

Defining the modelling region: extent and resolution

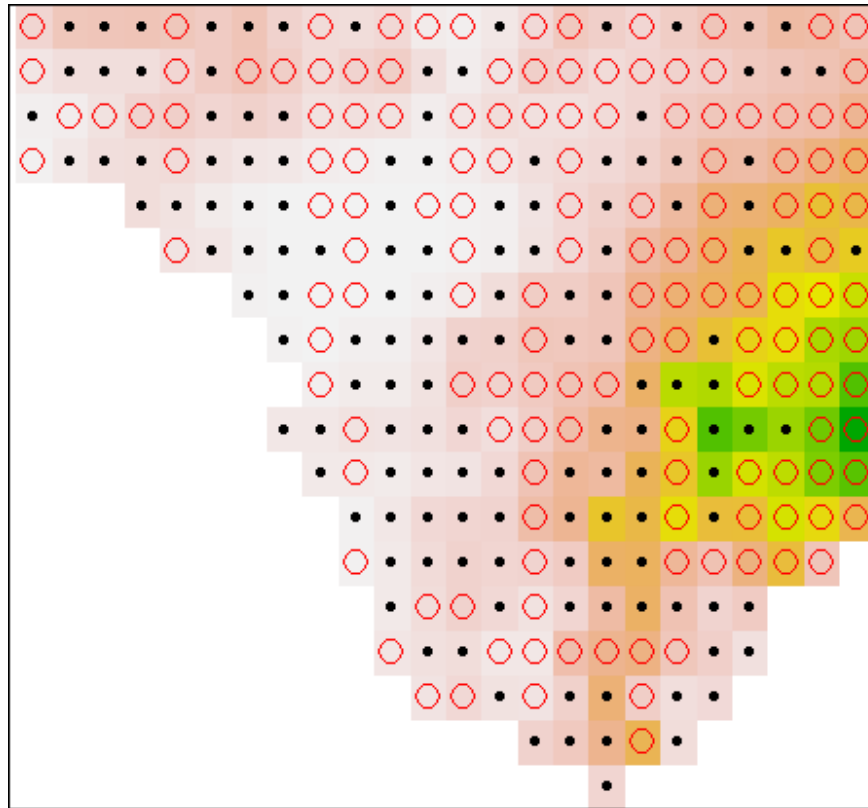
A. Márcia Barbosa

<https://modtools.wordpress.com/barbosa>

Types of modelling methods

- **presence-only**: use environmental values only from presence localities
- **presence / background**: compare presence to available environments across study region
- **presence / (pseudo)absence**: compare presence to "absence" environments

Presence/(pseudo)absence data

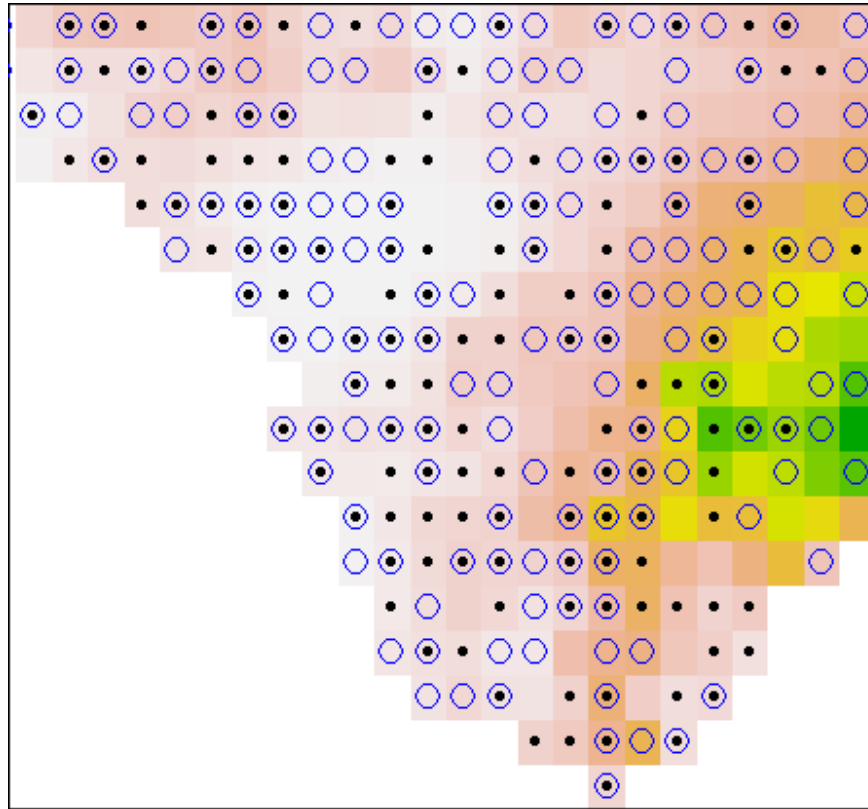


• presence

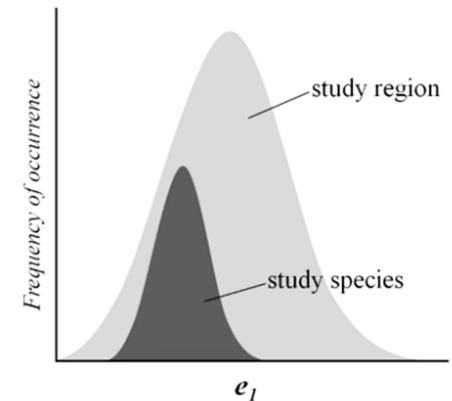
○ absence

- presence/absence of **records**! A surrogate for (sufficiently detectable) presence/absence of the species

Presence/background data

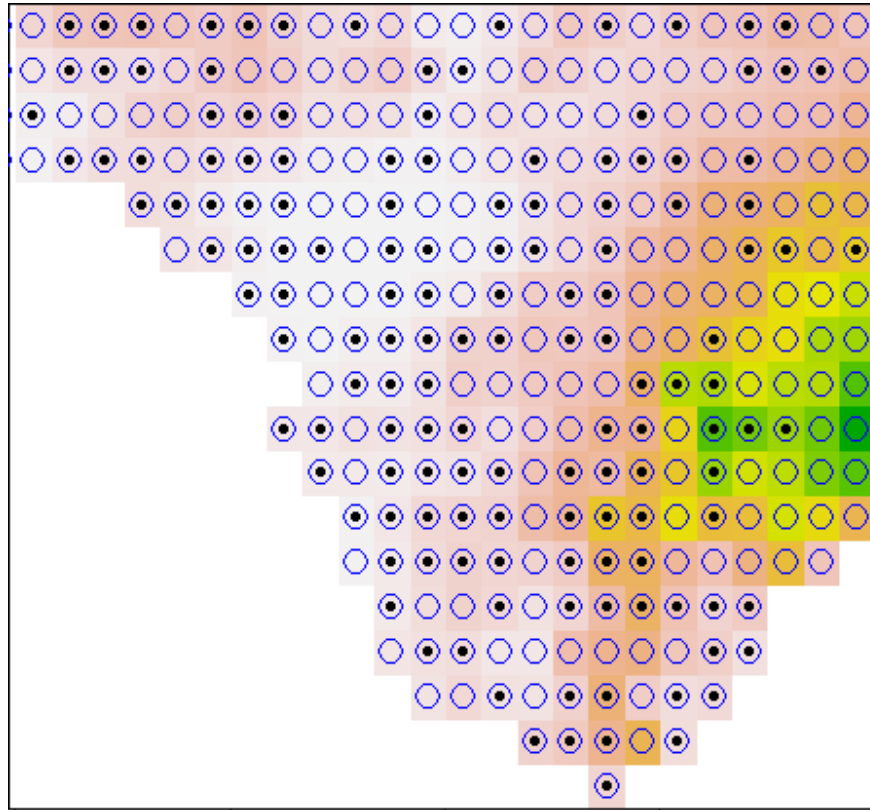


- presence
- background

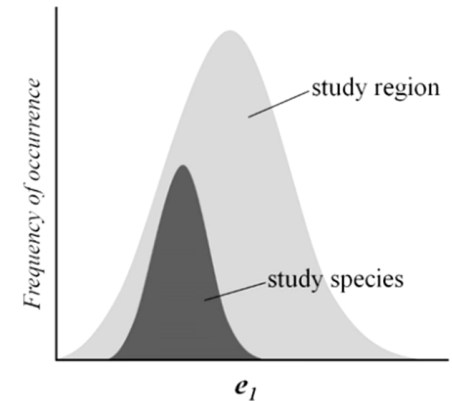


- background = (sample of) available environments
- presence = existing occurrence **records**

Presence/background data

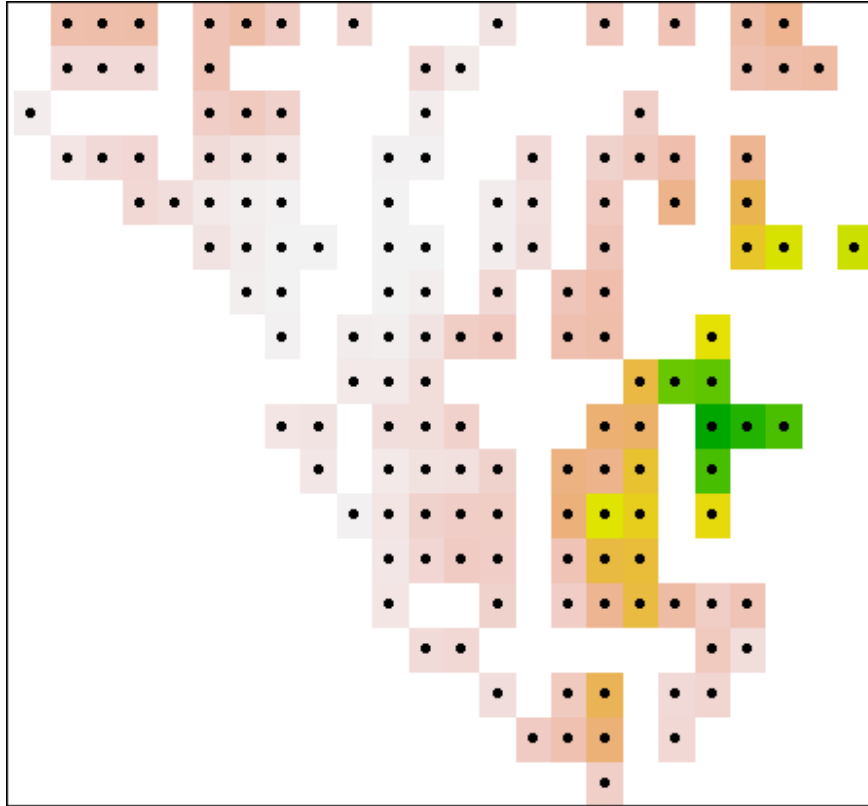


- presence
- background



- background = (sample of) available environments
- presence = existing occurrence **records**

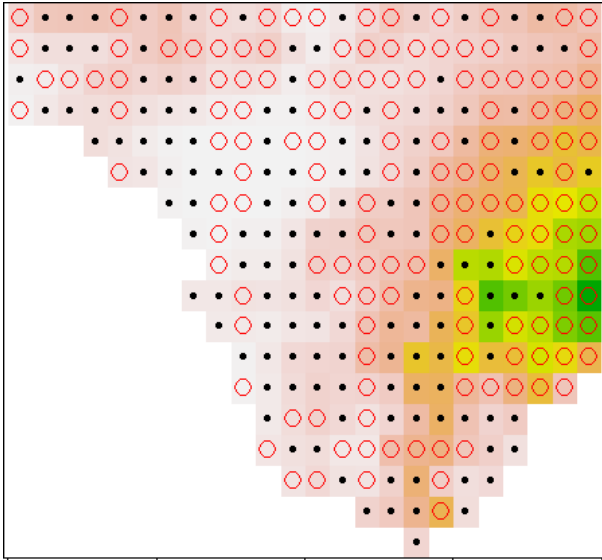
Presence-only data



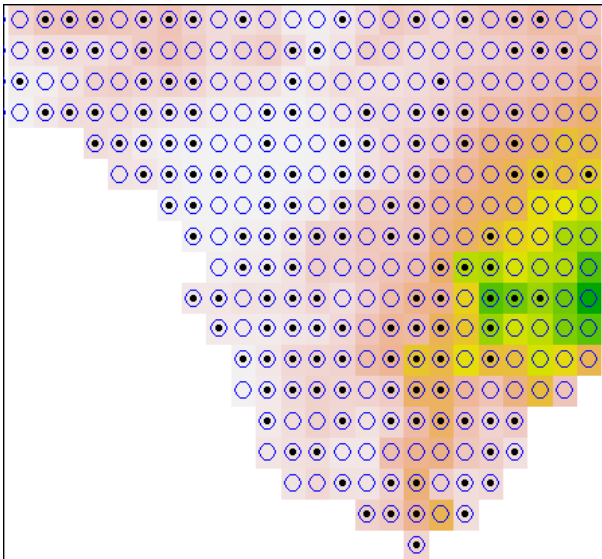
• presence

- presence **records** = environments where a species can live
- useful for very incomplete presence datasets

Presence/absence vs presence/background



presence-absence methods
penalize absences more than
presence-background methods



but the latter also consider the
unoccupied part of the
background, and are also
affected by missed presences

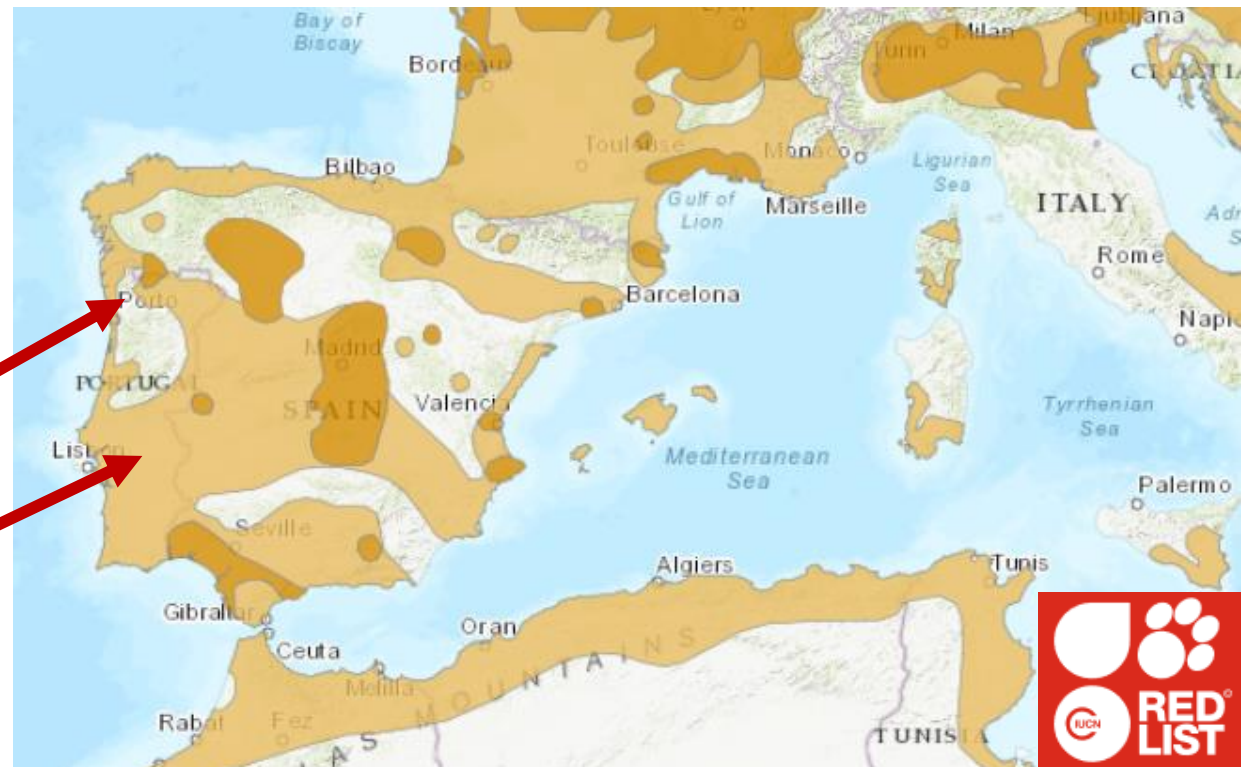
Species occurrence data

Northern lapwing (*Vanellus vanellus*)

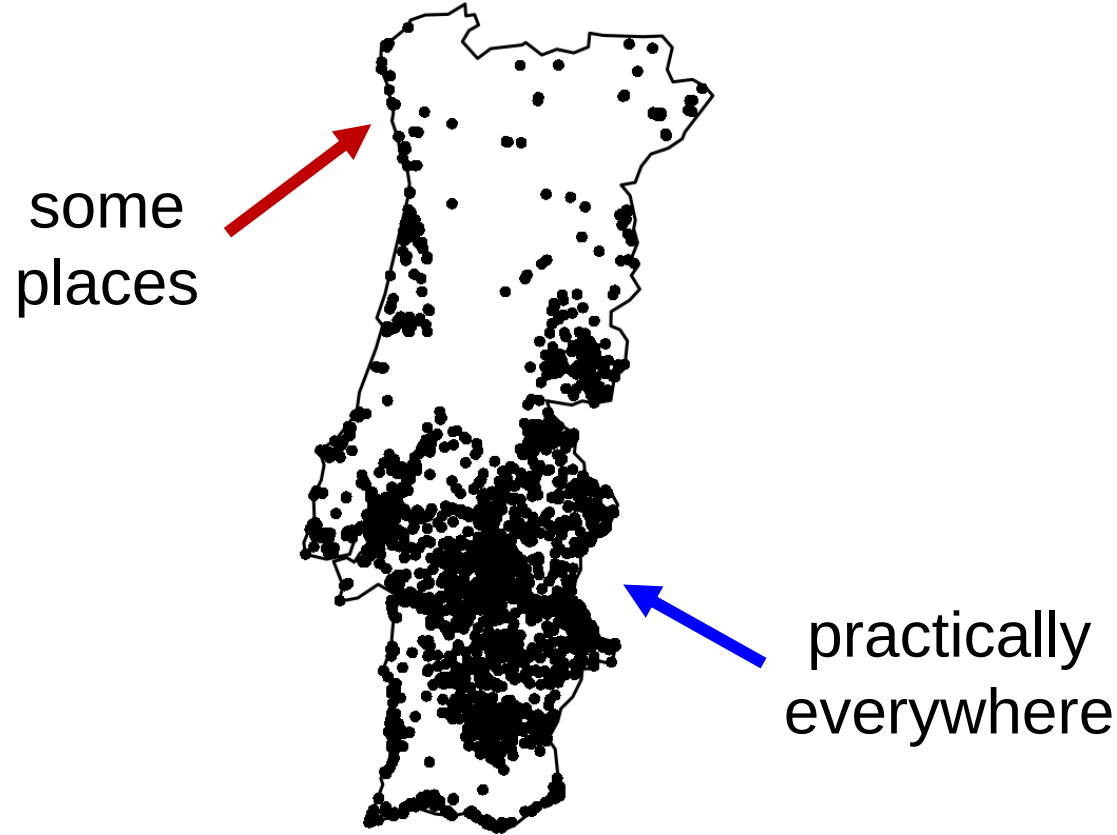


some winters

every winter



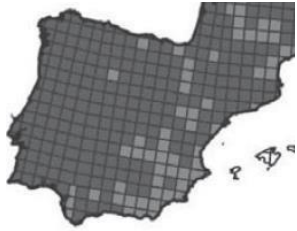
Species occurrence data



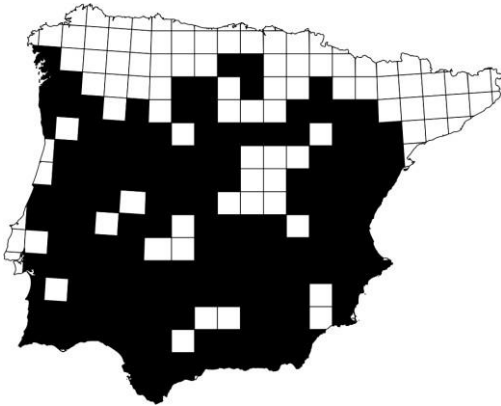
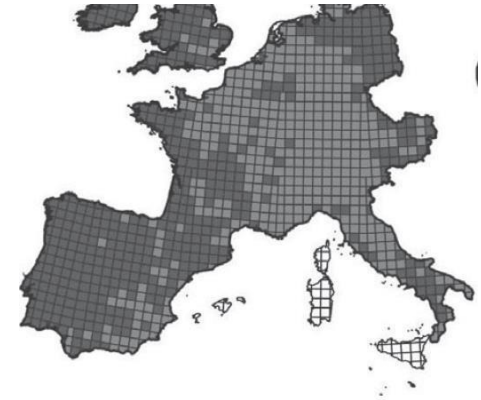
imperfect detection: it's OK!

as long as it reflects **lower occurrence frequency**
rather than **bias** in survey effort

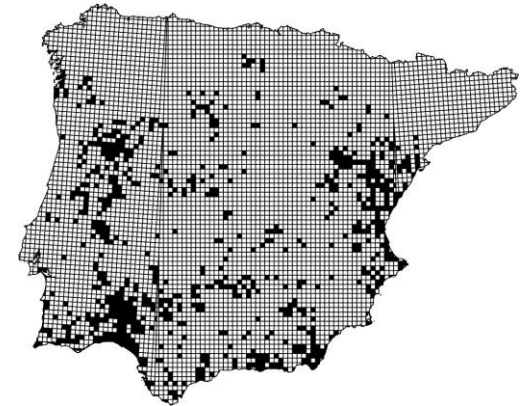
Extent and resolution



increasing **extent**



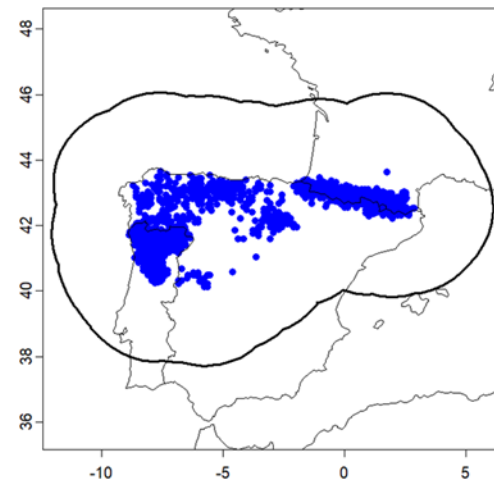
increasing **resolution**



both affect the model outcome!

Defining the modelling extent

- dispersal capacity
- available data (surveyed areas)
 - countries / regions with data
 - countries that appear fairly surveyed
 - minimum convex polygon
 - buffer (to follow survey bias)
 - other methods



Journal of Biogeography (J. Biogeogr.) (2012) **39**, 1383–1390

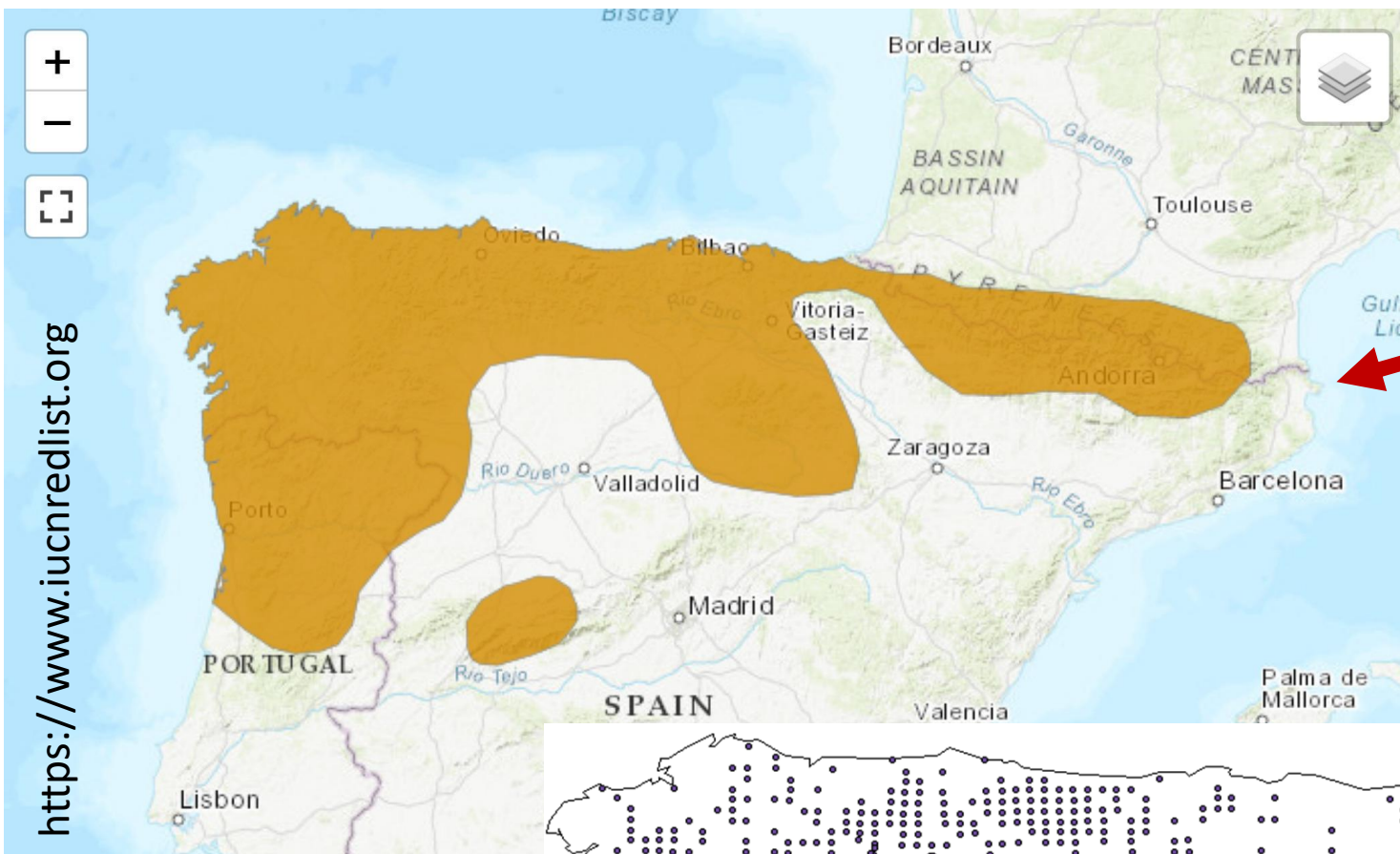


ORIGINAL
ARTICLE

Delimiting the geographical background in species distribution modelling

Pelayo Acevedo^{1,2,3*}, Alberto Jiménez-Valverde¹, Jorge M. Lobo⁴
and Raimundo Real¹

Defining the modelling extent



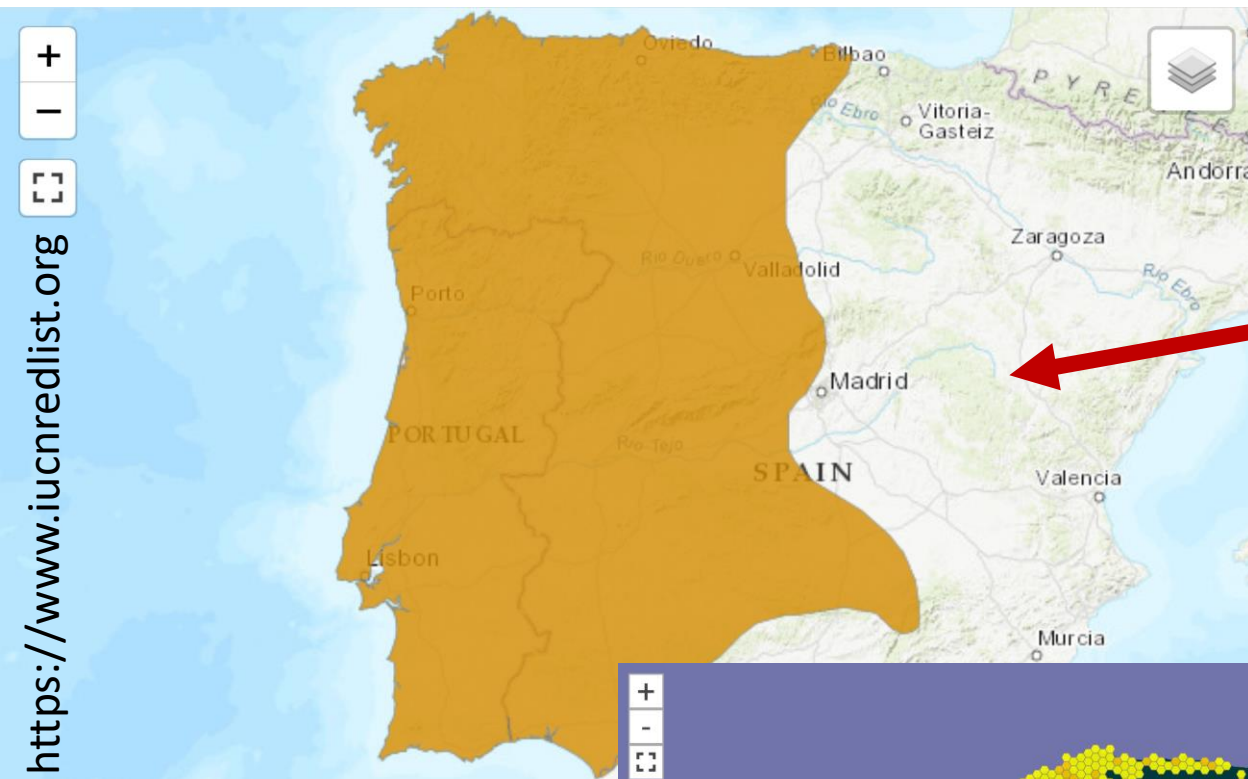
global
distribution
range

data missing from
part of the range



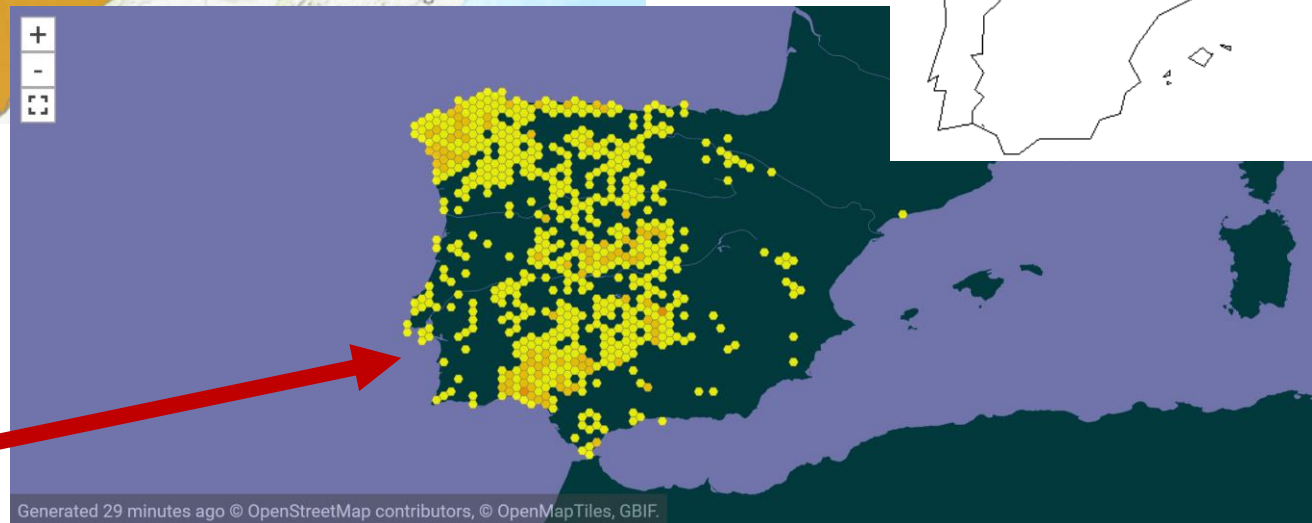
<https://www.gbif.org> (some time ago)

Defining the modelling extent



global
distribution
range

data scarcer in
part of the range;
administrative
border discernible
in the data



<https://www.gbif.org>

Defining the modelling extent

- **survey bias can strongly affect models!**
- regions that were clearly **under-surveyed** may mislead the model, so should be **left out** of both "**absence**" and "**background**" areas for model training
- choose a region that captures most of the species-environment relationships, but where survey was reasonably well distributed
- for real work, search different sources to **complete your dataset**

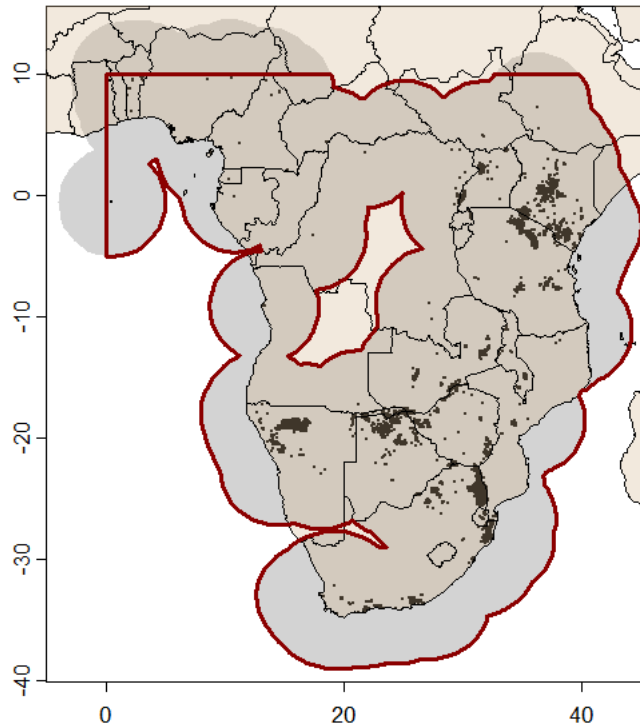
Delimit the study area: several options

```
# define a buffer around presence points:
```

```
pres_buff <- getRegion(pres_points, type = "width") # but explore  
other options!
```

```
mod_region <- crop(pres_buff, download_window)
```

(if you did use a limited
window to download
species data)



Delimit the study area: several options

```
# include / exclude particular polygons in the modelling  
area:
```

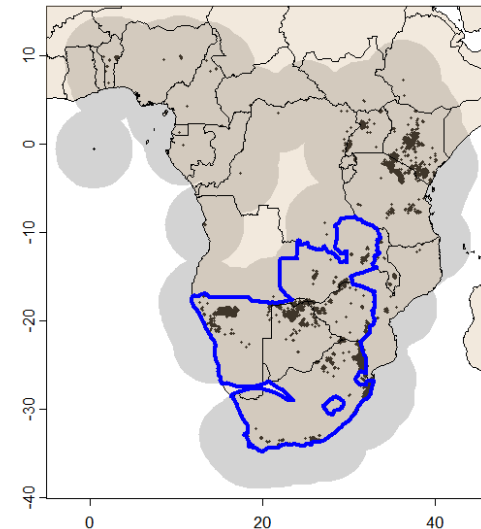
```
countries_with_points <- countries[presences, ] # terrestrial only!
```

```
include <- c("Namibia", "Botswana", "South Africa", "Zimbabwe",  
"Zambia")
```

```
mod_region <- crop(mod_region, subset(countries_with_points,  
countries_with_points$NAME %in% include))
```

```
exclude <- c("Angola", "Mozambique")
```

```
mod_region <- crop(subset(countries, !(countries$NAME  
%in% exclude)), mod_region)
```



Defining the modelling resolution

Depends on:

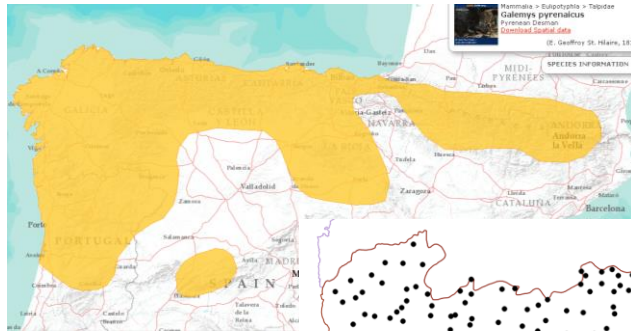
- modelling aims (scale of analysis)
- species' ecology / mobility
- available data

At overly fine resolution, behavioural interactions (territoriality, social attraction) can override environmental affinity.

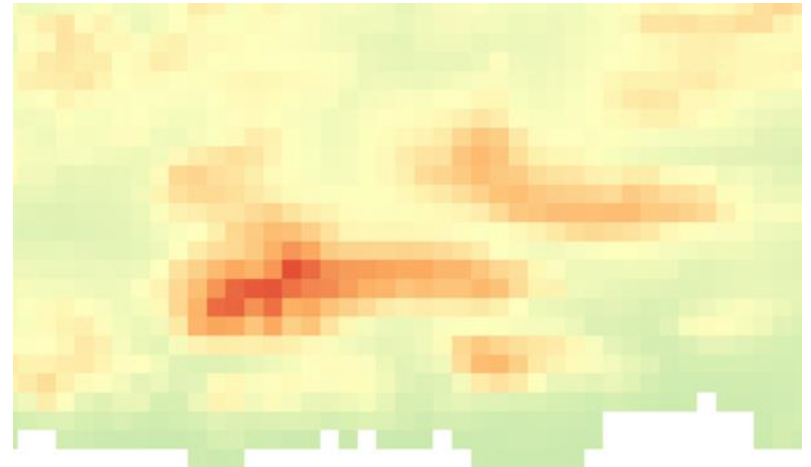
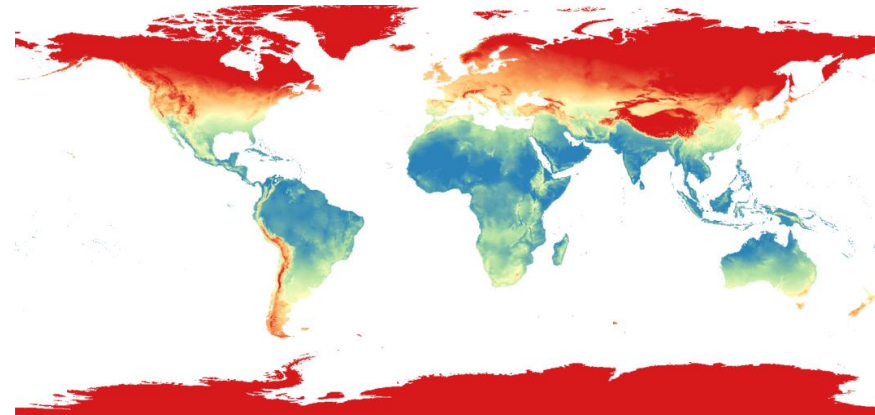
If occurrence data are at a coarse resolution, predictor variables should be rescaled / aggregated accordingly!

Species occurrence + predictor variables

- range maps
- point occurrences
- gridded data



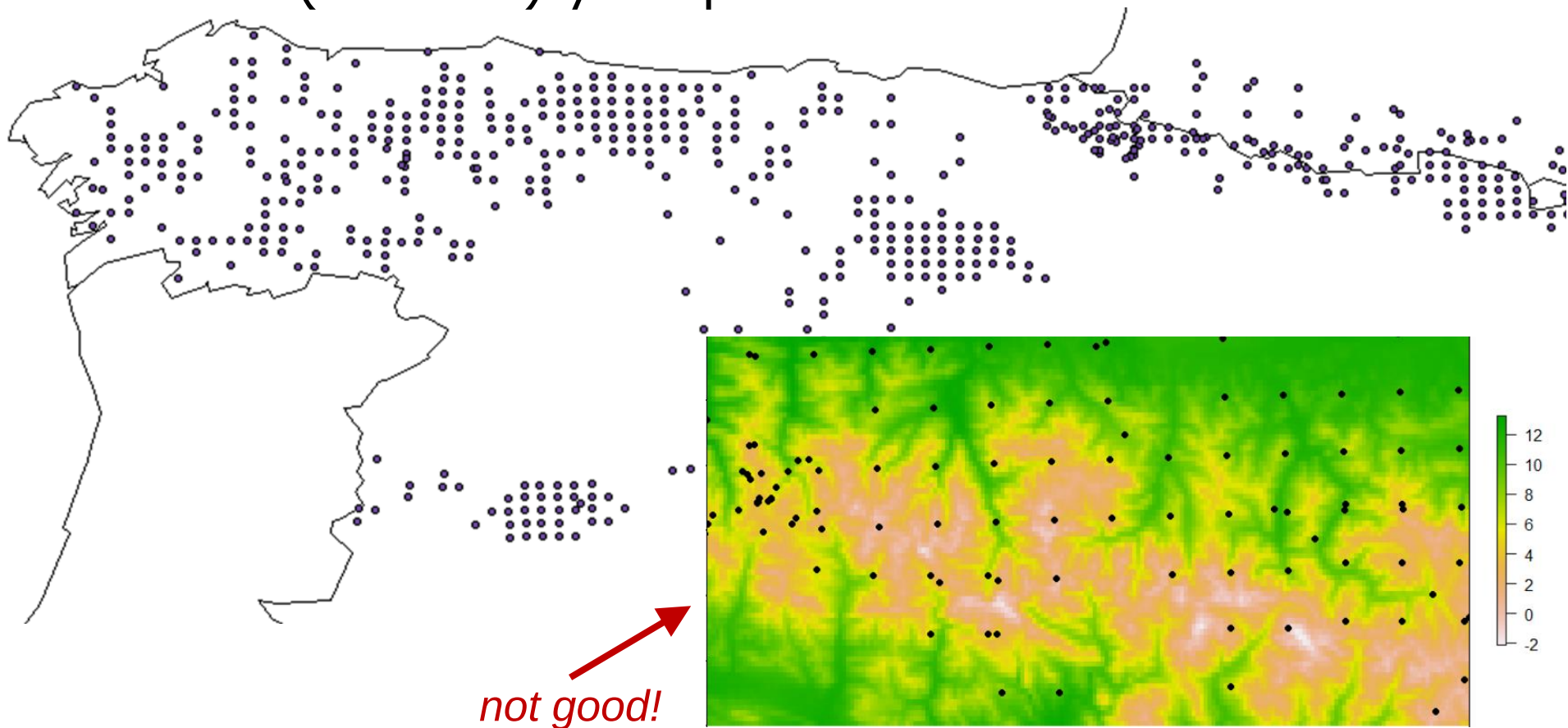
gridded data



Species occurrence + predictor variables

Spatial resolution should match!

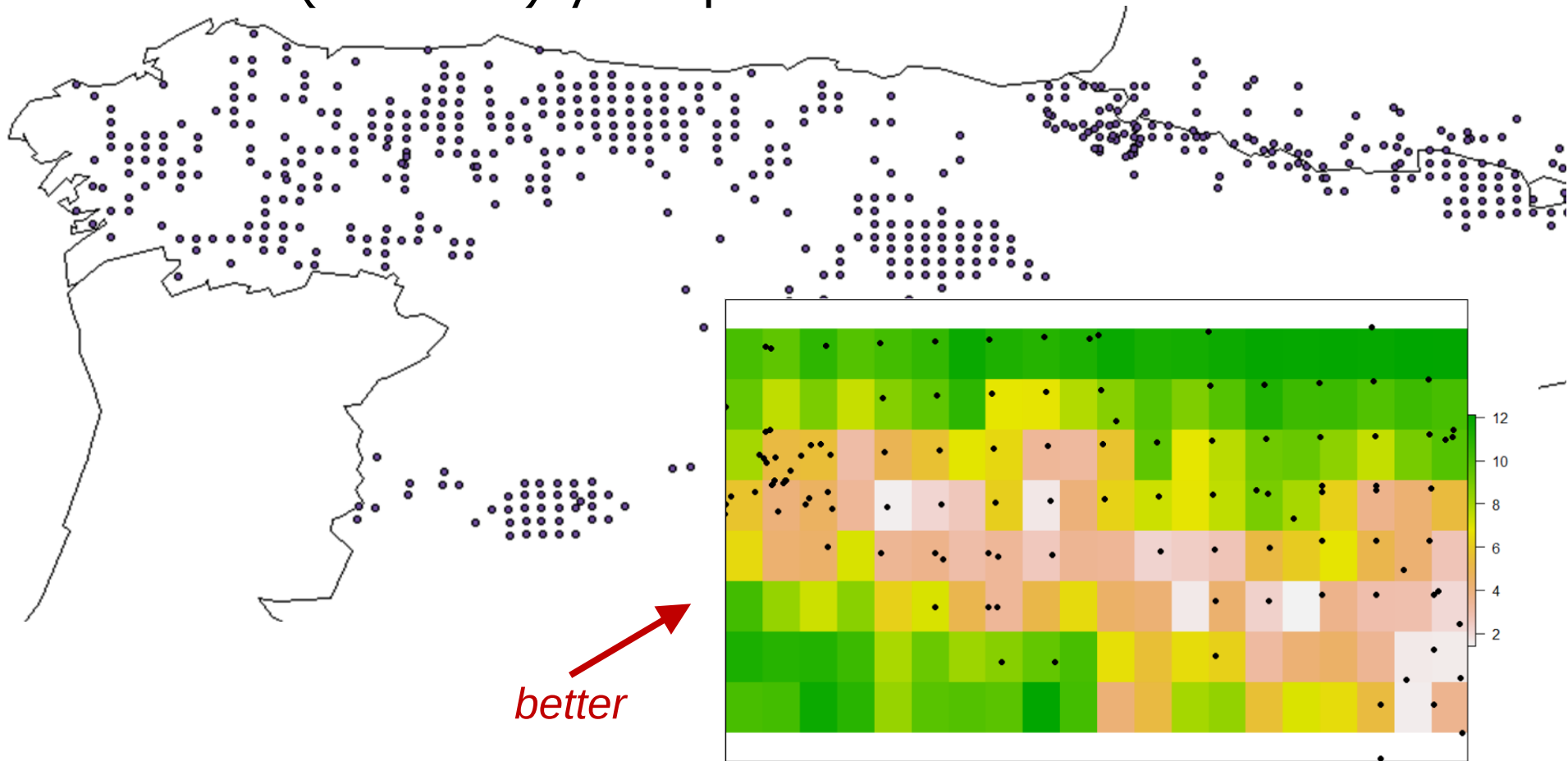
Don't use pixels finer than the regular distance between (most of) your points



Species occurrence + predictor variables

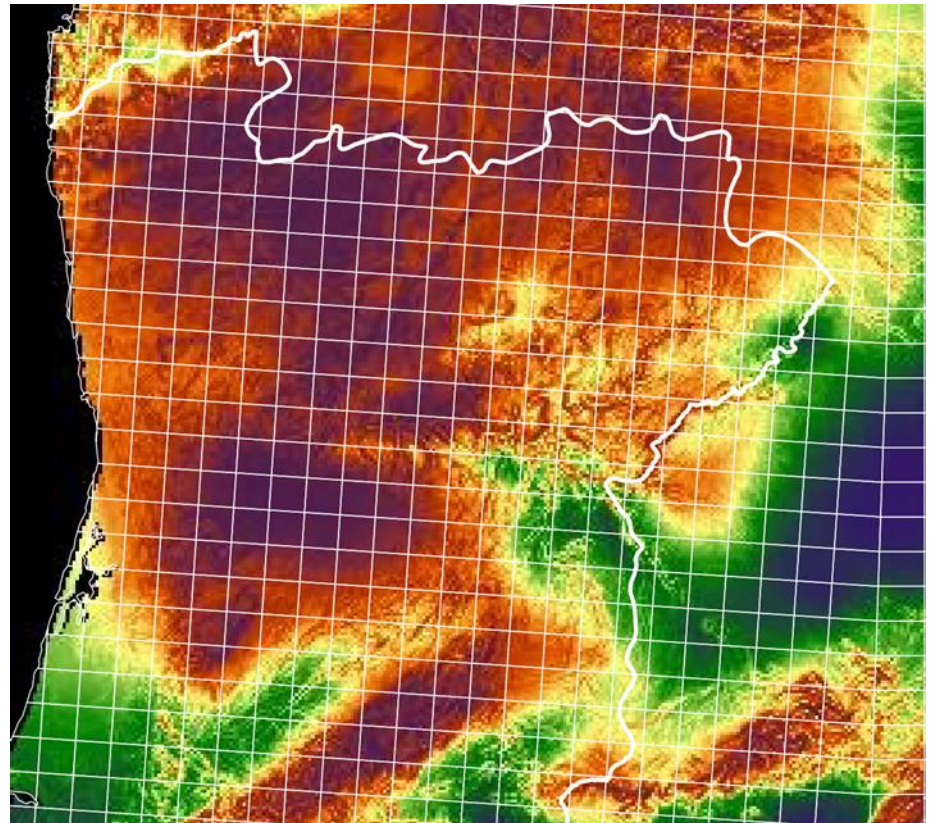
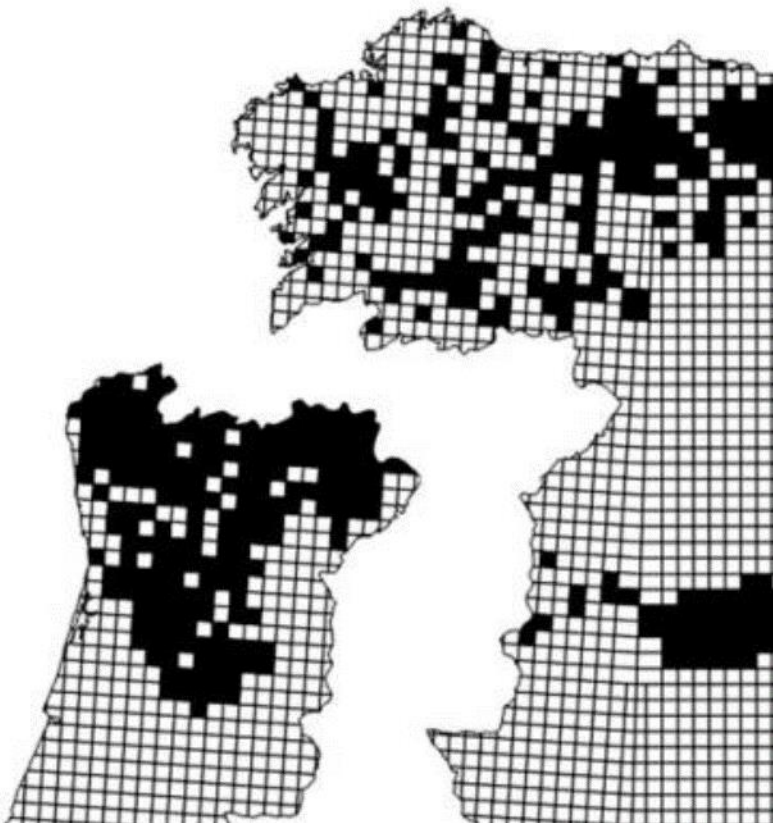
Spatial resolution should match!

Don't use pixels finer than the regular distance between (most of) your points



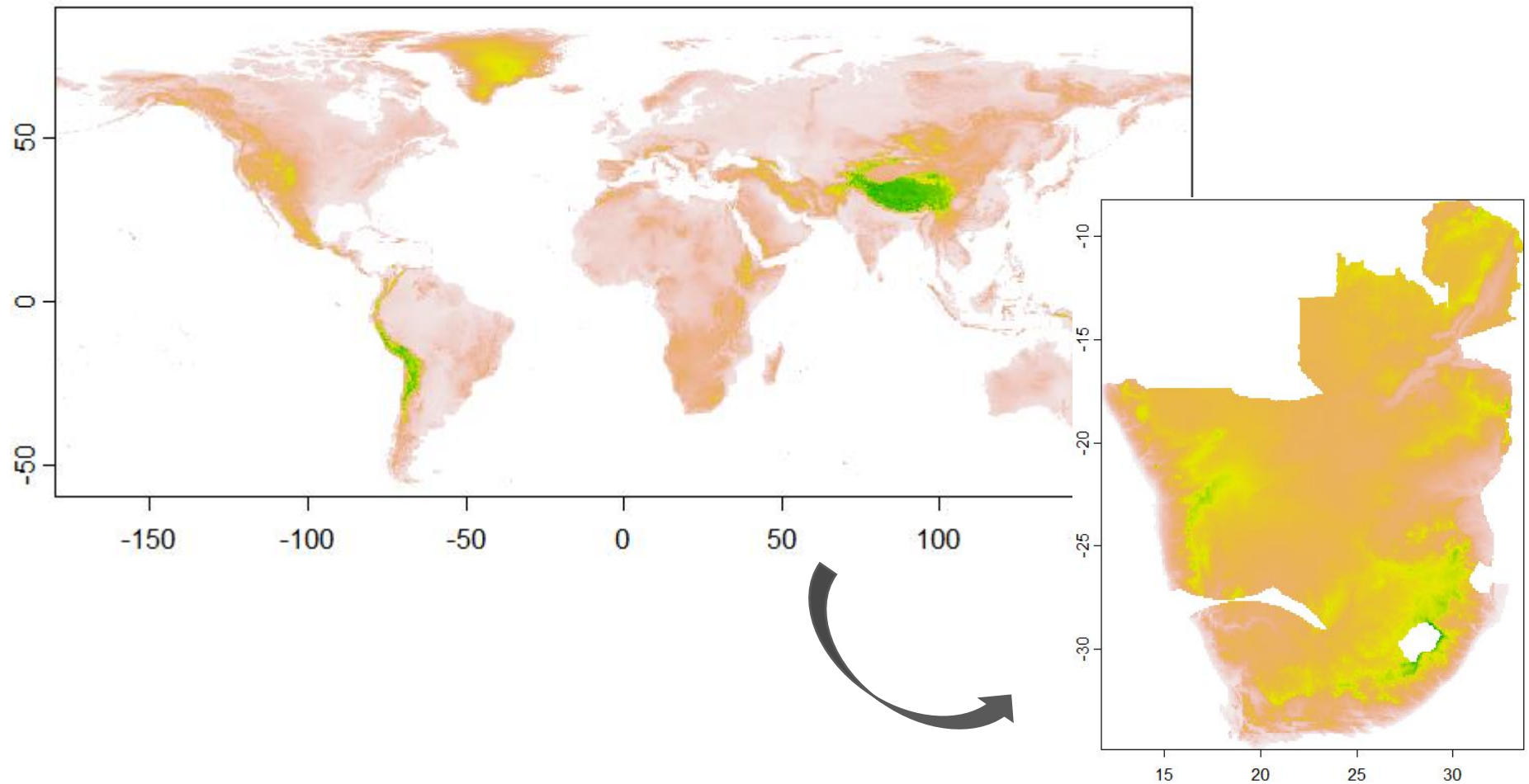
Downscaling model predictions

- compute the model on a coarse resolution
- apply predictions to finer-resolution versions of the same variables



Cut to desired extent

```
# crop the variables to the modelling region:  
layers_cut <- crop(layers, mod_region, mask = TRUE)
```

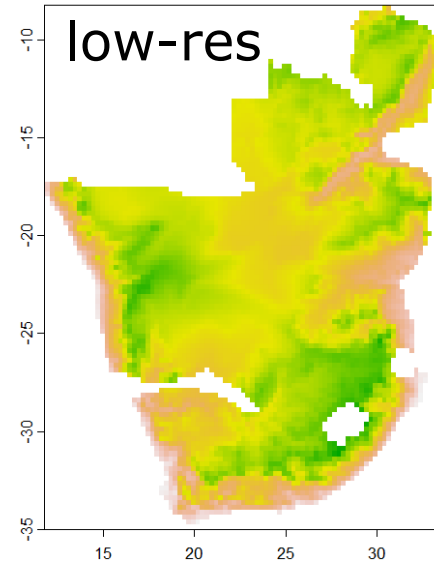
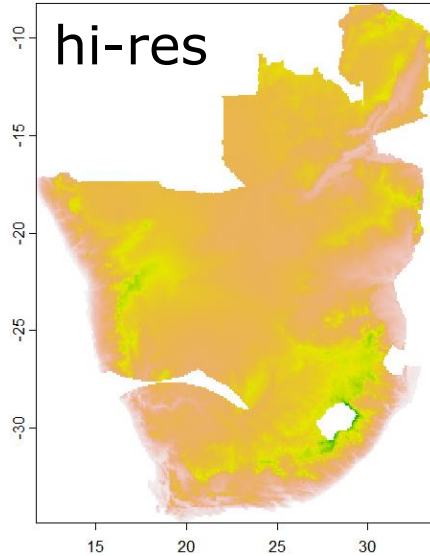


Aggregate to desired resolution

```
# e.g. to a 3-times coarser resolution (if needed!):
```

```
res(layers_cut)
```

```
layers_cut <- aggregate(layers_cut,  
                          fact = 3, fun = mean)
```



Grid records to a data frame

```
dat <- gridRecords(rst = layers_cut,  
                    pres.coords = presences,  
                    abs.coords = NULL)
```

optional

```
head(dat)
```

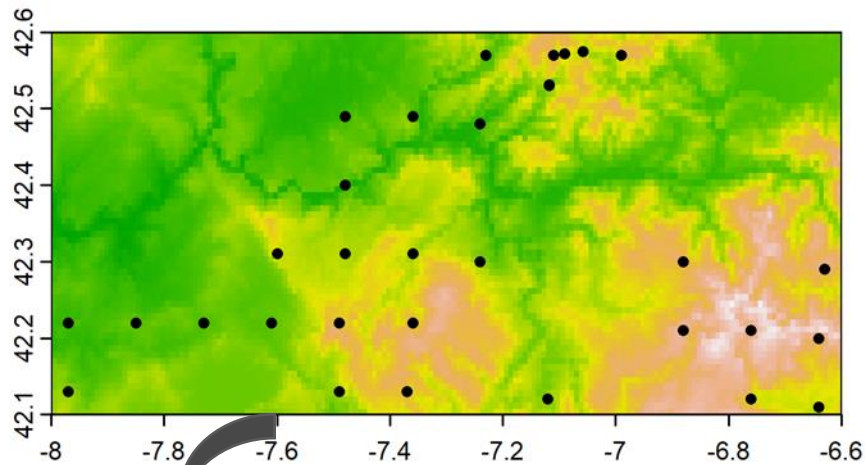
presence	x	y	cells	bio_1	bio_2	bio_3	bio_4
0	26.125	-25.125	5954	19.64711	16.02806	55.41713	482.3851
0	26.375	-25.125	5955	20.27168	16.56423	55.74219	487.0897
0	26.625	-25.125	5956	20.01786	16.05557	55.92709	470.6123
1	26.875	-25.125	5957	19.65874	15.92198	56.15314	465.0204
1	27.125	-25.125	5958	19.52998	15.71729	55.63602	465.7927
0	27.375	-25.125	5959	19.54671	15.34325	54.76509	470.6145

*species occurrence data: 1 = presence,
0 = absence (of records!)*

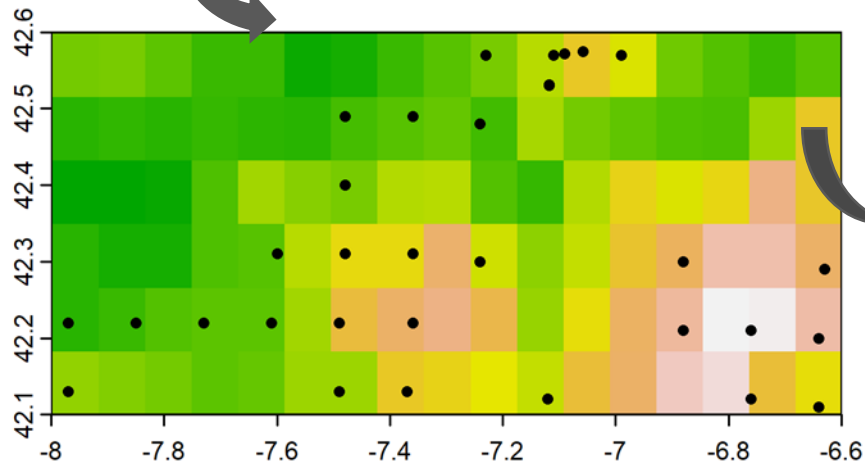
*predictor
variables*

this is the kind of data frame we need for modelling

Grid records to a data frame



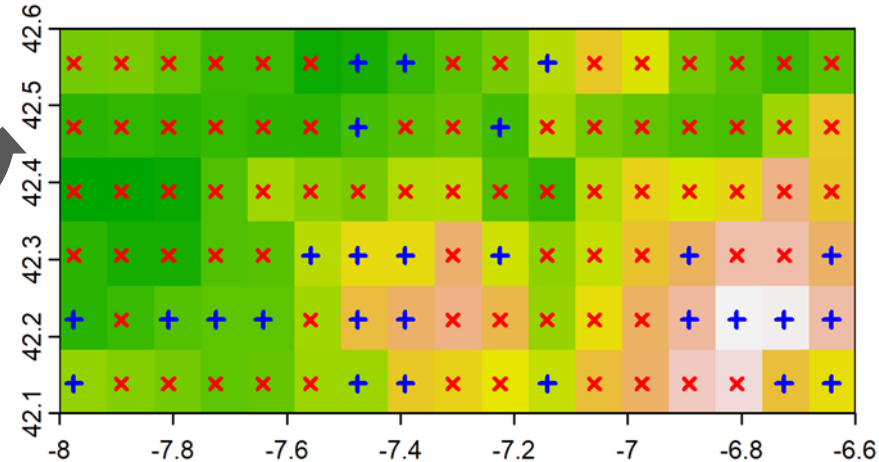
`aggregate()`



+ presence
x absence

*of observation
records!*

`gridRecords()`



gridding = thinning without removing presences
and without introducing false absences

Select a subset of the absence pixels

```
# you can select a subsample of the absence rows:

bias_layer <- footprint(path = "../outputs") # just a
suggestion, terrestrial only, see elsewhere for marine

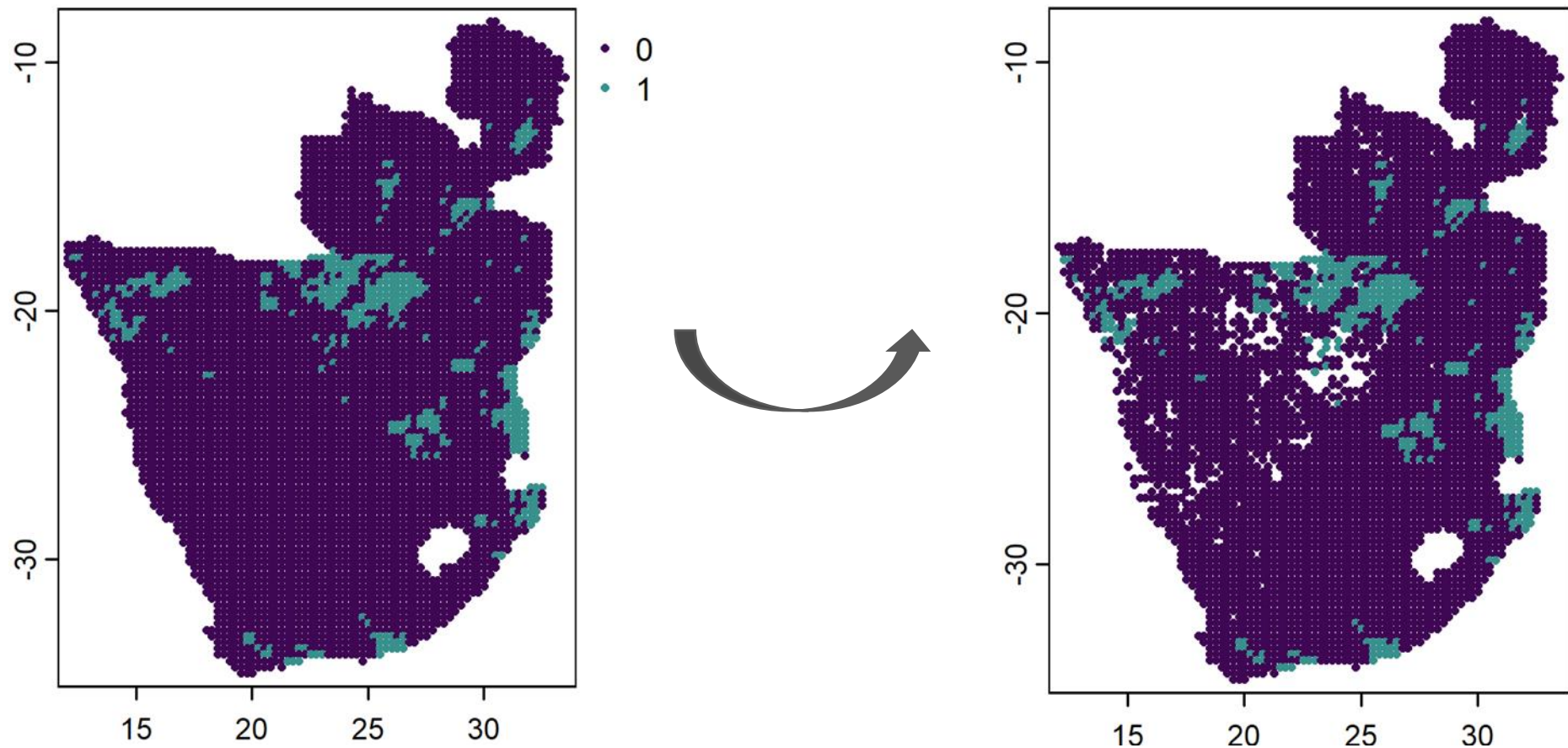
bias_layer <- project(bias_layer, layers_cut)

unsurveyed <- subset(countries, countries$NAME_0 %in%
c("Lesotho", "Eswatini"))

bias_layer[unsurveyed] <- NA # optional

dat$selected <- selectAbsences(data = dat, sp.cols =
"presence", coord.cols = c("x", "y"), mult.p = 10,
bias = bias_layer)
```

Select a subset of the absence pixels



PRACTICAL

- download script "**1b_region.R**" and save it in the "*practicals/R*" folder
- run one command at a time and understand each result before moving on
- turn in the **assignment**
- **ask me** anything!