

# Running the first ENM/SDM

Emilio Berti

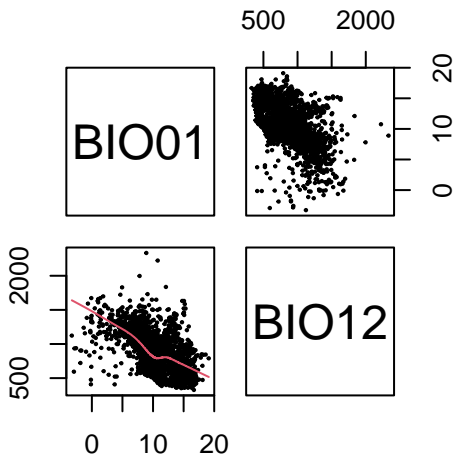
ENM

## Data

```
d <- read.csv("../data/occurrences.csv")  
head(d, n = 3)
```

```
  wc2.1_10m_bio_1 wc2.1_10m_bio_12 wc2.1_10m_bio_13 wc2.1_10m_bio_14  
1      14.709969           510           60  
2      11.454646           775           89  
3       8.854438           933          106  
  wc2.1_10m_bio_15 wc2.1_10m_bio_4 wc2.1_10m_bio_5 wc2.1_10m_bio_6  
1      25.68323           686.3660           32.04650  
2      21.44033           589.9808           25.82825  
0.33375 621014      1  
3      18.08941           578.0707           22.92975  
2.12900 618853      1
```

## Data from previous example



- ▶ Bioclimatic variables are highly correlated with each others.
- ▶ A subset of them should be used for ENM/SDM.

# Selecting variables

## Variable selection

- ▶ The usual statistical tricks ...
- ▶ Informed by the **biology** of the species.

Because I do not know the biology of the species, I will use a statistical trick: Akaike information criterion (AIC).

Average temperature (BI001) and total precipitation (BI012).

```
enm_01_12 <- glm(  
  occ ~ poly(wc2.1_10m_bio_1, 2, raw = TRUE) +  
    poly(wc2.1_10m_bio_12, 2, raw = TRUE),  
  data = d,  
  family = "binomial"  
)
```

Minimum temperature of the coldest month (BI006) and the precipitation of the driest month (BI014).

```
enm_06_14 <- glm(  
  occ ~ poly(wc2.1_10m_bio_6, 2, raw = TRUE) +  
    poly(wc2.1_10m_bio_14, 2, raw = TRUE),  
  data = d,  
  family = "binomial"  
)
```

## ENM-1 vs ENM-2

We can compare the two models by AIC, with the best model having the lowest AIC.

```
AIC(enm_01_12, enm_06_14)
```

	df	AIC
enm_01_12	5	4328.299
enm_06_14	5	4075.476

ENM-2 is more supported than ENM-1.



## Always inspect the inferred niche

Looking at AIC (or other stats) is only one side of the story.

Always **always** inspect the inferred niche.

The coefficients of the quadratic terms for the two ENMs are:

bio_1	bio_12
-1.709950e-02	-4.630913e-06

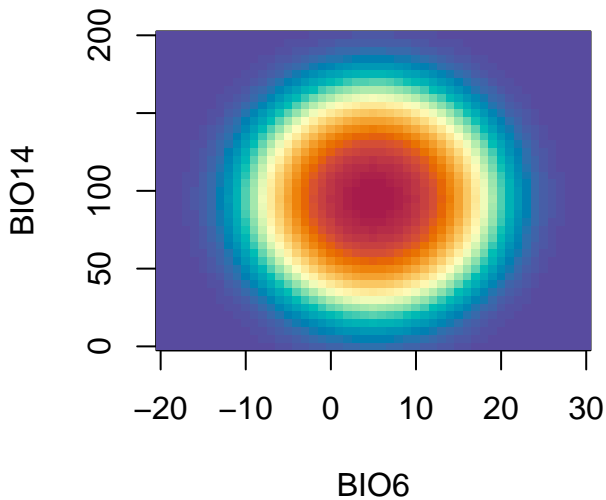
bio_6	bio_14
-0.0133533817	-0.0005297197

ENM-2 is both statistically supported and theoretically valid.

## Inferred niche of ENM-2

```
newd <- expand.grid(
  wc2.1_10m_bio_6 = seq(-20, 30),
  wc2.1_10m_bio_14 = seq(0, 200, by = 5)
)
z <- predict(enm_06_14, newdata = newd, type = "response")
z <- matrix(
  z,
  nrow = length(unique(newd$wc2.1_10m_bio_6)),
  ncol = length(unique(newd$wc2.1_10m_bio_14))
)
image(
  x = sort(unique(newd$wc2.1_10m_bio_6)),
  y = sort(unique(newd$wc2.1_10m_bio_14)),
  z = z,
  col = hcl.colors(100, "Spectral", rev = TRUE),
  xlab = "BI06", ylab = "BI014"
)
```

## Inferred niche of ENM-2



SDM

## Prepare rasters

Load layers.

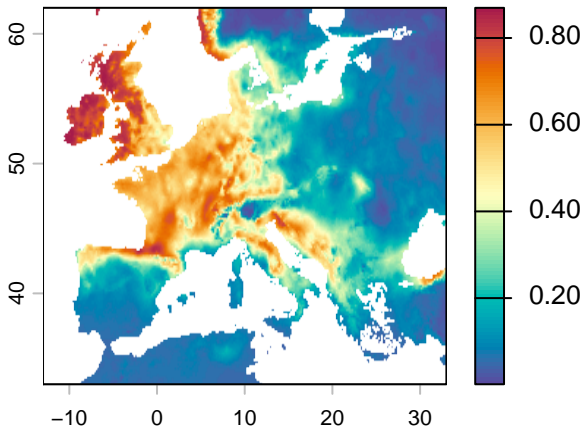
```
library(terra)

ff <- list.files("../data", pattern = ".tif") # all files w
r <- rast(file.path("../", "data", ff))
roi <- ext(-13, 33, 33, 62) # roi of Europe
r <- crop(r, roi) # crop to Europe
```

## Project distribution

```
terra function predict(<raster>, model).
```

```
sdm <- predict(r, enm_06_14, type = "response")  
plot(sdm, col = hcl.colors(100, "Spectral", rev = TRUE))
```



## Binary projection

If we are interested in a binary map, e.g. showing the climatic range of the species, we need to binarize this continuous value into 0/1.

True skill statistics (TSS).

1. Pick a threshold value  $th$ .
2. Set all cells  $< th$  to 0.
3. Set all cells  $\geq th$  to  $w$ .
4. Calculate the TSS.
5. Repeat for a different value of  $th$ .
6. Report maximum TSS and use  $th$  for which  
 $TSS = \max(TSS)$ .

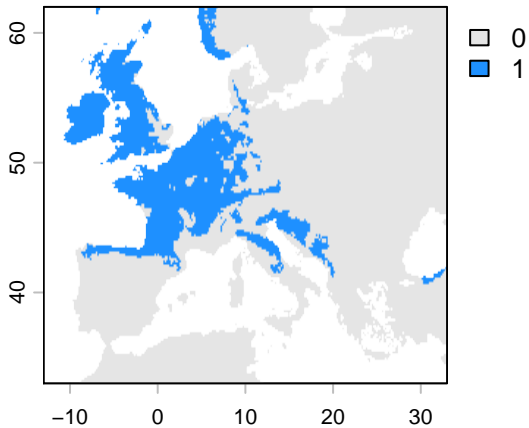
$$TSS = \frac{TP \times TN - FP \times FN}{(TP + FN)(TN + FP)}$$

## Binary projection

```
suit <- extract(sdm, d[, c("x", "y")], ID = FALSE)[, 1]
threshold <- seq(0.1, 0.9, by = 0.001)
tss <- rep(NA, length(threshold)) # empty vector for storage
for (i in seq_along(threshold)) {
  p <- ifelse(suit > threshold[i], 1, 0)
  TP <- sum(p == 1 & d$occ == 1)
  FP <- sum(p == 1 & d$occ == 0)
  FN <- sum(p == 0 & d$occ == 1)
  TN <- sum(p == 0 & d$occ == 0)
  tss[i] <- TP / (TP + FN) + TN / (TN + FP) - 1
}
th <- threshold[which.max(tss)] # best threshold
sdm_bin <- ifel(sdm >= th, 1, 0) # binarize map
```



# Binary projection



## Binary projection

Our map is quite incorrect for this species:

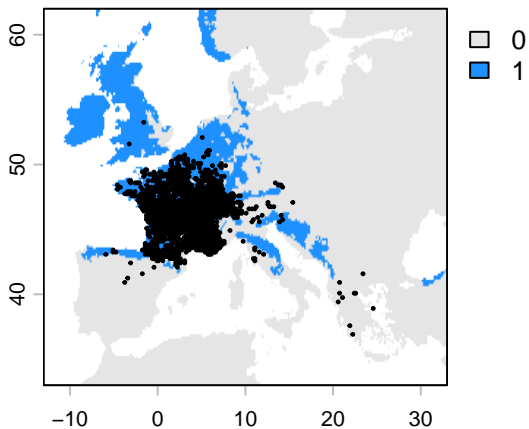
<https://www.iucnredlist.org/species/61550/12514105>

Why do we get such bad projections compared to the known range from IUCN?

We will answer this in a next lecture.

# Binary projection

Plot the detection points.



Geographic biases.