

Species distribution modelling using MAXENT

Modelling your species habitat suitability under present & future climate conditions

Systematics & Biodiversity, 2020



**Universiteit
Leiden**
The Netherlands



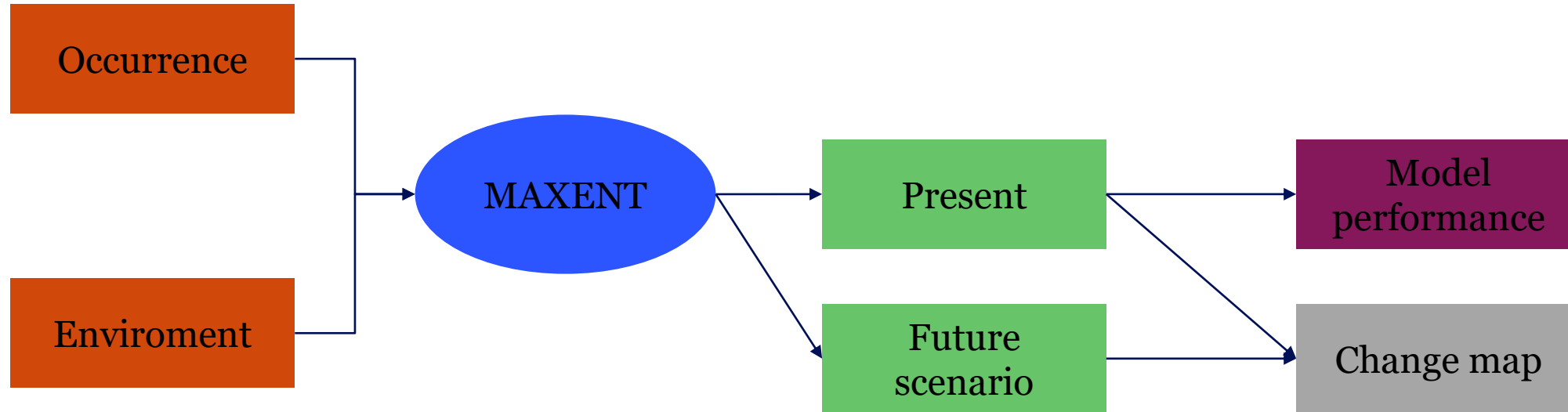
SDM Practical

- How will it work:
 - Different rooms for each group to work together
 - Talk with each other, help each other -> it's not a competition
 - Teaching staff will come in to help during the practical – People with mac, let us know!
- Possible Intermezzo's:
 1. Introducing the problem
 - And setting up your workspace (NOW)
 2. Setting up MAXENT and JAVA & Quick data download guide (Nuno - By group, if needed)
 3. Downloading & exploring occurrence data (Nuno - by group, if needed)
 4. Environmental data & selecting your Environmental variables (Nuno)
 5. Starting up MAXENT (Nuno)
 6. Exploring MAXENT outputs (Nuno)
 7. Creating change maps in ArcGIS (Maarten)
 8. My outputs (Nuno – if there is time)

Don't be afraid to



SDM Practical



- The practical tutorial is based on a bat species -> adapt it to your species
- There is an R script for each section -> you mostly have to adapt the folder paths

SDM Practical

- Think before starting!
 - What do you want to know about your species?
 - How do you experiment it?
 - What scenarios are you interested in testing?
- For your report, be explicit on the hypothesis:
 - Research objectives: How does climate change affect the habitat of X?
 - HO: There is no change in the suitable area of species X
 - HA: There is an increase/decrease of suitable area of species X
- You are here to learn, but for the report design experiment in function of the hypothesis



SDM Practical

- Be organized!
 - Saves time later for your report!
 - Saves time in adapting the code!
 - Think of your folders as a “database”
- Avoid “funky” characters and spaces
 - Programming languages do not like them and crash often because of it
- If you follow our suggested structure your code editing is mimized
- R Script: **OO_SettingUpWorkspace.R**

Nome	Data de modificação
Downloads	17/11/2020 00:28
EnvData	16/11/2020 21:20
FinalOutputs	18/11/2020 22:40
Maxent	18/11/2020 12:02
Occurrences	16/11/2020 21:20
R_Scripts	19/11/2020 16:28
.RData	05/12/2020 23:09
.Rhistory	05/12/2020 23:09
CorrelationTable_AOI	17/11/2020 18:28
CorrelationTable_AOI_AllVariables	17/11/2020 18:32
CorrelationTable_AOI_SelVariables	17/11/2020 18:32

Windows (C:) > Practical > EnvData

Nome	Data de modificação
AOI	18/11/2020 12:36
WLD	18/11/2020 11:59

Windows (C:) > Practical > Maxent

Nome
EXP01
EXP02

SDM Practical

```
2
3 dir.create("C:/Practical/")
4 setwd("C:/Practical/")
5 #from now on the base folder is c:/Practical
6
7 #first tier of folders
8 dir.create("./Downloads")
9 dir.create("./Occurrences")
10 dir.create("./EnvData")
11 dir.create("./Maxent")
12
13
14 #Second tier
15 dir.create("./Occurrences/Shapefiles")
16 dir.create("./Occurrences/Tables")
17
18 dir.create("./EnvData/AOI")
19 dir.create("./EnvData/WLD")
20
21 dir.create("./Maxent/EXP01")
22
23 #Third tier
24 dir.create("./EnvData/AOI/PresentAOI/")
25 dir.create("./EnvData/AOI/FutureAOI/") #NOTICE -
26
27 dir.create("./EnvData/WLD/PresentWLD/")
28 dir.create("./EnvData/WLD/FutureWLD/") #NOTICE -
29
```

The most important commands are the first two:

1. Dir.create - creates a new folder in c:\Practical
 2. Setwd – “set work directory”
- Once you do that – all your files have a “relative path” meaning:
 - Folders are an folder X, located in relation to c:\Practical
 - The rest create different folders
 - Does not overwrite if it is already existing
 - Notice:
 - You can have multiple scenarios but each has to be in its own folder, with an unique name:
 - MAXENT uses the folder name for the projections
 - But it needs that all variables inside have the same name as the files used for training

Setting up MAXENT & JAVA

And starting by Downloading the
climate data during lunch



Universiteit
Leiden
The Netherlands

Setting up Java

1. For Java:

1. Go to https://www.java.com/en/download/help/download_options.xml
2. Follow the steps to download and install the software for your operating system
3. Unsure about your OS? In windows, go to your system properties:

2. Check Java installation:

1. Go to the search box and write “Command line” (or in your own OS language)
2. On the DOS box that opens, type: java – version

```
C:\> Linha de comandos

Microsoft Windows [Version 10.0.19041.630]
(c) 2020 Microsoft Corporation. Todos os direitos reservados.

C:\Users\Nuno>java -version
java version "1.8.0_271"
Java(TM) SE Runtime Environment (build 1.8.0_271-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.271-b09, mixed mode)

C:\Users\Nuno>
```

3. If all is well, you will have a similar output

Ver informações básicas sobre o computador

Edição do Windows

Windows 10 Home

© 2020 Microsoft Corporation. Todos os direitos reservados.

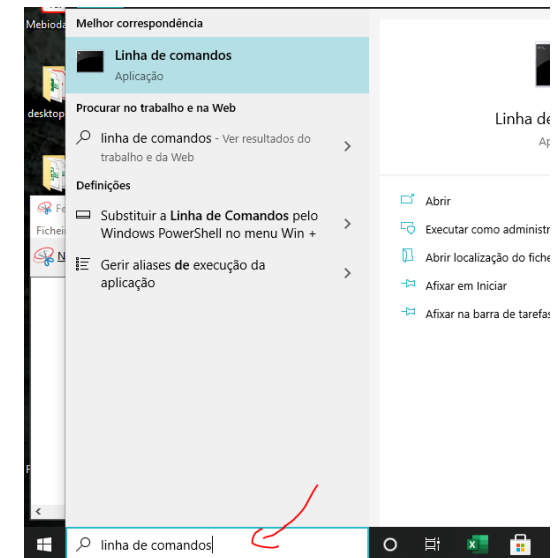
Sistema

Processador: Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz 2.50 GHz

Memória instalada (RAM): 16,0 GB

Tipo de sistema: Sistema Operativo de 64 bits, processador baseado em x64

Caneta e Toque: Não está disponível Introdução por Caneta ou Toque para este Ecrã



Setting up MAXENT:

1. Go to https://biodiversityinformatics.amnh.org/open_source/maxent/
2. Tell them a bit about yourself
 - You can lie, but remember it is a sin
3. Download the software and unzip it to a folder
4. Click on the .Jar file to activate
5. If you see the screen, should be fine

PS: Would be nice to cite them on your report

Current version 3.4.4

Please tell us a little about yourself!

Name:

Institution:

Email:

Comment/Intended Use*:

*Optional

I prefer to download without providing this information

Citation

If you use the application for analyses that result in a publication, report, or online posting, the following represents a proper citation of the software itself:

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2020-12-10.

**For information about earlier versions, please refer to the readme file on github or contact the developers mmmaxent@gmail.com

Maximum Entropy Species Distribution Modeling, Version 3.4.1

Samples **Environmental layers**

☒ Linear features ☐ Create response curves ☐

☒ Quadratic features ☒ Make pictures of predictions ☒

☒ Product features ☐ Do jackknife to measure variable importance ☐

☐ Threshold features ☐ Output format

☒ Hinge features ☐ Output file type

☒ Auto features

Downloading the climate data

1. Go to the manual

- It's already in brightspace!
- Follow the steps to download the climate data detailed there

In summary:

1. Go to <https://www.worldclim.org/>

2. Download:

- Historical bioclimatic data (bio 5m) at 5 minutes resolution
- Future scenario bioclimatic data (bc) data: IPSL-CM6A-LR / SSP370

Occurrence data



Universiteit
Leiden
The Netherlands

SDM Practical

- Occurrence data:
 - <https://www.gbif.org/>
 - It is supposed that you already downloaded and prepared the occurrence data
- No need to repeat this
 - If you are satisfied, move to the next step
 - Remember though, we would like to know that you explored your data
- Notice – everything is useful for the report:
 - Did you find synonyms and choose to use them?
 - Did you find errors on your report and corrected them?
 - Did you find spatial biases?



Global Names resolution tools and services

Resolve lists of scientific names against known sources. This service parses incoming names, executes exact or fuzzy matching as required, and displays a confidence score for each match along with its identifier.

Paste Scientific Names, one on each line

Rhinolophys euryale

<https://resolver.globalnames.org/>

SDM Practical

- Important explanation about .csv files:
 - .csv files are usually stored in different format depending on your operating system
- Please keep a US Version – **MAXENT requires it**
- How to correct: Explained in detail in the manual and the classes
- MAXENT expects:
 - Species name, longitude, latitude

Tab delimited

US

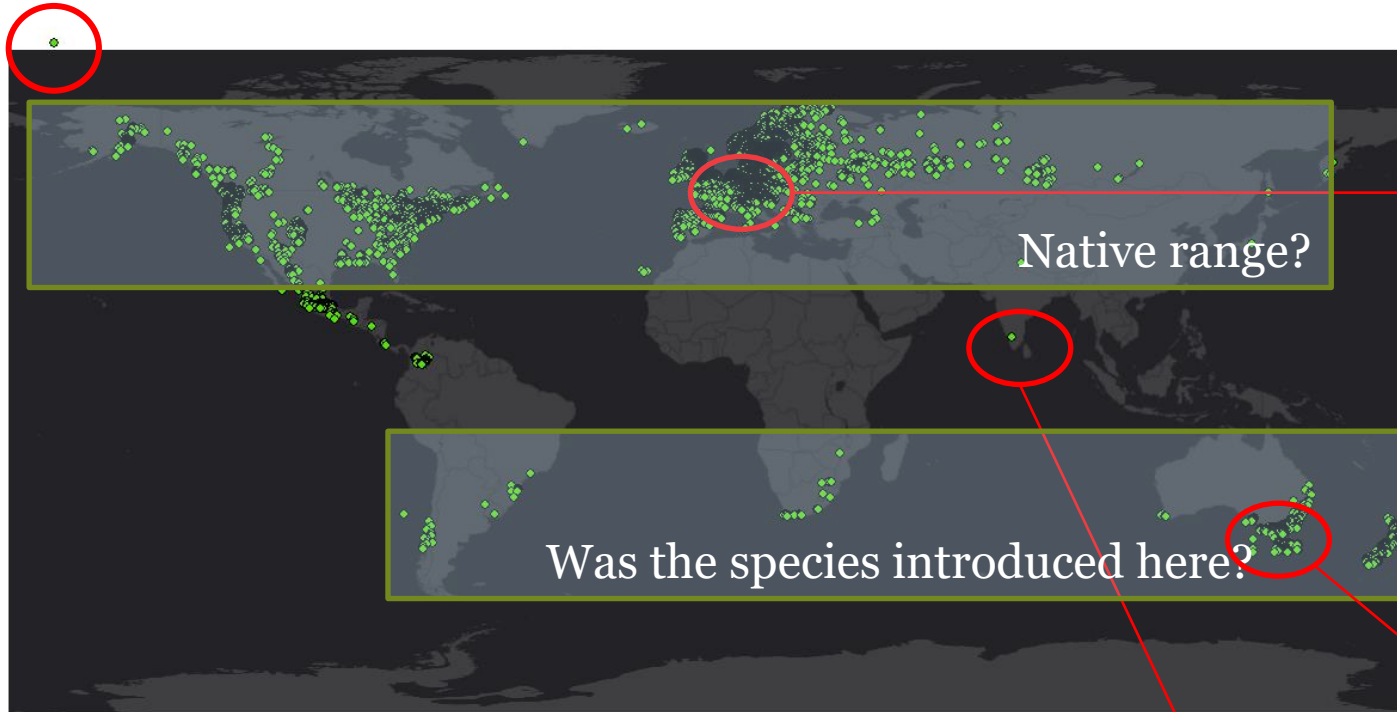
EU(general)

Rhinolophus_euryale_csv0 - Bloco de notas				Rhinolophus_euryale_csv1 - Bloco de notas				Rhinolophus_euryale_csv2 - Bloco de notas			
Ficheiro	Editar	Formatar	Ver Ajuda	Ficheiro	Editar	Formatar	Ver Ajuda	Ficheiro	Editar	Formatar	Ver Ajuda
"species"		"longitude"	"latitude"	"species";"longitude";"latitude"				species;longitude;latitude			
"Rhinolophus euryale"		10.261719	51.193676	"Rhinolophus euryale";10.261719;51.193676				Rhinolophus euryale;10.261719;51.193676			
"Rhinolophus euryale"		5.566667	50.633333	"Rhinolophus euryale";5.566667;50.633333				Rhinolophus euryale;5.566667;50.633333			
"Rhinolophus euryale"		5.566667	50.633333	"Rhinolophus euryale";5.566667;50.633333				Rhinolophus euryale;5.566667;50.633333			
"Rhinolophus euryale"		20.167	48.617	"Rhinolophus euryale";20.167;48.617				Rhinolophus euryale;20.167;48.617			
"Rhinolophus euryale"		20.75	48.516666	"Rhinolophus euryale";20.75;48.516666				Rhinolophus euryale;20.75;48.516666			
"Rhinolophus euryale"		20.887191	48.495193	"Rhinolophus euryale";20.887191;48.495193				Rhinolophus euryale;20.887191;48.495193			
"Rhinolophus euryale"		20.506927	48.467712	"Rhinolophus euryale";20.506927;48.467712				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"		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"		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"		-0.29907	48.20928	"Rhinolophus euryale";-0.29907;48.20928				Rhinolophus euryale;-0.29907;48.20928			
"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";16.917221;48.133888				Rhinolophus euryale;16.917221;48.133888			
"Rhinolophus euryale"		20.749859	48.056655	"Rhinolophus euryale";20.749859;48.056655				Rhinolophus euryale;20.749859;48.056655			
"Rhinolophus euryale"		20.749859	48.056655	"Rhinolophus euryale";20.749859;48.056655				Rhinolophus euryale;20.749859;48.056655			

- **Recommended:**
 - Keep all versions, saved with different names
 - E.g. Species_csv1.csv or species_csv2.csv

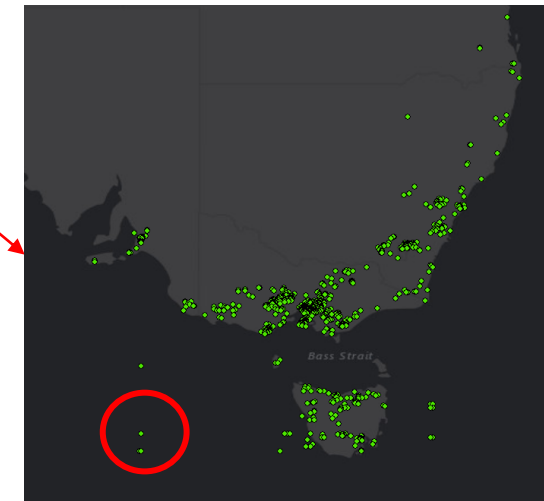
SDM Practical

- *Amanita muscaria*



Different sampling strategies

- In a GIS you can easily clean the obvious errors
- Further explore if the data “makes sense”
- When satisfied, export back to csv (check GIS notes)



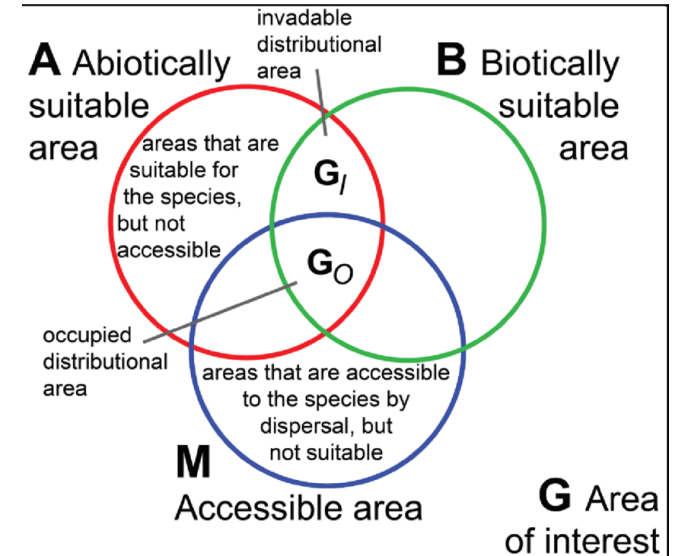
SDM Practical

Example problems:

- What is the real ecological niche of the species?
- Can the native range predict the invaded distribution?
- How does the sampling design affect the model prediction?

You are free to explore these for your case

- *Amanita muscaria*



Downloading and preparing environmental data



Universiteit
Leiden
The Netherlands

SDM Practical

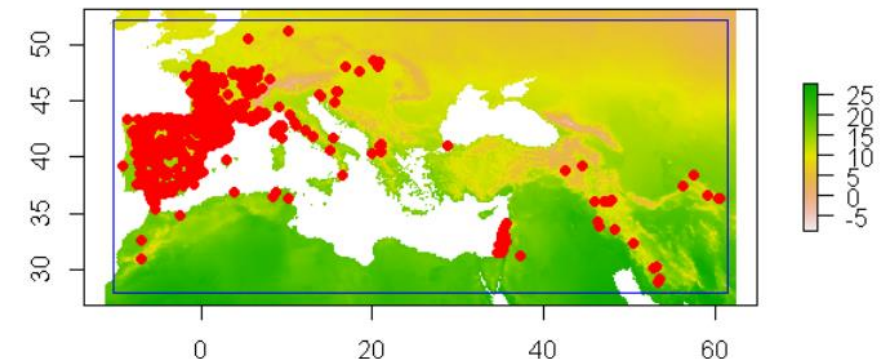
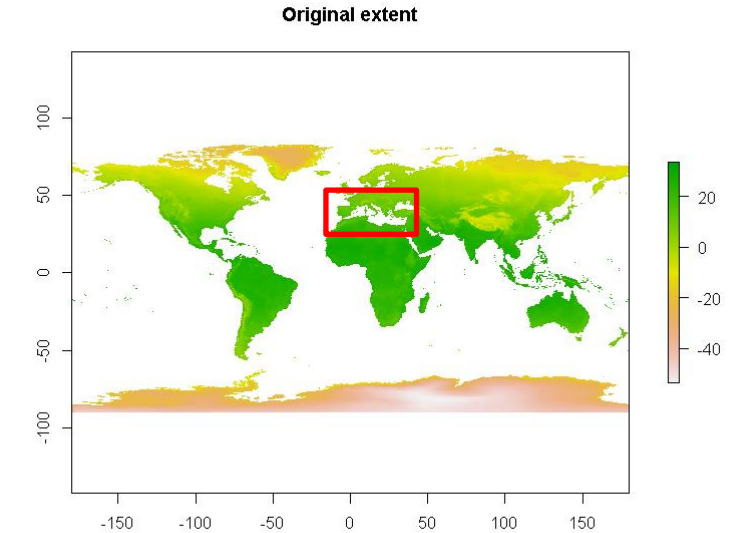
1. Downloading files from Wordclim

- Present and future scenario

2. Within R: `o1_CroppingEnvVariables.R`

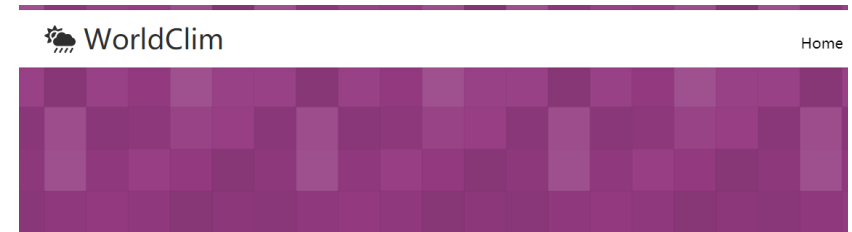
1. Fetching the environmental variables
2. Cropping to study area
3. Exporting files in .asc format for MAXENT

- When do the variables do not have to be cropped?
 - When your species has a global distribution.. It becomes moot



SDM Practical

- Worldclim data
 - World Climate Research Programme (1980 - Present)
 - Method: [Hijmans, 2005](#) (Also [Raster package R](#))
- Provides historical data (present)
 - 1970 ~ 2000
 - Represents the “recent trend” in climate
- Shared Socio-economic pathways ([Riahi, 2016](#))
 - Represent different future scenarios regarding CO2 emissions
- Available scenarios:
 - SSP1: Sustainability – “taking the green road”
 - SSP2: “Middle of the road”
 - **SSP3: Regional rivalry – “A rocky road”**
 - SSP5: Fossil-fueled Development – “Taking the highway”

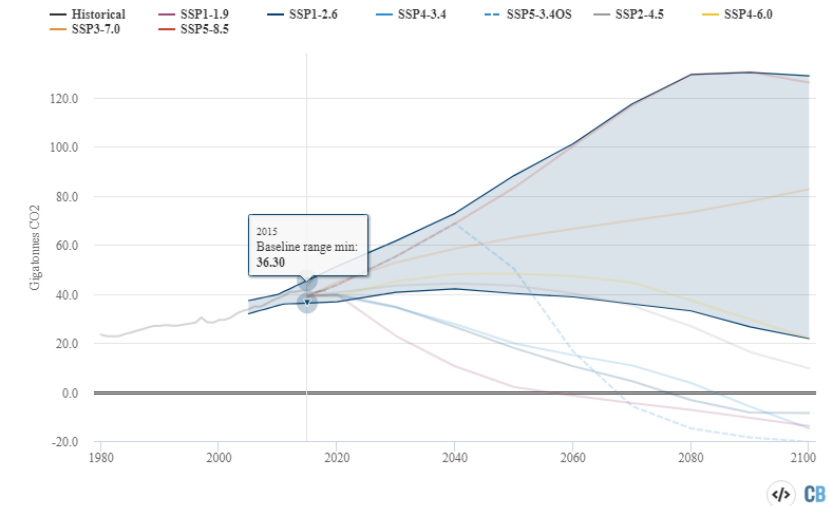


Global climate and weather data

Historical climate data
Historical monthly weather data
Future climate data

Worldclim

CO2 emissions in CMIP6 scenarios



Carbon brief

SDM Practical

$$SideLength_{meters} \approx \left(WclimRes_{seconds} \cdot \frac{EquatorEarthRadius_{meters}}{360^\circ} \right)$$

- On historical data:
 - The 10, 5, 2.5 minutes and 30s represent spatial resolution
 - Minutes of degrees in DMS geographical coordinates format
- We will use:
 - 5 minutes resolution (~10km in the equator)
 - Bioclimatic variables
- Select the Bioclimatic variables
 - Representative of seasonal trends that are somehow significant to the physiological constraints of species

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

For reference, here is the elevation data that was used to produce WorldClim 2.1. These were derived from the SRTM elevation data.

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
Elevation	elev 10m	elev 5m	elev 2.5m	elev 30s

SDM Practical

- 19 different variables
 - Represent relevant ecophysiological quantities for temperature and precipitation
- Keep note of these variable names!
- You'll have to select variables based on:
 - Ecological knowledge (first and foremost!)
 - Statistical tests

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

SDM Practical

- Future scenarios
 - On the tutorial we use: **SSP3 Regional Rivalry** - “rocky road”
- Global Circulation Model:
 - Predictions generated from complex physical simulations of climate models
 - We used [IPSL-CMC6-LR](#) – use this GCM also
 - **5 minute resolution!**
- Tutorial time interval: **2061-2080**
 - ~50 years from today.
- **Use at least this scenario** but you are free (and encouraged) to try more

Time interval used

2061-2080

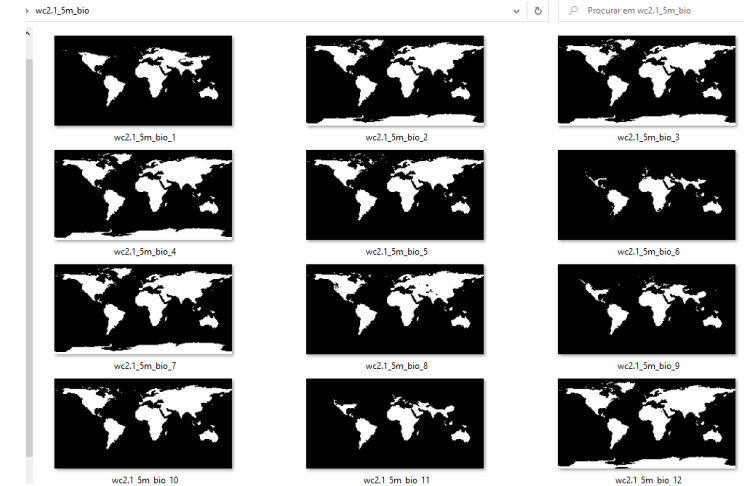
GCM	ssp126	ssp245	ssp370	ssp585
BCC-CSM2-MR	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
CNRM-CM6-1	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
CNRM-ESM2-1	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
CanESM5	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
GFDL-ESM4	tn, tx, pr, bc	--, --, --, --	tn, tx, pr, bc	--, --, pr, --
IPSL-CM6A-LR	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
MIROC-ES2L	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
MIROC6	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc
MRI-ESM2-0	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc	tn, tx, pr, bc

bc – Bioclimatic variables

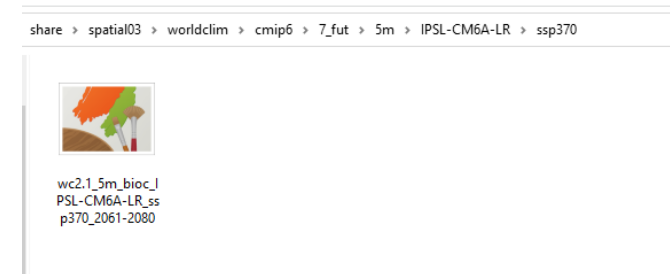
SDM Practical

- Once you download, and unzip the files you'll notice:
 - Historical data/Present – 19 files in .tif file
 - Scenario/Future – A single file in .tif format
 - The “order” of variables is “weird”: **Bio1,bio11 ... bio19,bio2,bio4**
- The scenario/future data has 19 bands:
 - Each band one of the bioclimatic variables, in order
- **MAXENT requires:**
 - One separate file per variable
 - Ascii format (.asc)
- R code will be adapted for both cases:
01_CroppingEnvVariables.R

Present:



Future:



SDM Practical

1

```
8
9 #wordclim version: 2.0
10 #future variables scenario:IPSL-CM6A-LR - ssp370 - 61-80
11
```

2

```
12 #set work directory
13 setwd("C:/Practical")
14
15 #lists all historical Bioclimatic variables into two objects: one for the names and one for the p
16 list.files("./Downloads/wc2.1_5m_bio",pattern=".tif")
17 #unfortunately the names are not in numerical order, we can fix this when we list the files
18 list.files("./Downloads/wc2.1_5m_bio",pattern=".tif")[c(1,12:19,2:11)]
19
```

3

```
20 #fetching historical data
21 rst.nms <- list.files("./Downloads/wc2.1_5m_bio",pattern=".tif")[c(1,12:19,2:11)]
22 rst.fld <- list.files("./Downloads/wc2.1_5m_bio",pattern=".tif",full.names = T)[c(1,12:19,2:11)]
23
```

4

```
24 #loading all rasters into a single multi-band raster:
25 rst.stk <- stack(rst.fld)
26
```

5

```
27 #Renaming bioclimatic layers:
28 names(rst.stk)
29 names(rst.stk) <- c("Bio01","Bio02","Bio03","Bio04",
30                    "Bio05","Bio06","Bio07","Bio08",
31                    "Bio09","Bio10","Bio11","Bio12",
32                    "Bio13","Bio14","Bio15","Bio16",
33                    "Bio17","Bio18","Bio19")
34
```


SDM Practical

```
49 sp <- read.csv2("./Occurrences/Tables/Rhinolophys_csv2_clean.csv",header=T)
50 head(sp) #check table looks correct
51 sp_shp <- sp #rename table
52 coordinates(sp_shp) <- ~longitude+latitude #convert table to points shapefile
53 proj4string(sp_shp) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
54
55 #Create bounding box around points
56 bbox <- extent(sp_shp) #create bounding box of points
57 bbox <- bbox+2 #increase border so we do not truncate data
58 plot(rst.stk$Bio01,ext=bbox+2)
59 plot(bbox, col='blue',add=T) #check if box surrounds points
60 plot(sp_shp,add=T,pch=19,col='red') #add
61
62 #cropping the present data
63 stk.present.AOI.crop <- crop(rst.stk,bbox) #clip to training area
64 #plotting the example
65 par(mfrow=c(1,2)) #sets the plotting area to a 1 line 2 columns set up
66 plot(rst.stk$Bio01,main="Original extent")
67 plot(stk.present.AOI.crop $Bio01,main="Cropped extent")
68 par(mfrow=c(1,1)) #sets it back to 1 image per plot
69
70 #cropping the future data
71 stk.ssp370.AOI.crop <- crop(rst.ssp370,bbox)
72
```


SDM Practical

1

```
73 #writing the images in ascii format
74
75 #first the uncropped data:
76 #now we can save them to another folder in a format
77 #that maxent can read
78 #saving the cropped historical data data in .asc format
79 writeRaster(rst.stk, #multilayer raster
80             "./EnvData/WLD/PresentWLD/.asc", #output folder plus extension .asc
81             overwrite=T, #overwrites any files with the same name in the folder
82             bylayer=T, #saves each variable with the layer name that we set before
83             suffix="names") #uses the band names instead of the band number
84
```

2

```
85 #Same for the future data
86 writeRaster(rst.ssp370, #multilayer raster
87             "./EnvData/WLD/FutureWLD/.asc", #output folder plus extension .asc
88             overwrite=T, #overwrites any files with the same name in the folder
89             bylayer=T, #saves each variable with the layer name that we set before
90             suffix="names") #uses the band names instead of the band number
91
```

3

```
92 #now we do the same, for the AOI
93 writeRaster(stk.present.AOI.crop, #multilayer raster
94             "./EnvData/AOI/PresentAOI/.asc", #output folder plus extension .asc
95             overwrite=T, #overwrites any files with the same name in the folder
96             bylayer=T, #saves each variable with the layer name that we set before
97             suffix="names") #uses the band names instead of the band number
98
```

4

```
99 #Same for the future data
100 writeRaster(stk.ssp370.AOI.crop, #multilayer raster
101             "./EnvData/AOI/FutureAOI/.asc", #output folder plus extension .asc
102             overwrite=T, #overwrites any files with the same name in the folder
103             bylayer=T, #saves each variable with the layer name that we set before
104             suffix="names") #uses the band names instead of the band number
105
```

Selecting environmental variables



Universiteit
Leiden
The Netherlands

SDM Practical

- First step:

- Considering your ecological knowledge of the species, select a number of bioclimatic variables
- 5 ~ 10 variables



All statistical tests have to be done on the AOI

- Second step:

- Spearman's rank autocorrelation test (r) between each pair of variables
- Measures statistical association/dependence between two variables
- If $r > 0.7$ between multiple variables, it is a problem



Helps identify which variables might be a problem

- Last step:

- Variance Inflation Factor (VIF) using only the environmental variable
- Measures the amount of multicollinearity in the model
- If $VIF > 10$ for any of the variables (normally pairs) – you have to remove



Tells you if there is a problem

SDM Practical

- Spearmans Rank correlation coefficient – R

$$r_s = \rho_{rg_X, rg_Y} = \frac{\text{cov}(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}},$$

where

ρ denotes the usual Pearson correlation coefficient, but applied to the rank variables,

$\text{cov}(rg_X, rg_Y)$ is the covariance of the rank variables,

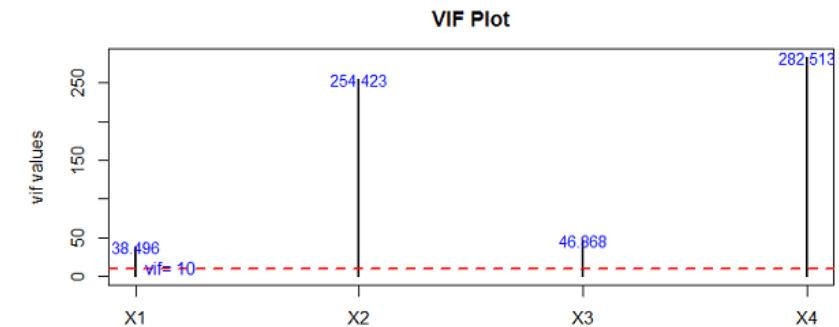
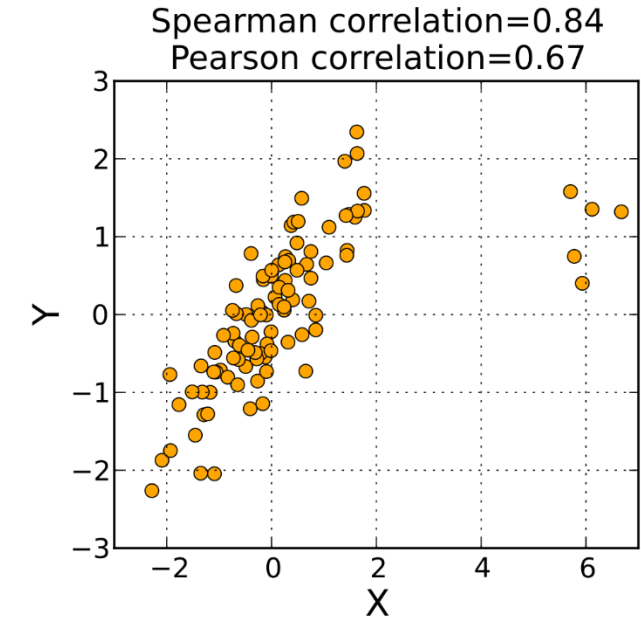
σ_{rg_X} and σ_{rg_Y} are the standard deviations of the rank variables.

- Variance of Inflation Factor (VIF)

1. OLS of Variable X ~ of all others
2. Calculate VIF of variable X in function of all the other variables

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

where R_i^2 is the coefficient of determination of the regression equation



SDM Practical

- Script: 02_VariableSelection.R

1

```
9  ###first load all the env data of your AOI in the present
10 setwd("c:/Practical")
11 path2presentData <- list.files("./EnvData/AOI/PresentAOI/",pattern=".asc",full.names = T)
12 stk.present.AOI.crop <- stack(path2presentData)
13 names(stk.present.AOI.crop) <- c("Bio01","Bio02","Bio03","Bio04",
14                                "Bio05","Bio06","Bio07","Bio08",
15                                "Bio09","Bio10","Bio11","Bio12",
16                                "Bio13","Bio14","Bio15","Bio16",
17                                "Bio17","Bio18","Bio19")
18
```

2

```
19 ### the autocorrelation testing is important ONLY for the areas where the
20 ### model is trained, so, for this section, we use only the cropped enviromental data
21
22 ### pairwise testing |
23 #first we convert the cropped raster to a data.frame
24 stk.present.AOI.crop <- na.omit(as.data.frame(stk.present.AOI.crop)) #we also remove NA's
25 #now this stores the pearson correlation in a matrix
26
```

3

```
27 #here we can already select only the variables that we are interested on - or - you can just calculate
28 #using all the variables. It is the same, the pairwise correlation does not change since each one is a different comparision
29 stk.present.AOI.crop.sel <- stk.present.AOI.crop[,c("Bio01","Bio04","Bio07","Bio12","Bio15","Bio19")]
30
```

4

```
31 #cor function calculates the pairwise r correlation
32 cor.tab.allvariables <-cor(stk.present.AOI.crop)
33 cor.tab.selvariables <-cor(stk.present.AOI.crop.sel)
34
35 cor(stk.present.AOI.crop.sel,method="spearman")
36 cor(stk.present.AOI.crop.sel,method="spearman")
37
```

SDM Practical

1. Some of my variables have high r value
2. There is only need to remove if VIF > 10

	A	B	C	D	E	F	G
1		Bio01	Bio04	Bio07	Bio12	Bio15	Bio19
2	Bio01	1	-0,18395	0,08421	-0,5858	0,656659	-0,30598
3	Bio04	-0,18395	1	0,91608	-0,40059	-0,12086	-0,4368
4	Bio07	0,08421	0,91608	1	-0,57283	0,104595	-0,5289
5	Bio12	-0,5858	-0,40059	-0,57283	1	-0,40992	0,851161
6	Bio15	0,656659	-0,12086	0,104595	-0,40992	1	-0,06551
7	Bio19	-0,30598	-0,4368	-0,5289	0,851161	-0,06551	1
8							

```
46 #multicollinearity testing
47 library(usdm)
48
49 #e.g. i select Bio01; Bio04; Bio07; Bio 12; Bio 15 and bio 19
50 stk.present.AOI.crop.sel <- stk.present.AOI.crop[,c("Bio01","Bio04","Bio07","Bio12","Bio15","Bio19")]
51 #and the VIF test
52 vif(stk.present.AOI.crop.sel, maxobservations=nrow(stk.present.AOI.crop.sel))
53
54 df.stk.AOI <- stk.present.AOI.crop[,c("Bio01","Bio04","Bio12","Bio15","Bio19")] #minus the temperature range
55 vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
56
```

```
Variables      VIF
1      Bio01  3.302447
2      Bio04 11.499434
3      Bio07 11.401577
4      Bio12  9.375736
5      Bio15  2.338738
6      Bio19  5.847520
>
```

Indeed, we opted to remove Bio07 – “Temperature Annual Range” because we considered seasonality more important in comparision

SDM Practical

- You have to **repeat VIF test and remove variables until $VIF < 10$**
- If you have multiple variables contributing to the VIF
 - Choose the most important ones, check if VIF is acceptable
 - You can also try different configuration
- The pair comparisons of the **Spearman's r do not need to be repeated** as nothing changes
- Once you finalized the final “structure” of your model, remember the names because you will need them for MAXENT

Starting up MAXENT – Finally!



Universiteit
Leiden
The Netherlands

SDM Practical:

- Check-up:
 - Do you have Java installed and downloaded MAXENT?
 - Checking java: Open “Command line” and type “java -version”.
 - Checking maxent: open the .bat file
- Regarding occurrence data:
 - Is it in NA format (decimals as points and separators as commas)
 - Is the table ordered: Species, longitude, latitude?
 - Do you have errors in the rows? Different species names, missing values?
- Regarding the environmental data:
 - Do you have the list of the selected variables?
 - Are all variables (present and scenarios) in .asc format?
 - Are they saved to different folders so that each last subfolder has a unique name?
 - Do all the files (not the folders) have precisely the same name?

SDM Practical:

Load the occurrences
(beware of the type of CSV)



Maximum Entropy Species Distribution Modeling, Version 3.4.1

Samples

File:

☒ Rhinolophus_euryale

Environmental layers

Directory/File:

Variable	Type
<input type="checkbox"/> _Bio09	Continuous
<input type="checkbox"/> _Bio10	Continuous
<input type="checkbox"/> _Bio11	Continuous
<input checked="" type="checkbox"/> _Bio12	Continuous
<input type="checkbox"/> _Bio13	Continuous
<input type="checkbox"/> _Bio14	Continuous
<input checked="" type="checkbox"/> _Bio15	Continuous
<input type="checkbox"/> _Bio16	Continuous
<input type="checkbox"/> _Bio17	Continuous
<input type="checkbox"/> _Bio18	Continuous
<input checked="" type="checkbox"/> _Bio19	Continuous

☒ Linear features
☒ Quadratic features
☒ Product features
☐ Threshold features
☒ Hinge features
☒ Auto features

☒ Create response curves
☒ Make pictures of predictions
☒ Do jackknife to measure variable importance

Output format:
Output file type:

Output directory:
Projection layers directory/file:

Load the enviromental variables –
point to the folder



De-select all and then select
your variables



Activate everything and
choose logistic output



Press settings for extra
options



SDM Practical

The screenshot shows the 'Maximum Entropy Parameters' dialog box with three tabs: 'Basic', 'Advanced', and 'Experimental'. The 'Basic' tab is selected. It contains a list of checkboxes and several input fields. Annotations with arrows point from specific UI elements to explanatory text on the right.

UI Element	Description
<input checked="" type="checkbox"/> Random seed	Tells the algorithm to randomize the sample generation before each validation
<input checked="" type="checkbox"/> Give visual warnings	Pop-up warnings and interface info
<input checked="" type="checkbox"/> Show tooltips	
<input checked="" type="checkbox"/> Ask before overwriting	
<input type="checkbox"/> Skip if output exists	Tests and removes presences that fall in the same pixel – keeps only 1 per pixel
<input checked="" type="checkbox"/> Remove duplicate presence records	
<input checked="" type="checkbox"/> Write clamp grid when projecting	Saves a clamp file when reprojecting and does the MESS analysis
<input checked="" type="checkbox"/> Do MESS analysis when projecting	
Random test percentage: 10	Presence data % used for validation (e.g. AUC etc)
Regularization multiplier: 1	A scalar value multiplier to apply to the regularization parameters
Max number of background points: 10000	Number of samples to use for estimating the background distribution; AKA Pseudo-absences
Replicates: 10	
Replicated run type: Crossvalidate	Number of replicates (repetitions) and training validation method
Test sample file: [empty] [Browse]	Optional external validation dataset

SDM Practical

Maximum Entropy Parameters

Basic Advanced Experimental

- ☒ Add samples to background → Equivalent to sampling with replacement or not
- ☒ Add all samples to background
- ☒ Write plot data → Save the data used on the plots (e.g. Response curves)
- ☒ Extrapolate → Allow the model to extrapolate outside the training range
- ☒ Do clamping → Apply clamping to extraopolation areas
- ☐ Write output grids → If on, all models will be saved onto the disk in .asc files
- ☒ Write plots → Write the plots used in the model report onto the disk
- ☐ Append summary results to maxentResults.csv file → Append the new run to a previous run or start over
- ☒ Cache ascii files

Maximum iterations	500	Run the machine learning 500 times model or until the convergence is below 0,00001
Convergence threshold	0,00001	
Adjust sample radius	0	→ Cosmetics, changes the size of background data points on the output map
Log file	maxent.log	→ Log file and assumed prior prevalence
Default prevalence	0,5	
Apply threshold rule		→ Generate a presence/absence map based on the rule
Bias file	<input type="text"/> Browse	→ Spatial sampling bias file

SDM Practical

The screenshot shows the 'Maximum Entropy Parameters' dialog box with the 'Experimental' tab selected. It contains several checkboxes and a list of numerical parameters.

Parameter	Value
Threads	1
Lq to lqp threshold	80
Linear to lq threshold	10
Hinge threshold	15
Beta threshold	-1
Beta categorical	-1
Beta lqp	-1
Beta hinge	-1
Default nodata value	-9999

Mostly cosmetic, and for exploring the outputs.

Fade by clamping alters your outputs

Here: Number of CPU cores the software can use

$$\text{RE}(\tilde{\pi} \parallel q_{\lambda}) + \sum_j \beta_j |\lambda_j|$$

All these options relate to this equation and within the algorithm parameters
Sets default nodata value

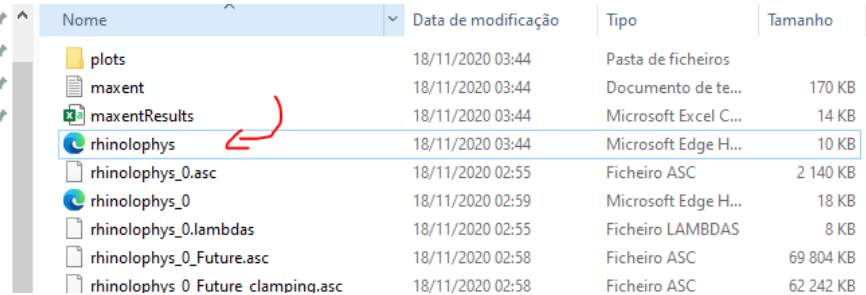
MAXENT – Exploring the outputs



Universiteit
Leiden
The Netherlands

SDM Practical

- Open the MAXENT output folder
 - C:/Practical/Maxent/<etc>
- You see a huge number of files
 - Click the .html with your species name
- The one without any number will provide a summary of all the replicates
 - Each specis_number is the summary of that individual run
- The “original” output files are also here, ready to be opened in a GIS
 - They are the .asc files
 - Notice, the outputs are in the same coordinate system as the inputs BUT the images do not have any projection information associated with them

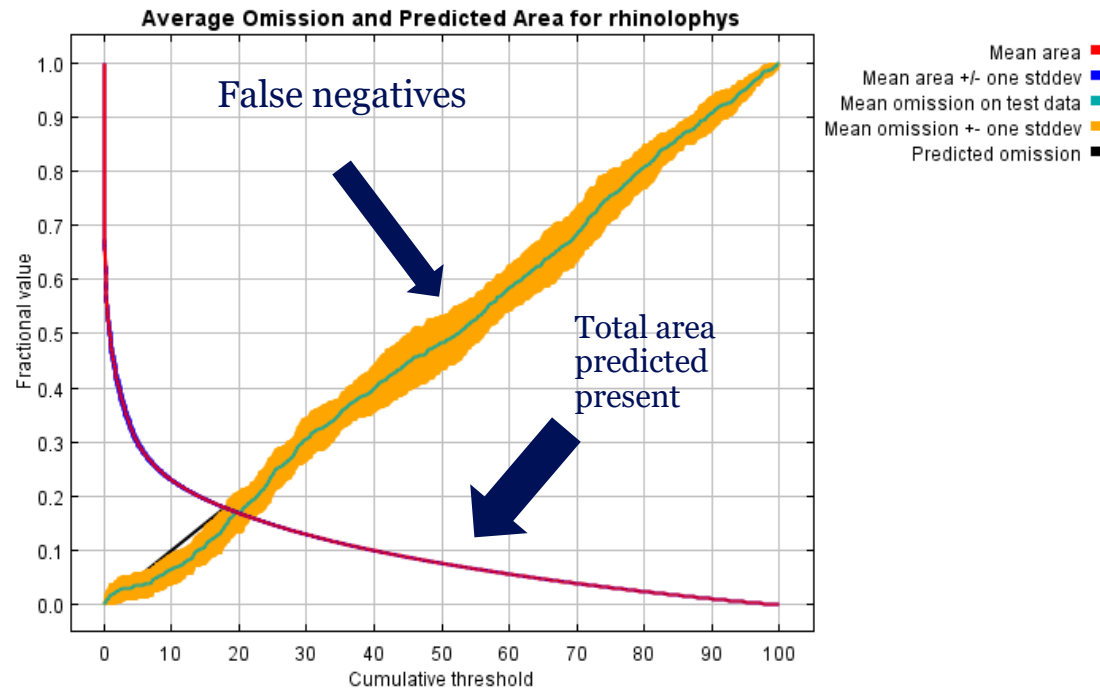


Nome	Data de modificação	Tipo	Tamanho
plots	18/11/2020 03:44	Pasta de ficheiros	
maxent	18/11/2020 03:44	Documento de te...	170 KB
maxentResults	18/11/2020 03:44	Microsoft Excel C...	14 KB
rhinolophys	18/11/2020 03:44	Microsoft Edge H...	10 KB
rhinolophys_0.asc	18/11/2020 02:55	Ficheiro ASC	2 140 KB
rhinolophys_0	18/11/2020 02:59	Microsoft Edge H...	18 KB
rhinolophys_0.lambdas	18/11/2020 02:55	Ficheiro LAMBDAS	8 KB
rhinolophys_0_Future.asc	18/11/2020 02:58	Ficheiro ASC	69 804 KB
rhinolophys 0 Future clamping.asc	18/11/2020 02:58	Ficheiro ASC	62 242 KB

Replicated maxent model for rhinolophys

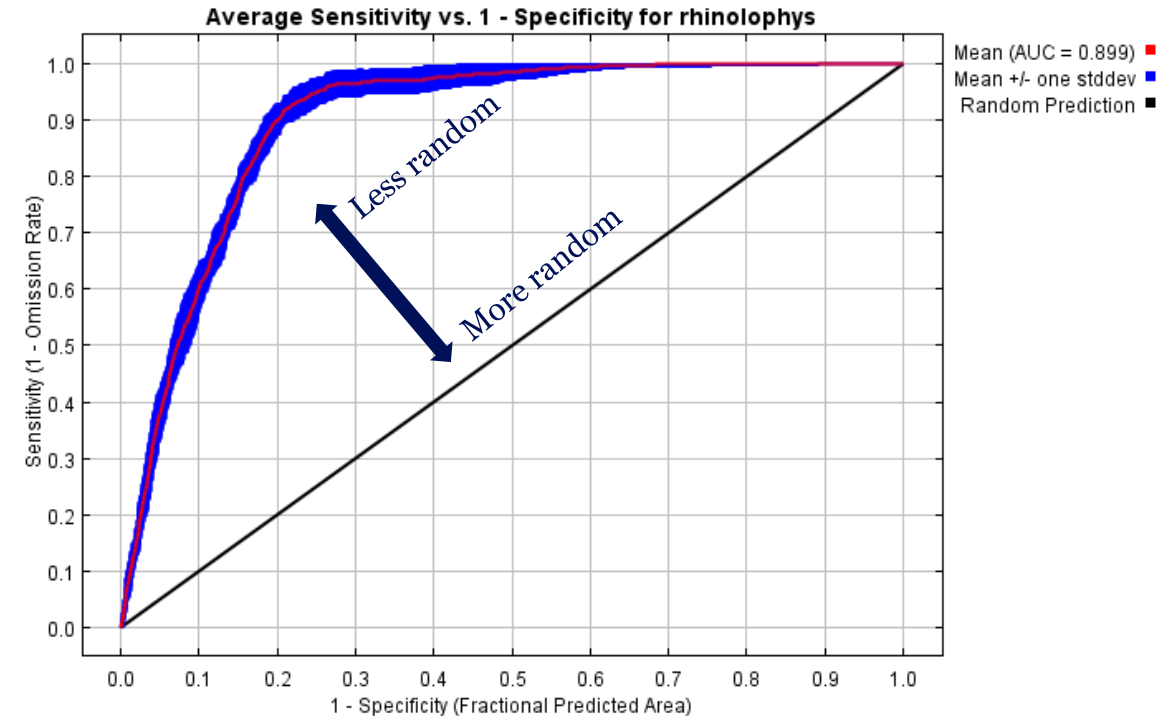
v 18 03:32:58 CET 2020 using Maxent version 3.4.1. The individual models are here: [\[0\]](#) [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#)

SDM Practical



The closer the omission rate (yellow range) is to the predicted omissions (black line) -> the better!

The other curve represents the total area predicted as presence, given the threshold.

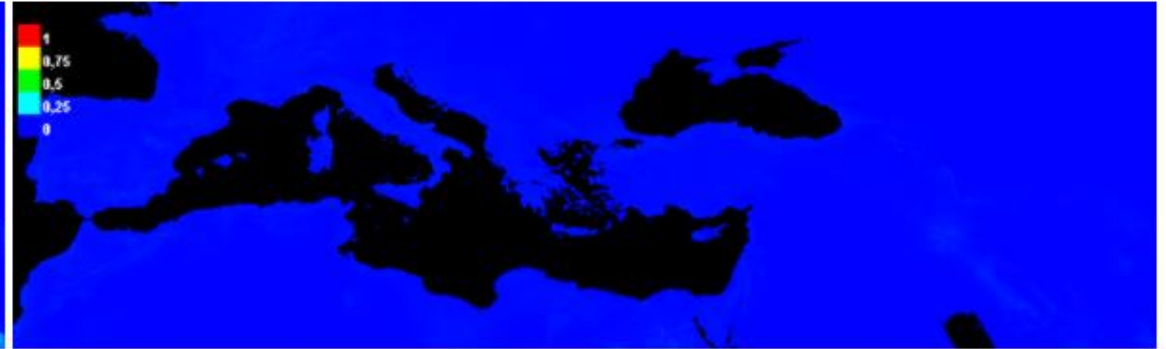
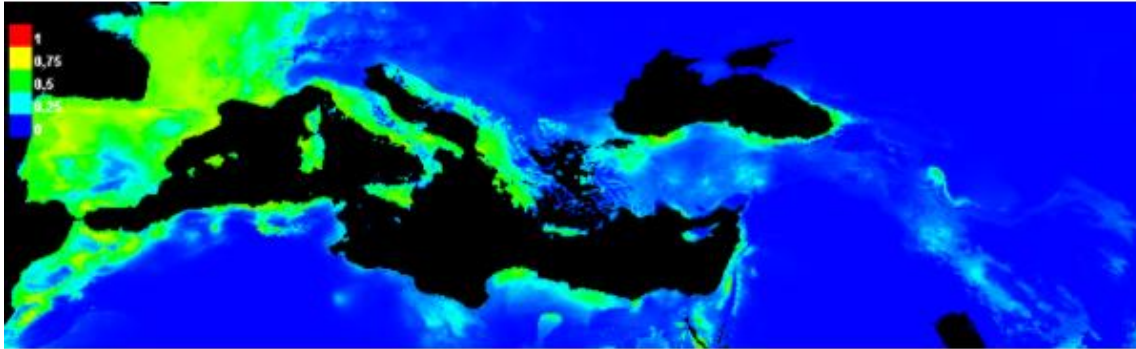


AUC-ROC curve for increasing threshold values. The closer to 1, the better (remember the theory!).

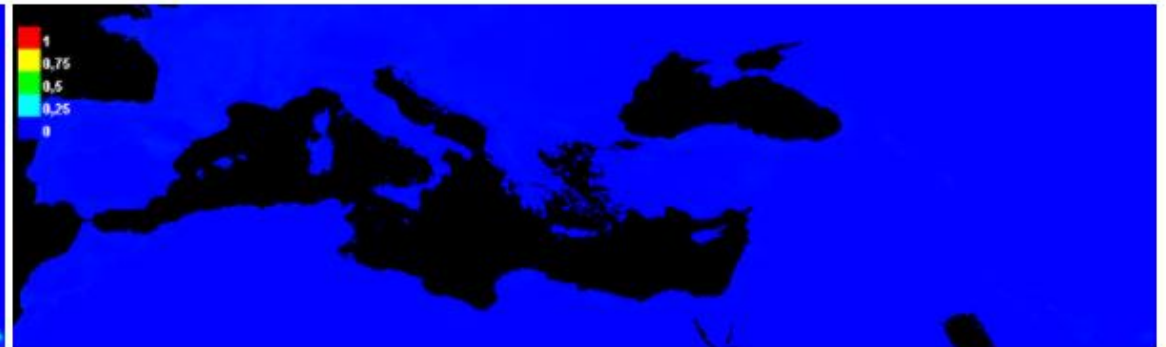
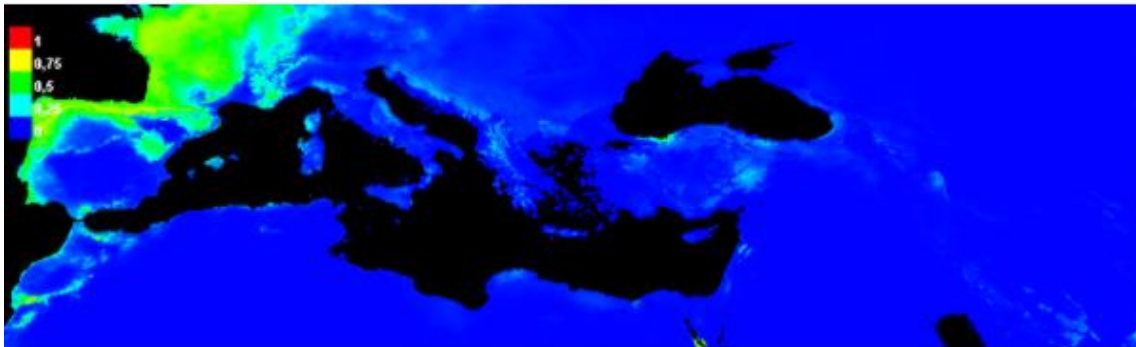
The closer to the 1:1 line, the more “random” your prediction

SDM Practical

The following two pictures show the point-wise mean and standard deviation of the 5 output grids. Other available summary grids are [min](#), [max](#) and [median](#).

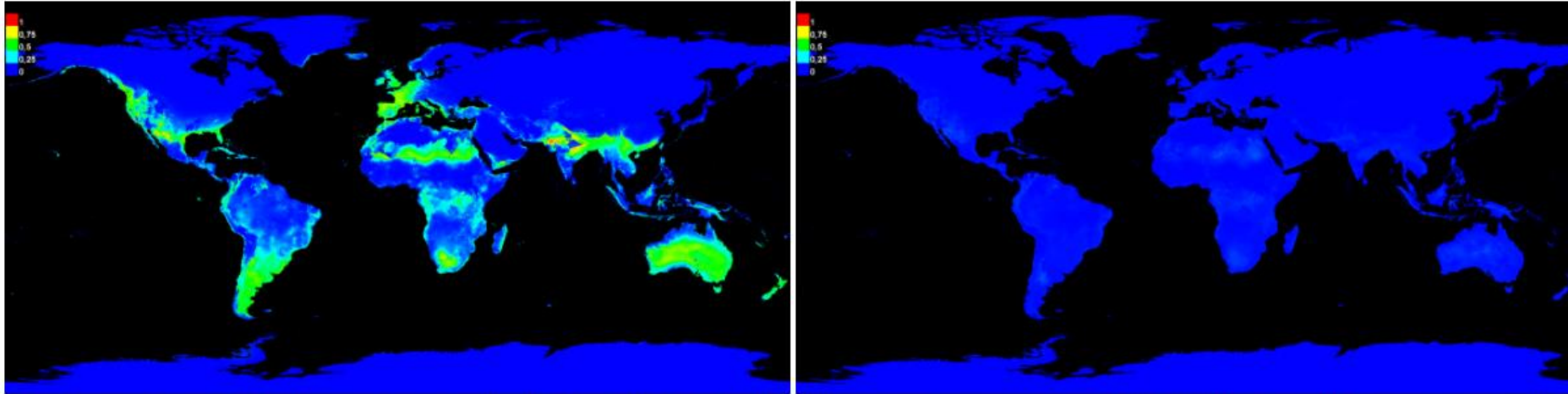


The following two pictures show the point-wise mean and standard deviation of the 5 models applied to the environmental layers in Future. Other available summary grids are [min](#), [max](#) and [median](#).

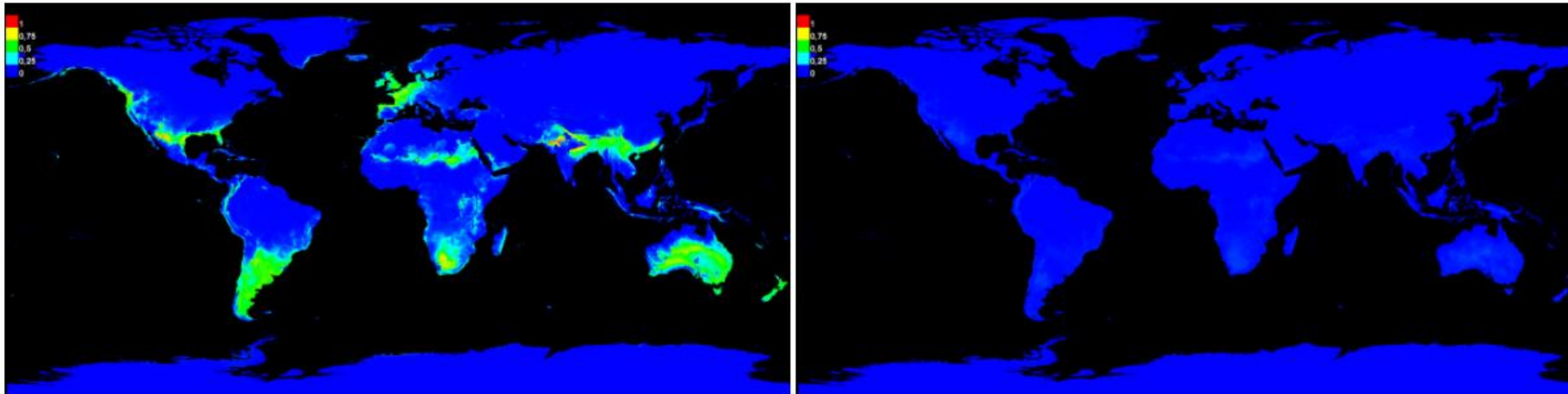


SDM Practical

The following two pictures show the point-wise mean and standard deviation of the 5 models applied to the environmental layers in PresentWLD. Other available summary grids are [min](#), [max](#) and [median](#).

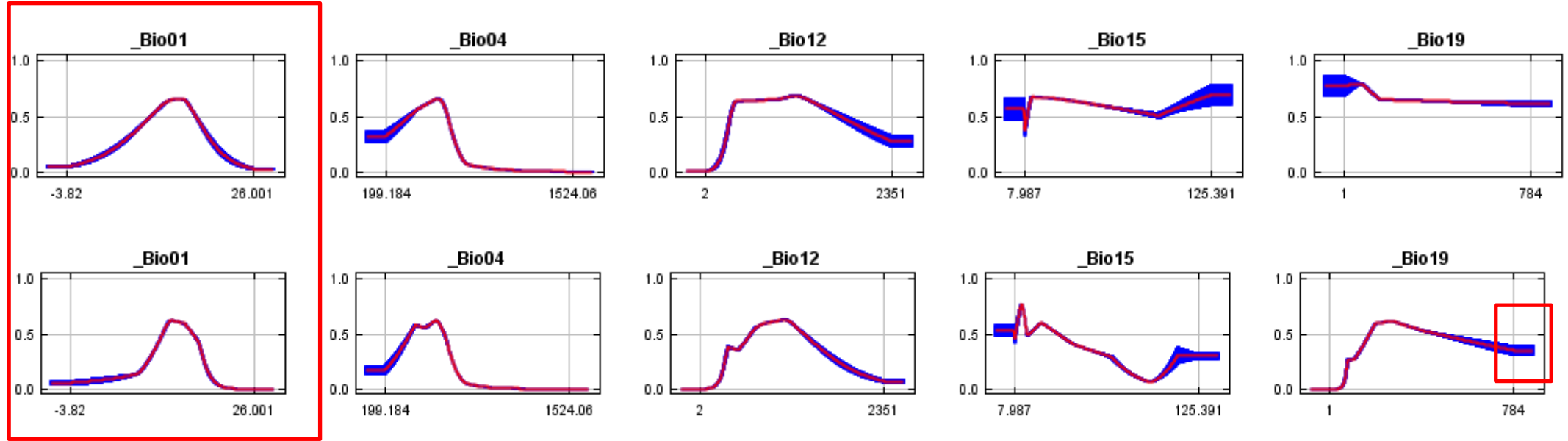


The following two pictures show the point-wise mean and standard deviation of the 5 models applied to the environmental layers in FutureWLD. Other available summary grids are [min](#), [max](#) and [median](#).



SDM Practical

1)



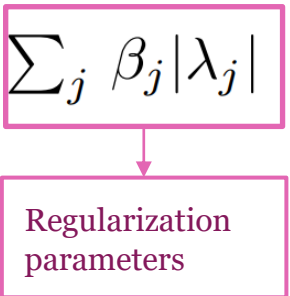
1. P when varying X variable while
keeping all others at the mean value

2. P when using only variable X

- Bio 1: Annual mean temperature
- Bio 4: Temperature seasonality
- Bio 12: Annual precipitation
- Bio 15: Precipitation seasonality
- Bio 19: Precipitation of the coldest quarter

SDM Practical

- Percent contribution:
 - Changes in regularization parameters (PS: bit unclear)
 - Normalized to [0 – 100]%
- Permutation importance:
 - Changes in training AUC by excluding/including the given variable
 - Normalized to [0-100%]
- Take note:
 - % contribution \approx Permutation importance \therefore **Good model**
 - These estimates are **highly affected by autocorrelation between variables**
- Very important for ecologists!
 - In my case, temperature seasonality was most important

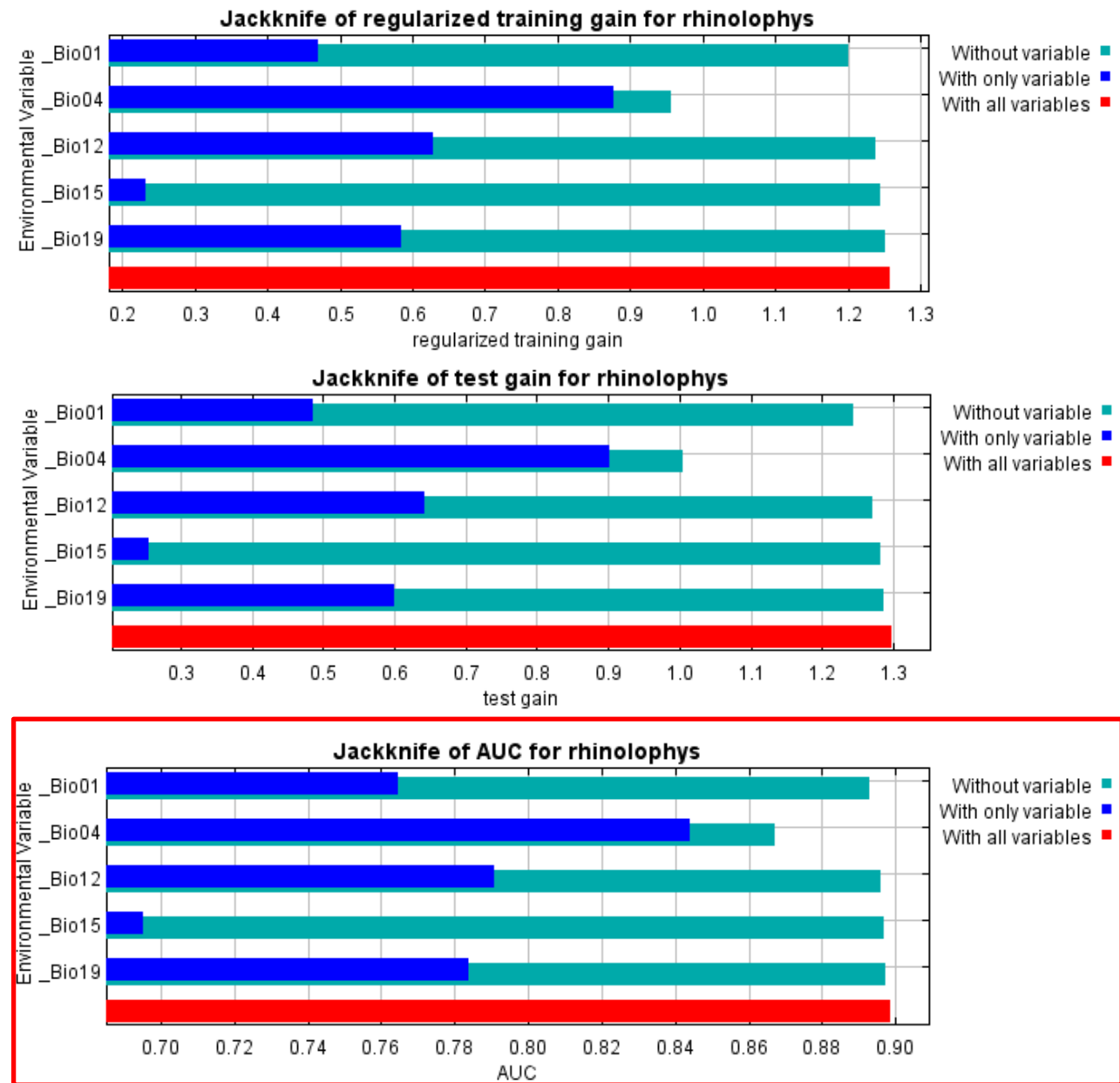
$$\text{RE}(\tilde{\pi} \parallel q_{\lambda}) + \sum_j \beta_j |\lambda_j|$$


Regularization parameters

Variable	Percent contribution	Permutation importance
_Bio04	52.9	50.8
_Bio12	36.3	31.9
_Bio01	6.6	12.9
_Bio15	3.2	2.6
_Bio19	1.1	1.9

SDM Practical

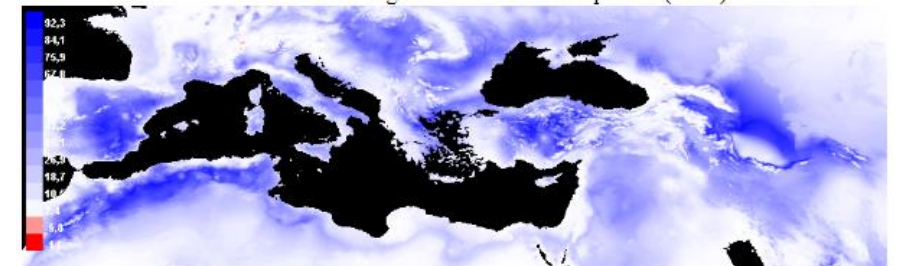
- Three types of jackknife testing:
- First two:
 - Gain in regularization parameters (not clear!)
 - Using training (first) or test data (second)
- **3rd:**
 - Changes in AUC when variable is included
 - Clearest test, focus on this one
- Very similar to Variable importance
 - And should be used to help conclude which variables were most important



SDM Practical

- If you focus on one of the replicates
 - On the top of the page, click on one of the **[nr]**
- Mostly is the same summaries but related only with each specific run
- With the exception of two important sets of information:
 - A table showing different possible thresholds
 - The outcomes of the Multivariate Environmental Similarity Surfaces (MESS) and Most Dissimilar Variables (MOD)

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.028	Fixed cumulative value 1	0.484	0.014	0.020	4.383E-23
5.000	0.107	Fixed cumulative value 5	0.299	0.035	0.031	7.121E-48
10.000	0.241	Fixed cumulative value 10	0.230	0.057	0.082	2.018E-59
0.008	0.000	Minimum training presence	0.861	0.000	0.000	3.531E-5
14.887	0.325	10 percentile training presence	0.194	0.099	0.143	5.351E-62
20.452	0.382	Equal training sensitivity and specificity	0.166	0.165	0.184	0E0
12.176	0.282	Maximum training sensitivity plus specificity	0.212	0.065	0.102	3.58E-62
19.388	0.372	Equal test sensitivity and specificity	0.171	0.151	0.173	0E0
7.651	0.185	Maximum test sensitivity plus specificity	0.254	0.046	0.041	4.677E-58
2.124	0.050	Balance training omission, predicted area and threshold value	0.405	0.023	0.020	2.16E-31
5.769	0.131	Equate entropy of thresholded and original distributions	0.283	0.036	0.031	1.228E-51



SDM Practical

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.028	Fixed cumulative value 1	0.484	0.014	0.020	4.383E-23
5.000	0.107	Fixed cumulative value 5	0.299	0.035	0.031	7.121E-48
10.000	0.241	Fixed cumulative value 10	0.230	0.057	0.082	2.018E-59
0.008	0.000	Minimum training presence	0.861	0.000	0.000	3.531E-5
14.887	0.325	10 percentile training presence	0.194	0.099	0.143	5.351E-62
20.452	0.382	Equal training sensitivity and specificity	0.166	0.165	0.184	0E0
12.176	0.282	Maximum training sensitivity plus specificity	0.212	0.065	0.102	3.58E-62
19.388	0.372	Equal test sensitivity and specificity	0.171	0.151	0.173	0E0
7.651	0.185	Maximum test sensitivity plus specificity	0.254	0.046	0.041	4.677E-58
2.124	0.050	Balance training omission, predicted area and threshold value	0.405	0.023	0.020	2.16E-31
5.769	0.131	Equate entropy of thresholded and original distributions	0.283	0.036	0.031	1.228E-51

- Here are possible different “Thresholds” you can use to define presence/absence
 - If Probability > 0.372 then, consider the species to be present
- The reasoning behind each threshold varies:
 - Recommended: **Equal test sensitivity and specificity** (Generally applicable, but, in some cases might not be reasonable)
 - Also notice the significance value, should be very close to 0 (< .005)

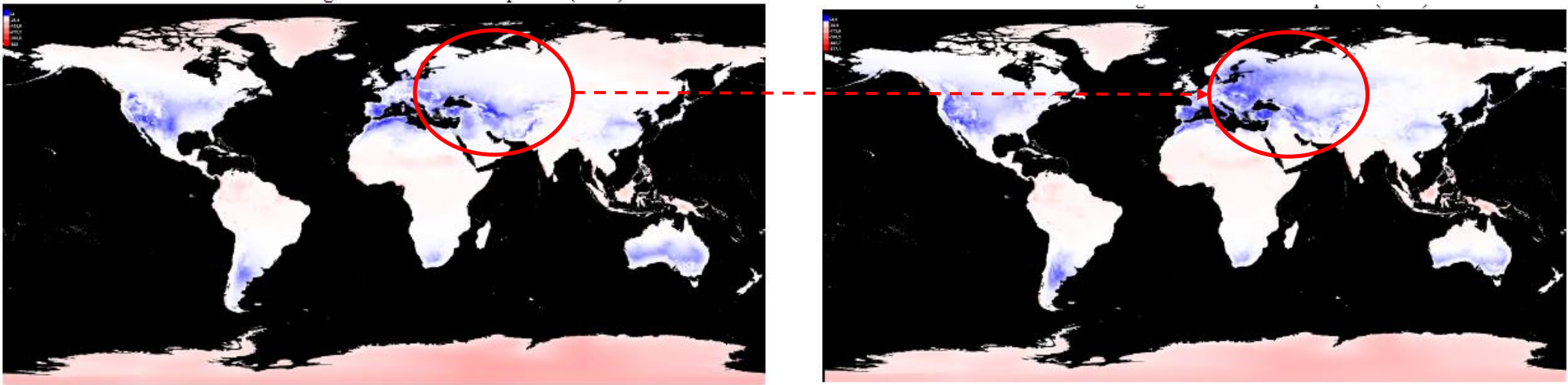
SDM Practical

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.028	Fixed cumulative value 1	0.484	0.014	0.020	4.383E-23
5.000	0.107	Fixed cumulative value 5	0.299	0.035	0.031	7.121E-48
10.000	0.241	Fixed cumulative value 10	0.230	0.057	0.082	2.018E-59
0.008	0.000	Minimum training presence	0.861	0.000	0.000	3.531E-5
14.887	0.325	10 percentile training presence	0.194	0.099	0.143	5.351E-62
20.452	0.382	Equal training sensitivity and specificity	0.166	0.165	0.184	0E0
12.176	0.282	Maximum training sensitivity plus specificity	0.212	0.065	0.102	3.58E-62
19.388	0.372	Equal test sensitivity and specificity	0.171	0.151	0.173	0E0
7.651	0.185	Maximum test sensitivity plus specificity	0.254	0.046	0.041	4.677E-58
2.124	0.050	Balance training omission, predicted area and threshold value	0.405	0.023	0.020	2.16E-31
5.769	0.131	Equate entropy of thresholded and original distributions	0.283	0.036	0.031	1.228E-51

- These results are given “per replicate”, so how do I select the threshold?
 - Go through the various replicates and select the mean value of the threshold.
- 1. Notice, in your report, you can even let us know your mean value +/- sd -> but it's up to you how to find out how to do this

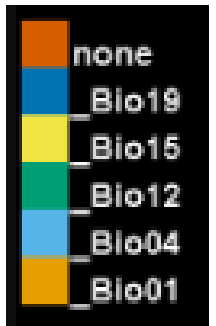
SDM Practical

- Multivariate Environmental Similarity Surfaces (MESS)
 - Show areas where your model is most dissimilar (aka different)
 - The data is also available as a .asc file which you can explore, edit or use in GIS to for example mask regions of high uncertainty



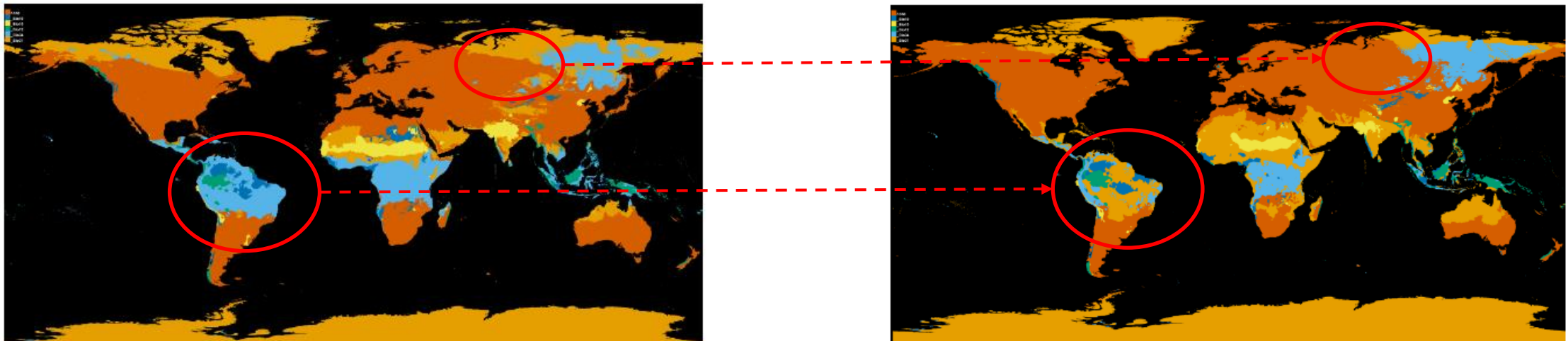
- Similar (Blue) <-> Dissimilar (red)
 - In the “rocky road” scenario, more similar areas seem to be found -> meaning an increase of the available habitat for the species

SDM Practical



- Most Dissimilar Variable (MOD)

- Show areas which specific variable was found to be most dissimilar for every location
- The data is also available as a .asc file which you can explore, edit or use in GIS to for example, to identify which variable is the most limiting or to identify changes in these between the present and the future scenario



- In most ranges and especially in the areas where the habitat was predicted -> there was no significant problem
 - Some changes from None -> Bio 15 (Precipitation seasonality) in the artic regions
 - Various changes in the “amazon” (Bio14 Precipitation -> various others) -> **could it be because of less precipitation due to climate change?**

MAXENT – My outputs



Universiteit
Leiden
The Netherlands

SDM Practical

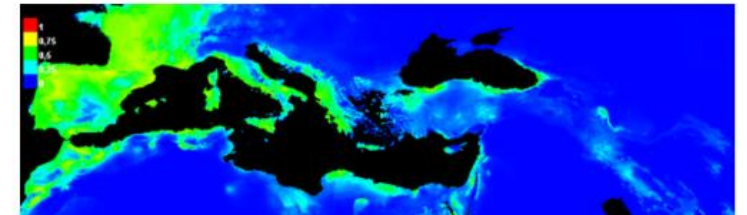
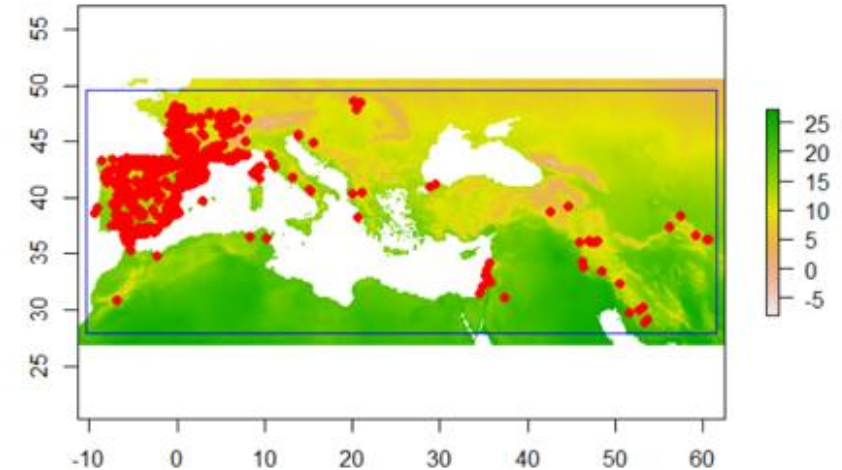
In summary:

1. AUC = .89

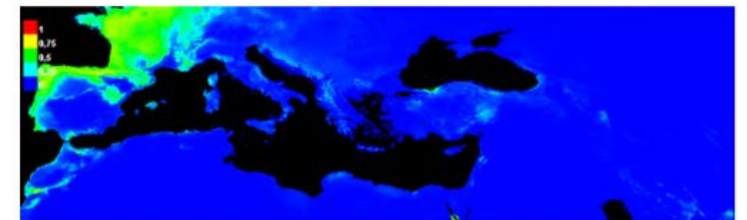
- ~ moderate (I general) -> but remember my species distribution is somewhat specialist
- Can be that I would need more samples in the middle east to have a better picture

2. There are indications of a decrease in habitats in the training region

- Which is problematic for the conservation of the species
- Can be quantified by change maps



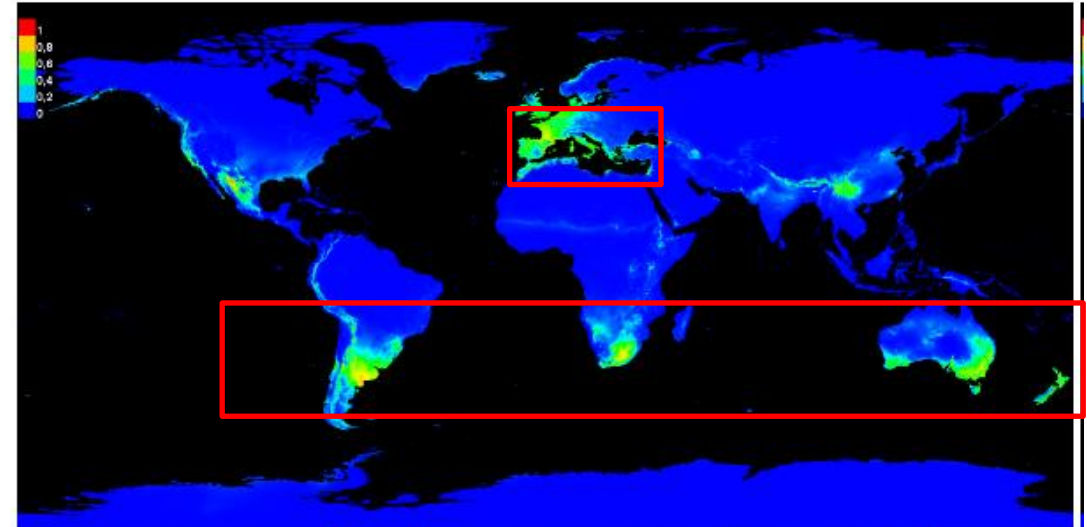
The following two pictures show the point-wise mean and standard deviation of the 5 models



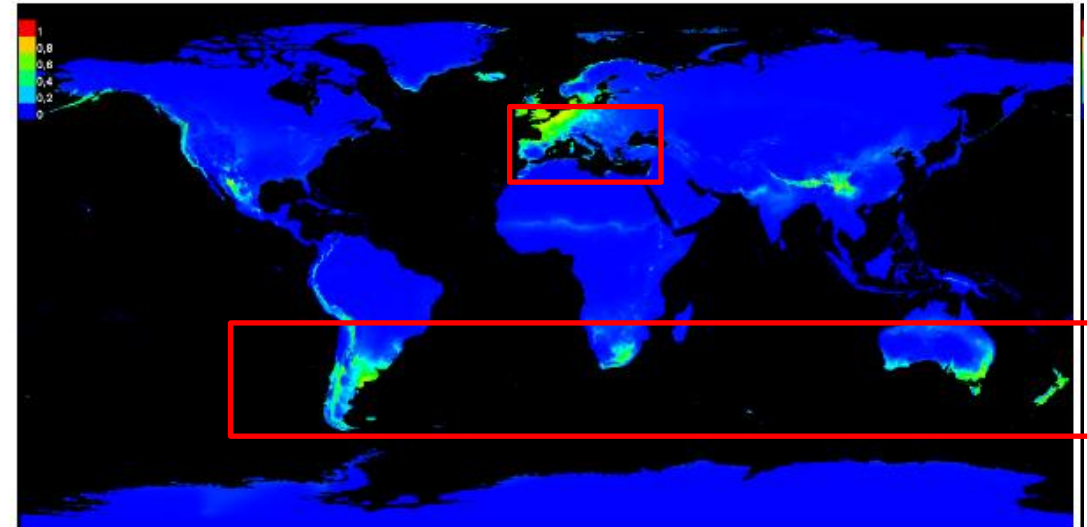
SDM Practical

In summary:

- This indication of habitat loss seems to be in-line with the global distribution
- But further quantification with the change map will confirm (or not) this visual inspection
- Interestingly, there are many regions likely invadable
 - Global south, regions with similar Mediterranean climates
 - California and Mexico/Texas (also similar climates)
 - Near the Tibetan/Chinese border



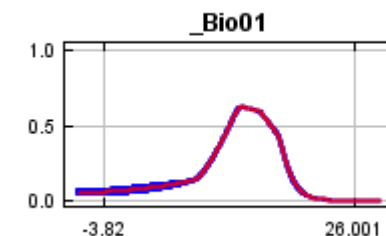
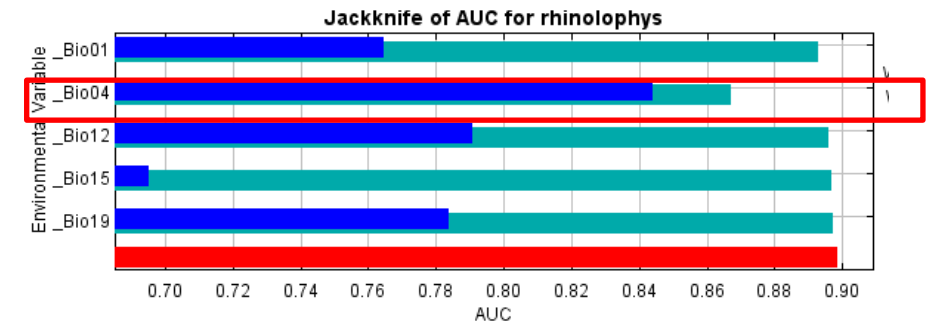
The following two pictures show the point-wise mean and standard deviation of the 10 models



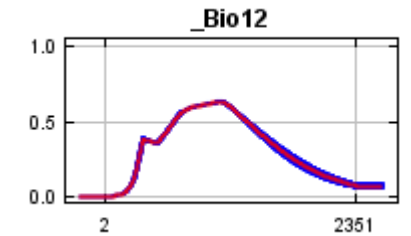
SDM Practical

- Most important variable for the species distribution was:
 - Bio 04 - Temperature seasonality
 - Both on the permutation tests as well on the jackknife testing.
- Other highly significant variable was:
 - Bio 12 – annual precipitation
- If I look into the response variables I can get an idea of the “ideal” ranges of the species:

Variable	Percent contribution	Permutation importance
_Bio04	52.9	50.8
_Bio12	36.3	31.9
_Bio01	6.6	12.9
_Bio15	3.2	2.6
_Bio19	1.1	1.9



- Not “too much” seasonality



- Mid-high precipitation

SDM Practical

- Mediterranean horseshow bat

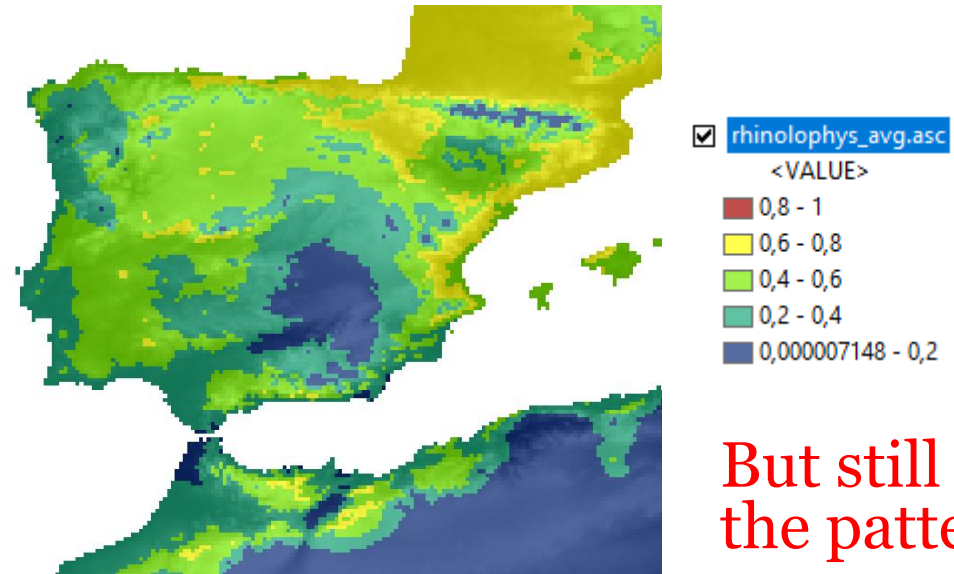
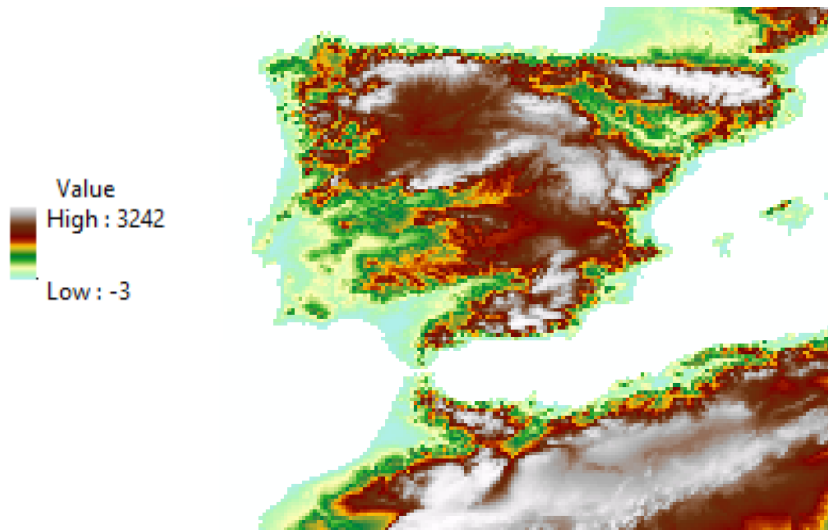
- According to wikipedia:

Habitat [\[edit \]](#)

R. euryale tends to live in warm, wooded areas in foothills and mountains, preferring limestone areas with numerous caves and nearby water. Summer roosts and nurseries are in caves, although sometimes in warm attics in the north. Roosts are frequently shared with other horseshoe bat species, although without any kind of intermingling.



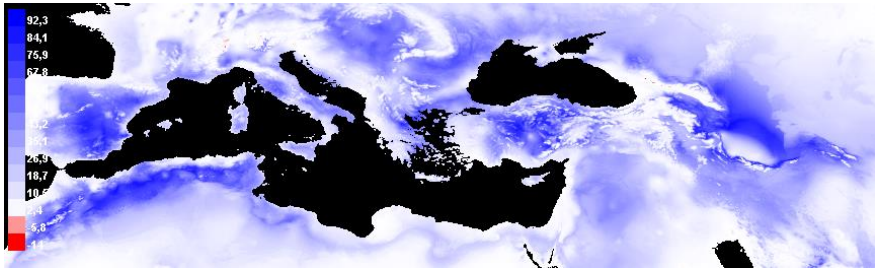
Notice: i did NOT use elevation



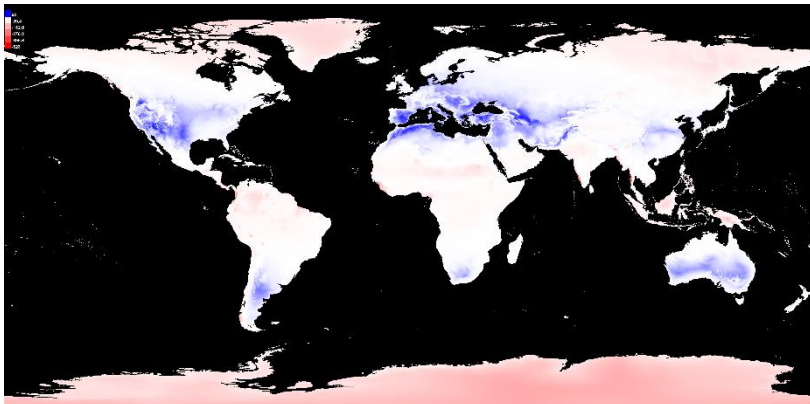
But still “learned” the pattern

SDM Practical

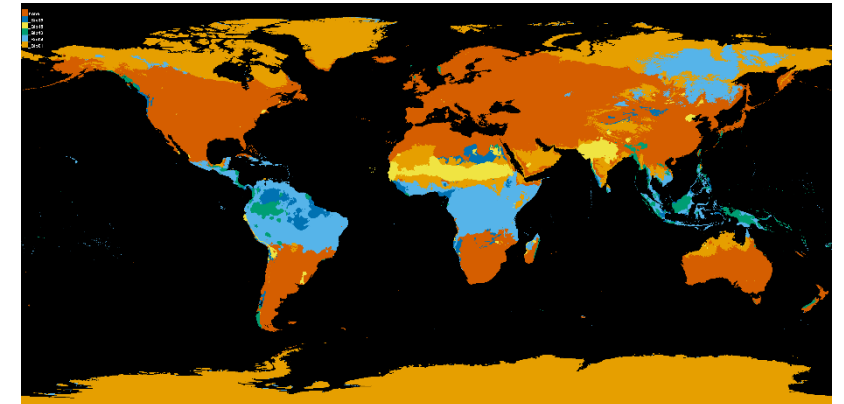
- MESS & MOD in training area:
 - Absolutely no reason not to be confident in my models when predicting to the future (~50years)



- MESS & MOD for the globe (Present):

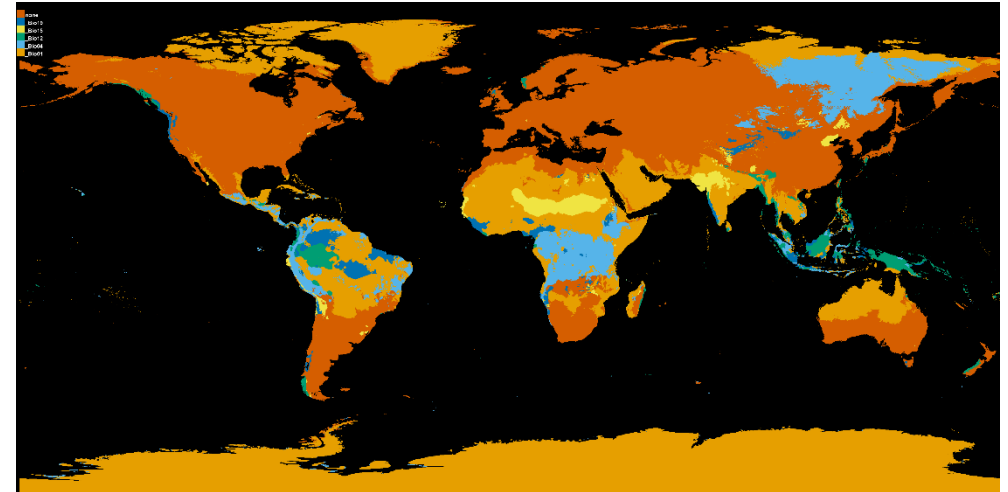
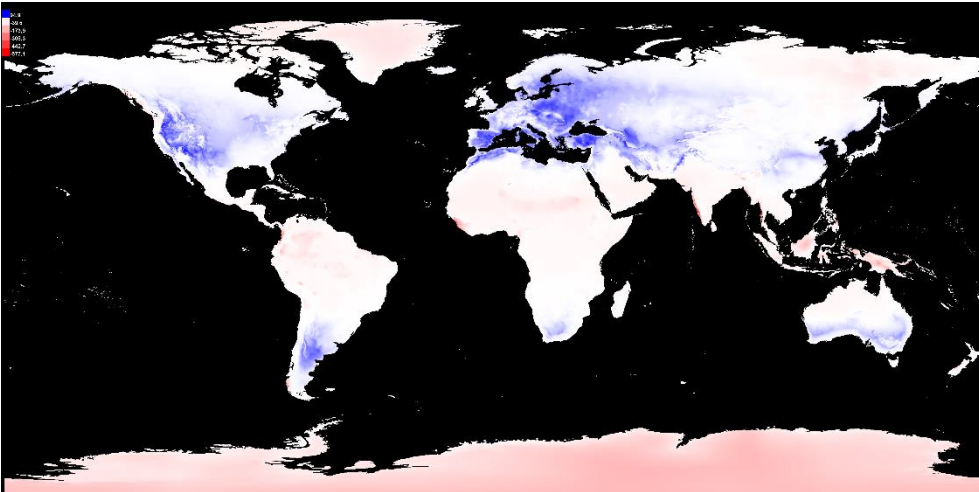


- Some risk in the tropical regions & the poles



SDM Practical

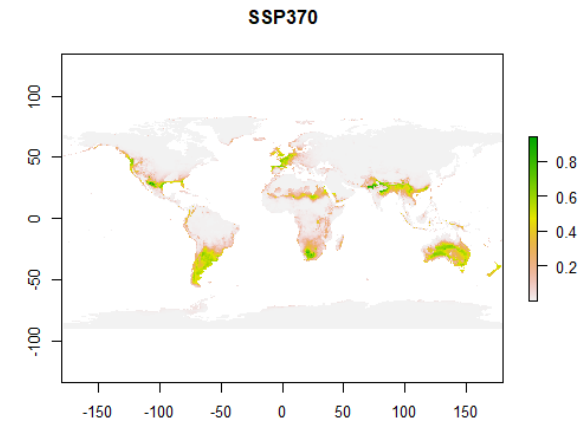
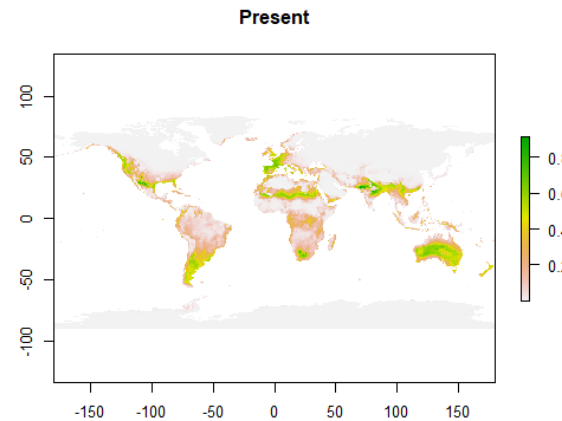
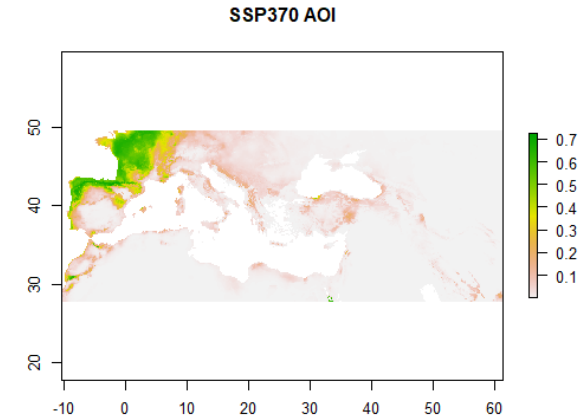
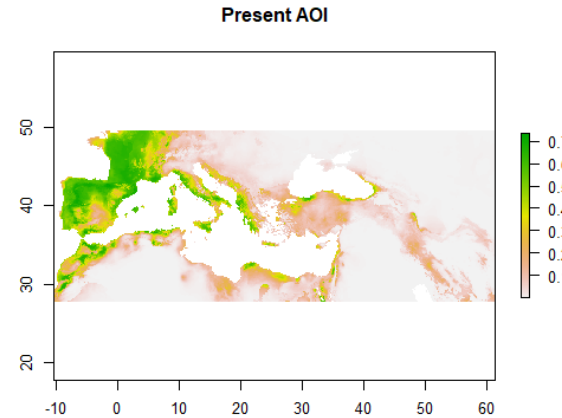
- MESS & MOD for the globe (Future):



- More risk in the tropical regions & the poles
- More risk near the actual current distribution (north of the sahara)

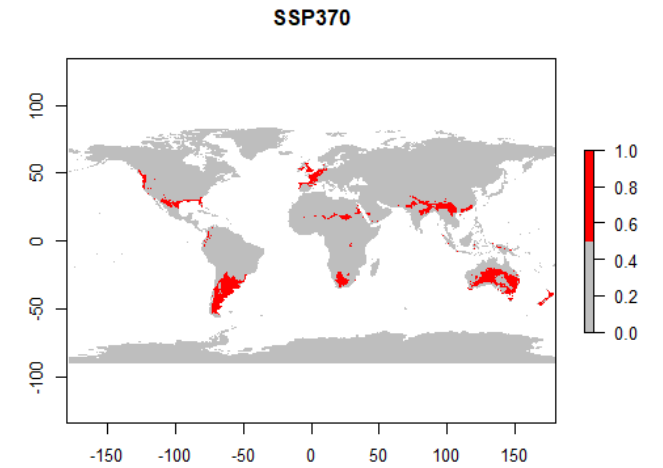
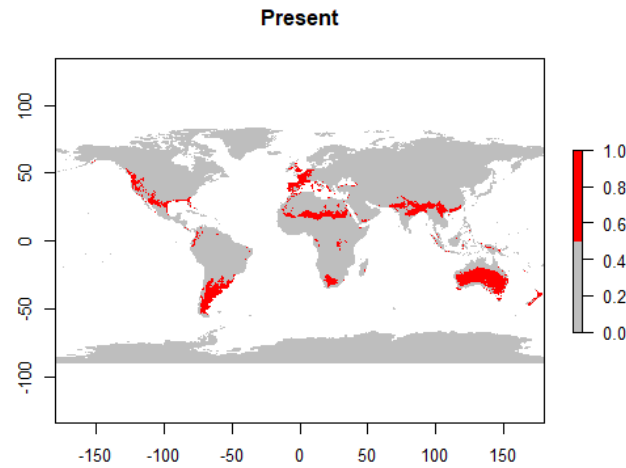
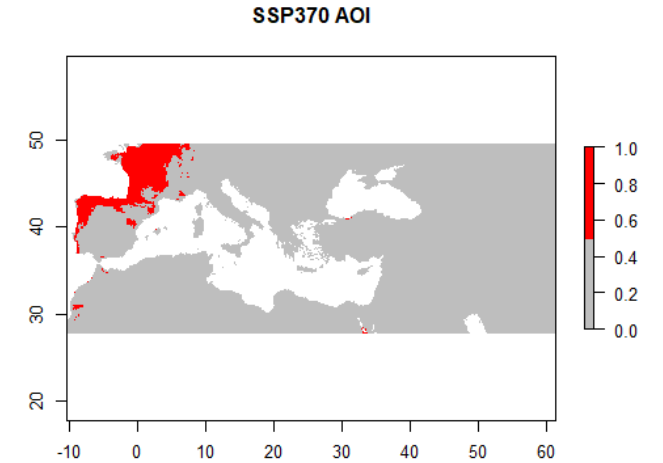
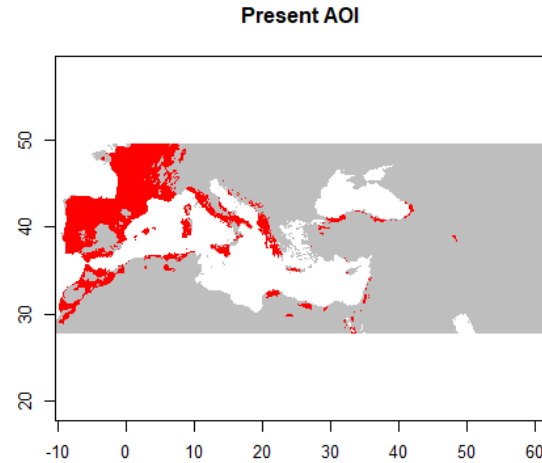
SDM Practical

- Now, using R outputs (ugly maps!)
 - You can do better using GIS
- Probability maps
 - This is equivalent to the maxent results
- We can already see that there are very big differences in probability
 - Habitat suitability seems to displace northwards
 - Inline, with increasing temperatures and changes in precipitation



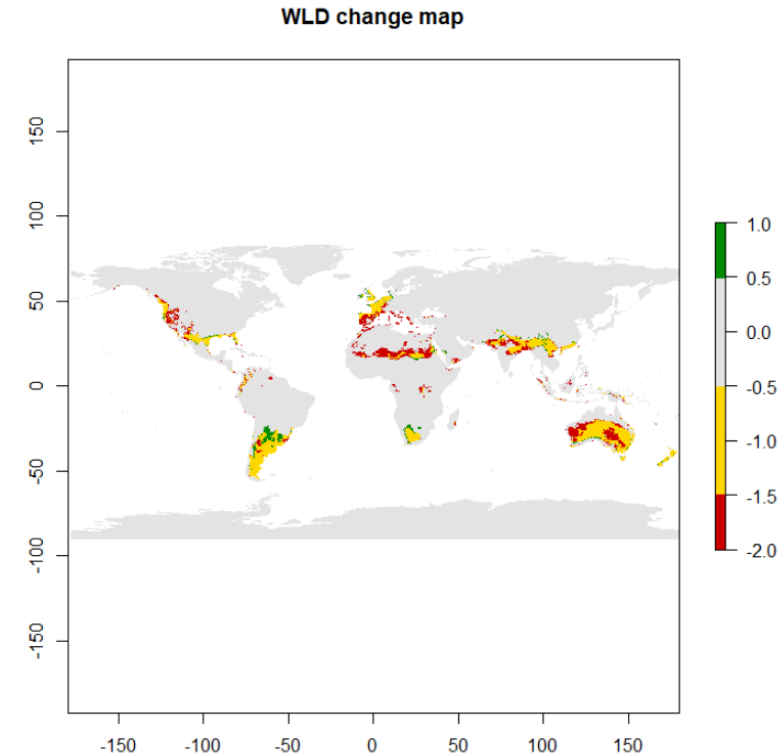
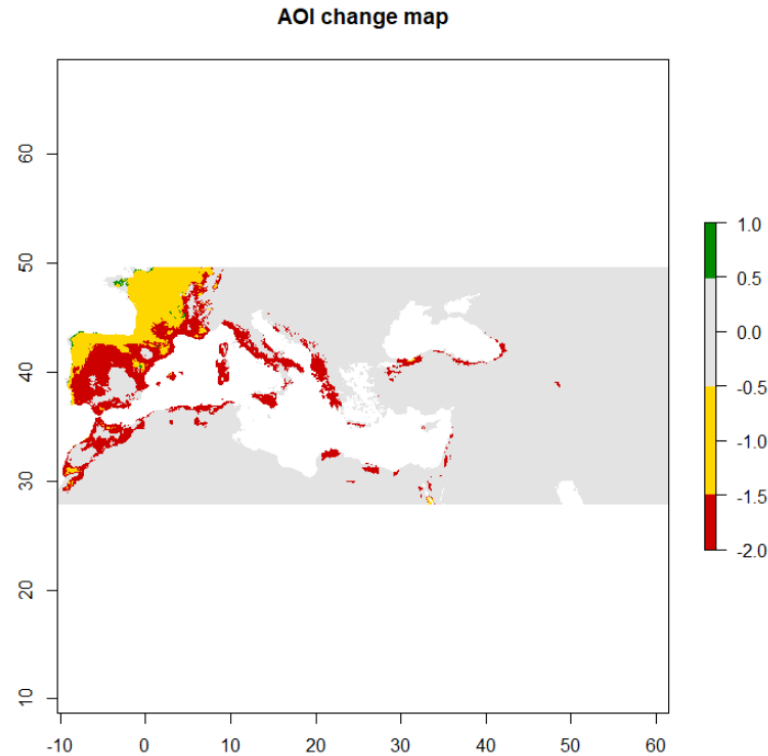
SDM Practical

- Presence/Absence maps
 - But now using the threshold
- More easily visible the displacement of suitable habitat
- Again, you can make nicer maps using GIS
 - And following the 10 golden rules!



SDM Practical

- Legend:
 - **Green:** Gain
 - **Gray:** Maintained (where it was absent)
 - **Yellow:** Maintained (where it was present)
 - **Red:** Loss



SDM Practical

	R variables	Description	PixelCounts	Metrics	Percentage	Formula	R variables
Habitat change	Gain	Gain	23367	Habitat gain	9,743%	Gain/(Disa+Stable1)	PercLoss
	Stable0	Maintained (Absent -> Absent)	2905195				
	Stable1	Maintained (Present -> Present)	151248	Habitat loss	36,939%	Disa/(Disa+Stable1)	PercGain
	Disa	Loss	88596				
	CurrentRangeSize	Current range size	239844	Species range change	-27,196%	Habitat Gain - Habitat loss	SpeciesRangeChange
	FutureRangeSize0Disp	Future range size (no migration)	151248				
	FutureRangeSize1Disp	Future range size (migration)	174615				

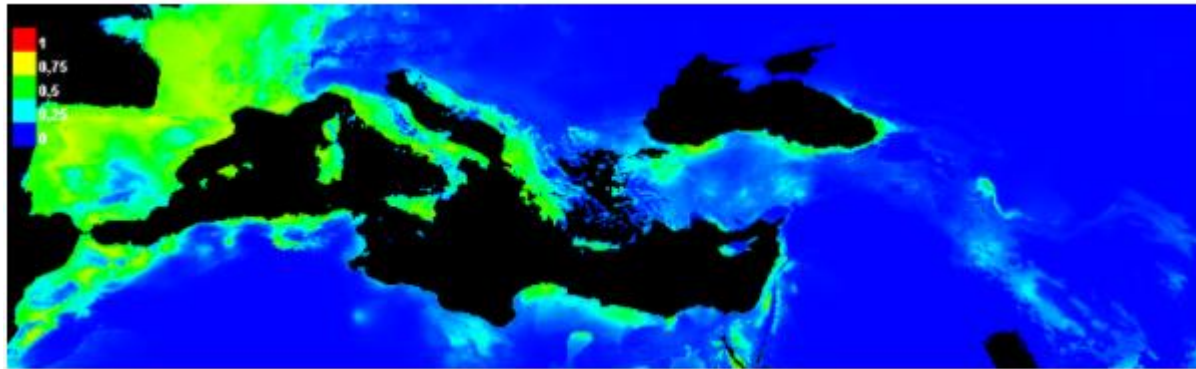
- The biomod function BIOMOD_RangeSize summarizes everything for us
- And thus, its confirmed a loss of ~27% of the habitat

A curve ball:

- Two models trained with different variables
- Extremely similar (almost equal)
- **Why do you think that is?**

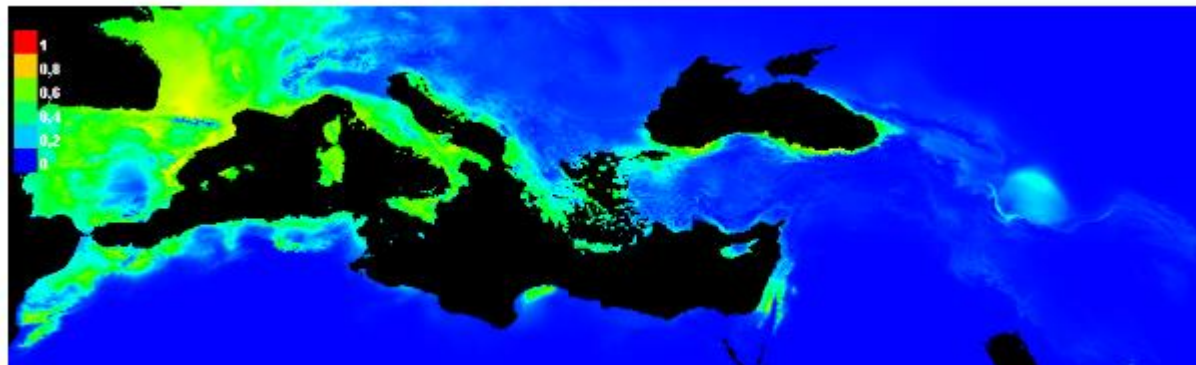


- “Correct” variables



Variable	Percent contribution	Permutation importance
_Bio04	52.9	50.8
_Bio12	36.3	31.9
_Bio01	6.6	12.9
_Bio15	3.2	2.6
_Bio19	1.1	1.9

- “Incorrect” variables



Variable	Percent contribution	Permutation importance
Var_Bio04	61.7	60.6
Var_Bio09	31.9	21.2
Var_Bio02	4	4.7
Var_Bio07	2.4	13.5