Running the first ENM/SDM

ENM

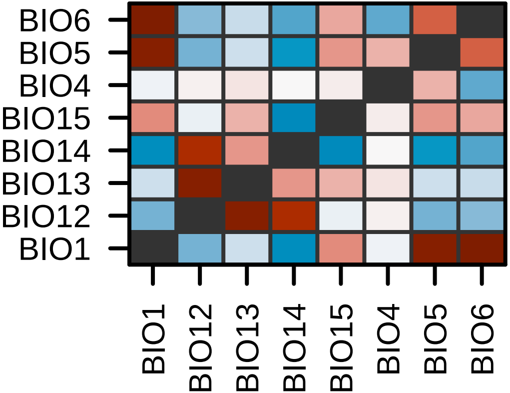
Data

```r
library(mgcv)
library(gratia)

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               20
2       11.454646              775               89               44
3        8.854438              933              106               60
  wc2.1_10m_bio_15 wc2.1_10m_bio_4 wc2.1_10m_bio_5 wc2.1_10m_bio_6   cell
1         25.68323        686.3660        32.04650         2.11375 620998
2         21.44033        589.9808        25.82825        -0.33375 621014
3         18.08941        578.0707        22.92975        -2.12900 618853
          x        y occ
1 -0.4166667 42.08333   1
2  2.2500000 42.08333   1
```

Correlation among variables



- 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.

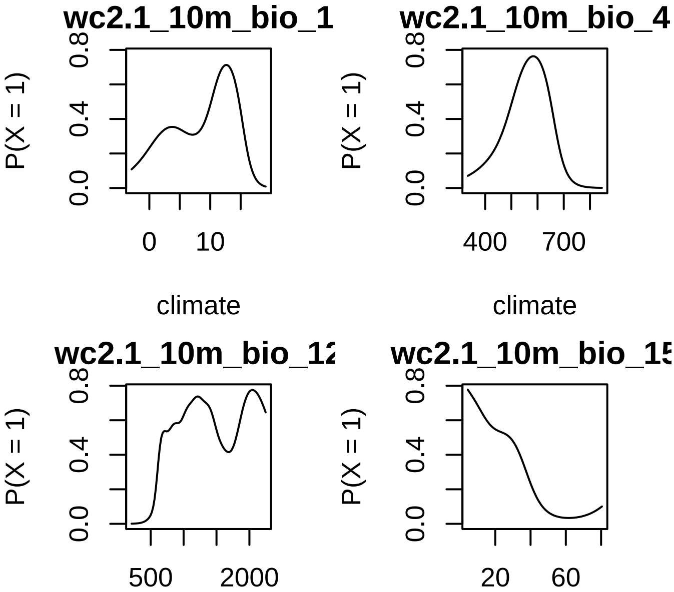
Because I do not know the biology of the species, I will use a statistical trick: Akaike information criterion (AIC).

Full ENM

```r
enm <- gam(
  occ ~ s(wc2.1_10m_bio_1, k = 5) +
    s(wc2.1_10m_bio_4, k = 5) +
    s(wc2.1_10m_bio_12, k = 20) +
    s(wc2.1_10m_bio_15, k = 5),
  data = d,
  family = "binomial"
)
k.check(enm)
```

```
                    k'       edf   k-index p-value
s(wc2.1_10m_bio_1)   4  3.967871 0.9937998   0.360
s(wc2.1_10m_bio_4)   4  3.739801 0.9914376   0.315
s(wc2.1_10m_bio_12) 19 11.160664 0.7681976   0.000
s(wc2.1_10m_bio_15)  4  3.846506 0.9902516   0.285
```

Inspect ENM



wc2.1_10m_bio_1
wc2.1_10m_bio_4
wc2.1_10m_bio_12
wc2.1_10m_bio_15
P(X = 1)
0.0
0.4
0.8
0
10
400
700
500
2000
20
60
climate

SDM

Prepare rasters

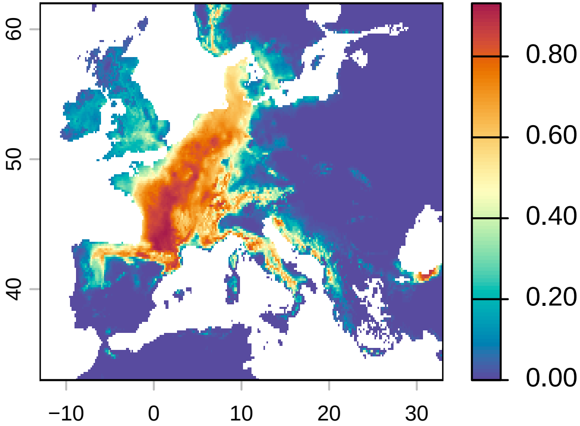
Load layers.

```
library(terra)

ff <- paste0("../data/", unlist(smooth_terms(enm)), ".tif")
r <- rast(ff)
roi <- ext(-13, 33, 33, 62) # roi of Europe
r <- crop(r, roi) # crop to Europe
```

Project distribution

```
sdm <- predict(r, enm, type = "response")
writeRaster(sdm, "../data/hist-sdm.tif", overwrite = TRUE)
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).

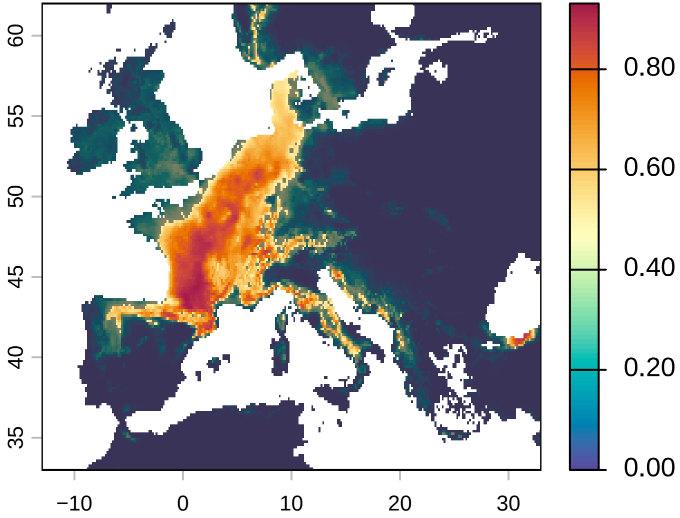
1. Pick a threshold value `th`.
2. Set all cells $< th$ to `0`.
3. Set all cells $\geq th$ to `w`.
4. Calculate the TSS.
5. Repeat for a different value of `th`.
6. Report maximum TSS and use `th` for which $TSS = \max(TSS)$.

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

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.

Hint: geographic biases in the detections