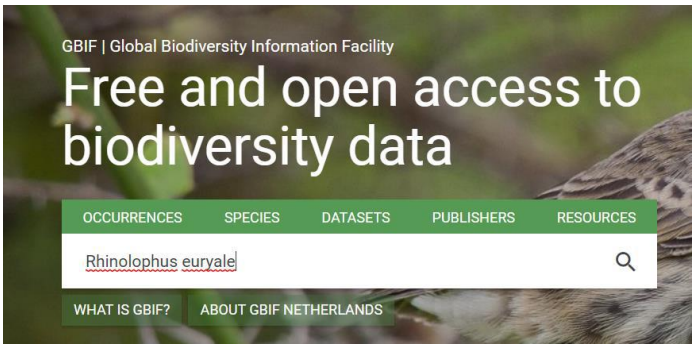
And unzip it to the maxent folder

Step 1: Species occurrence data

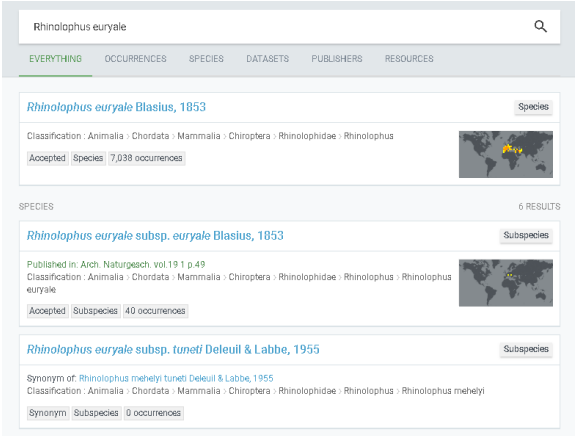
1. Choose your species and go to the GBIF www.gbif.org.
    a. Create an account and login as user to be allowed to download

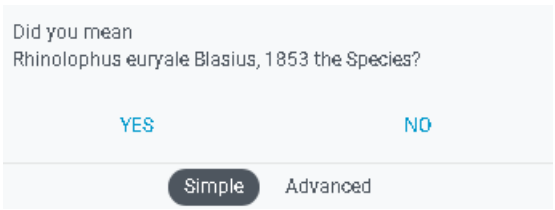


2. Search for the species of interest: this example – *Rhinolophys euryale*



3. It will show you multiple options, including some other potential synonyms





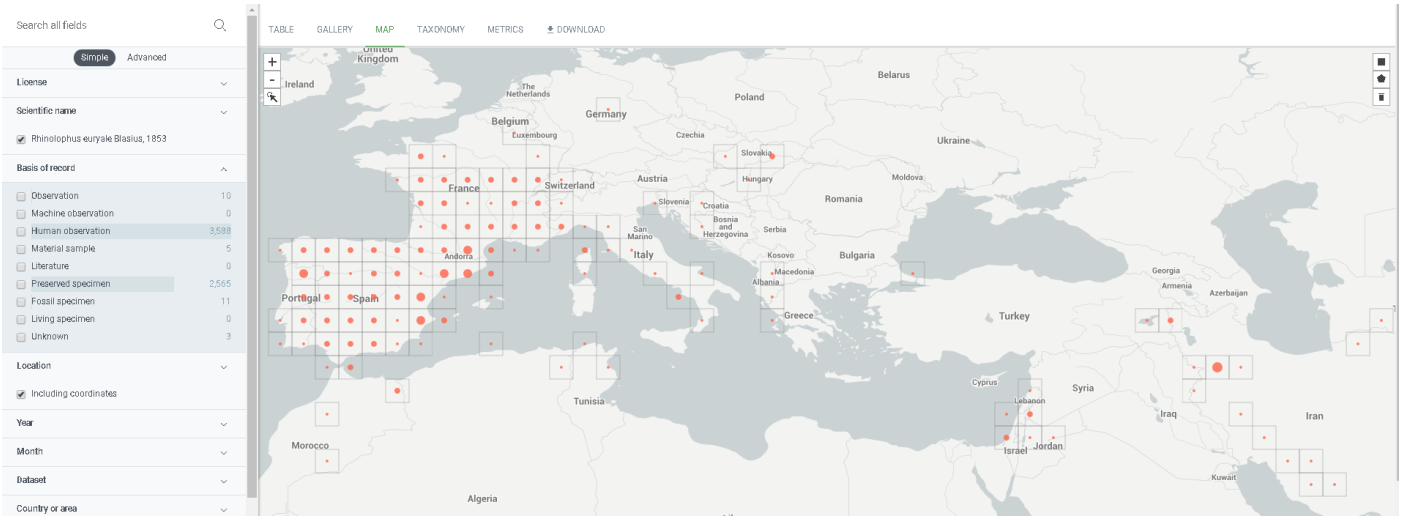
    a. Keep note of these, just in case
        i. Accept the suggestions given.

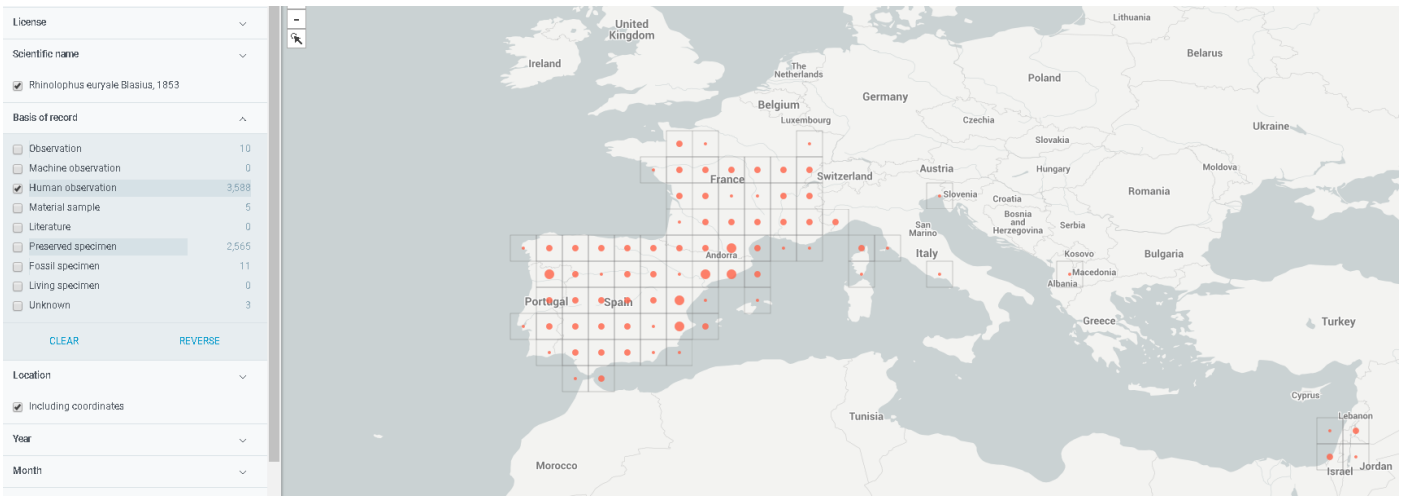
b. Choose “Occurrences” and then “map” on the top to see a summary of the distribution

i. Select occurrences which include coordinates and are based on human observations

1. Should we use human observations only?

a. I used both for this example.

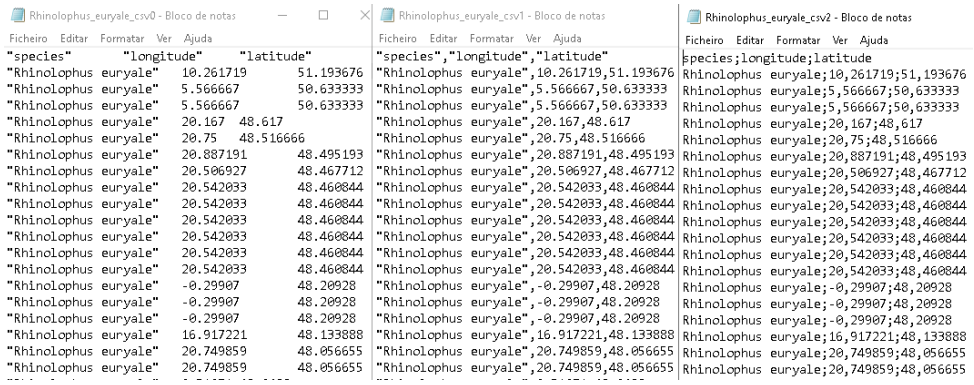




4. Press download on the upper tab and select the simple tab delimited CSV.

a. Once you download and unzip the CSV (Comma separated variable) you might run into formatting problems:

i. North American system: decimals are points and commas separate variables
ii. European system: decimals are commas and semicolon separate variables. E.g.
iii. Tab separated system: tabular spaces separate variables while decimals are points (or commas sometimes).



Rhinolophus_euryale_csv0 - Bloco de notas

```
"species"       "longitude"     "latitude"
"Rhinolophus euryale"   10.261719       51.193676
"Rhinolophus euryale"   5.566667        50.633333
"Rhinolophus euryale"   5.566667        50.633333
"Rhinolophus euryale"   20.167  48.617
"Rhinolophus euryale"   20.75   48.516666
"Rhinolophus euryale"   20.887191       48.495193
"Rhinolophus euryale"   20.506927       48.467712
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   20.542033       48.460844
"Rhinolophus euryale"   -0.29907        48.20928
"Rhinolophus euryale"   -0.29907        48.20928
"Rhinolophus euryale"   -0.29907        48.20928
"Rhinolophus euryale"   16.917221       48.133888
"Rhinolophus euryale"   20.749859       48.056655
"Rhinolophus euryale"   20.749859       48.056655
```

Rhinolophus_euryale_csv1 - Bloco de notas

```
"species","longitude","latitude"
"Rhinolophus euryale",10.261719,51.193676
"Rhinolophus euryale",5.566667,50.633333
"Rhinolophus euryale",5.566667,50.633333
"Rhinolophus euryale",20.167,48.617
"Rhinolophus euryale",20.75,48.516666
"Rhinolophus euryale",20.887191,48.495193
"Rhinolophus euryale",20.506927,48.467712
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",20.542033,48.460844
"Rhinolophus euryale",-0.29907,48.20928
"Rhinolophus euryale",-0.29907,48.20928
"Rhinolophus euryale",-0.29907,48.20928
"Rhinolophus euryale",16.917221,48.133888
"Rhinolophus euryale",20.749859,48.056655
"Rhinolophus euryale",20.749859,48.056655
```

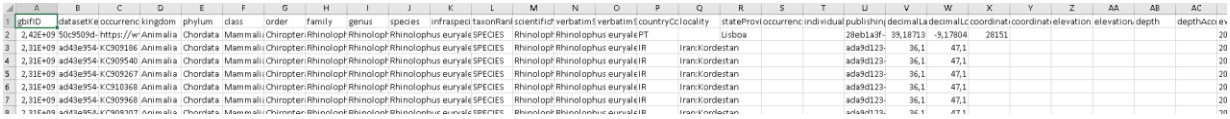
Rhinolophus_euryale_csv2 - Bloco de notas

```
species;longitude;latitude
Rhinolophus euryale;10,261719;51,193676
Rhinolophus euryale;5,566667;50,633333
Rhinolophus euryale;5,566667;50,633333
Rhinolophus euryale;20,167;48,617
Rhinolophus euryale;20,75;48,516666
Rhinolophus euryale;20,887191;48,495193
Rhinolophus euryale;20,506927;48,467712
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;20,542033;48,460844
Rhinolophus euryale;-0,29907;48,20928
Rhinolophus euryale;-0,29907;48,20928
Rhinolophus euryale;-0,29907;48,20928
Rhinolophus euryale;16,917221;48,133888
Rhinolophus euryale;20,749859;48,056655
Rhinolophus euryale;20,749859;48,056655
```

b. All formats can be useful, so keep examples of all of them. It's possible that each software reads them differently

5. Importing in CSV Tab delimited format:
   a. Microsoft Excel
      i. Open a black workbook
      ii. Go to data tab
      iii. Click button from "text" in the *General External data section*
      iv. Select your CSV file
      v. Follow the Text import wizard (in step 2, you can select the delimiter you want)
   b. Its possible to correct these in both notepad or in R but that is not shown here. You can opt to find it yourself.

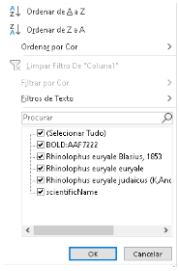
6. The file you've downloaded has a lot of extra fields that provide more info than what you need for modelling.



   a. This is because GBIF's function is not only provide occurrence data but collect biodiversity data into a single standard:
      i. https://www.gbif.org/darwin-core
   b. Maxent expects a .csv file, in NA style, with the following fields: species, longitude, latitude
      i. Thus, you must create a new CSV from the original GBIF file using excel with these 3 fields, in the given order and with the name as given.

	A	B	C
1	species	longitude	latitude
2	Rhinoloph	10,26172	51,19368
3	Rhinoloph	5,566667	50,63333
4	Rhinoloph	5,566667	50,63333
5	Rhinoloph	20,167	48,617
6	Rhinoloph	20,75	48,51667
7	Rhinoloph	20,88719	48,49519
8	Rhinoloph	20,50693	48,46771
9	Rhinoloph	20,54203	48,46084
10	Rhinoloph	20,54203	48,46084
11	Rhinoloph	20,54203	48,46084
12	Rhinoloph	20,54203	48,46084
13	Rhinoloph	20,54203	48,46084
14	Rhinoloph	20,54203	48,46084

      ii. Notice also that it's likely you have multiple species names in the species name column:

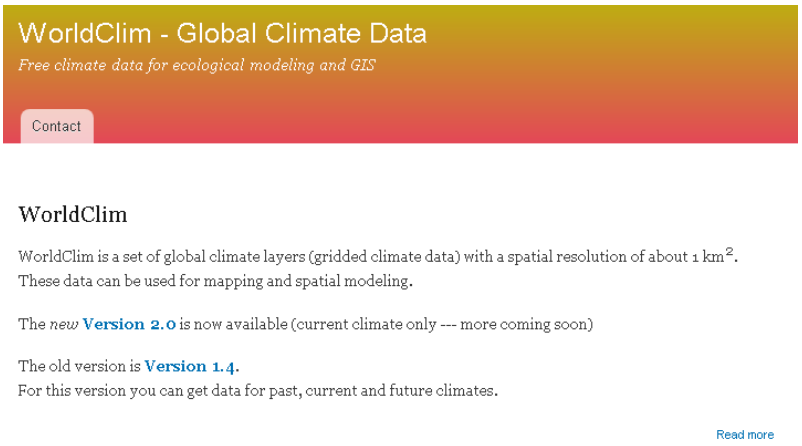


      iii. The simplest solution is to just change all the values of the column species to a single species name – otherwise maxent will consider them different species

7. In this step, you should explore your occurrences in ArcGIS to identify any sampling biases, non-sensical occurrences or overlapping data.
    a. Important also consider if the occurrence data you have is representative of the species niche: e.g. are they representative of the known niche? Are they in locations where they are considered invasive? Etc. These are important questions to explore as an ecologist.
    b. When you are finished, export the final data into a .csv file in NA style (GIS manual explains these steps) to your "Occurrences" folder.

Step 2: Preparing your environmental data:

1. Download climate variables: https://www.worldclim.org/



2. Select the version 1.4.

WorldClim Version2

WorldClim version 2 has average monthly climate data for minimum, mean, and maximum temperature and for precipitation for 1970-2000.

You can download the variables for different spatial resolutions, from 30 seconds (~1 km^2) to 10 minutes (~340 km^2). Each download is a "zip" file containing 12 GeoTiff (.tif) files, one for each month of the year (January is 1; December is 12).

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
solar radiation (kJ m^{-2} day^{-1})	srad 10m	srad 5m	srad 2.5m	srad 30s
wind speed (m s^{-1})	wind 10m	wind 5m	wind 2.5m	wind 30s
water vapor pressure (kPa)	vapr 10m	vapr 5m	vapr 2.5m	vapr 30s

Below you can download the standard (19) WorldClim Bioclimatic variables for WorldClim version 2. They are the average for the years 1970-2000. Each download is a "zip" file containing 19 GeoTiff (.tif) files, one for each month of the variables.

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
Bioclimatic variables	bio 10m	bio 5m	bio 2.5m	bio 30s

a. There are various variables available, the bioclimatic variables represent a "standard" obtained from each of the above. Use these ones for your modelling exercise.
    i. 10m, 5m, 2.5m, 30s refers to the spatial resolution at the equator in the WGS84 projection
        1. 30s is 1 Km data
        2. 5m is 10 Km resolution (download this one)

3. What are the bioclimatic variables?

Bioclimatic variables

Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in species distribution modeling and related ecological modeling techniques. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

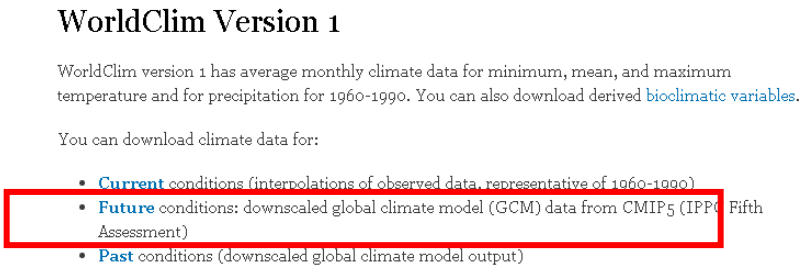
They are coded as follows:

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of 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 (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

This scheme follows that of ANUCLIM, except that for temperature seasonality the standard deviation was used because a coefficient of variation does not make sense with temperatures between -1 and 1).

To create these values yourself, you can use the 'biovars' function in the R package dismo

4. Download the future scenario data
    a. The future scenario is only available on section 1.4



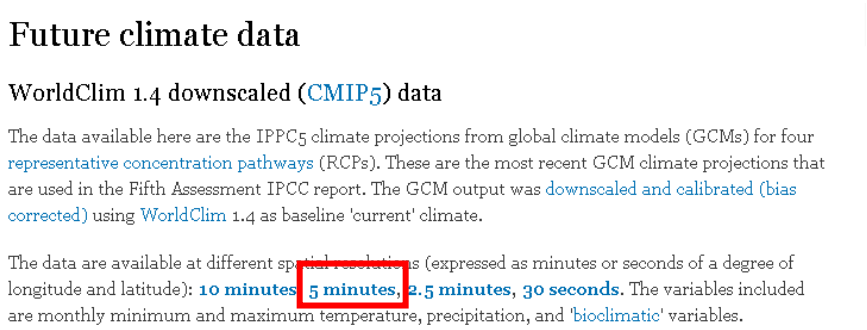
WorldClim Version 1

WorldClim version 1 has average monthly climate data for minimum, mean, and maximum temperature and for precipitation for 1960-1990. You can also download derived bioclimatic variables.

You can download climate data for:

- **Current** conditions (interpolations of observed data, representative of 1960-1990)
- **Future** conditions: downscaled global climate model (GCM) data from CMIP5 (IPCC Fifth Assessment)
- **Past** conditions (downscaled global climate model output)

b. Select the same resolution as before (5 minutes in the case of this tutorial)



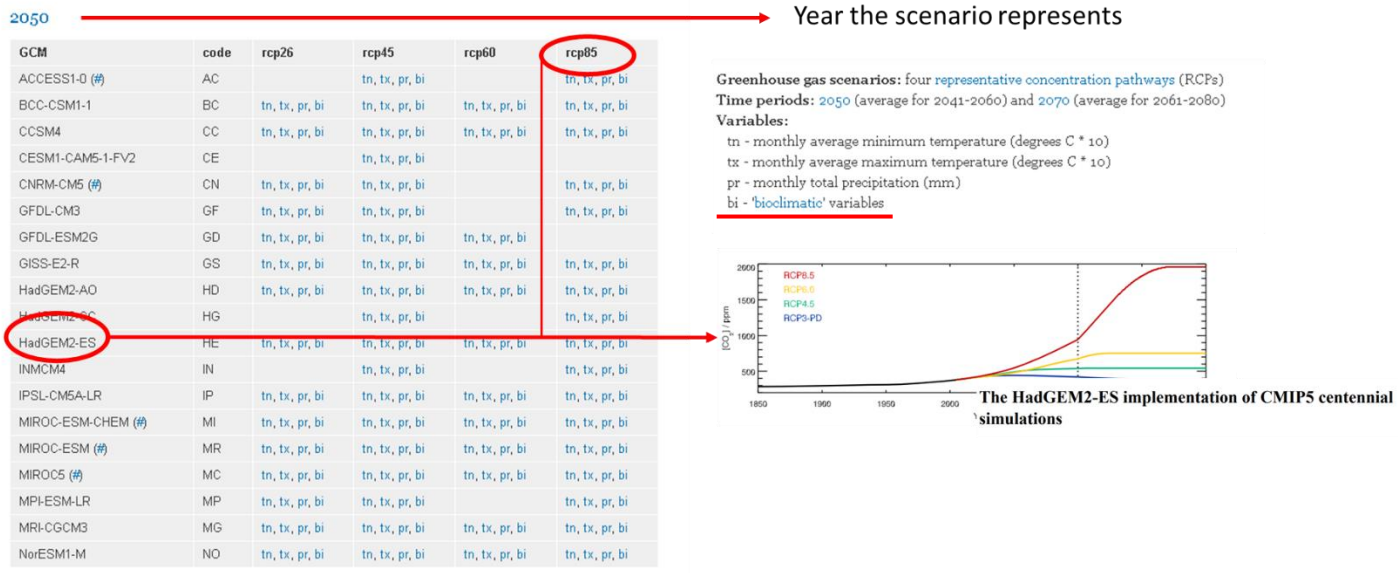
Future climate data

WorldClim 1.4 downscaled (CMIP5) data

The data available here are the IPPC5 climate projections from global climate models (GCMs) for four representative concentration pathways (RCPs). These are the most recent GCM climate projections that are used in the Fifth Assessment IPCC report. The GCM output was downscaled and calibrated (bias corrected) using WorldClim 1.4 as baseline 'current' climate.

The data are available at different spatial resolutions (expressed as minutes or seconds of a degree of longitude and latitude): **10 minutes**, **5 minutes**, **2.5 minutes**, **30 seconds**. The variables included are monthly minimum and maximum temperature, precipitation, and 'bioclimatic' variables.

5. Future climate scenarios:
   a. Each scenario will have different assumptions
   b. For this tutorial we downloaded the HadGEM2-ES (code HE)
      i. More info @ https://portal.enes.org/models/earthsystem-models/metoffice-hadley-centre/hadgem2-es



2050 → Year the scenario represents

GCM	code	rcp26	rcp45	rcp60	rcp85
ACCESS1-0 (#)	AC		tn, tx, pr, bi		tn, tx, pr, bi
BCC-CSM1-1	BC	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
CCSM4	CC	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
CESM1-CAM5-1-FV2	CE		tn, tx, pr, bi		
CNRM-CM5 (#)	CN	tn, tx, pr, bi	tn, tx, pr, bi		tn, tx, pr, bi
GFDL-CM3	GF	tn, tx, pr, bi	tn, tx, pr, bi		tn, tx, pr, bi
GFDL-ESM2G	GD	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	
GISS-E2-R	GS	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
HadGEM2-AO	HD	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
HadGEM2-CC	HG		tn, tx, pr, bi		tn, tx, pr, bi
HadGEM2-ES	HE	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
INMCM4	IN		tn, tx, pr, bi		tn, tx, pr, bi
IPSL-CM5A-LR	IP	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
MIROC-ESM-CHEM (#)	MI	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
MIROC-ESM (#)	MR	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
MIROC5 (#)	MC	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
MPI-ESM-LR	MP	tn, tx, pr, bi	tn, tx, pr, bi		tn, tx, pr, bi
MRI-CGCM3	MG	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
NorESM1-M	NO	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi

Greenhouse gas scenarios: four representative concentration pathways (RCPs)
Time periods: 2050 (average for 2041-2060) and 2070 (average for 2061-2080)
Variables:
tn - monthly average minimum temperature (degrees C * 10)
tx - monthly average maximum temperature (degrees C * 10)
pr - monthly total precipitation (mm)
bi - 'bioclimatic' variables

The HadGEM2-ES implementation of CMIP5 centennial simulations

c. To download the bioclimatic variables, select “bi” on each of the scenarios you want to use. We downloaded all scenarios and prepared all for maxent but only used: rcp45 for this example.
    i. Download them to your downloads folder

6. Unzip all the data (Present and future scenarios) to appropriate folders created in the beginning:
    a. Present
    b. Future
        i. One subfolder per each scenario
    c. Note: if you decide to explore this data in ArcGIS, the software will create some auxiliary files which might make your R scripts stop working. Just erase these auxiliary files or create another folder for exploring the environmental data to avoid this.

7. Cropping variables
    a. Notice:
        i. The code in this section refers to the script: 01_CroppingEnvVariables.R and 02_VariableSelection.R
        ii. The paths for the folders are as given earlier (and set up by the script 00_SettingUpWorkEnvironment.R). If you opted for a different file path, then, you must adapt the code accordingly.
        iii. Highly recommended that you use the latest version of R, R-studio and RTools to avoid any problem loading packages

    b. Open R-studio and open the script: “01_CroppingEnvVariables.R”
    c. Install the packages that used in the script (if you haven’t already)
        i. And then load them:

```
1
2  gc()
3  #Packages
4  library(sp)
5  library(rgdal)
6  library(raster)
7  library(biomod2)
```

ii. You will likely be prompted to install RTools also: https://cran.r-project.org/bin/windows/Rtools/

d. First step is to load the environmental variables:

i. The next code snippet:

1. Sets ups your R working directory to C:/Practical and loads all the data downloaded from wordclim for the present and future scenarios
2. Beware that the bio files are loaded in the wrong order so those numbers: c(1,12:19,2:11) correct for that.

```
14 #checks where your R is currently working
15 getwd()
16 #sets a new working directory
17 setwd("C:/Practical/")
18
19 #lists all file names and paths to the files, saves it to a list
20 rst.names <- list.files("./Present/",pattern=".bil")[c(1,12:19,2:11)]
21 rst.fld <-  list.files("./Present/",pattern=".bil",full.names = T)[c(1,12:19,2:11)]
22 #explore your list to see if you only have rasters on it
23 rst.fld
24 #loads all rasters into a stack object (a multi-dimensional raster)
25 stk.present <- stack(rst.fld)
26
27 #loading future variables is a bit more complicated due to the subfolders
28 path.fut.26 <- list.files("./Future/he26bi50/",pattern=".tif",full.names = T)
29 path.fut.45 <- list.files("./Future/he45bi50/",pattern=".tif",full.names = T)
30 path.fut.60 <- list.files("./Future/he60bi50/",pattern=".tif",full.names = T)
31 path.fut.85 <- list.files("./Future/he85bi50/",pattern=".tif",full.names = T)
32
33 stk.fut.26 <- stack(path.fut.26)
34 stk.fut.45 <- stack(path.fut.45)
35 stk.fut.60 <- stack(path.fut.60)
36 stk.fut.85 <- stack(path.fut.85)
```

ii. First, we need to rename all the files to have more meaningful names. It's also important to ensure that both the future scenario data and the present data has the same name (otherwise maxent won't be able to recognize what is what)

```
#checks the order of the layers loaded and renames them to bioXX, each XX represents
#a bioclimatic layer
names(stk.present)
names(stk.present) <- c("Bio01","Bio02","Bio03","Bio04",
                        "Bio05","Bio06","Bio07","Bio08",
                        "Bio09","Bio10","Bio11","Bio12",
                        "Bio13","Bio14","Bio15","Bio16",
                        "Bio17","Bio18","Bio19")

```

iii. Now we need to do the same to the future scenario data (which is a bit more complicated in this case because the filenames are in not proper order)

```
#this shows us the order is "messed up", so we need to fix it.
names(stk.fut.26)
names(stk.fut.45)
names(stk.fut.60)
names(stk.fut.85)

#the easiest way to  is just to re-load the variables again with the proper order
path.fut.26 <- list.files("./Future/he26bi50/",pattern=".tif",full.names = T)[c(1,12:19,2:11)]
path.fut.45 <- list.files("./Future/he45bi50/",pattern=".tif",full.names = T)[c(1,12:19,2:11)]
path.fut.60 <- list.files("./Future/he60bi50/",pattern=".tif",full.names = T)[c(1,12:19,2:11)]
path.fut.85 <- list.files("./Future/he85bi50/",pattern=".tif",full.names = T)[c(1,12:19,2:11)]

stk.fut.26 <- stack(path.fut.26)
stk.fut.45 <- stack(path.fut.45)
stk.fut.60 <- stack(path.fut.60)
stk.fut.85 <- stack(path.fut.85)

#check if they loaded fine
names(stk.fut.26)
names(stk.fut.45)
names(stk.fut.60)
names(stk.fut.85)
```

iv. Once we have loaded them in the proper order [1,2, 3, …, 19], we can just rename them

```
#and then rename them easily
list.of.names <- c("Bio01","Bio02","Bio03","Bio04",
                   "Bio05","Bio06","Bio07","Bio08",
                   "Bio09","Bio10","Bio11","Bio12",
                   "Bio13","Bio14","Bio15","Bio16",
                   "Bio17","Bio18","Bio19")

names(stk.fut.26) <- list.of.names
names(stk.fut.45) <- list.of.names
names(stk.fut.60) <- list.of.names
names(stk.fut.85) <- list.of.names
```

e. In the next part we will crop the Area of Interest in both the present and future scenario layers.

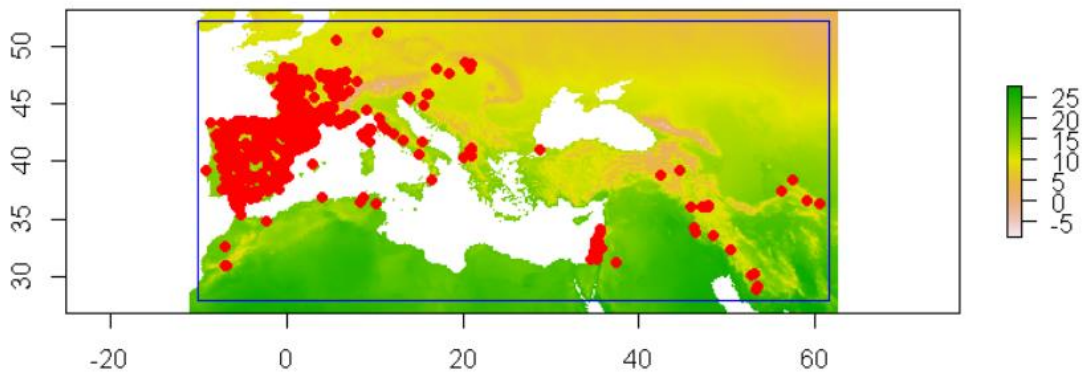
i. The next snippet loads the occurrence data and plots them on top of bioclimate 01 – mean annual temperature

1. First, we load the csv (notice the path should be adapted to your case)
    a. Read.csv opens NA style tables, while read.csv2 opens EU style tables
2. Then we create a point “shapefile” with the coordinates command
3. And we set it to the WGS84 coordinate system

```
# Load species occurrence file
# notice im using read.csv2, which expects a EU type of table. If you want to use the NA style,
#then you must switch read.csv2 with read.csv
#You can also use custom delimitrs
sp <- read.csv2("./Occurrences/Rhinolophus_euryale_csv2.csv",header=T) #load csv of occurrence
head(sp) #check table looks correct
sp_shp <- sp #rename table
coordinates(sp_shp) <- ~longitude+latitude #convert table to points shapefile
proj4string(sp_shp) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
```

ii. On the second part we create an extent object from the point shapefile and plot it against the bioclimatic variable 1 (Mean annual temperature)

```
#Create bounding box around points
bbox <- extent(sp_shp) #create bounding box of points
bbox <- bbox+2 #increase border so we do not truncate data
plot(stk.present$Bio01,ext=bbox+2)
plot(bbox, col='blue',add=T) #check if box surrounds points
plot(sp_shp,add=T,pch=19,col='red') #add points
```



f. In the next section we will crop the present data and then plot an example. On the second part of the script we crop the future scenario data

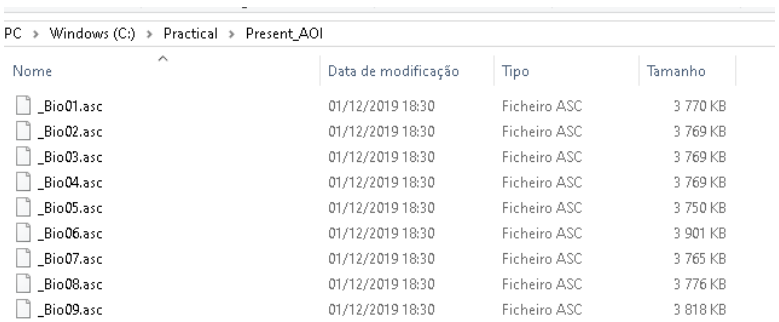
```
#cropping the present data
stk.present.AOI.crop <- crop(stk.present,bbox) #clip to training area
#plotting the example
par(mfrow=c(1,2)) #sets the plotting area to a 1 line 2 columns set up
plot(stk.present$Bio01,main="Original extent")
plot(stk.present.AOI.crop $Bio01,main="Cropped extent")
par(mfrow=c(1,1)) #sets it back to 1 image per plot
#cropping the future data
stk.fut.26.AOI.crop <- crop(stk.fut.26,bbox)
stk.fut.45.AOI.crop <- crop(stk.fut.45,bbox)
stk.fut.60.AOI.crop <- crop(stk.fut.60,bbox)
stk.fut.85.AOI.crop <- crop(stk.fut.85,bbox)
```

g. In the next section, we will save all these outputs to the folders designated previously.

   i. First the example for the present data:

```
#now we can save them to another folder in a format
#that maxent can read
#saving the cropped present data in .asc format
writeRaster(stk.present.AOI.crop,
            "./Present_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")
```

1. Uses writeRaster function to save to folder: “./Present_AOI/”
2. With file format .asc (ASCII)
3. Overwrites any pre-existing rasters
4. Bylayer=T means each layer is saved independently
5. Suffix= ”names” tells R to use the layer names as a filename
   a. It will add an underscore, but that isn’t a problem (e.g. _Bio01.asc)
   b. Should look like the following example:



ii. Now we do the same for the Present data (uncropped, world data) and the future scenarios (both uncropped and cropped versions)
    1. This will take quite some time if you have multiple scenarios like in this example
    2. Make sure all naming are proper before running

```
#now we can save them to another folder in a format
#that maxent can read
#saving the cropped present data in .asc format
writeRaster(stk.present.AOI.crop,
            "./Present_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

#saving present data uncropped - notice, these files might be extra large
writeRaster(stk.present,
            "./Present_WLD/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

#saving scenario he26bi50 uncropped and cropped
writeRaster(stk.fut.26,
            "./Future_WLD/he26bi50_WLD/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

writeRaster(stk.fut.26.AOI.crop,
            "./Future_AOI/he26bi50_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

#saving scenario he45bi50 uncropped and cropped
writeRaster(stk.fut.45,
            "./Future_WLD/he45bi50_WLD/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

writeRaster(stk.fut.45.AOI.crop,
            "./Future_AOI/he45bi50_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

#saving scenario he50bi50 uncropped and cropped
writeRaster(stk.fut.60,
            "./Future_WLD/he60bi50_WLD/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

writeRaster(stk.fut.60.AOI.crop,
            "./Future_AOI/he60bi50_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

#saving scenario he85bi50 uncropped and cropped
writeRaster(stk.fut.85,
            "./Future_WLD/he85bi50_WLD/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")

writeRaster(stk.fut.85.AOI.crop,
            "./Future_AOI/he85bi50_AOI/.asc",
            overwrite=T,
            bylayer=T,
            suffix="names")
```

iii. Once the previous steps are done, we have prepared all data to be used in maxent. The benefit of doing it this way is that any small change doesn't imply coming back to R to fix it. All the data is ready to use, on call.

8. Selecting environmental variables:

a. First criteria: Ecology
    i. Which variables are more ecologically significant for the species? <u>Think about it before using a statistical justification</u>!
    ii. In my case I selected:
        1. Bio 01 – Mean annual temperature

2. Bio 04 – Temperature seasonality
3. Bio 07 – Temperature Annual range
4. Bio 12 – Annual precipitation
5. Bio 15 – Precipitation seasonality
6. Bio 19 – Precipitation of the coldest quarter

Notice:

Statistical interpretation of autocorrelation

Prior to modelling, there are two core problems that can occur: Environmental autocorrelation and spatial autocorrelation (of model or data). The second we will not explore in this exercise but the first we will address it with two tests:

- Pearson's pairwise correlation: r
  - Measures the statistical association or dependence between two variables

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\,n\Sigma x^2 - (\Sigma x)^2\,]\,[\,n\Sigma y^2 - (\Sigma y)^2\,]}}$$

- Variance inflaction factor: VIF
  - Measures the statistical dependence of one variable to a combination of N other variables

$$VIF_i = \frac{1}{1 - R_i^2}$$

- As a rule of thumb, models should have less than 0.7 pearson correlation and less than 10 VIF for any given variable
  - You can accept r > .7 if and only if VIF < 10
- The autocorrelation testing is important only for the areas where the model is trained.

9. Testing for autocorrelation:
    a. From here onwards, it refers to script: 02_VariableSelection.R
        i. Pairwise autocorrelation test:
            1. This is the R2 between two sets of data
            2. If R2 is > 0.7, then, you should consider excluding one of the variables
            3. The following code generates a table you can explore in Excel
                a. If needed, use write.csv instead of write.csv2

```
### pairwise testing
#first we convert the cropped raster to a data.frame
stk.present.AOI.crop <- na.omit(as.data.frame(stk.present.AOI.crop)) #we also remove NA's
#now this stores the pearson correlation in a matrix
cor.tab <-cor(stk.present.AOI.crop)
#remember to change to write.csv if needed
write.csv2(cor.tab,"CorrelationTable_AOI.csv")
```

4. Open the file: “CorrelationTable_AOI.csv”

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Bio01	Bio02	Bio03	Bio04	Bio05	Bio06	Bio07	Bio08	Bio09	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
2	Bio01	1	0,686055	0,660273	-0,23315	0,875827	0,820743	0,071517	0,025992	0,844474	0,891814	0,913634	-0,56364	-0,44576	-0,63376	0,685389	-0,45453	-0,63032	-0,68202	-0,26329
3	Bio02	0,686055	1	0,659006	0,111116	0,81381	0,293616	0,509326	-0,10757	0,650355	0,730932	0,506533	-0,60923	-0,51346	-0,63433	0,645142	-0,52374	-0,644	-0,64656	-0,37722
4	Bio03	0,660273	0,659006	1	-0,64641	0,409402	0,724167	-0,28811	-0,22129	0,608793	0,356148	0,803072	-0,15062	-0,07727	-0,28211	0,561358	-0,07839	-0,26755	-0,35339	0,075049
5	Bio04	-0,23315	0,111116	-0,64641	1	0,243434	-0,71377	0,909285	0,18432	-0,18764	0,228713	-0,60605	-0,40359	-0,41508	-0,25636	-0,14273	-0,42122	-0,28258	-0,19758	-0,44821
6	Bio05	0,875827	0,81381	0,409402	0,243434	1	0,452432	0,539321	0,068535	0,764088	0,989195	0,614411	-0,75876	-0,64563	-0,76733	0,639536	-0,65728	-0,77509	-0,78853	-0,47539
7	Bio06	0,820743	0,293616	0,724167	-0,71377	0,452432	1	-0,50698	-0,05363	0,682023	0,494753	0,971028	-0,14168	-0,05604	-0,27523	0,507805	-0,05796	-0,25494	-0,34915	0,090673
8	Bio07	0,071517	0,509326	-0,28811	0,909285	0,539321	-0,50698	1	0,116884	0,094506	0,488915	-0,32306	-0,59959	-0,57111	-0,48176	0,13863	-0,58056	-0,50842	-0,43246	-0,5451
9	Bio08	0,025992	-0,10757	-0,22129	0,18432	0,068535	-0,05363	0,116884	1	-0,28637	0,10579	-0,06057	-0,15776	-0,19362	0,020961	-0,22475	-0,19881	0,004904	0,186607	-0,36424
10	Bio09	0,844474	0,650355	0,608793	-0,18764	0,764088	0,682023	0,094506	-0,28637	1	0,76456	0,774035	-0,48841	-0,36614	-0,64124	0,672729	-0,37387	-0,62398	-0,75172	-0,12991
11	Bio10	0,891814	0,730932	0,356148	0,228713	0,989195	0,494753	0,488915	0,10579	0,76456	1	0,635333	-0,74352	-0,63235	-0,75114	0,608282	-0,64363	-0,75762	-0,77618	-0,46092
12	Bio11	0,913634	0,506533	0,803072	-0,60605	0,614411	0,971028	-0,32306	-0,06057	0,774035	0,635333	1	-0,28708	-0,18652	-0,41157	0,611413	-0,19089	-0,39586	-0,47981	-0,0184
13	Bio12	-0,56364	-0,60923	-0,15062	-0,40359	-0,75876	-0,14168	-0,59959	-0,15776	-0,48841	-0,74352	-0,28708	1	0,941355	0,843861	-0,42529	0,95245	0,871175	0,804919	0,844799
14	Bio13	-0,44576	-0,51346	-0,07727	-0,41508	-0,64563	-0,05604	-0,57111	-0,19362	-0,36614	-0,63235	-0,18652	0,941355	1	0,651168	-0,17839	0,996067	0,683382	0,667598	0,896504
15	Bio14	-0,63376	-0,63433	-0,28211	-0,25636	-0,76733	-0,27523	-0,48176	0,020961	-0,64124	-0,75114	-0,41157	0,843861	0,651168	1	-0,67518	0,668518	0,992771	0,909616	0,511984
16	Bio15	0,685389	0,645142	0,561358	-0,14273	0,639536	0,507805	0,13863	-0,22475	0,672729	0,608282	0,611413	-0,42529	-0,17839	-0,67518	1	-0,19707	-0,67729	-0,61078	-0,07466
17	Bio16	-0,45453	-0,52374	-0,07839	-0,42122	-0,65728	-0,05796	-0,58056	-0,19881	-0,37387	-0,64363	-0,19089	0,95245	0,996067	0,668518	-0,19707	1	0,699409	0,680928	0,903838
18	Bio17	-0,63032	-0,644	-0,26755	-0,28258	-0,77509	-0,25494	-0,50842	0,004904	-0,62398	-0,75762	-0,39586	0,871175	0,683382	0,992771	-0,67729	0,699409	1	0,903737	0,550295
19	Bio18	-0,68202	-0,64656	-0,35339	-0,19758	-0,78853	-0,34915	-0,43246	0,186607	-0,75172	-0,77618	-0,47981	0,804919	0,667598	0,909616	-0,61078	0,680928	0,903737	1	0,399641
20	Bio19	-0,26329	-0,37722	0,075049	-0,44821	-0,47539	0,090673	-0,5451	-0,36424	-0,12991	-0,46092	-0,0184	0,844799	0,896504	0,511984	-0,07466	0,903838	0,550295	0,399641	1

5. Highlight the content of the table - > Styles section -> Conditional Formatting - > highlight cells rules -> Larger than 0.7 .

> Also highlight negative correlations: less than -0.7 (you can choose a different color).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Bio01	Bio02	Bio03	Bio04	Bio05	Bio06	Bio07	Bio08	Bio09	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
2	Bio01	1	0,686055	0,660273	-0,23315	0,875827	0,820743	0,071517	0,025992	0,844474	0,891814	0,913634	-0,56364	-0,44576	-0,63376	0,685389	-0,45453	-0,63032	-0,68202	-0,26329
3	Bio02	0,686055	1	0,659006	0,111116	0,81381	0,293616	0,509326	-0,10757	0,650355	0,730932	0,506533	-0,60923	-0,51346	-0,63433	0,645142	-0,52374	-0,644	-0,64656	-0,37722
4	Bio03	0,660273	0,659006	1	-0,64641	0,409402	0,724167	-0,28811	-0,22129	0,608793	0,356148	0,803072	-0,15062	-0,07727	-0,28211	0,561358	-0,07839	-0,26755	-0,35339	0,075049
5	Bio04	-0,23315	0,111116	-0,64641	1	0,243434	-0,71377	0,909285	0,18432	-0,18764	0,228713	-0,60605	-0,40359	-0,41508	-0,25636	-0,14273	-0,42122	-0,28258	-0,19758	-0,44821
6	Bio05	0,875827	0,81381	0,409402	0,243434	1	0,452432	0,539321	0,068535	0,764088	0,989195	0,614411	-0,75876	-0,64563	-0,76733	0,639536	-0,65728	-0,77509	-0,78853	-0,47539
7	Bio06	0,820743	0,293616	0,724167	-0,71377	0,452432	1	-0,50698	-0,05363	0,682023	0,494753	0,971028	-0,14168	-0,05604	-0,27523	0,507805	-0,05796	-0,25494	-0,34915	0,090673
8	Bio07	0,071517	0,509326	-0,28811	0,909285	0,539321	-0,50698	1	0,116884	0,094506	0,488915	-0,32306	-0,59959	-0,57111	-0,48176	0,13863	-0,58056	-0,50842	-0,43246	-0,5451
9	Bio08	0,025992	-0,10757	-0,22129	0,18432	0,068535	-0,05363	0,116884	1	-0,28637	0,10579	-0,06057	-0,15776	-0,19362	0,020961	-0,22475	-0,19881	0,004904	0,186607	-0,36424
10	Bio09	0,844474	0,650355	0,608793	-0,18764	0,764088	0,682023	0,094506	-0,28637	1	0,76456	0,774035	-0,48841	-0,36614	-0,64124	0,672729	-0,37387	-0,62398	-0,75172	-0,12991
11	Bio10	0,891814	0,730932	0,356148	0,228713	0,989195	0,494753	0,488915	0,10579	0,76456	1	0,635333	-0,74352	-0,63235	-0,75114	0,608282	-0,64363	-0,75762	-0,77618	-0,46092
12	Bio11	0,913634	0,506533	0,803072	-0,60605	0,614411	0,971028	-0,32306	-0,06057	0,774035	0,635333	1	-0,28708	-0,18652	-0,41157	0,611413	-0,19089	-0,39586	-0,47981	-0,0184
13	Bio12	-0,56364	-0,60923	-0,15062	-0,40359	-0,75876	-0,14168	-0,59959	-0,15776	-0,48841	-0,74352	-0,28708	1	0,941355	0,843861	-0,42529	0,95245	0,871175	0,804919	0,844799
14	Bio13	-0,44576	-0,51346	-0,07727	-0,41508	-0,64563	-0,05604	-0,57111	-0,19362	-0,36614	-0,63235	-0,18652	0,941355	1	0,651168	-0,17839	0,996067	0,683382	0,667598	0,896504
15	Bio14	-0,63376	-0,63433	-0,28211	-0,25636	-0,76733	-0,27523	-0,48176	0,020961	-0,64124	-0,75114	-0,41157	0,843861	0,651168	1	-0,67518	0,668518	0,992771	0,909616	0,511984
16	Bio15	0,685389	0,645142	0,561358	-0,14273	0,639536	0,507805	0,13863	-0,22475	0,672729	0,608282	0,611413	-0,42529	-0,17839	-0,67518	1	-0,19707	-0,67729	-0,61078	-0,07466
17	Bio16	-0,45453	-0,52374	-0,07839	-0,42122	-0,65728	-0,05796	-0,58056	-0,19881	-0,37387	-0,64363	-0,19089	0,95245	0,996067	0,668518	-0,19707	1	0,699409	0,680928	0,903838
18	Bio17	-0,63032	-0,644	-0,26755	-0,28258	-0,77509	-0,25494	-0,50842	0,004904	-0,62398	-0,75762	-0,39586	0,871175	0,683382	0,992771	-0,67729	0,699409	1	0,903737	0,550295
19	Bio18	-0,68202	-0,64656	-0,35339	-0,19758	-0,78853	-0,34915	-0,43246	0,186607	-0,75172	-0,77618	-0,47981	0,804919	0,667598	0,909616	-0,61078	0,680928	0,903737	1	0,399641
20	Bio19	-0,26329	-0,37722	0,075049	-0,44821	-0,47539	0,090673	-0,5451	-0,36424	-0,12991	-0,46092	-0,0184	0,844799	0,896504	0,511984	-0,07466	0,903838	0,550295	0,399641	1

6. Remove all the rows and columns of variables you are not interested in.
    a. Remember, I choose Bio 1, 4,7,12,15 and 19 beforehand
    b. Save to a new file (e.g. same name_sel.csv)

	A	B	C	D	E	F	G
1		Bio01	Bio04	Bio07	Bio12	Bio15	Bio19
2	Bio01	1	-0,23315	0,071517	-0,56364	0,685389	-0,26329
3	Bio04	-0,23315	1	0,909285	-0,40359	-0,14273	-0,44821
4	Bio07	0,071517	0,909285	1	-0,59959	0,13863	-0,5451
5	Bio12	-0,56364	-0,40359	-0,59959	1	-0,42529	0,844799
6	Bio15	0,685389	-0,14273	0,13863	-0,42529	1	-0,07466
7	Bio19	-0,26329	-0,44821	-0,5451	0,844799	-0,07466	1

    c. If 2 or more variables are having high pairwise correlation, its likely you have to remove one of them.
    d. We can confirm it with a VIF test

ii. Testing for the multicollinearity using the VIF test:
    1. It can be done by “hand” but package usdm has implemented it.
    2. Each variable must be tested against the linear combination of all other variables
    3. VIF is then 1/(1-R^2), where R is the correlation of the model
    4. If VIF > 10, then variable X must be removed

a. More conservative: > 5

5. If any variable is removed, then the VIF test has to performed all over
6. We first exclude all variables we do not want to be part of the final model and then calculate the VIF for all variables using the usdm package
    a. Maxobservations is necessary because by default the function vif() uses only 5000 data points.

```
33 #multicollinearity testing
34 library(usdm)
35
36 #e.g. i select Bio01; Bio04; Bio07; Bio 12; Bio 15 and bio 19
37 head(stk.present.AOI.crop)
38 #select only the variables i am interested on
39 df.stk.AOI <- stk.present.AOI.crop[,c(1,4,7,12,15,19)]
40 #confirm the selection
41 head(df.stk.AOI)
42 #and the VIF test
43 vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
```

```
> vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
  Variables       VIF
1     Bio01  3.623841
2     Bio04 13.534628
3     Bio07 13.308972
4     Bio12  9.459776
5     Bio15  2.617598
6     Bio19  5.691570
>
```

7. This implies that either Bio07 or Bio 04 must be removed. I choose Bio07
    a. Although the pairwise Pearson correlation doesn't change when a variable is removed, the VIF does (it is model dependent).
    b. Therefore, you must repeat the test

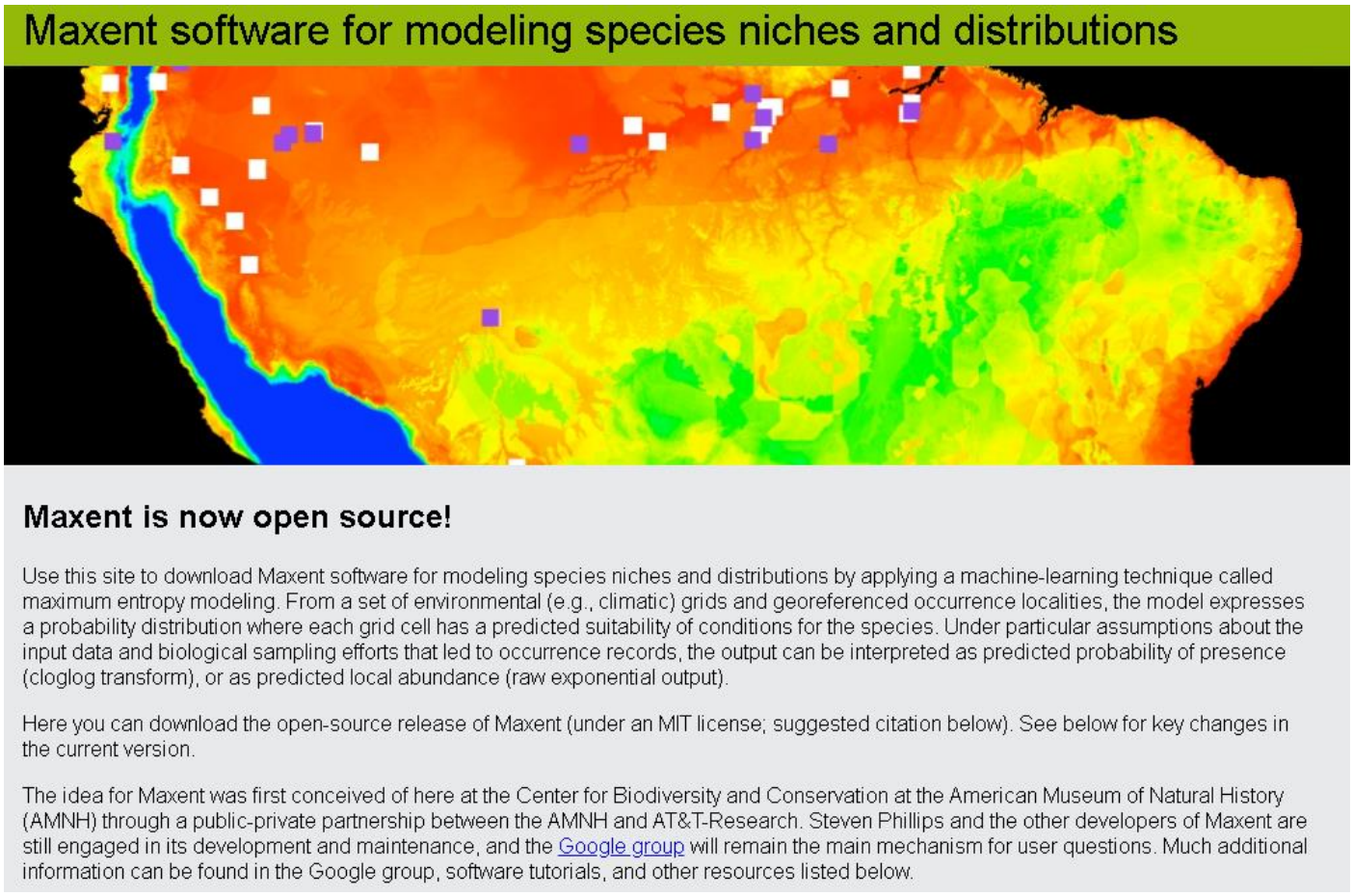
```
> df.stk.AOI <- stk.present.AOI.crop[,c(1,4,12,15,19)] #minus the temperature range
> vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
  Variables      VIF
1     Bio01 3.285936
2     Bio04 1.934874
3     Bio12 9.448515
4     Bio15 2.331450
5     Bio19 5.659382
>
```

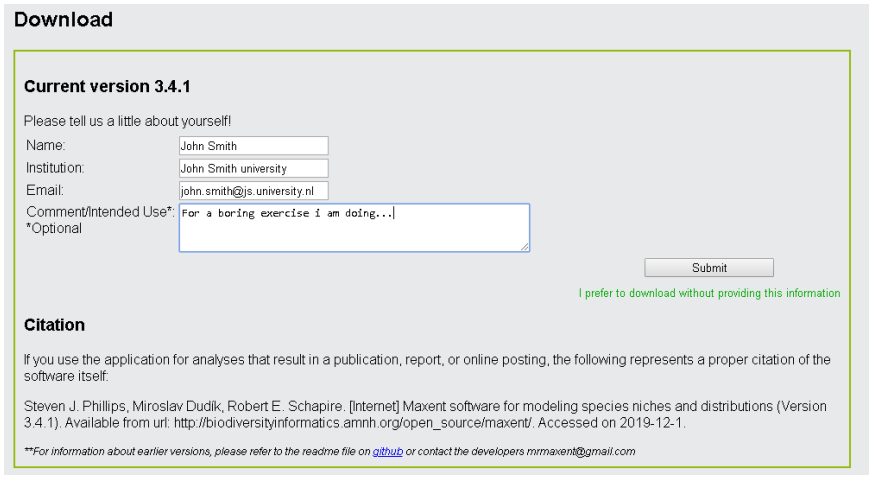
10. We have now everything ready to go for maxent:
    a. Before we start:
        i. Confirm you have a version of your occurrence data on NA csv style (tabular should work also)
        ii. Confirm your environmental data is in each of the folders, in .asc format

---

Downloading MAXENT:

- Go to https://biodiversityinformatics.amnh.org/open_source/maxent/

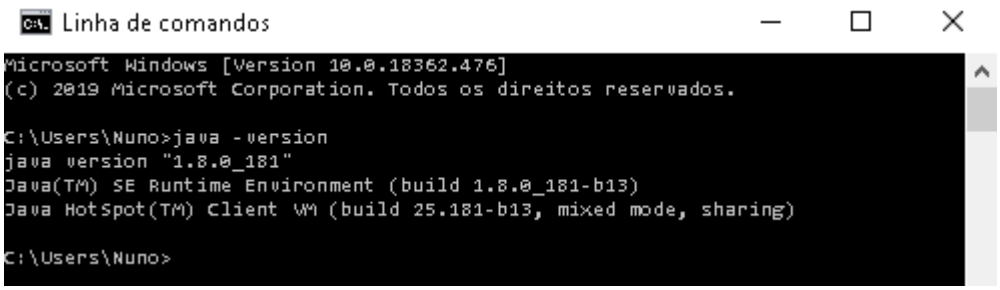




Downloading JAVA:

- Go to: https://www.java.com/en/download/help/download_options.xml
  - Follow the installation instructions
- Once it is installed, to confirm:
  - Open a command line window (use the search option on the bottom left)

- Type java -version



- If the installation was successful, you should see an output like the one above

Step 3: Setting up Maxent

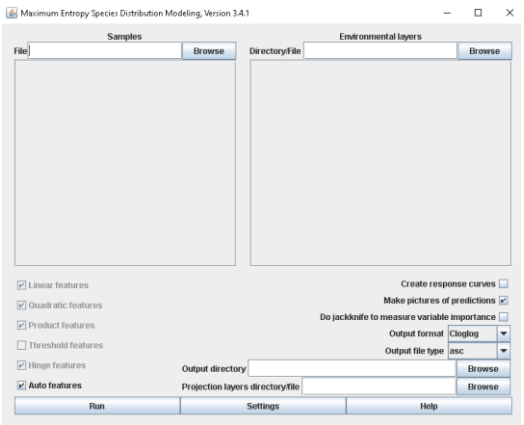
Species Distribution Models (SDMs), also known as Ecological Niche Models (ENMs), predict the presence and absence of species by interpolating identified relationships between collection data, stored in Natural History Museums and Herbaria, and environmental data. In this section we identify the relationships between species presence records and environmental data with a distribution modelling application, and subsequently interpolate the relationships to the research area of interest; here Borneo. We start with downloading the MaxEnt application from http://www.cs.princeton.edu/~schapire/maxent/. You can also download the MaxEnt tutorial from this website. Besides MaxEnt there are many other algorithms and applications such as GARP (http://nhm.ku.edu/desktopgarp/), BioMapper (http://www2.unil.ch/biomapper/), and Generalized Dissimilarity Modelling (http://www.biomaps.net.au/gdm/), amongst others! Most modelling algorithms are also available in R (R Development Core Team 2014). MaxEnt is a Java application, so you need Java to be installed on your computer (http://www.java.com). You open MaxEnt by clicking the Maxent.bat file.

Maxent uses the maximum entropy algorithm which is defined as follows:

> *MaxEnt, or the maximum entropy method for species' distribution modelling, estimates the most uniform distribution (''maximum entropy'') across the study area, given the constraint that the expected value of each environmental predictor variable under this estimated distribution matches its empirical average (average values for the set of species' presence records)* (Phillips et al. 2006

Step by step instructions:

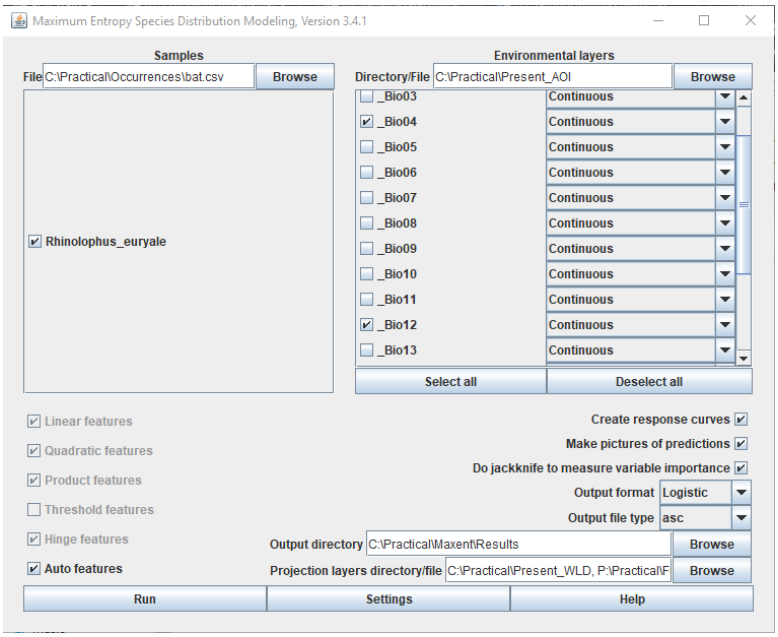
1. Open MaxEnt software
    a. Use the "Executable jar file"



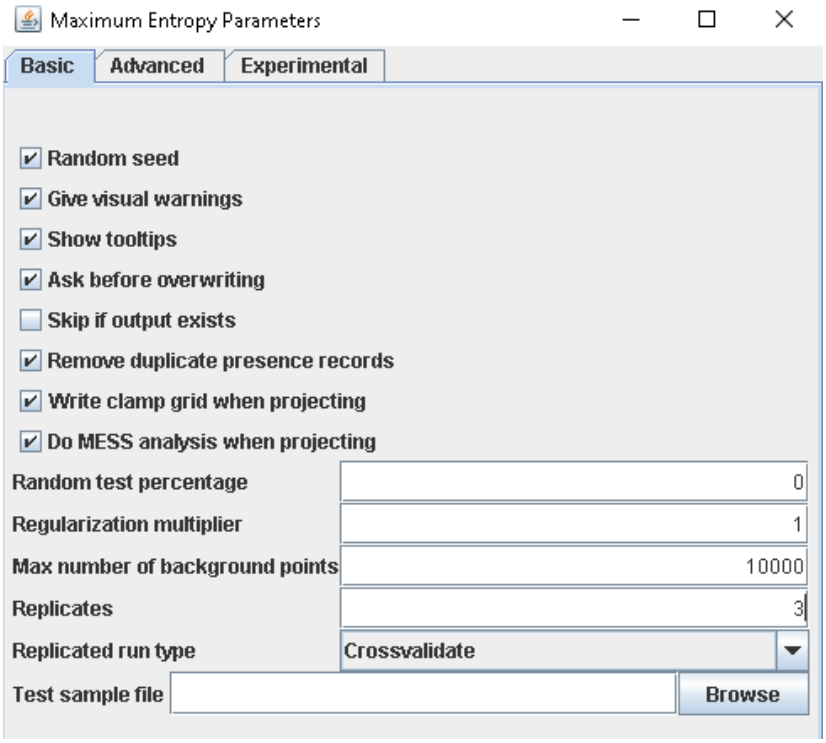
2. Use the browse buttons to load your occurrence and environmental data (Present_AOI):

    a. On the environmental layers side, press deselect all. And select only the environmental variables you already selected on the previous section: 1,4,12,15,19

b. Activate:
    - Create Response Curves
    - Make pictures of predictions
    - Do Jackknife to measure variable importance
c. Output directory:
    - Results folder (subfolder of the maxent) or any other folder you want to use
    - Maxent produces a lot of data so it's recommended that you have an independent folder for each time you run it (to avoid confusion)
d. Projections layers directory/file
    - Here you will add the path to:
        - Present_WLD
        - Future_WLD/ <scenarios>_WLD
        - Future_AOI/<scenarios>_AOI
    - This will make MAXENT produce predictions for all the different possible scenarios.
    - Use comma to separate each scenario. Choose only the scenarios you want. E.g:
        - C:\Practical\Present_WLD, C:\Practical\Future_WLD\he26bi50_WLD, C:\Practical\Future_WLD\he45bi50_WLD, C:\Practical\Future_WLD\he60bi50_WLD, C:\Practical\Future_WLD\he85bi50_WLD, C:\Practical\Future_AOI\he26bi50_AOI, C:\Practical\Future_AOI\he45bi50_AOI, C:\Practical\Future_AOI\he60bi50_AOI, C:\Practical\Future_AOI\he85bi50_AOI
            - My advice is writing all the paths in a notepad or word file and copy pasting it to maxent
e. Keep auto-features ticked
f. By now, maxent should look like this:

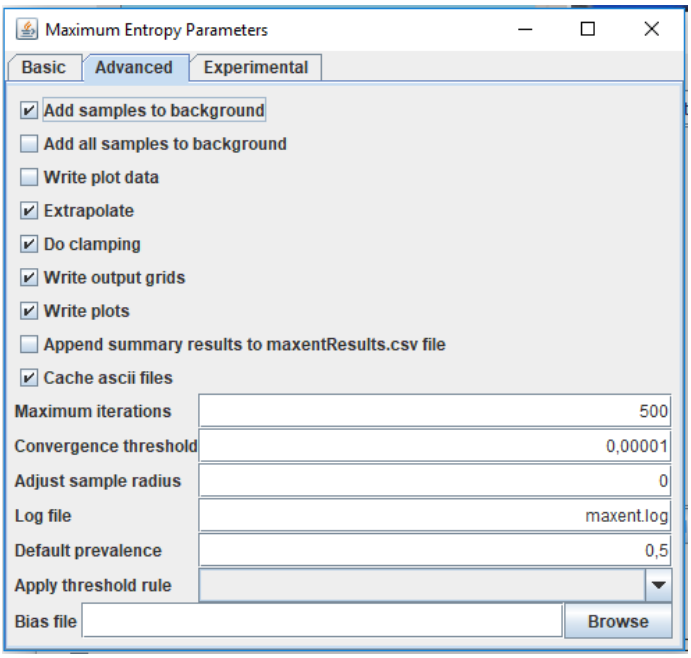


3. Click the button 'Settings' and set your maxent as follows:



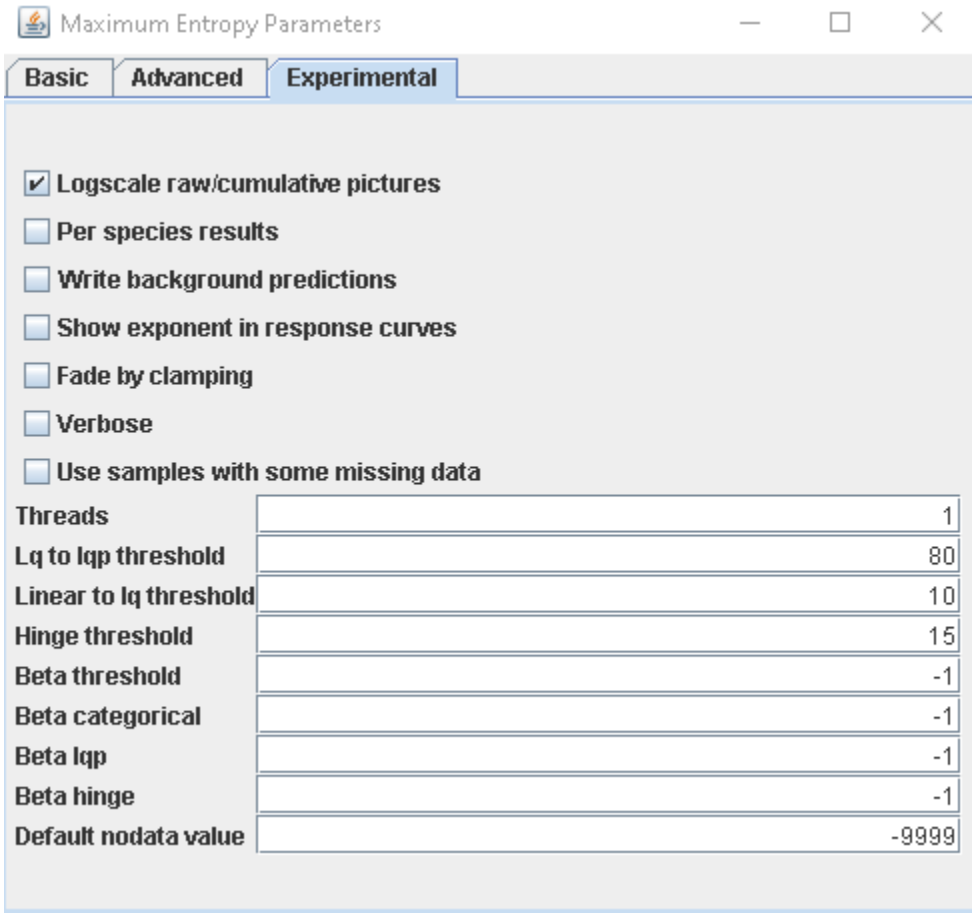
   a. This will perform 3 replicates of maxent to train the model and report the average results. For your report you should use at least 5 replicates.

4. The advanced tab should look like this:



a. Optional:
   - Test the impact of having “extrapolate” on – what do you expect happens?
   - Test the impact of having “Do clamping” off – what do you expect happens?

5. The experimental pane should look like this:

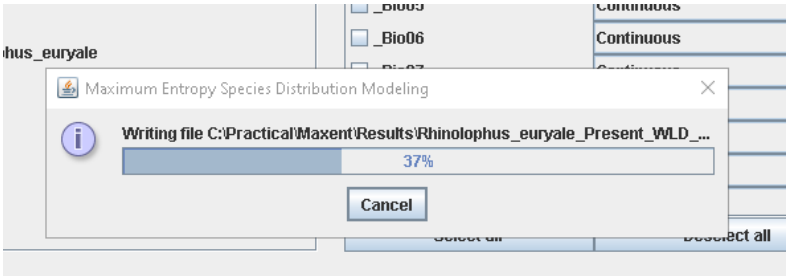


a. Optional:
   - What would be the impact of "Fade by clamping"?

6. Once all is set, go back to the main window and press run:
   a. Its ok to ignore the following warnings:

   b. Its also common to have a warning that occurrences overlap, and thus are being removed. It's ok to ignore those warnings
      - It's a verification to consider only one occurrence per cell.

7. Once you see a bar like this:



a. Then all is fine, and you must wait for the model to finish.

Validating SDMs

The most widely applied method to validate SDMs is the Area Under the Curve (AUC) of the Reciever Operator Curve (ROC) (Fielding and Bell 1997, McPherson et al. 2004, Raes and ter Steege 2007). The advantage of the AUC value over other measures of model accuracy (i.e. Cohen's kapa, sensitivity, specificity) is that it is a) threshold independent, and b) prevalence insensitive.

Setting a threshold means that continues MaxEnt values, running from 0-1, not have to be converted to discrete presence/absence values. There are several techniques to set thresholds (Liu et al. 2005), but this is not required for the AUC value. Prevalence is the proportion of the data representing species' presence, or presences /(presences + absences). The fact that the AUC value is relatively insensitive to prevalence is of special relevance because when absences are lacking, which is often the case, they are replaced by pseudo-absences, or background points. A sufficiently large sample of pseudo-absences is needed to provide a reasonable representation of the environmental variation exhibited by the geographical area of interest, typically 1,000- 10,000 points. These large numbers of pseudo-absences automatically result in low prevalence values. The number of records by which a species is represented in herbaria and natural history museums range from 1 to 150-200 records. Even when a species is represented by 200 unique presence-only records and 1000 pseudo- absences are used, prevalence is only 16.7% (200/1200).

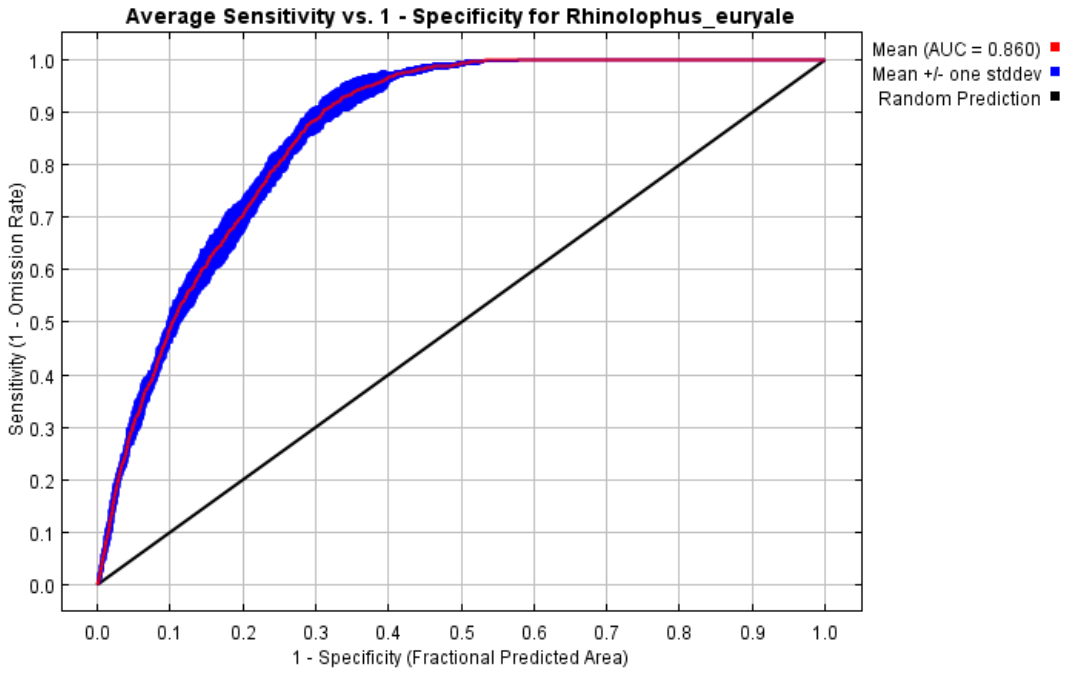
AUC values range from 0 to 1, with a value of 0.5 indicating model accuracy not better than random, and a value of 1.0 indicating perfect model fit (Fielding and Bell 1997). An AUC value can be interpreted as indicating the probability that, when a presence site (site where a species is recorded as present) and an absence site (site where a species is recorded as absent) are drawn at random from the population, the presence site has a higher predicted value than the absence site (Phillips et al. 2006). SDMs with an AUC value of 0.7 are considered to be reliable, values over 0.8 as good.

A major drawback of using pseudo-absences instead of true absences, however, is that the maximum achievable AUC value indicating perfect model fit, is no longer 1, but 1-a/2 (where a is the fraction of the geographical area of interest covered by a species' true distribution, which typically is not known). Nevertheless, random prediction still corresponds to an AUC value of 0.5. Therefore, standard

thresholds of AUC values indicating SDM accuracy (e.g. the threshold of AUC>0.7 that is often used), do NOT apply (Raes and ter Steege 2007). Therefore be very cautious when SDMs based on presence-only data are validated with AUC values. This problem can be solved by testing against a null-model. This procedure is described in detail in Raes and ter Steege 2007, but this goes beyond this practical.

1. Open your maxent/Results folder.
    a. You will find an .html file/s that summarize your results and includes the AUC evaluation. If you have repetitions, there is one html file per repetition, and one that summarizes the results of each replicate, reporting means and standard deviations. The folder also contains ascii files which you can open in arcgis, png files, and other separate files.

2. Check the average AUC value.



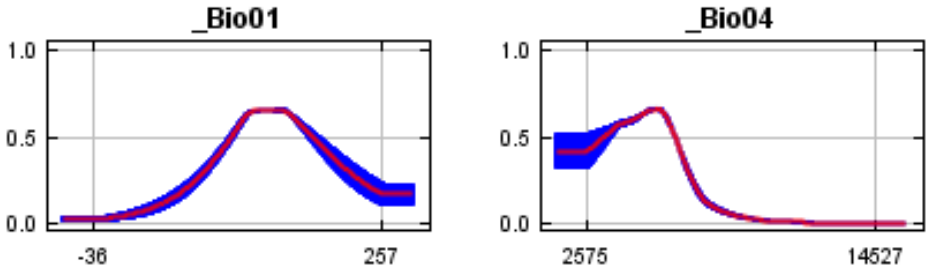
3. Examine your maps: take a look at the present distribution and what happens to suitable habitat with the future scenario/s that you have chosen. Compare the your clipped maps with the world maps. Do they future projections make sense? Why or why not?

4. Go to the threshold table in one of your individual models. This table contains a list of thresholds. Among the most widely used ones are: '10 percentile training presence', 'Equal training sensitivity and specificity', and 'Maximum training sensitivity plus specificity'.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate
1.000	0.045	Fixed cumulative value 1	0.549	0.003
5.000	0.162	Fixed cumulative value 5	0.404	0.032
10.000	0.259	Fixed cumulative value 10	0.331	0.081
0.797	0.038	Minimum training presence	0.568	0.000
11.844	0.282	10 percentile training presence	0.311	0.099
22.934	0.392	Equal training sensitivity and specificity	0.227	0.226
13.130	0.296	Maximum training sensitivity plus specificity	0.299	0.108
21.368	0.382	Equal test sensitivity and specificity	0.237	0.209
9.563	0.253	Maximum test sensitivity plus specificity	0.336	0.077
1.868	0.078	Balance training omission, predicted area and threshold value	0.497	0.006
5.732	0.180	Equate entropy of thresholded and original distributions	0.390	0.035

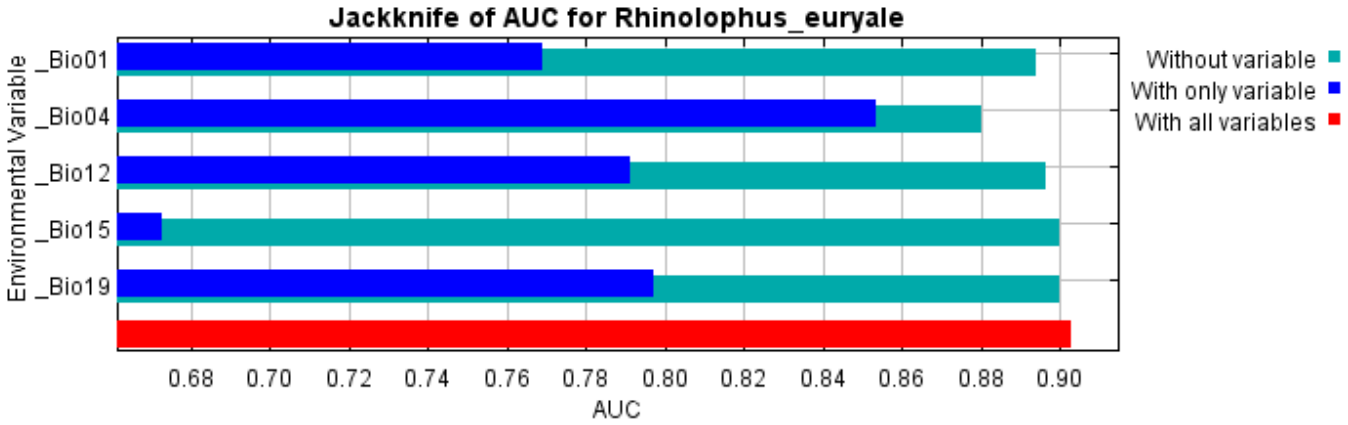
For the next steps (change detection map), you will need to choose a threshold, obtain the value for each of your replicates, and average it.

5. In your summary results, check the contribution of the different variables in the model and the response curves. The variables are scaled so do not have a clear ecological meaning, and you can only see the general broad patterns. We should return the variables to their original values but this is not necessary for this workshop.

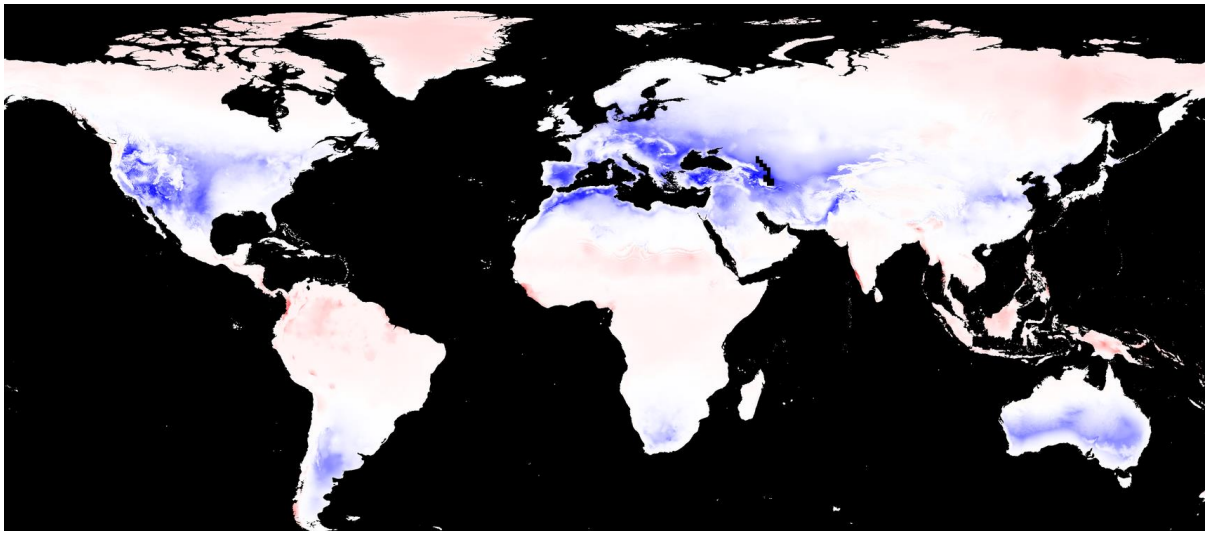


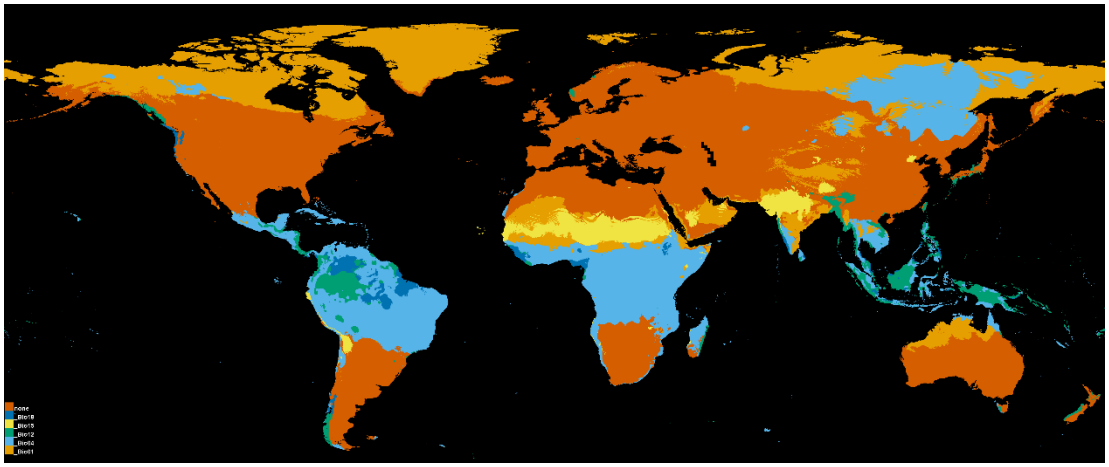
Variable	Percent contribution	Permutation importance
_Bio04	56.6	36.6
_Bio12	32.4	49.2
_Bio01	7.8	7.9
_Bio15	2	1.6
_Bio19	1.2	4.8

The jackknife tests test variable importance by measuring model performance using each variable individually (seeing how much is gained) and how much performance decreases when omitted.



If we select one of the models on the top e.g. [0] and scroll down, we can find information about the clamping problems:

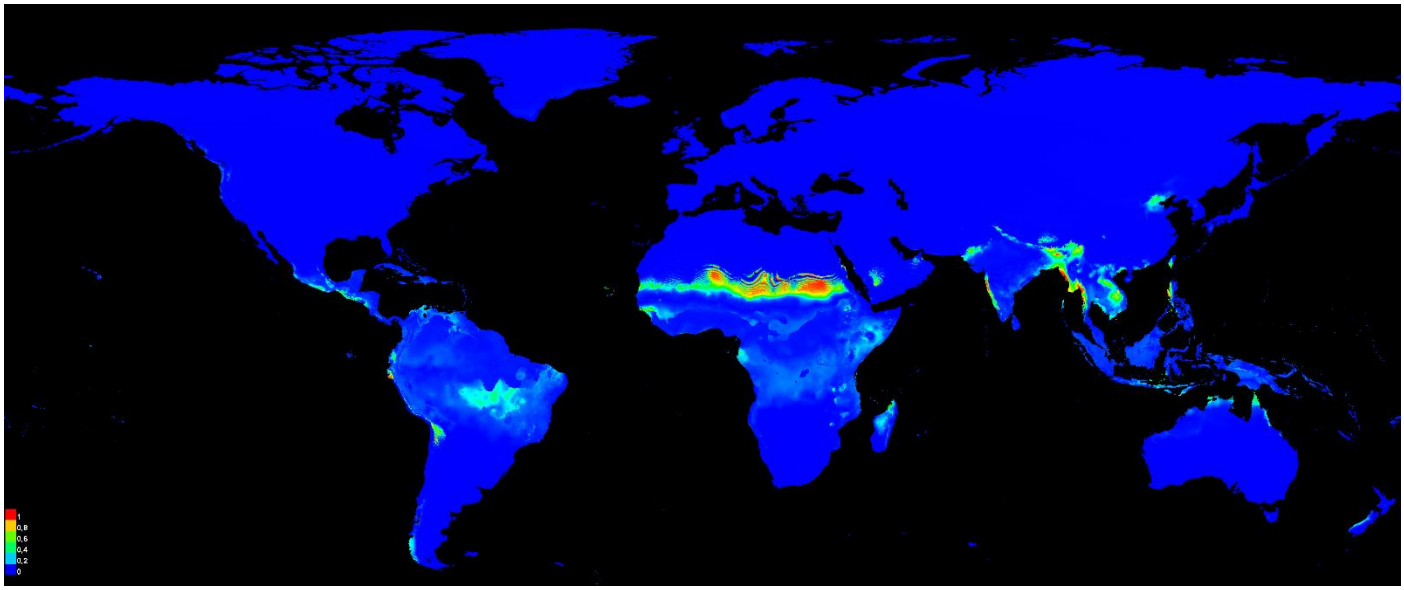




The blue areas show areas where most variables are within the “training range”, regarding the environmental space whilst red areas are locations where predictions should be treated with caution as some variables are outside the training range.

On the bottom map, we see which variables are outside the training area. These maps can be explored in GIS as they are produced as an output of each single model run.

Finally, another important map to look at is:



Which shows the regions where more clamping is having an effect.

Step 4: Making (Change) Maps

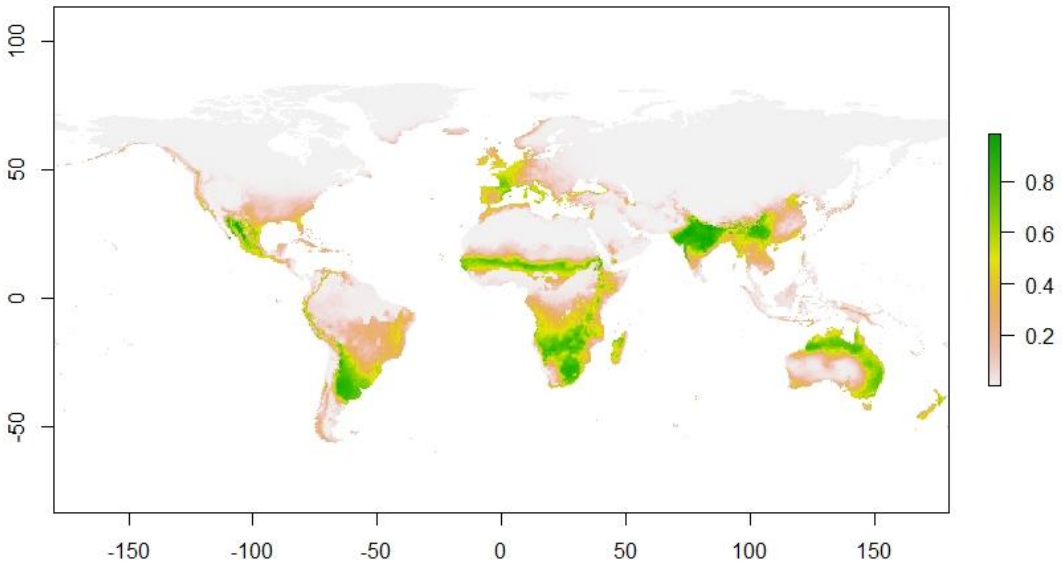
We use a new R-code: 03_Making_ChangeMaps.R

In this code we first have to

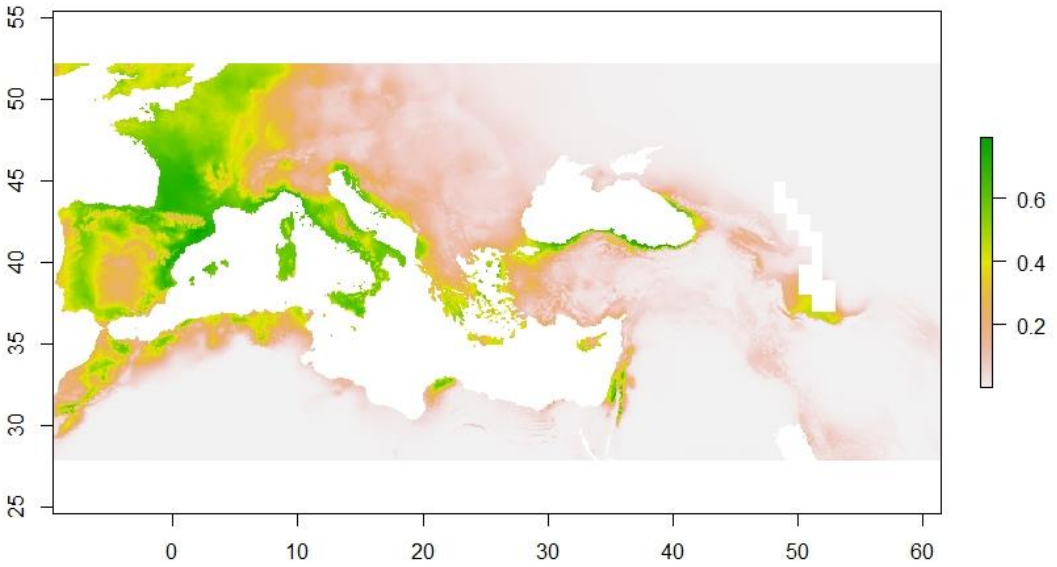
- load again the libraries
- set the working directory
- load the species occurrence file
- create the bounding box around the points

Then we are loading, plotting and clip projection rasters for present situation

```
> pres <- raster("./MaxEnt/Results/Rhinolophus_euryale_Present_WLD_avg.asc")
> # load present global distribution
> plot(pres) #visualize
```


100
50
0
-50
-150
-100
-50
0
50
100
150
0.8
0.6
0.4
0.2

```
> pres_clip <- crop(pres,bbox) #clip to training area
> plot(pres_clip) #visualize
```



Then we are loading, plotting and clip projection rasters for present situation

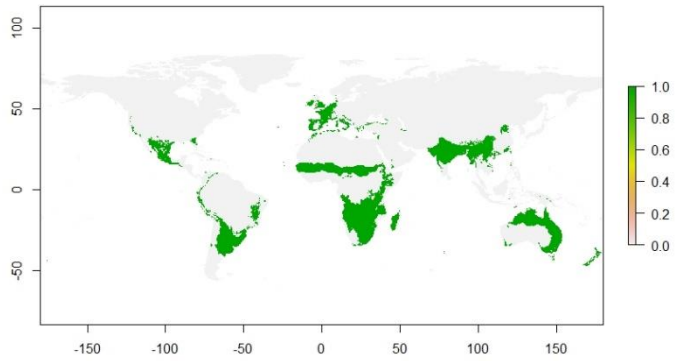
Here we take as an example an moderate scenario HE 4.5

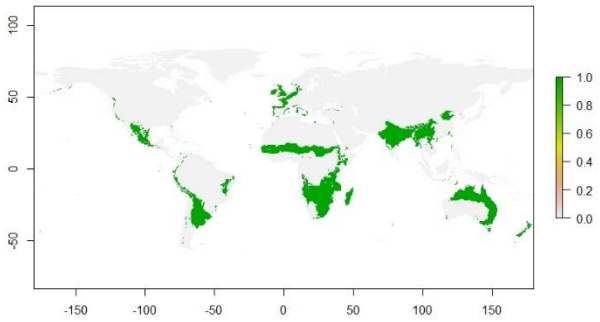
```
> fut <- raster("./MaxEnt/Results/Rhinolophus_euryale_he45bi50_WLD_avg.asc")
> # load future global distribution
> plot(fut)
> fut_clip <- crop(fut,bbox) #clip to training area
> plot(fut_clip) #visualize (not shown)
```

We can see from the global figure and the clip that when only looking at climate variables there is much greater suitable habitat when compared to the actual occurrence of the species. Why do you think this is the case? What other factors have we not modelled that mean the distribution of the species is more restricted?

Next step is to Convert habitat suitability maps to thresholded binary maps (0,1) using the threshold ‘ Maximum training sensitivity plus specificity’. Use the average threshold of the multiple runs

```
> th <- 0.362 #define threshold
> m <- c(0, th,0, th, 1, 1) #matrix everything before th as 0 and everything after th as 1
> bin_mat <- matrix(m, ncol=3, byrow=TRUE) #convert to correct matrix
> pres_bin <- reclassify(pres, bin_mat) #reclassify by matrix present
> fut_bin <- reclassify(fut, bin_mat) #reclassify by matrix future
> plot(pres_bin)
> plot(fut_bin)
```

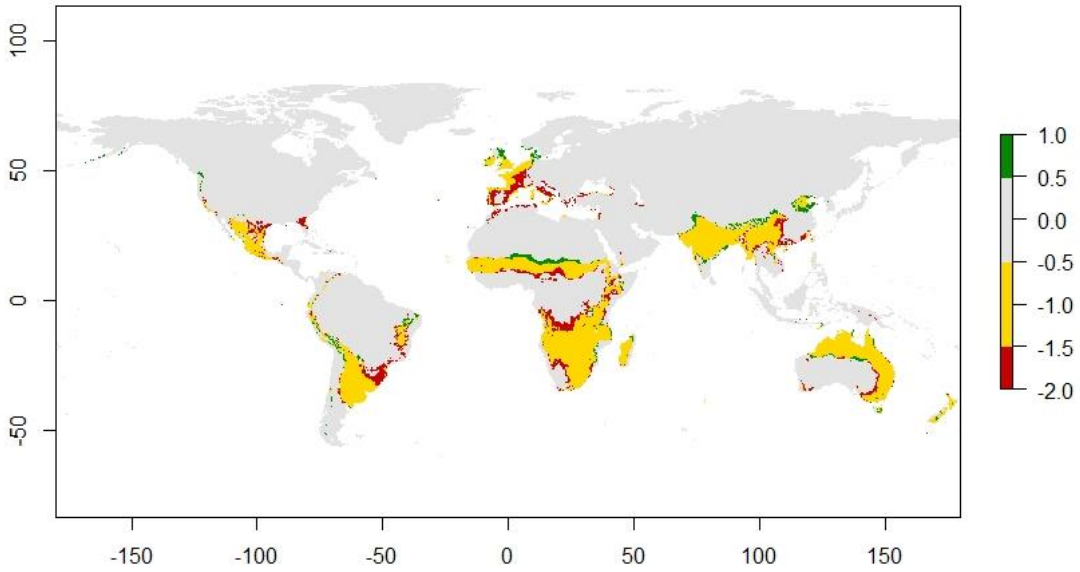




Very little difference between present and future distribution – suggests that under scenario RCP 4.5 for 2050 not a strong effect of climate change – the distribution of the species is more affected by other factors.

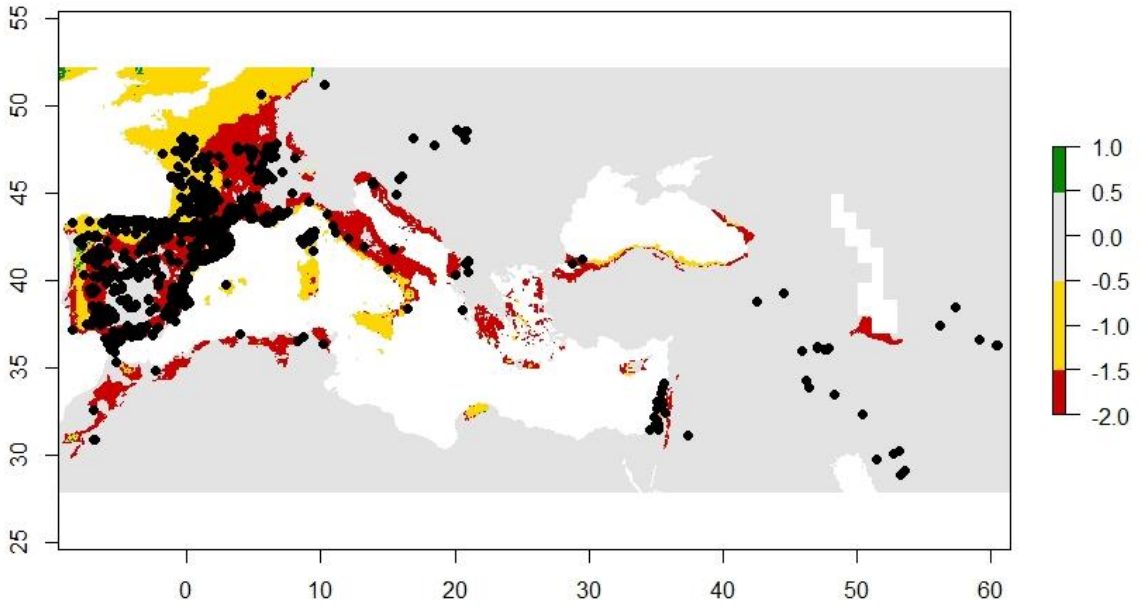
We will calculate differences between present and future in habitat suitability at Global scale. Grey is never suitable, yellow is remains suitable, red is lost, and green is gained

```
> #Calculate Change in Suitable Climate Conditions
> range_change<-BIOMOD_RangeSize(CurrentPred=pres_bin,FutureProj=fut_bin,SpChange.Save=NULL)
> #calculate range change between two maps
> col.lst <- c("red3" , "gold", "grey89" , "green4") #define plot colours
> plot(range_change$Diff.By.Pixel,col=col.lst) #plot range change
```



We also calculate differences between present and future in habitat suitability at occurrence scale. Grey is never suitable, yellow is remains suitable, red is lost, and green is gained.

```
> range_change_clip<-BIOMOD_RangeSize(CurrentPred=pres_bin_clip,FutureProj=fut_bin_cli
p,SpChange.Save=NULL)
> #calculate range change between two maps
> col.lst <- c("red3" , "gold", "grey89" , "green4") #define plot colours
> plot(range_change_clip$Diff.By.Pixel,col=col.lst) #plot range change
> plot(sp_shp,add=T, col="black", pch=19) #show points on top
```



As we expected the climatic envelope of the mediterranean horseshoe bat is much wider than the actual occurrence and therefore climate change does not appear to be the greatest threat. There are other factors as to why this species in threatened and on the redlist (see below).

Threats [top]

Major Threat(s):	Threats include loss of foraging habitat, and disturbance and loss of underground habitats. On a landscape scale, fragmentation and loss of linear elements such as hedgerows and riparian vegetation is a problem because such elements are used as landscape references for commuting. Foraging habitats are lost due to intensive agriculture, urbanization and large infrastructures. The species' strong dependence upon caves for roosting makes it particularly sensitive to cave disturbance, such as that from caving or tourism. Tourist disturbance of caves affects the species in a number of range states. The use of organochlorine pesticides is believed to have contributed to the earlier dramatic decline of the species in France (Brosset *et al.* 1988). In North Africa, threats include habitat loss due to agriculture (livestock) and human disturbance.

Are you predicting into unknown conditions?

The following two pictures compare the environmental similarity of variables in ClimateFuture to the environmental data used for training the model. In the first picture (MESS), areas in red have one or more environmental variables outside the range present in the training data, so predictions in those areas should be treated with strong caution. The second picture (MoD) shows the most dissimilar variable, i.e., the one that is furthest outside its training range. For details, see Elith et al., Method