Practical Introduction to Species Distribution Models

Blas M. Benito

EECRG

University of Bergen

Email: blasbenito@gmail.com

Blas M. Benito

PROGRAM

SESSIONS

- **Day 1**
  - Introduction to SDMs
  - The modelling workflow
  - Applications of SDMs
  - Limitations of SDMs
  - Commputer programs
  - Preparation of presence records and environmental predictors

SESSIONS

- **Day 2**
  - Calibrating SDMs with different algorithms: theory and practice
  - GLM, GAM, Random Forest, and Maxent
  - Evaluating SDMs
  - Projection of SDMs on space and time
- **Day 3**
  - Practical workshop
- **Day 4**
  - Presentation of results of the practical workshop
  - Brief introduction to Dynamic SDMs
  - Discussion and conclussions

INTRODUCTION

WHAT IS A MODEL?

Simplified destription of a physical system that still has properties appearing in the real system (Joly 1988).

Objectives:

- Offer a downsized version of the system
- Allows for a better understanding
- Hypothesis generation and testing
- Exploring “what-if” scenarios
- Forecasting and hindcasting
- Support decision making
- Have fun and learn stuff

PRESENCE/ABSENCE AND ECOLOGICAL NICHE

LOOKING AT THE SYSTEM

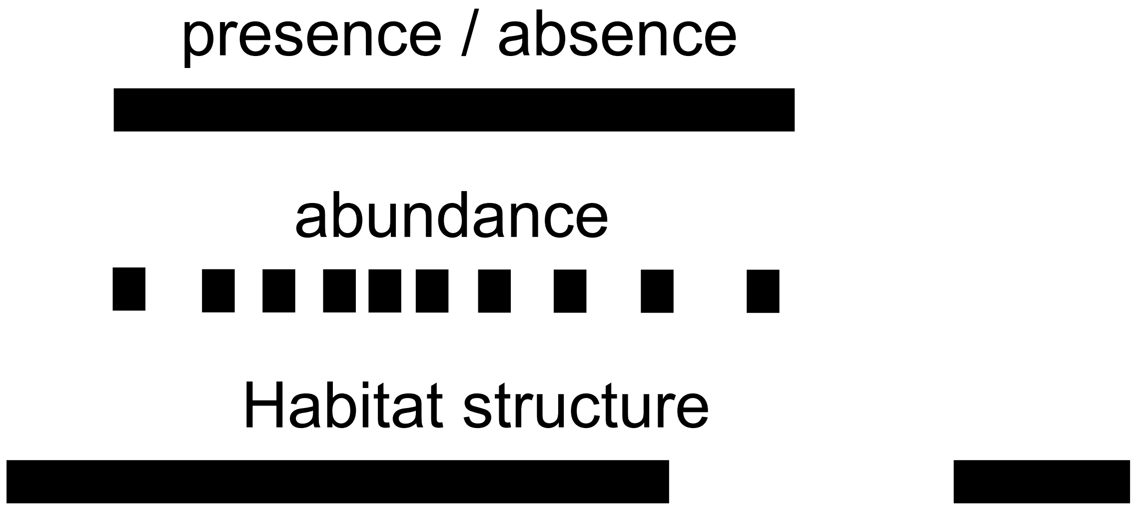
presence / absence

LOOKING AT THE SYSTEM

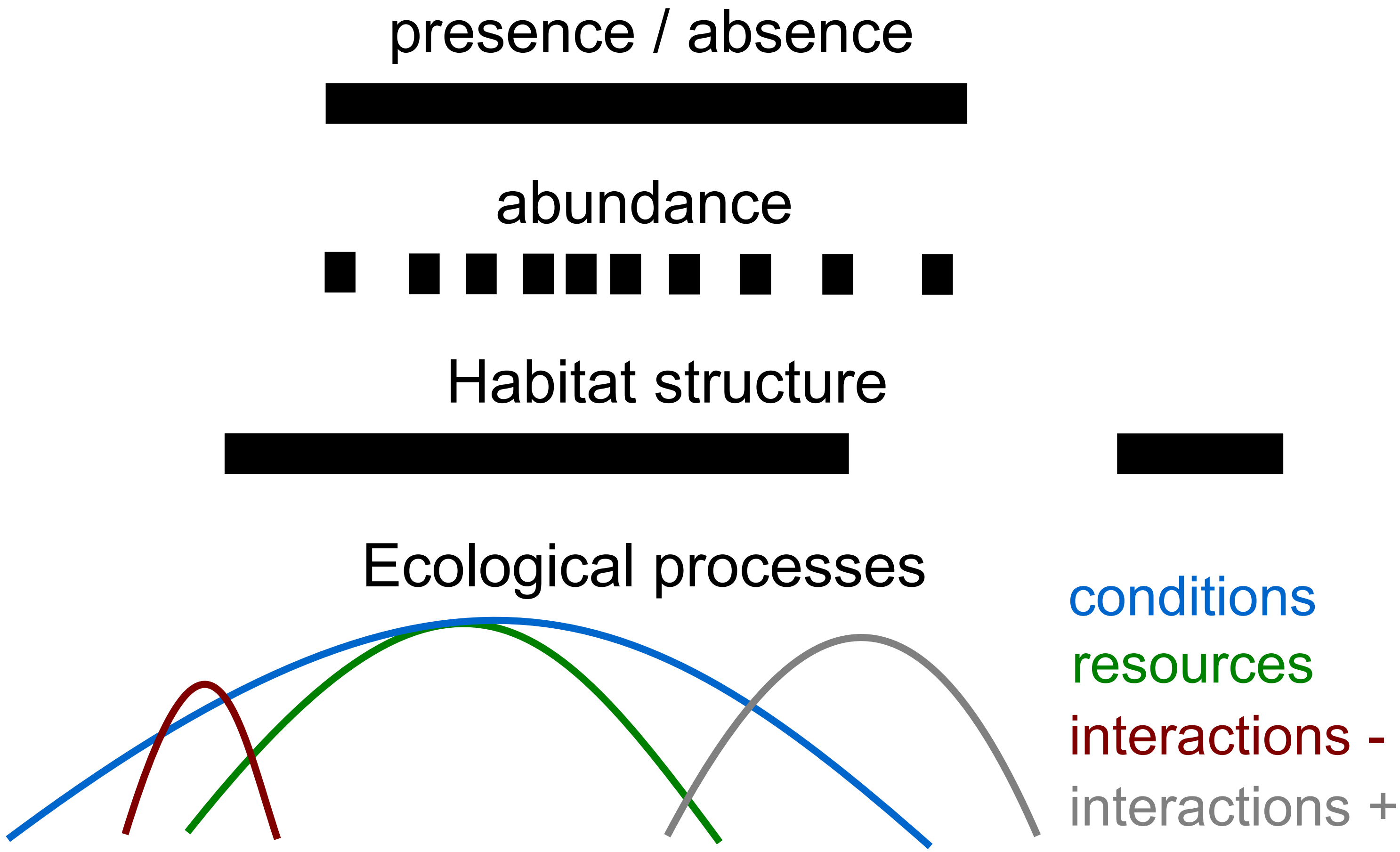
presence / absence

abundance

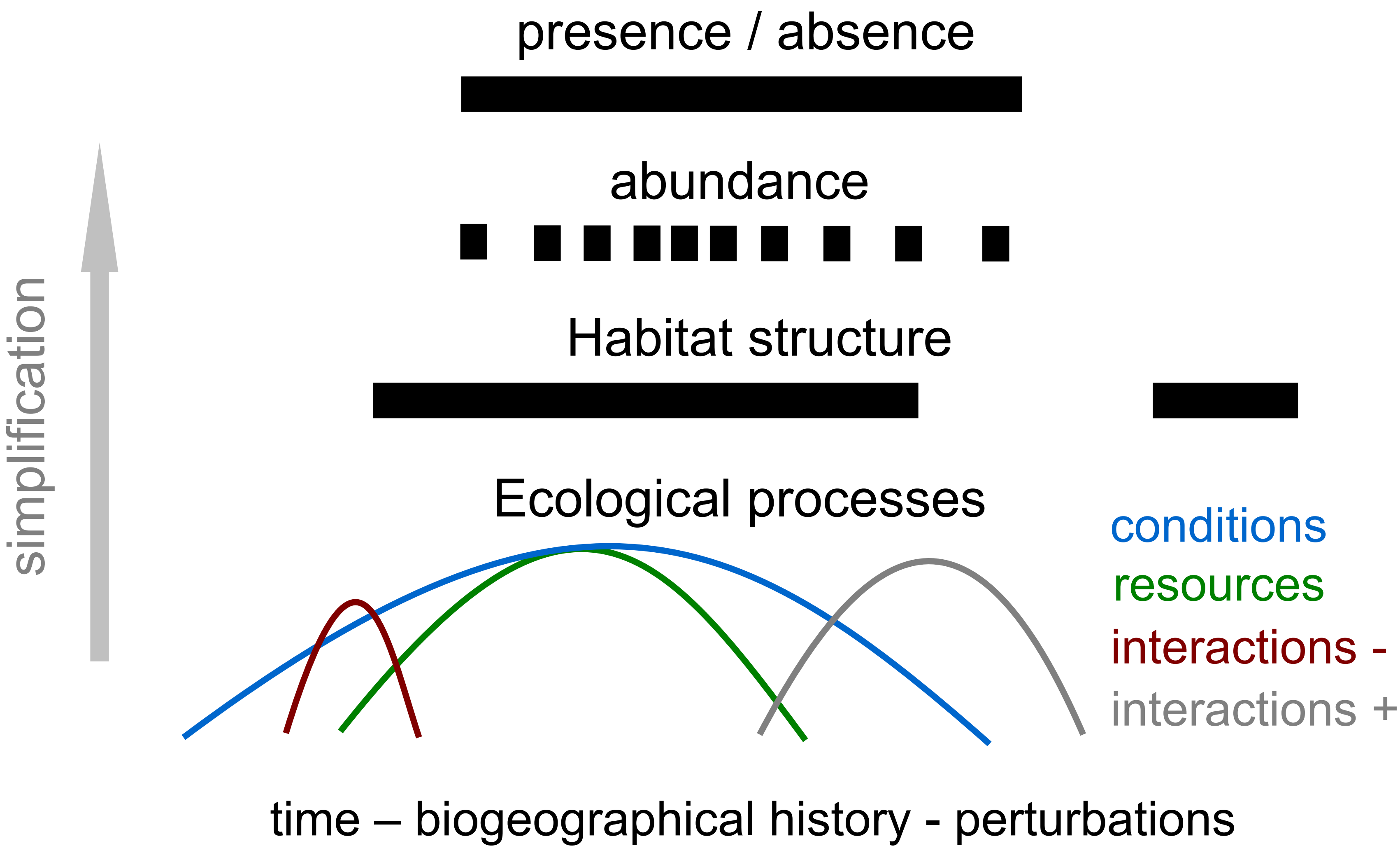
LOOKING AT THE SYSTEM



LOOKING AT THE SYSTEM



LOOKING AT THE SYSTEM



ECOLOGICAL NICHE STRUCTURE

Abiotic factors
- conditions
- resources

Population dynamics

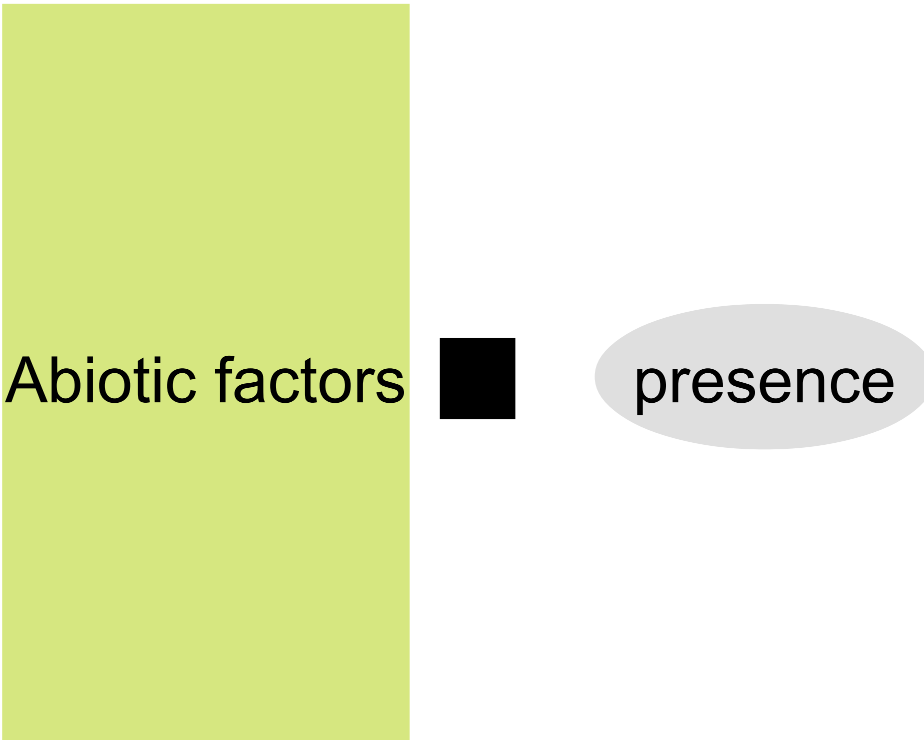
presence

Biotic factors
- Negative interactions
- positive interactions

ver McInerny y Etienne 2012 J Biogeogr

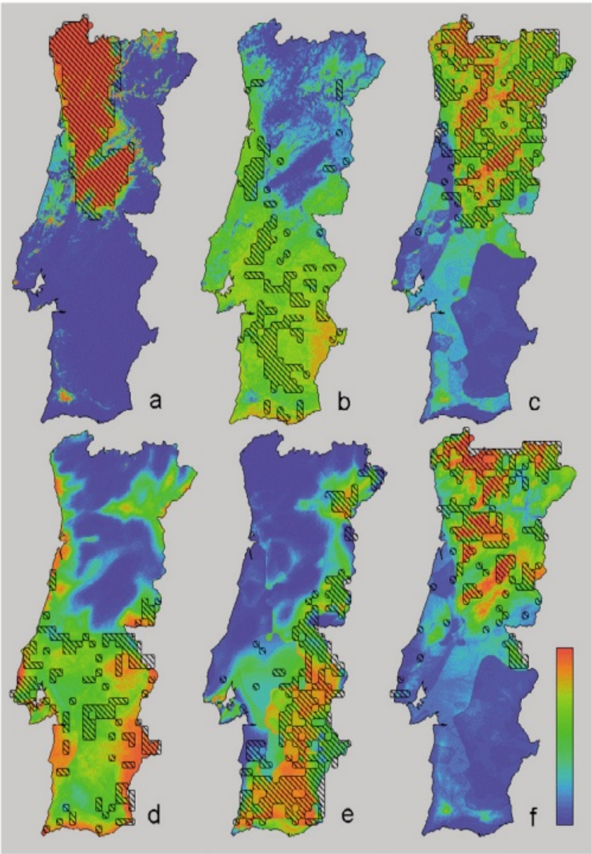
THERE ARE DIFFERENT WAYS TO MODEL ECOLOGICAL NICHES

CORRELATIVE MODELS



■ = we only know the pattern, not the process

MODELOS CORRELATIVOS



Frontiers in Zoology

BioMed Central

Research

Open Access

From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles

JW Arntzen*

Figure 1

Descriptive distribution models for six amphibian species across Portugal. Descriptive distribution models for six amphibian species across Portugal. Models are derived with stepwise logistic regression analysis of the dependent variable 'presence-absence of the target species' against 13 independent ecological variables (details see text and Table 1). The estimated probability of occurrence (g) ranges from 0 (blue) to 1 (red). Composite colours represent intermediate probabilities as in the colour scale bar. Species are: a) *Chioglossa lusitanica*, b) *Pleurodeles waltl*, c) *Triturus marmoratus*, d) *T. pygmaeus*, e) *Alytes cisternasii* and f) *A. obstetricans*. Recorded presences over the 10 × 10 km UTM-grid are shown by black shadings, after Godinho et al. [34].

CORRELATIVE MODELS WITH INTERACTIONS

Abiotic factors ■ presence ■ Abiotic factors

■ = we only know the pattern, not the process

MODELOS CORRELATIVOS CON INTERACCIONES

Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants

Tereza Cristina Giannini, Daniel S. Chapman, Antonio Mauro Saraiva, Isabel Alves-dos-Santos and Jacobus C. Biesmeijer

Ecography 36: 649–656, 2013

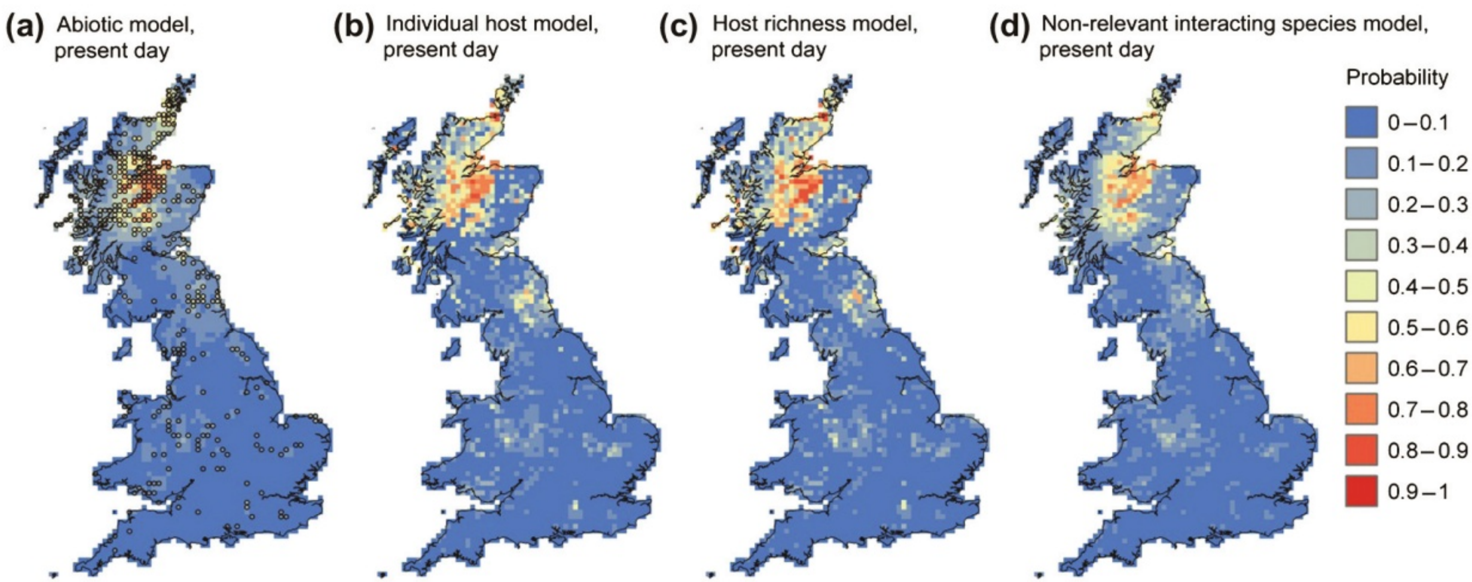
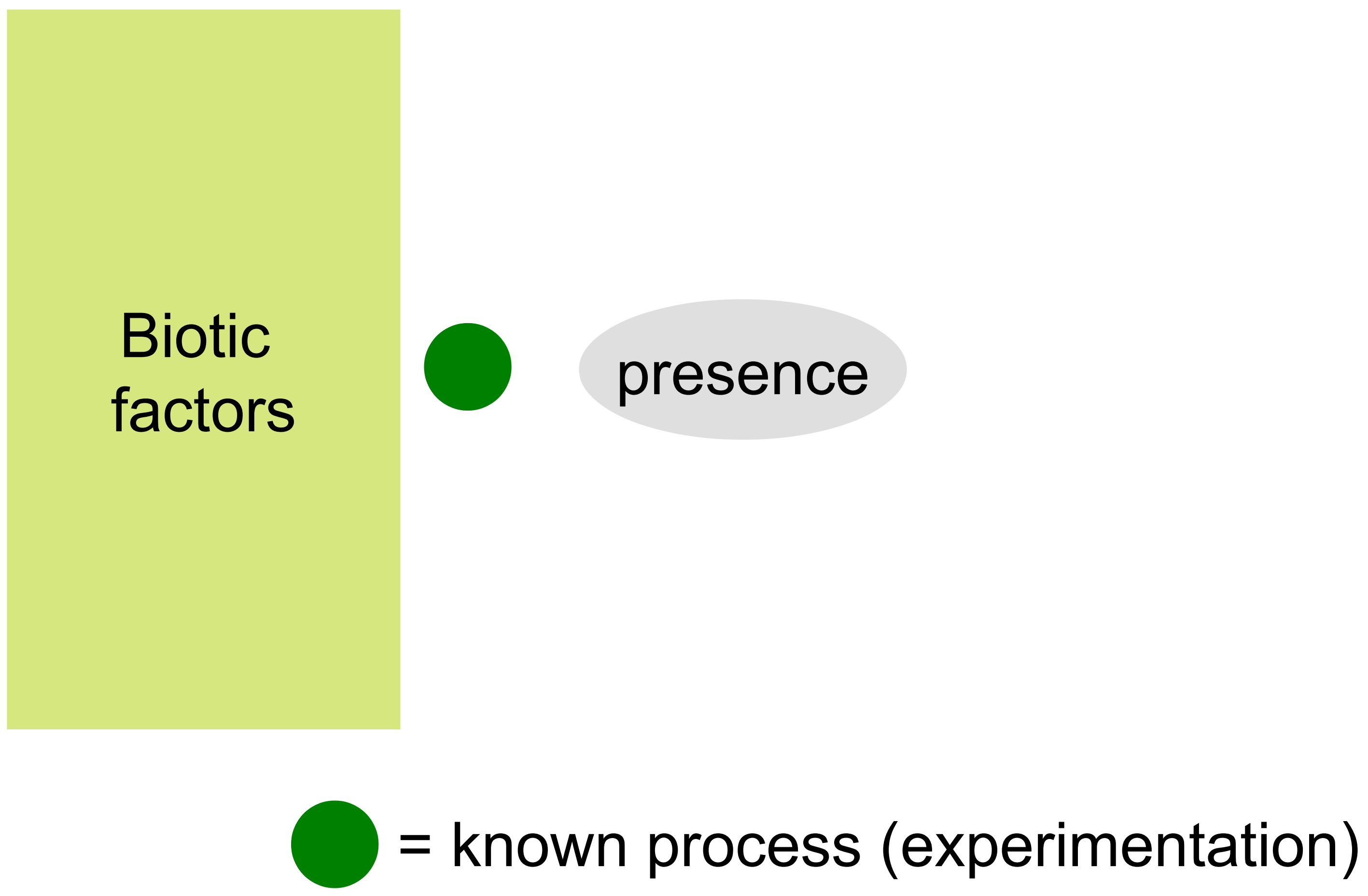


Figure 2. Projections of *Bombus bohemicus* occurrence areas. Top: present climatic conditions using (a) abiotic information only; (b) the most widespread host *B. lucorum*; (c) richness of its three host species (*B. lucorum*, *B. magnus* and *B. cryptarum*); (d) non-relevant interacting species (*B. terrestris*).

ECO-PHYSIOLOGICAL MODELS

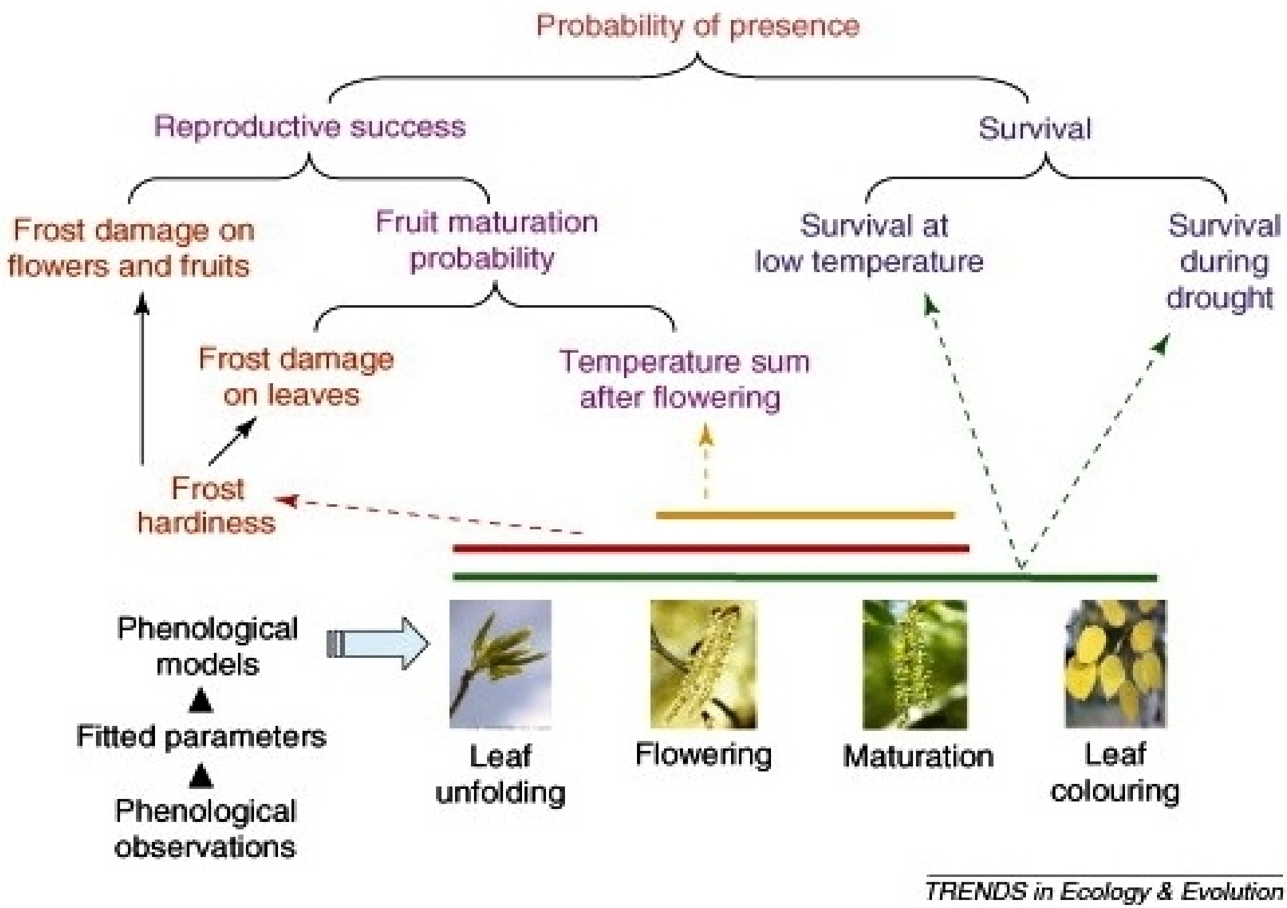


MODELOS ECO-FISIOLÓGICOS

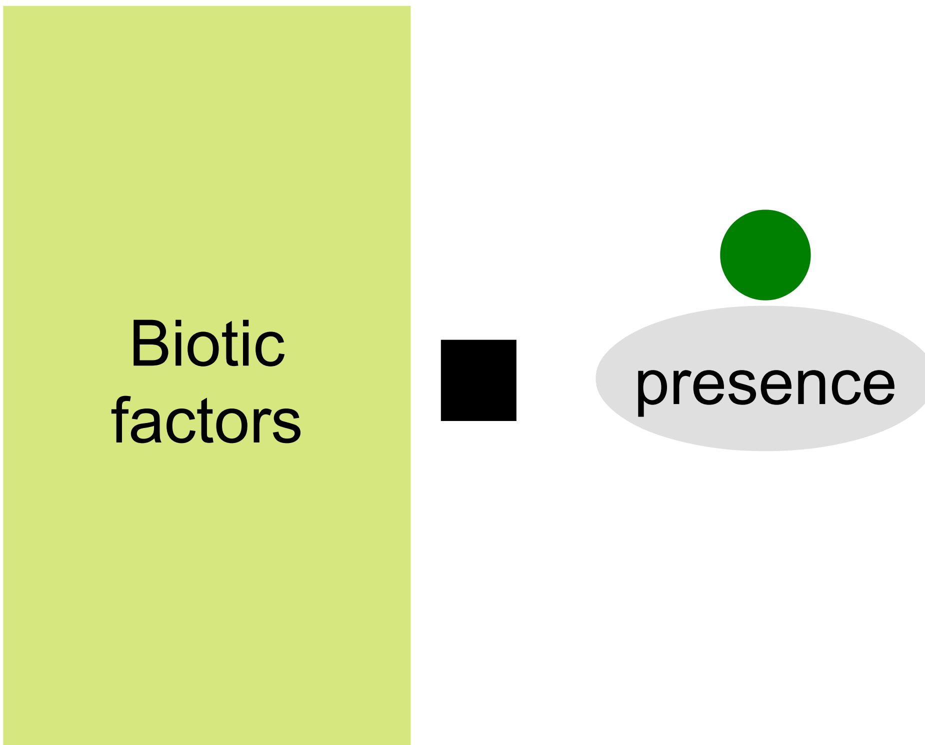
Review TRENDS in Ecology and Evolution Vol.22 No.7
Full text provided by www.sciencedirect.com ScienceDirect

Shifting plant phenology in response to global change

Elsa E. Cleland[1], Isabelle Chuine[2], Annette Menzel[3], Harold A. Mooney[4] and Mark D. Schwartz[5]



HYBRID MODELS



= known process (population viability)

= correlation, only pattern

MODELOS HÍBRIDOS

Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps

Damaris Zurell, Volker Grimm, Eva Rossmanith, Niklaus Zbinden, Niklaus E. Zimmermann and Boris Schröder

Ecography 34: 001–014, 2011

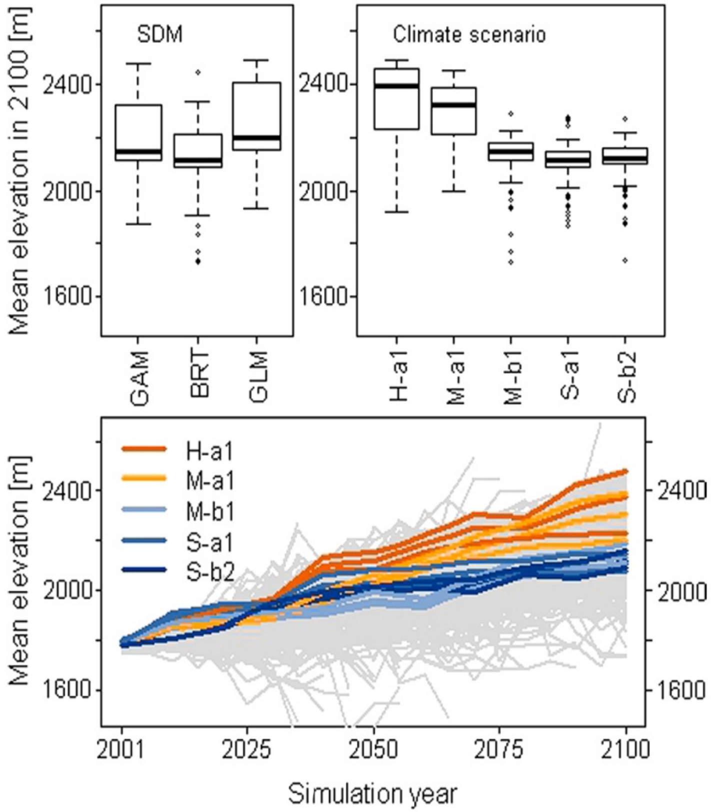
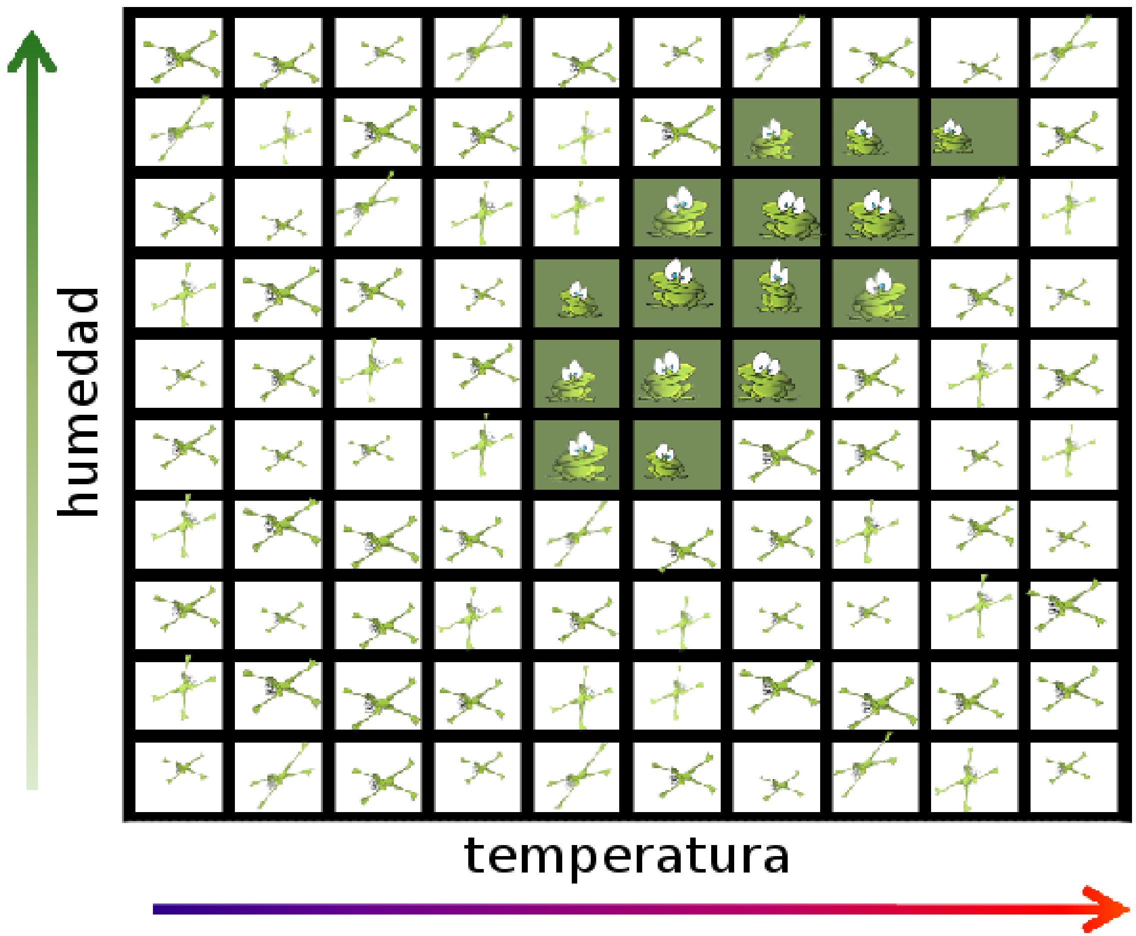


Figure 2. Mean elevation occupied by black grouse for scenarios of climate change. Bottom: grey lines show mean elevations across all simulations, coloured lines those for default IBM parameterisation (cf. Table 2) across different SDMs and climate scenarios. Top: box-plots depict variation of mean elevations predicted for the end of 21st century (2100) and for different SDMs and climate scenarios.

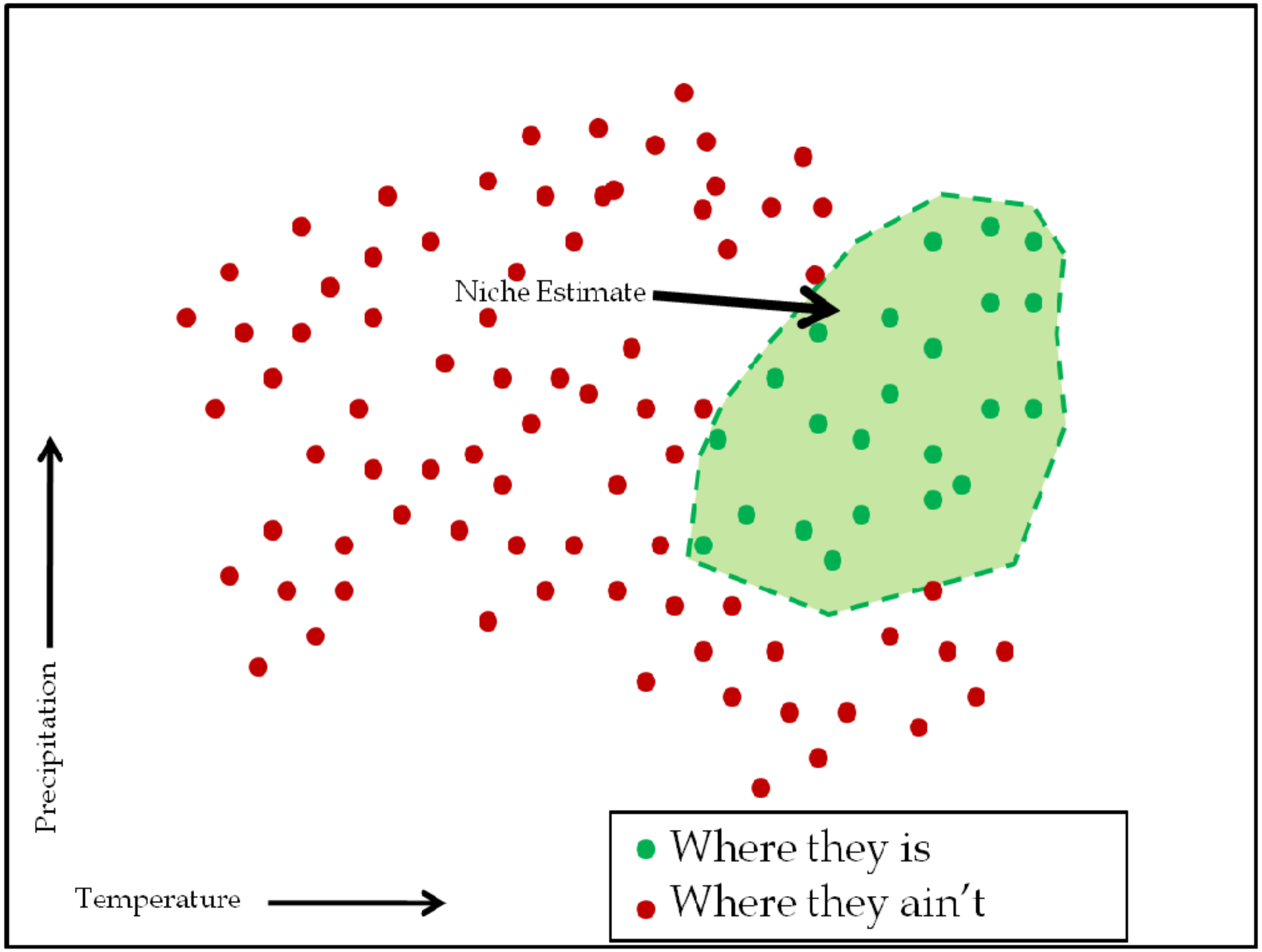
SPECIES DISTRIBUTION MODELLING

EXPERIMENTAL APPROACH



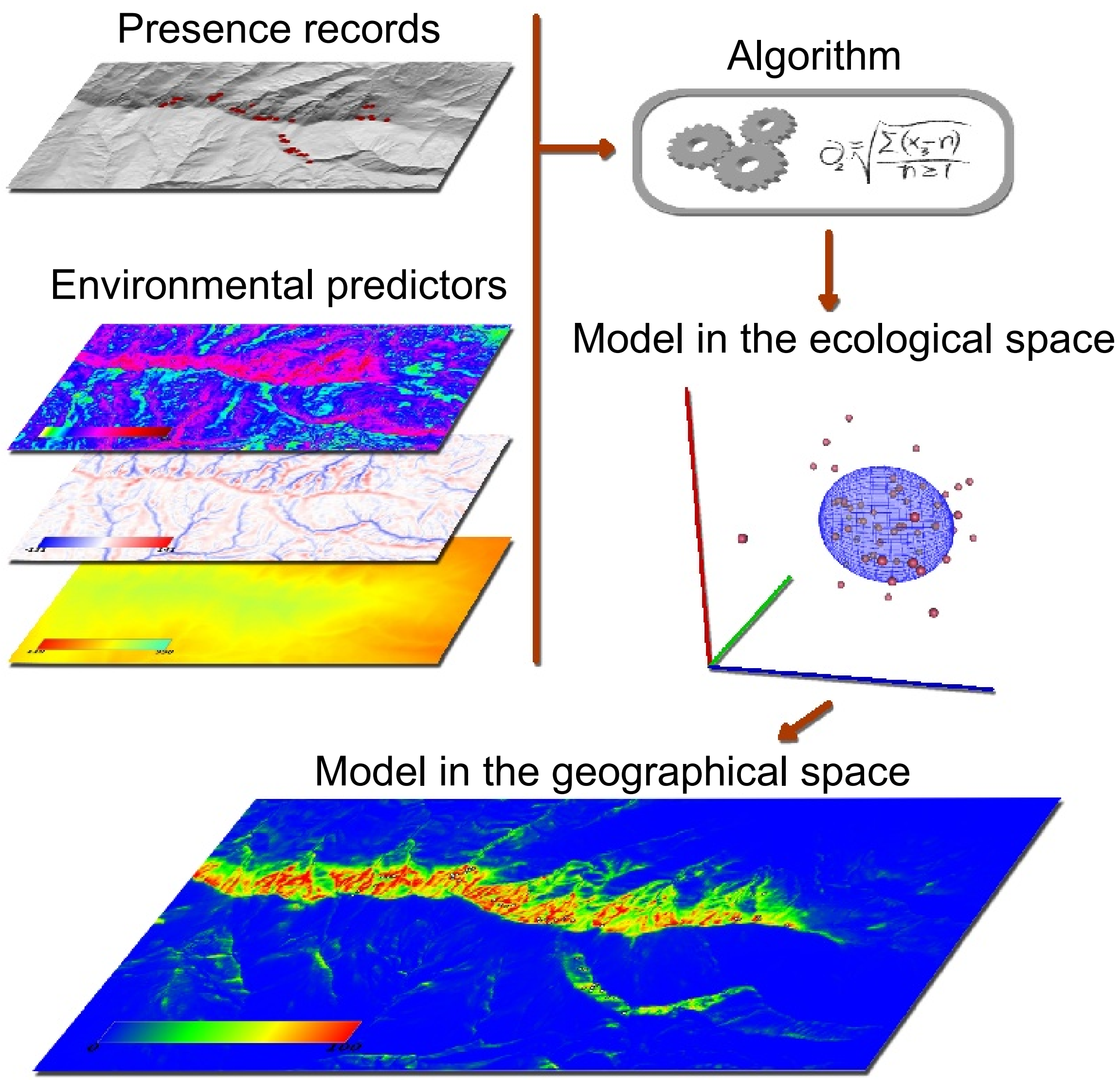
sciencesasaverb.wordpress.com

CORRELATIVE APPROACH



sciencesaverb.wordpress.com

CORRELATIVE APPROACH



KEY ASSUMPTIONS

SPECIES ARE IN EQUILIBRIUM WITH CLIMATE: The species is present in every climatically suitable area, and absent from every unsuitable one.

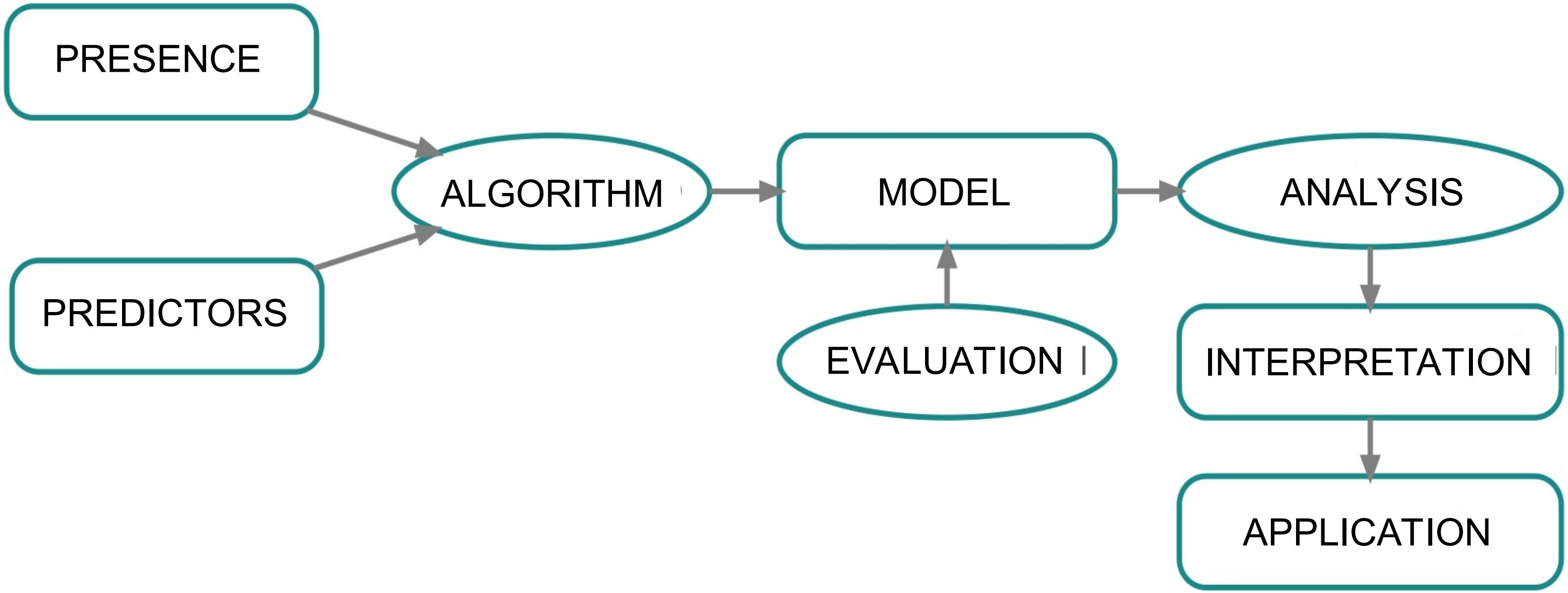
ECOLOGICAL NICHE IS STATIC IN TIME: Niche dimensions and hypervolume's shape remains constant over time.

THE MODELLING WORKFLOW

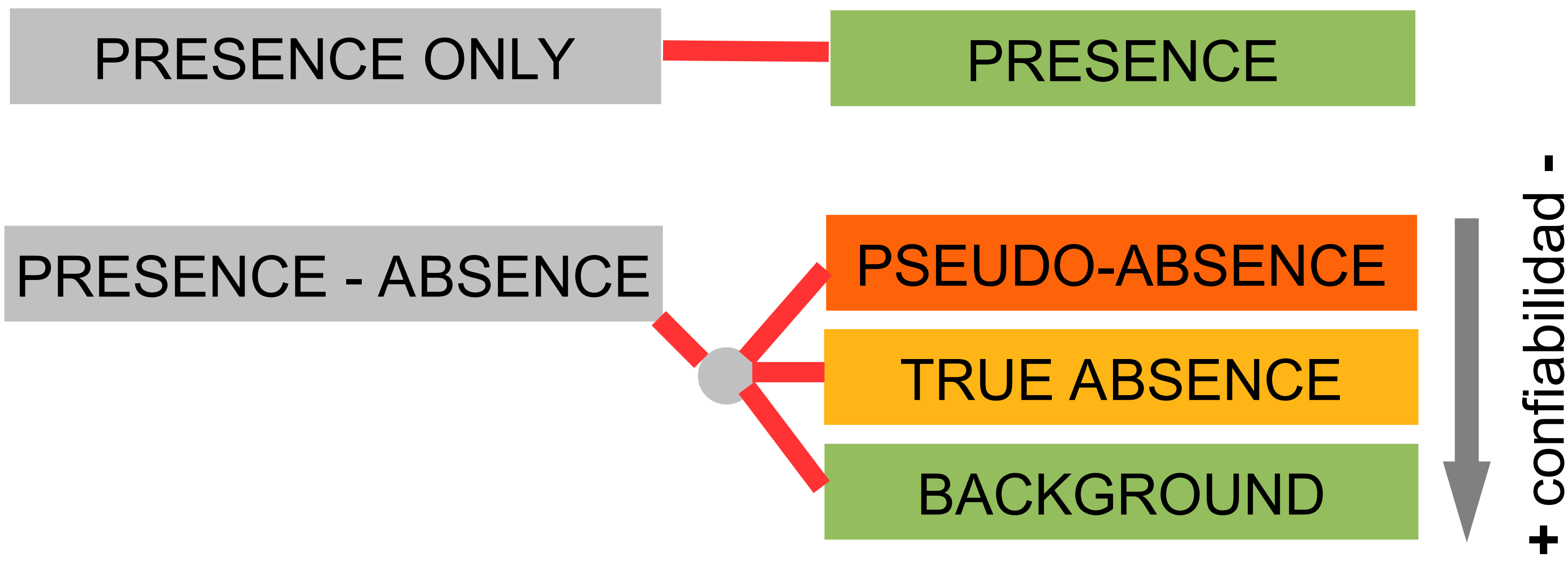
THEORETICAL FORMULATION

- Define objectives
- Gather knowledge on the targe species
- Select meaningful predictors
- Select the right tools
- Draw a workflow, have a plan!

FLUJO DE TRABAJO



TYPES OF PRESENCES



PRESENCE ONLY

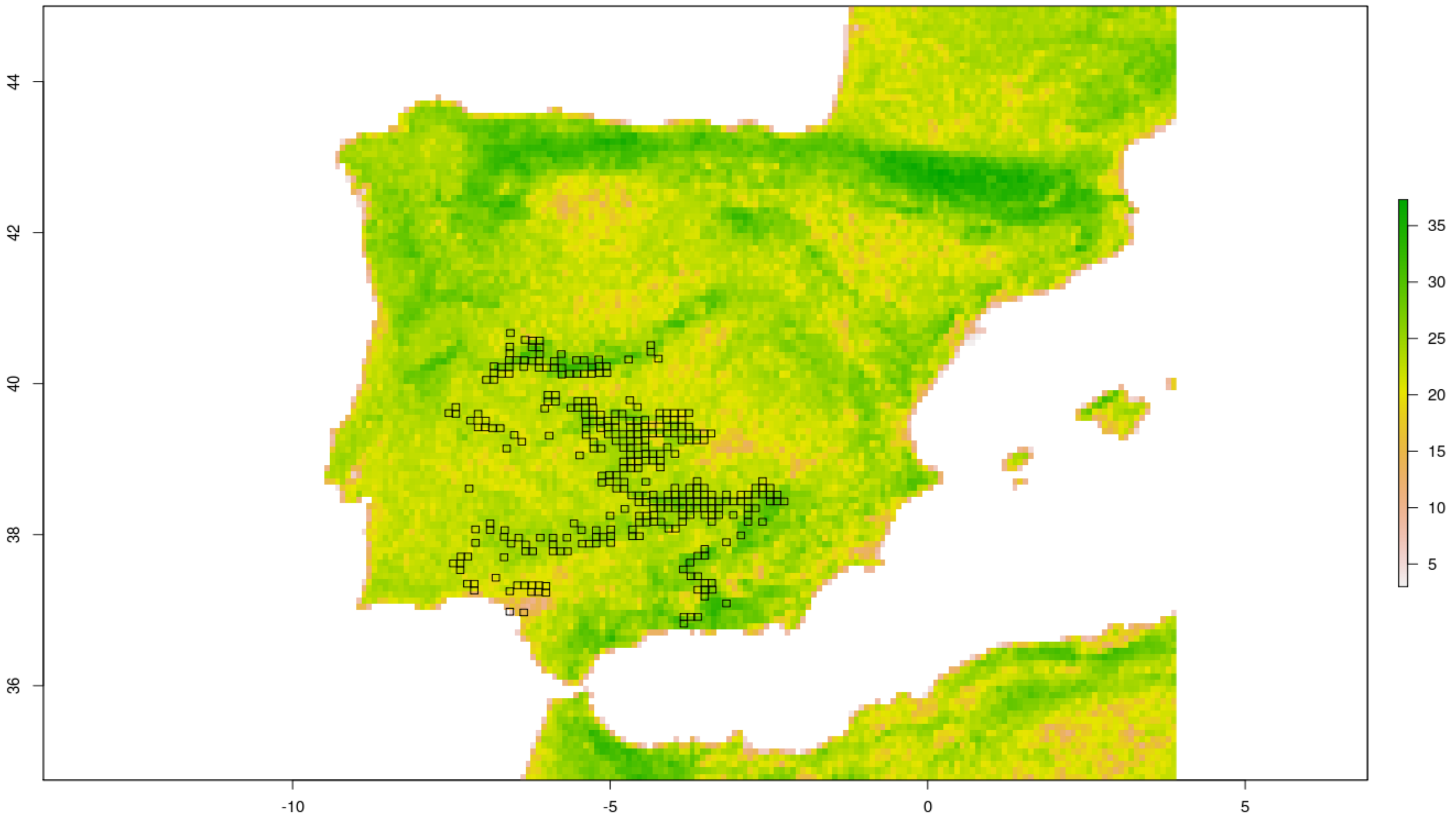
PROS

- High availability (GBIF and the likes)

CONS

- Difficult to compute probability of presence
- Sampling bias usually unknown
- Be careful with coordinate precision!

PRESENCE ONLY



PRESENCE - ABSENCE

PROS

- Allow to compute probability of presence
- Provide info on sampling bias

CONS

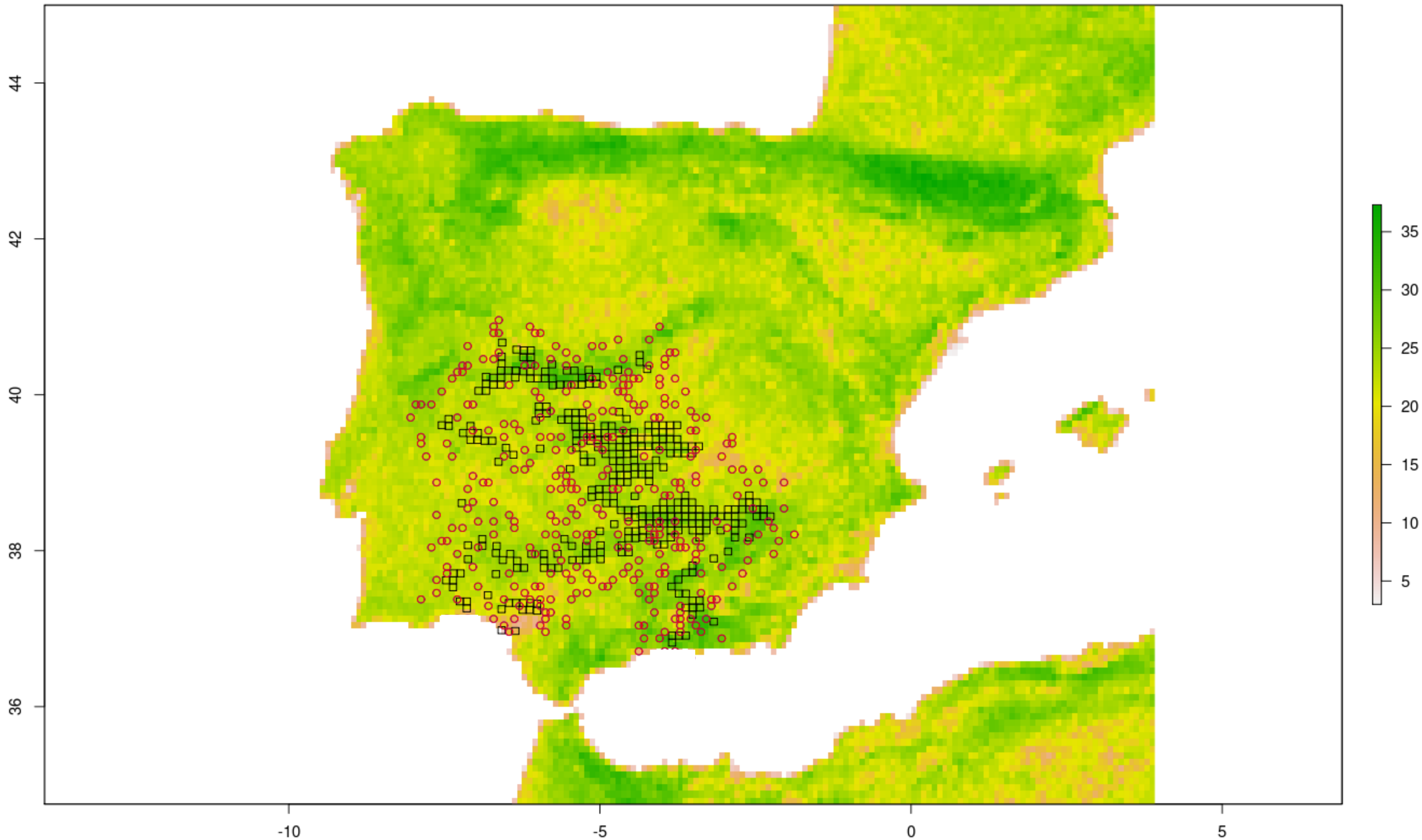
- How do you differentiate a real absence from false one (cryptic species, short lived individuals, etc)?
- Low availability

NATURE OF ABSENCES

- **Contingency**: area is sutiable, but there are dispersal restrictions, historical reasons, or temporal perturbations
- **Environmental**: absence of suitable habitat
- **Methodological**: due to sampling bias

Lobo et al. 2010

ABSENCE



PSEUDO-ABSENCE

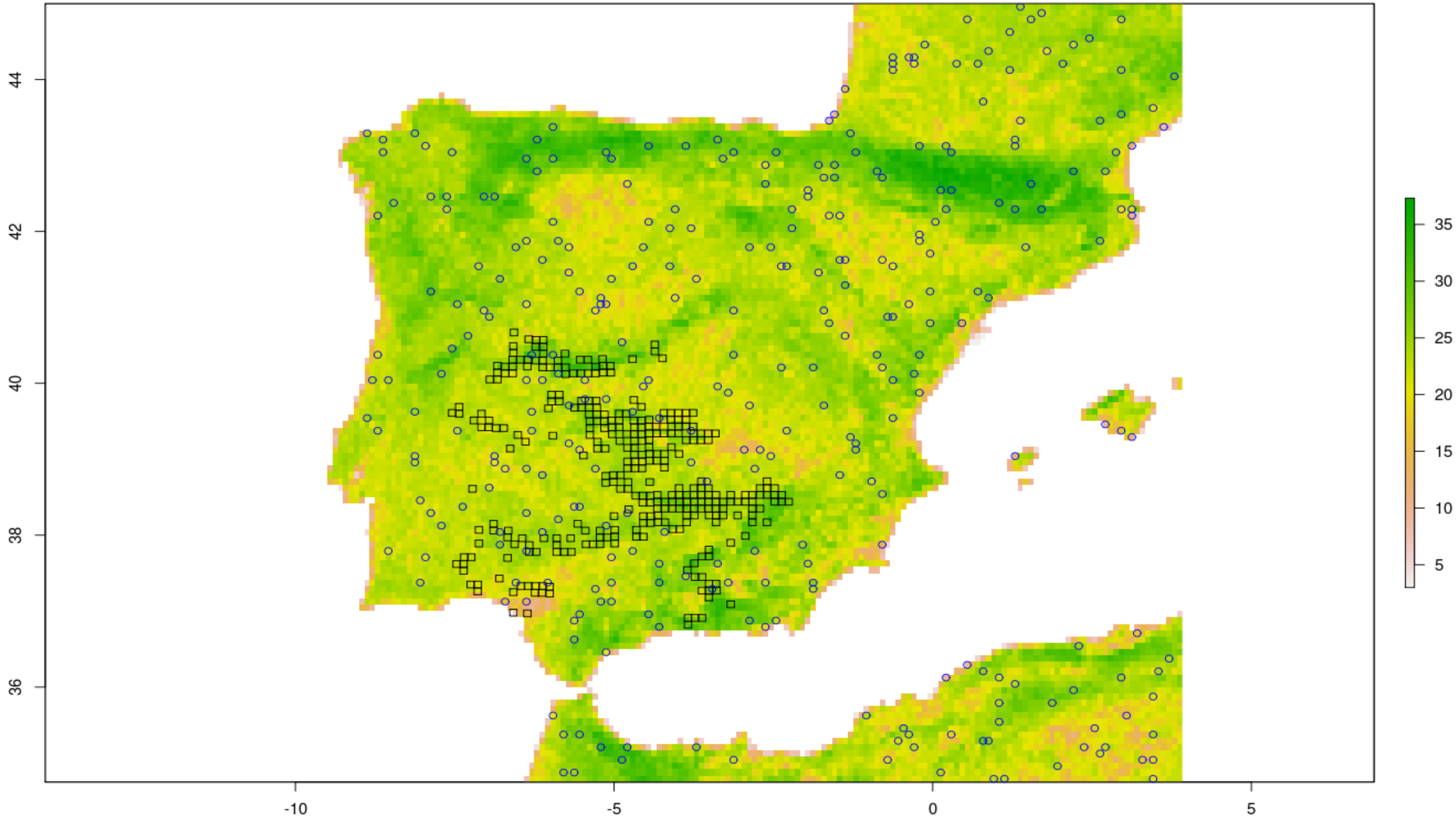
PROS

- Easy to generate

CONS

- They are not real absences
- What ecological criteria should we use?
- Don’t take sampling bias into account

PSEUDO-ABSENCE



BACKGROUND

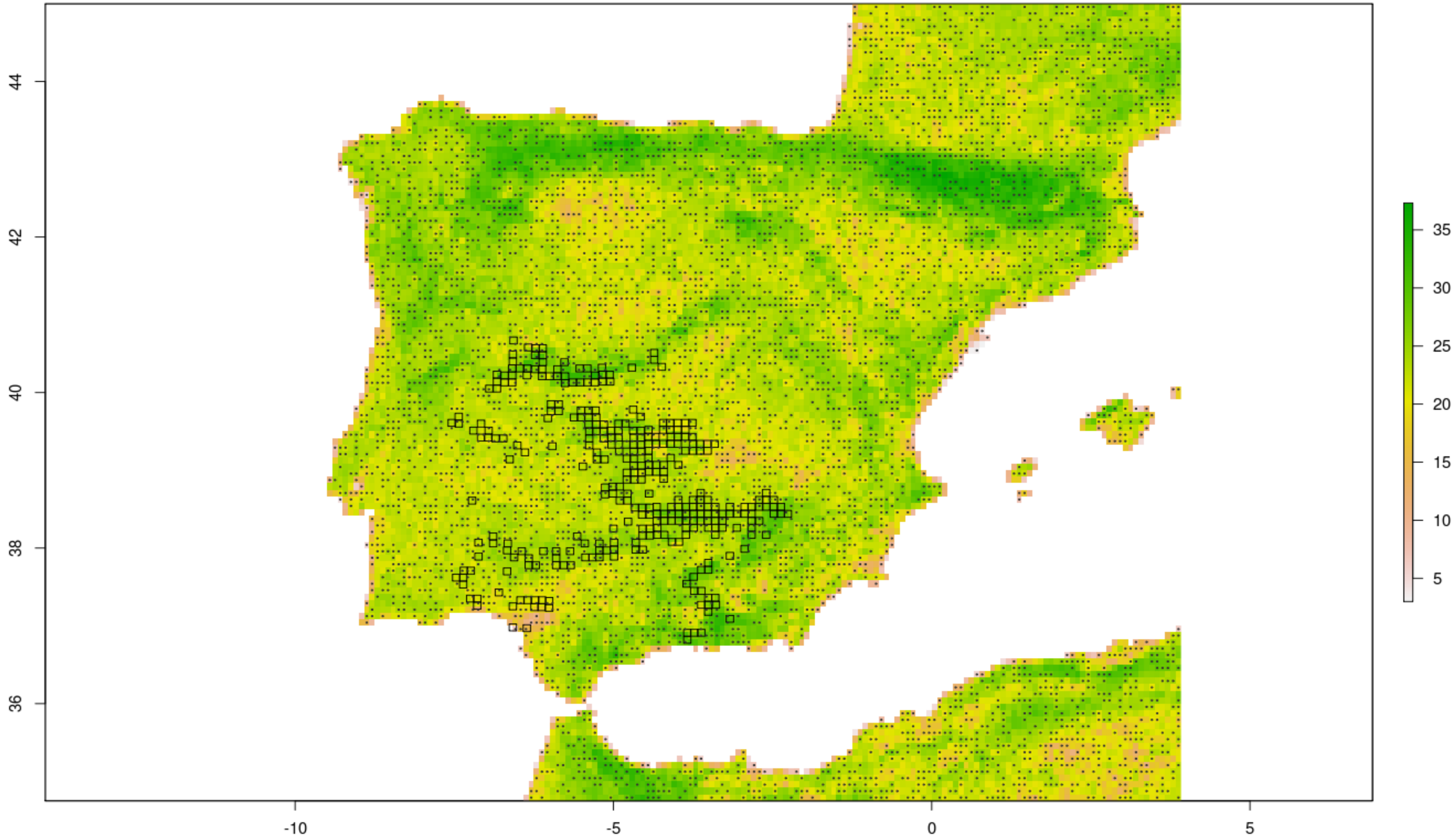
PROS

- Easy to generate
- No issues if they overlap with presences (in fact, they should overlap!)

CONS

- Some methods don’t work well with them (Random Forest, Support Vector Machines)
- Required weighting to be used in regression models

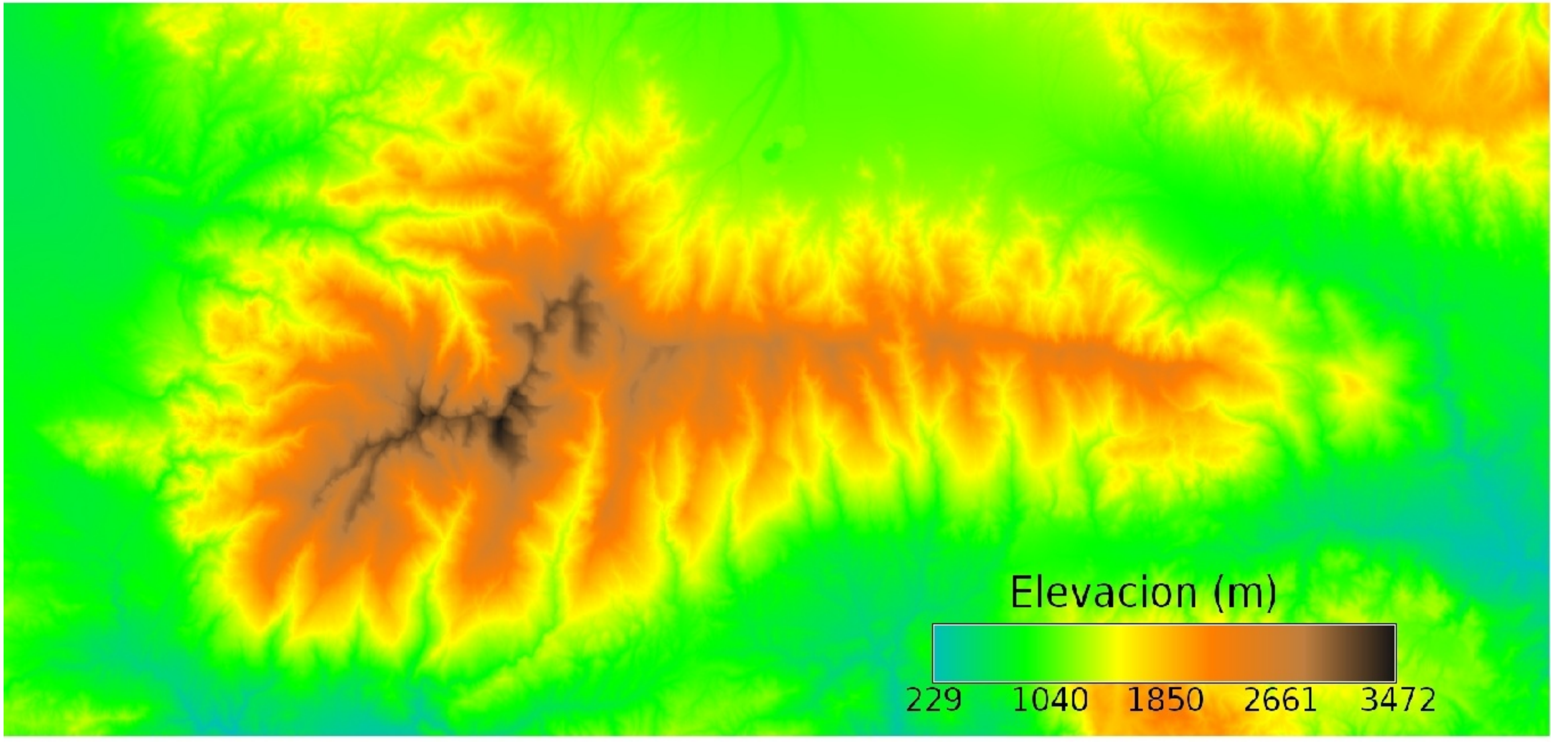
BACKGROUND:



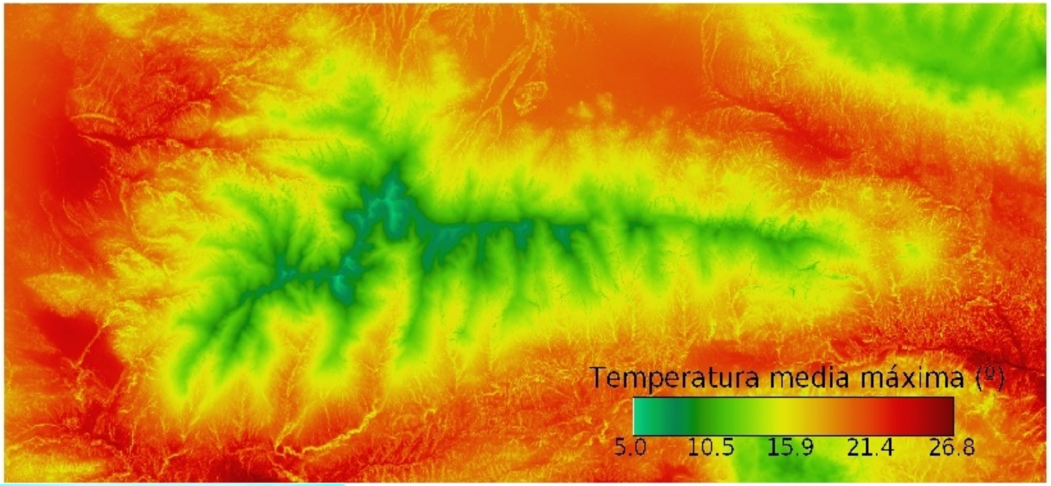
ENVIRONMENTAL PREDICTORS

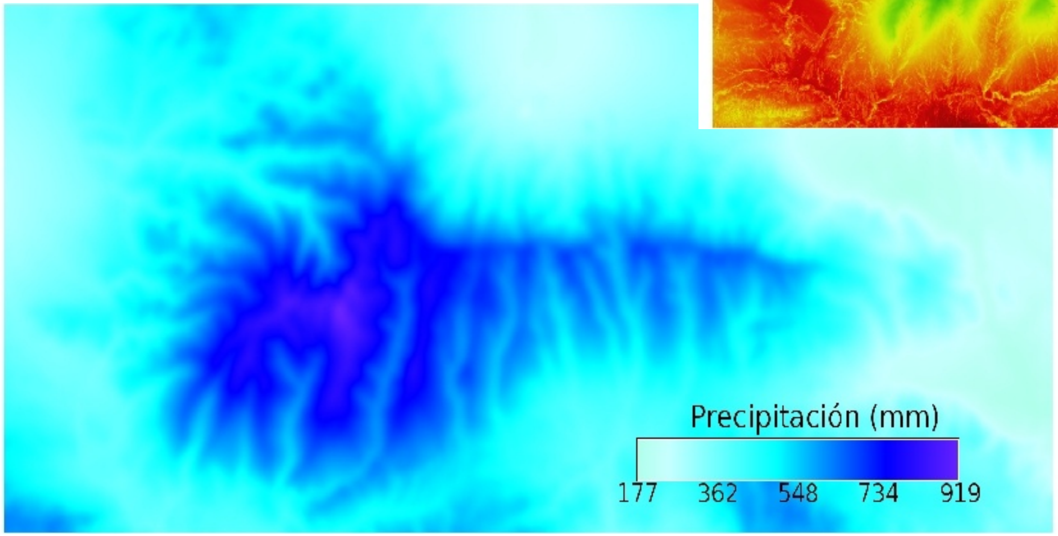
- Main features
  - Effect on species distribution
  - Available as a raster GIS map
  - Same extension and resolution
  - Low collinearity
- Matching resolution with the accuracy of presence data
- Watch for interactions among predictors

TOPOGRAPHY

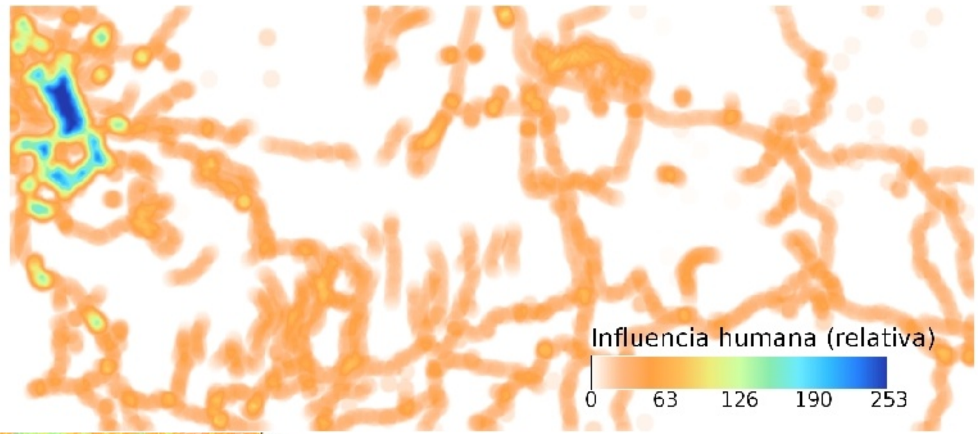


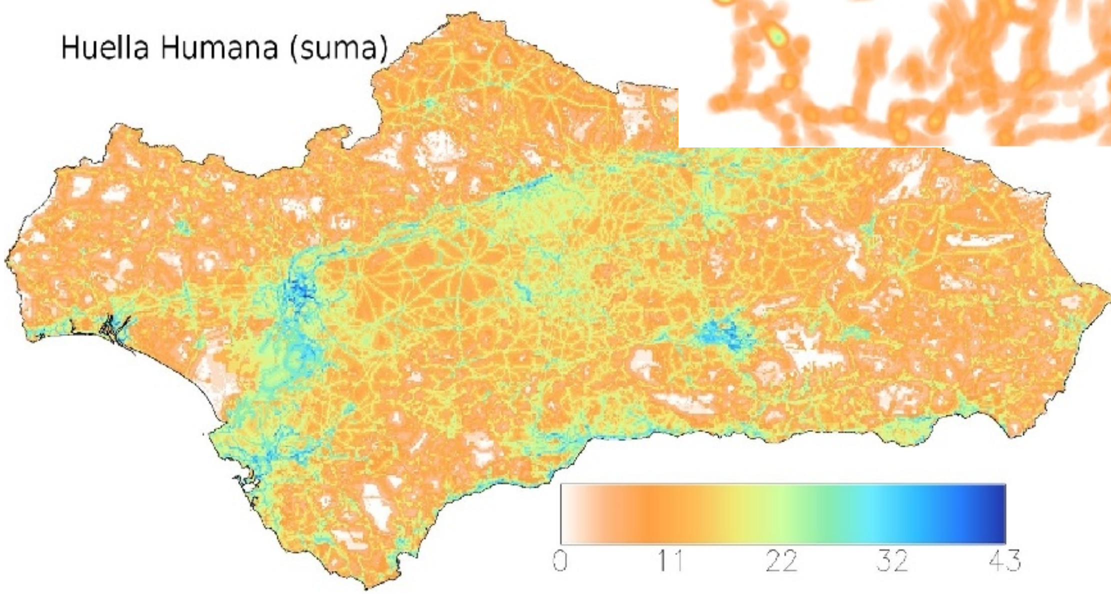
CLIMATE


Temperatura media máxima (º)
5.0
10.5
15.9
21.4
26.8

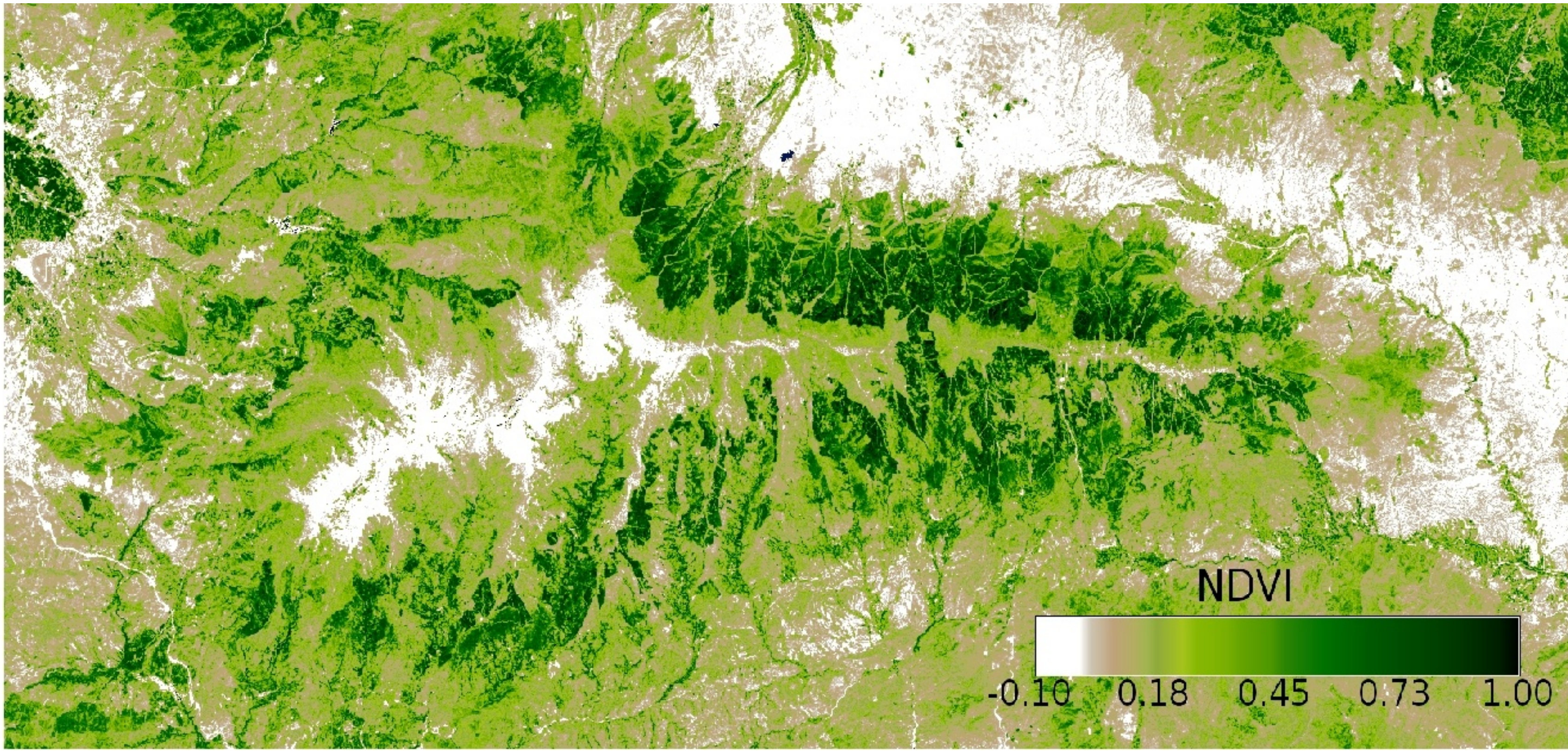

Precipitación (mm)
177
362
548
734
919

HUMAN FOOTPRINT

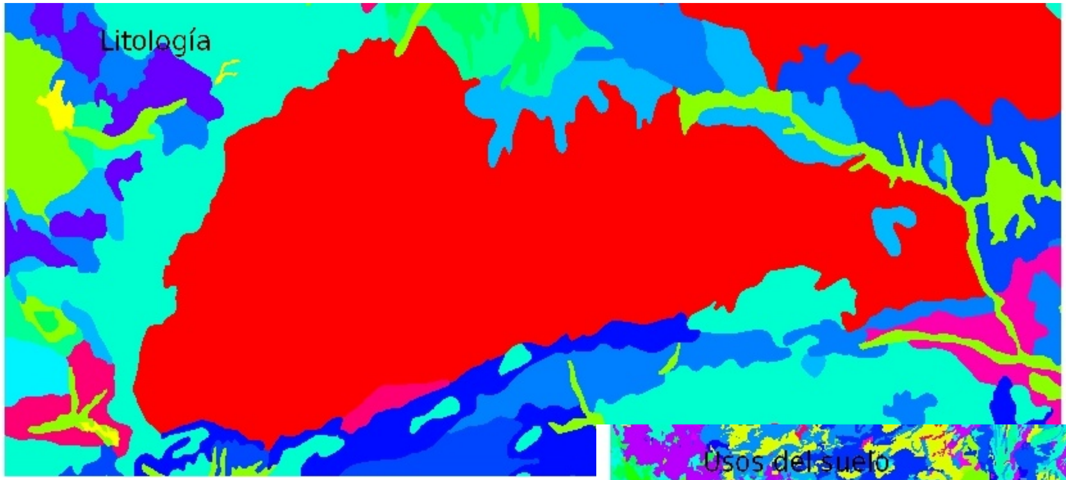


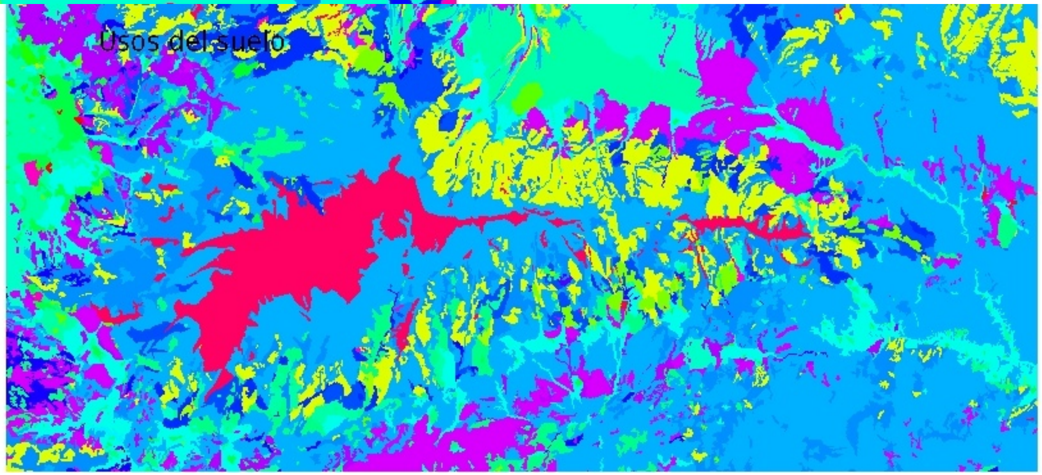


REMOTE SENSING

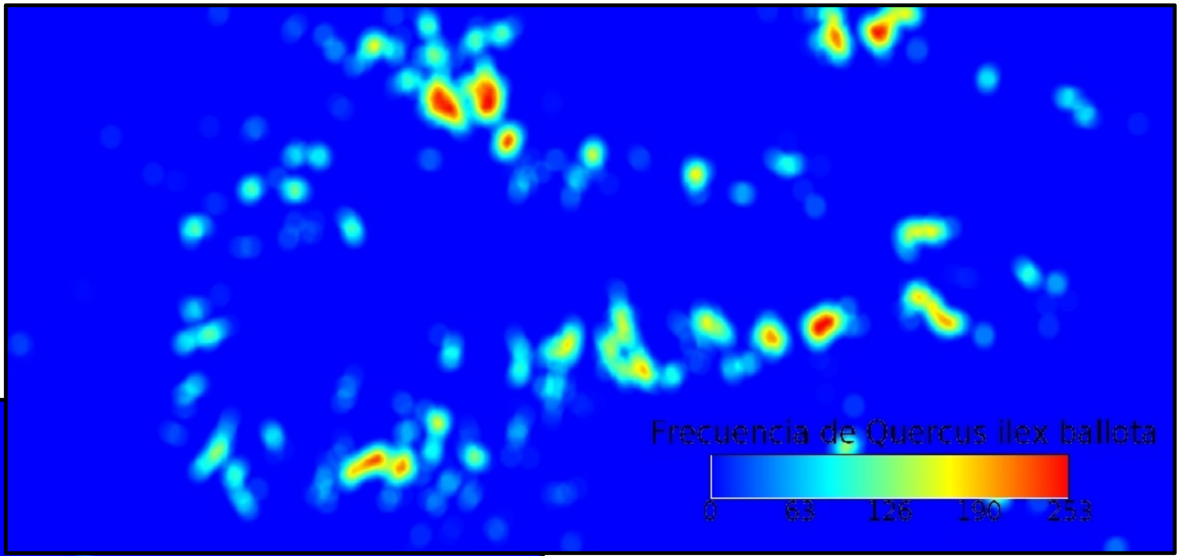


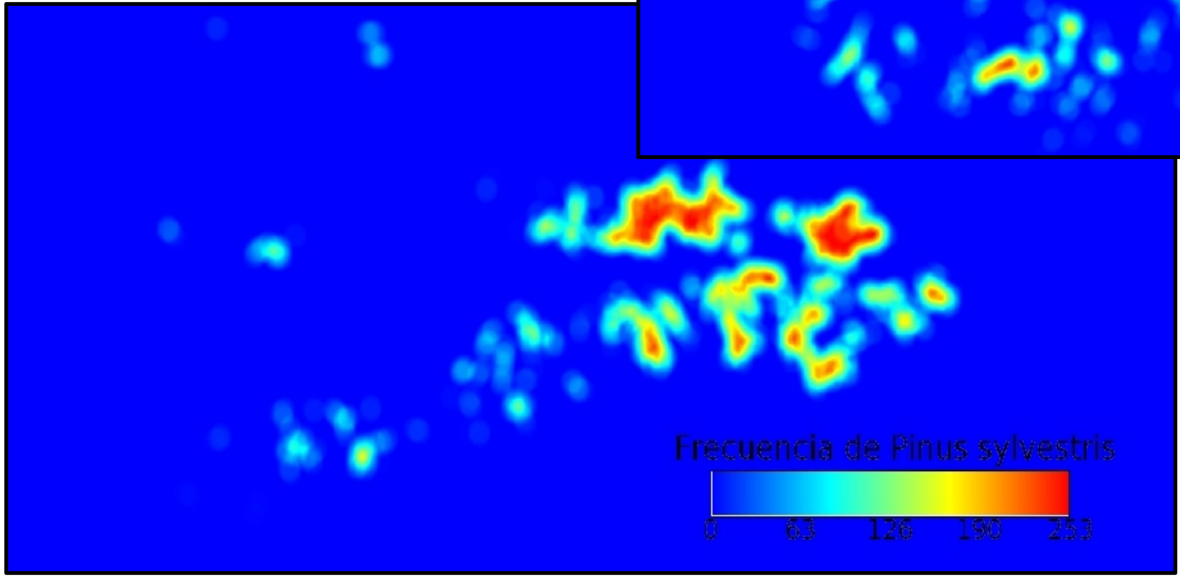
QUALITATIVE PREDICTORS



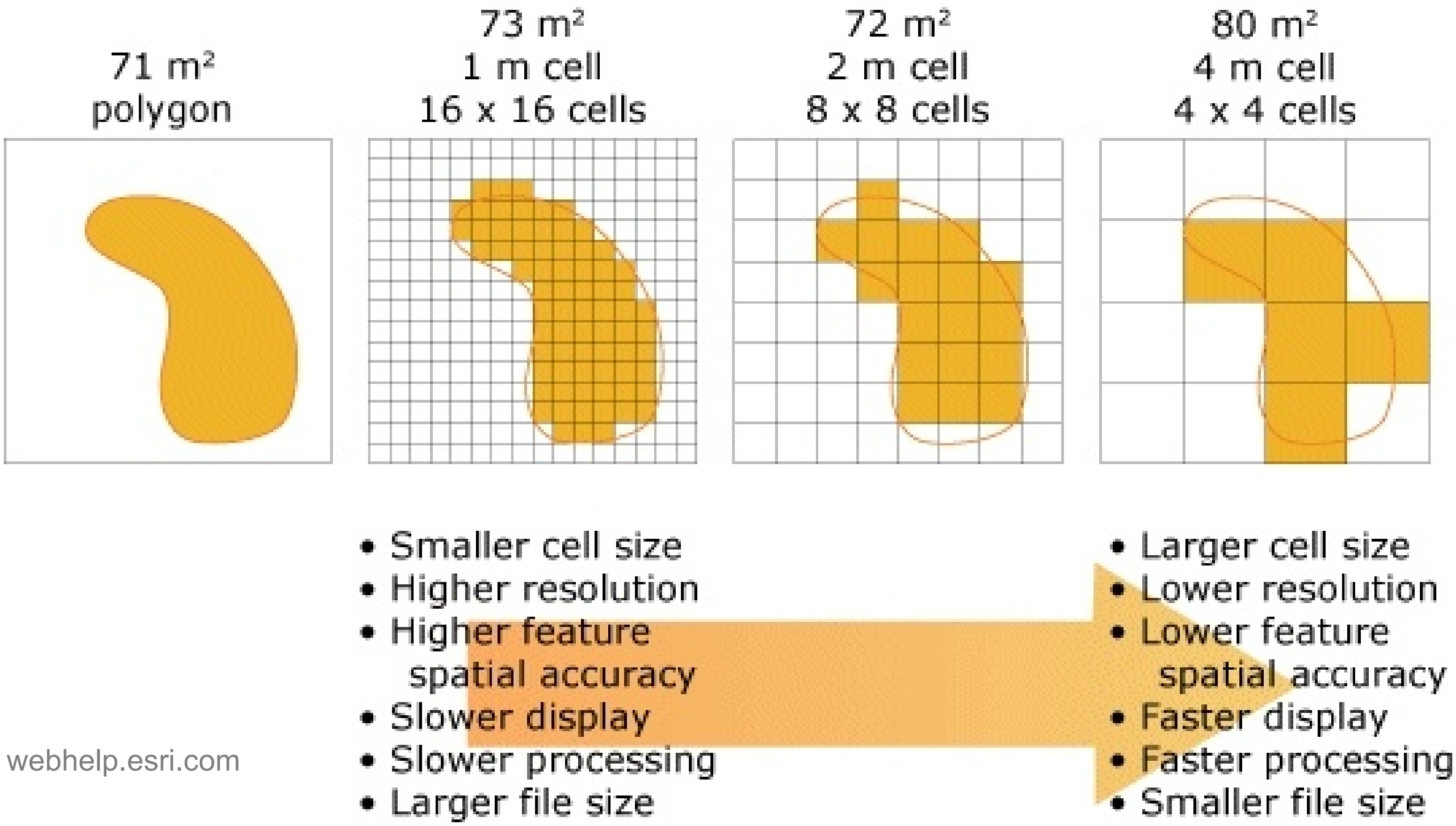


BIOLOGICAL PREDICTORS



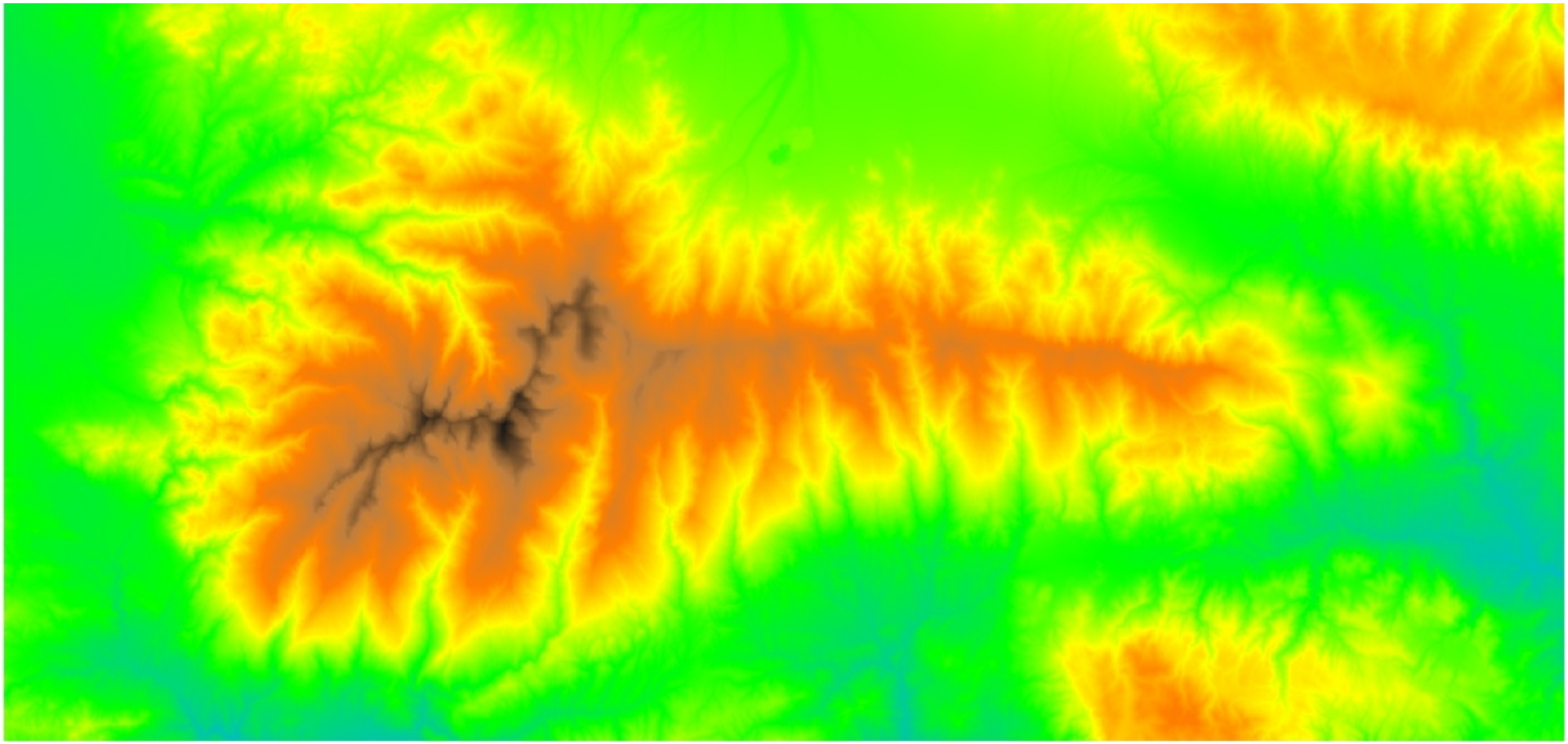


RESOLUTION

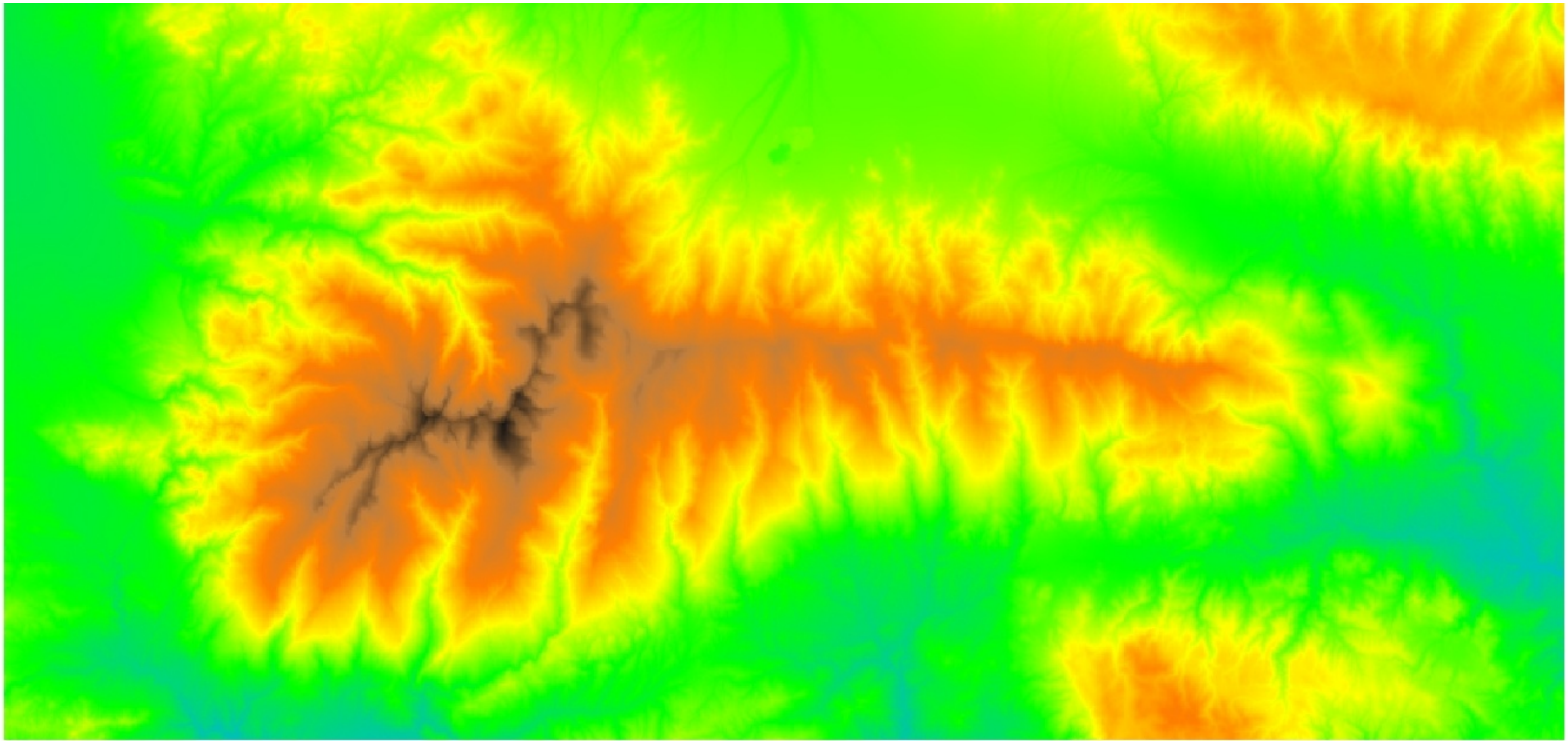


Fuente: webhelp.esri.com

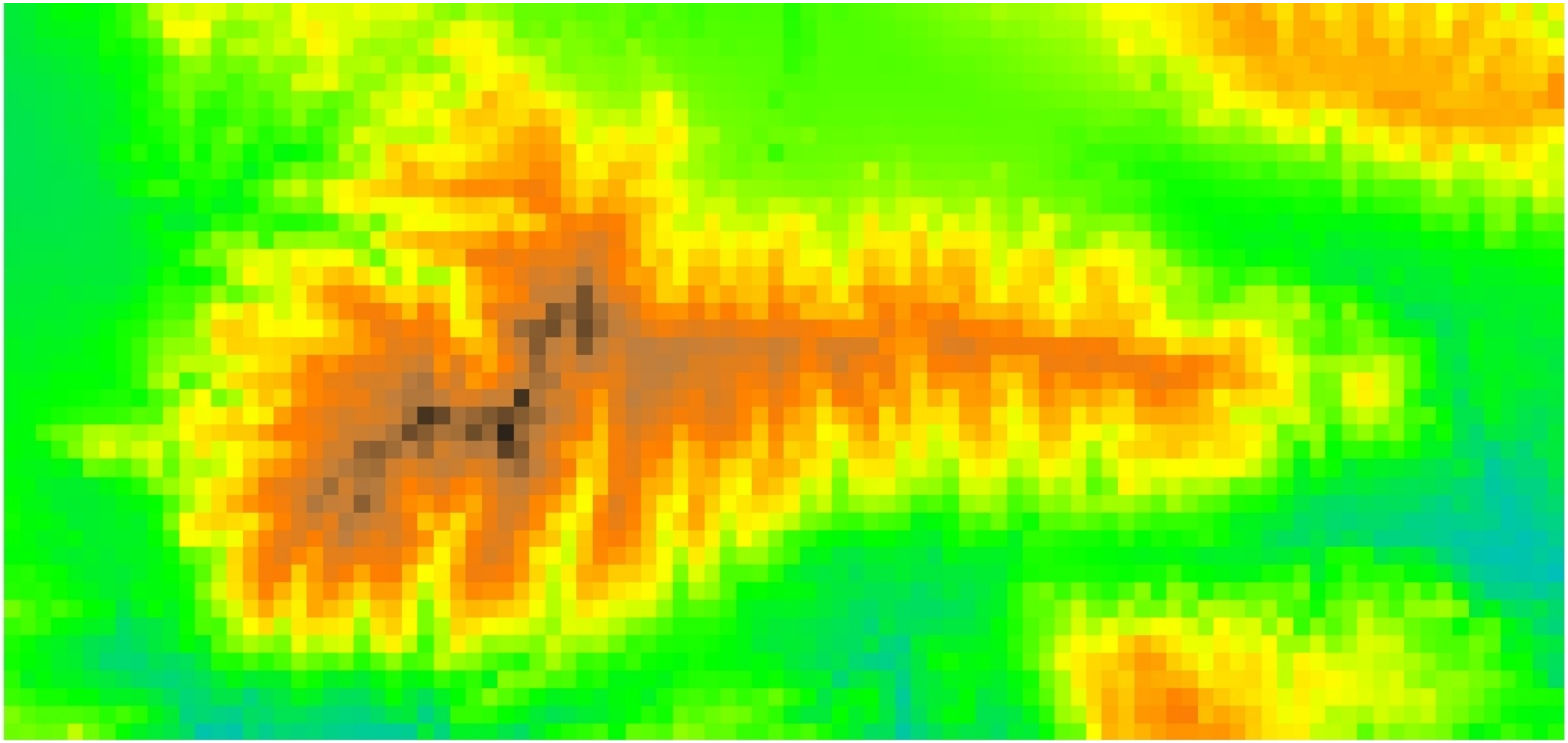
RESOLUTION (10 m)



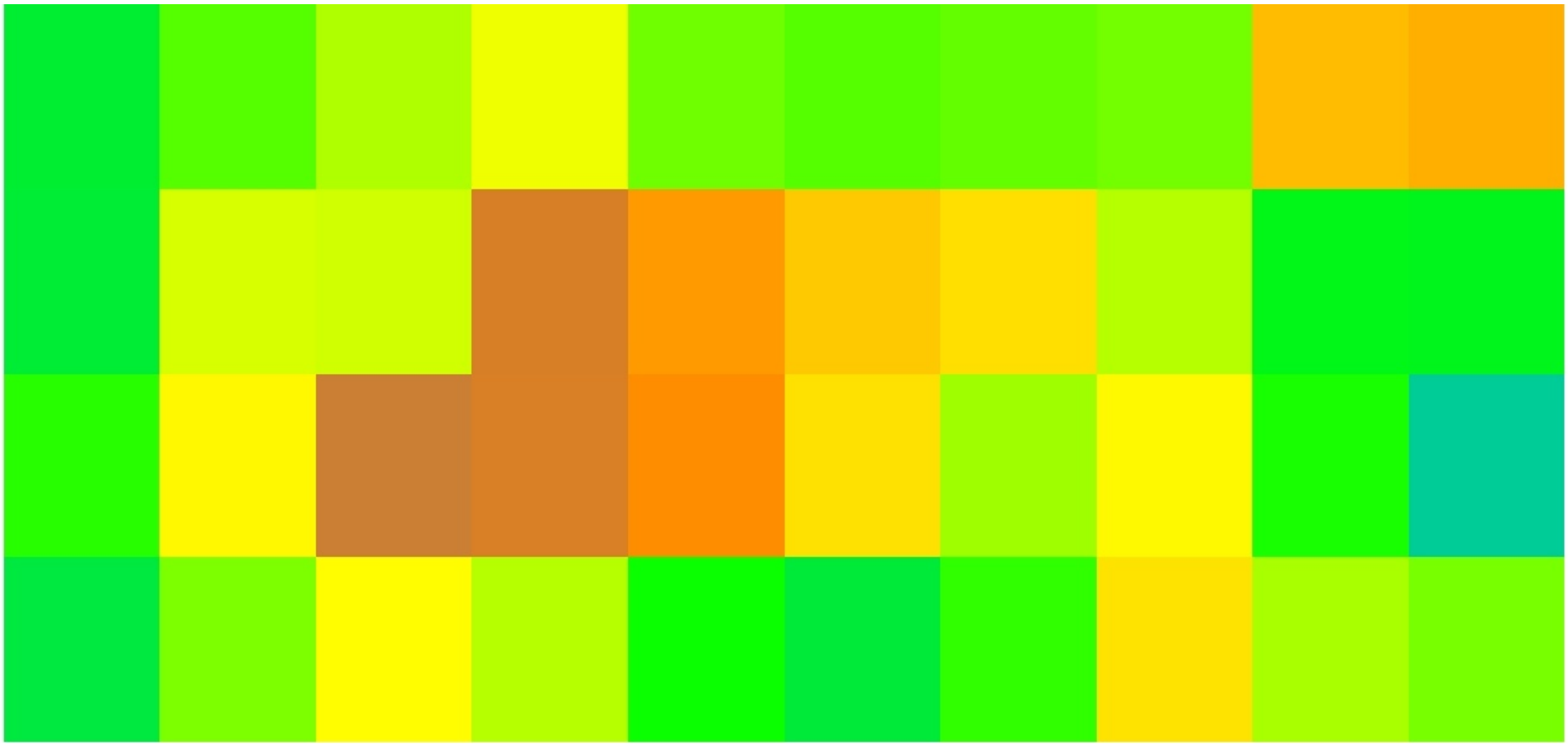
RESOLUTION (100 m)



RESOLUTION (1000 m)

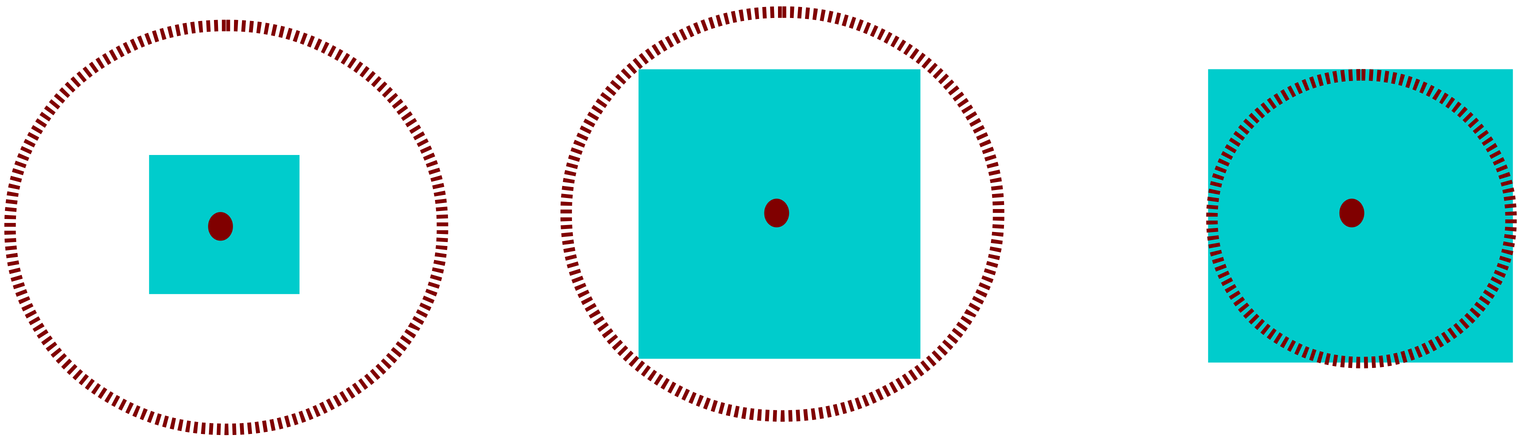


RESOLUTION (10000 m)



IMPORTANT

Resolution of predictors should match the coordinates precision of the presence records!



SOME DATA SOURCES

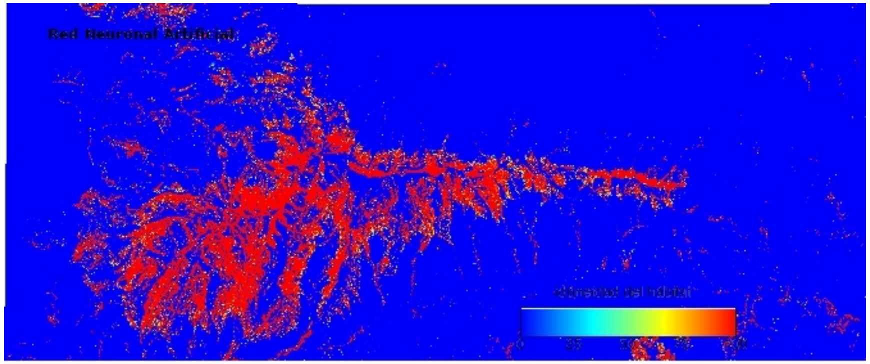
Tipo	Dataset	Sitio web
Climate	WORLDCLIM	
Topography	SRTM	
NDVI	GIMMS	
Vegetation	MODIS VCF	
Human footpr.	Human Footprint	
Land uses	GLOBCOVER	

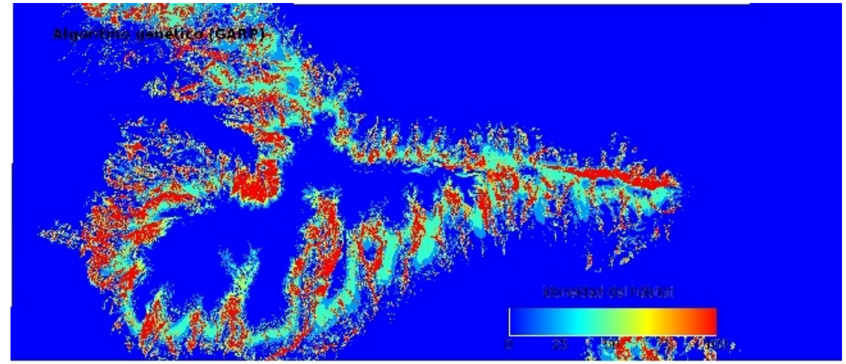
ALGORITHMS

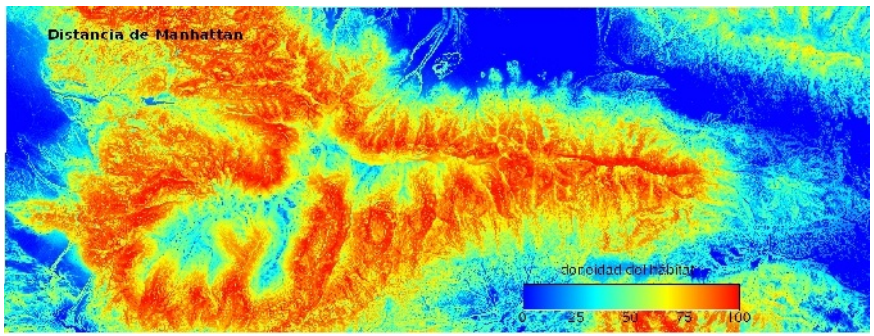
Main families of methods

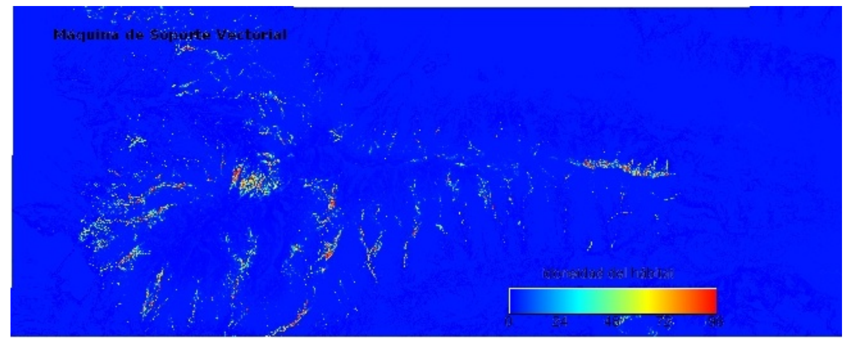
- Bioclimatic envelopes
- Ecological similarity
- Regression
- Machine learning

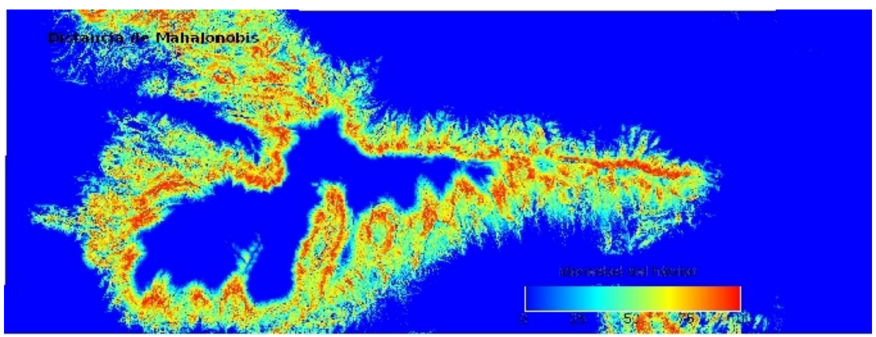
DIFF METHODS = DIFF RESULTS

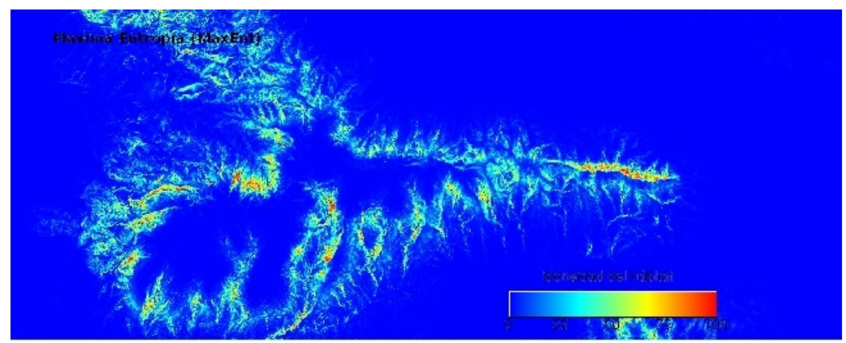




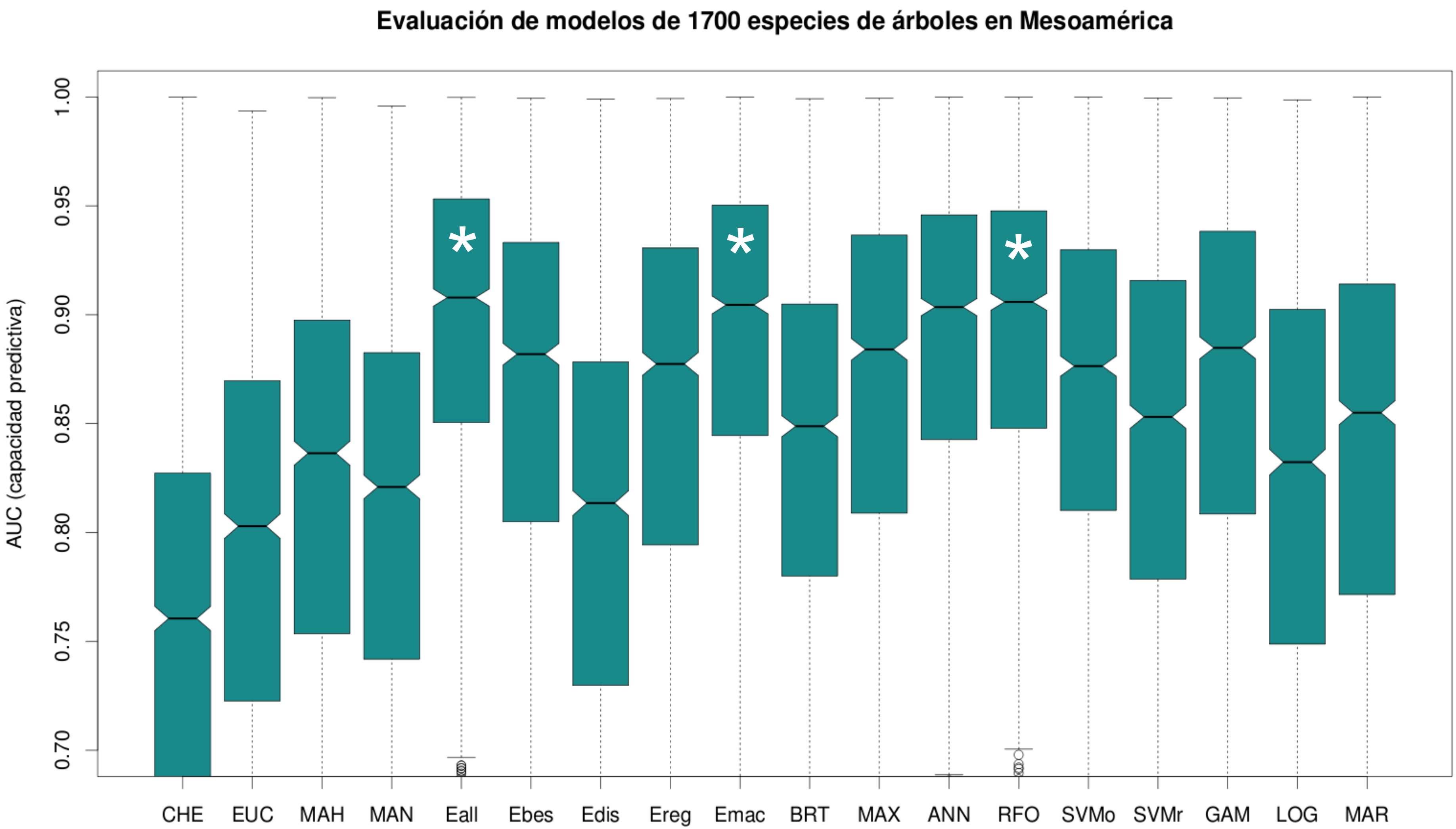




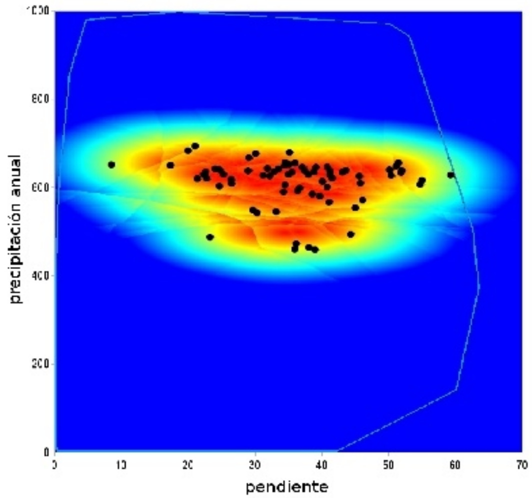




DIFFERENT PREDICTIVE ABILITIES

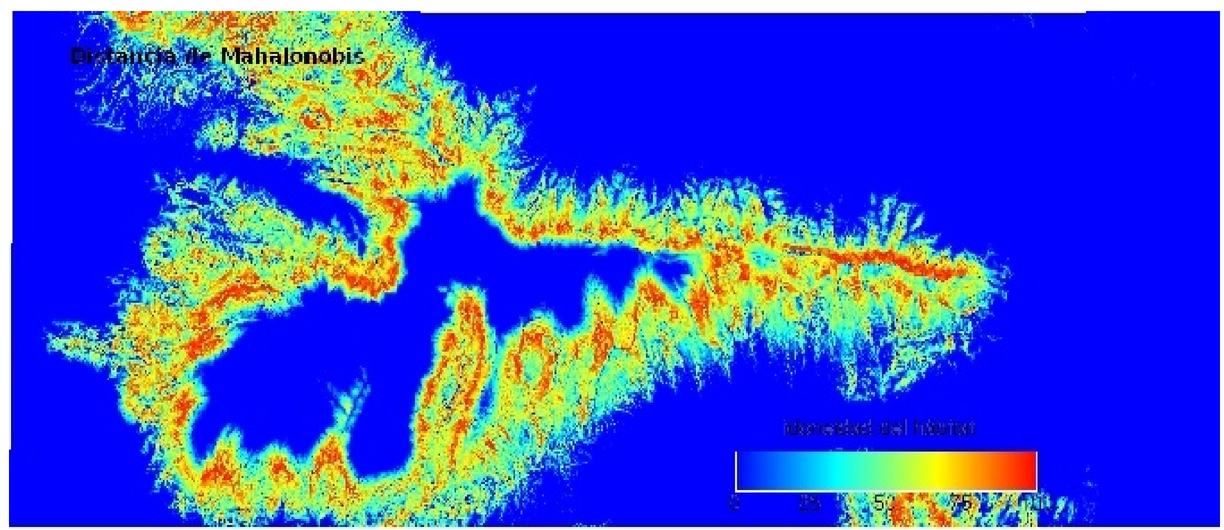


MODEL



Ecological space

Geographical space

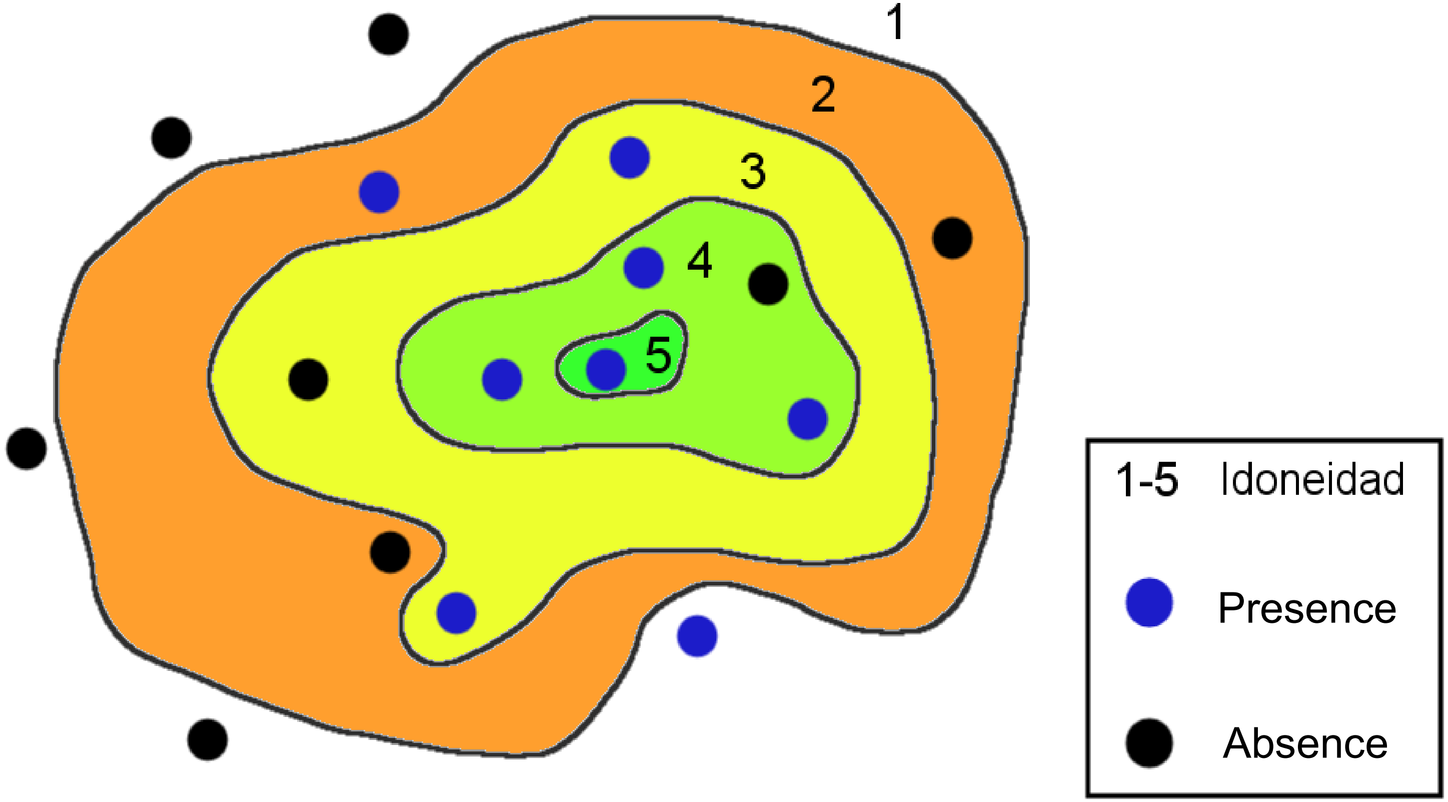


EVALUATION

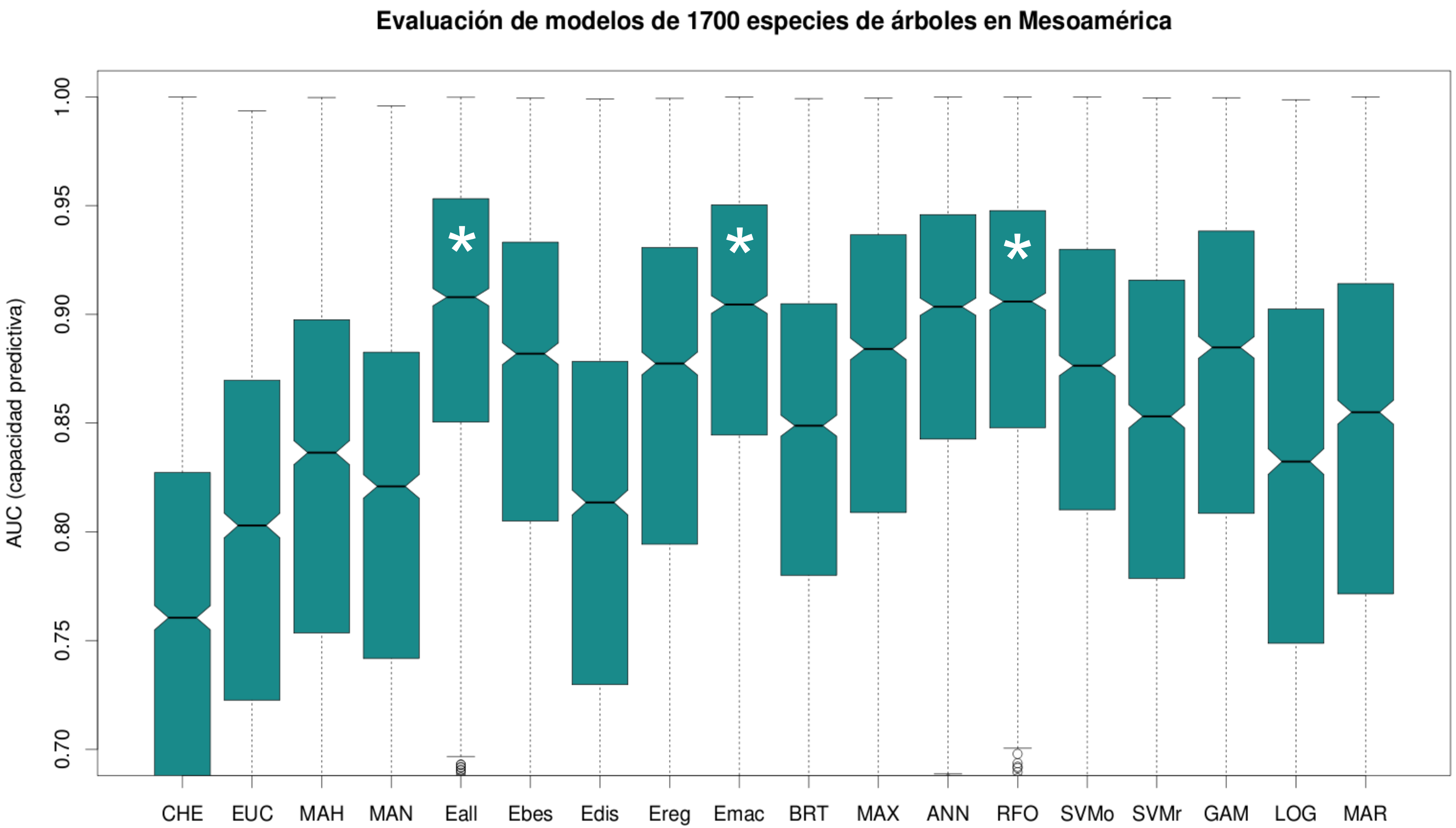
Key paper:

Fielding AH y Bell JF 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1), 38-49

EVALUATION



DIFFERENT PREDICTIVE ABILITIES



CONFUSION MATRIX

A → correctly predicted presences
D → correctly predicted absences
B → missed absences (false positives or commission error)
C → missed absences (false negatives or omission error)

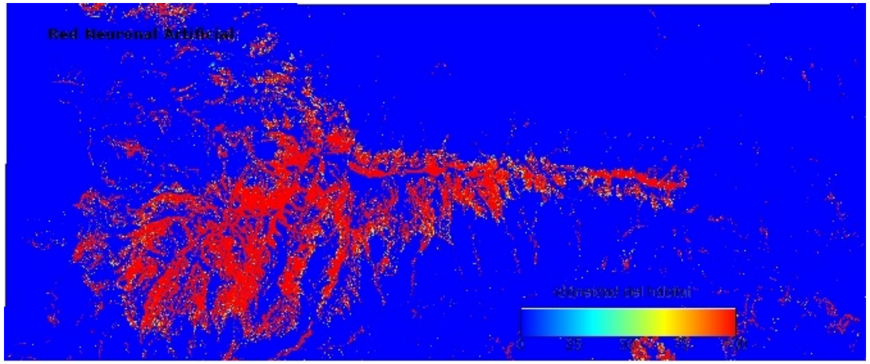
		PRESENCE DATA	
		Presence	Absence
MODEL PREDICTION	Presence	**A**	**B**
	Absence	**C**	**D**

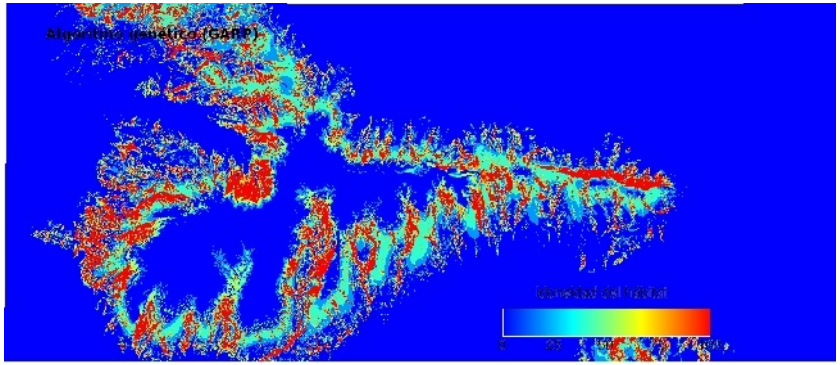
sensitivity → $S = A/(A+C)$

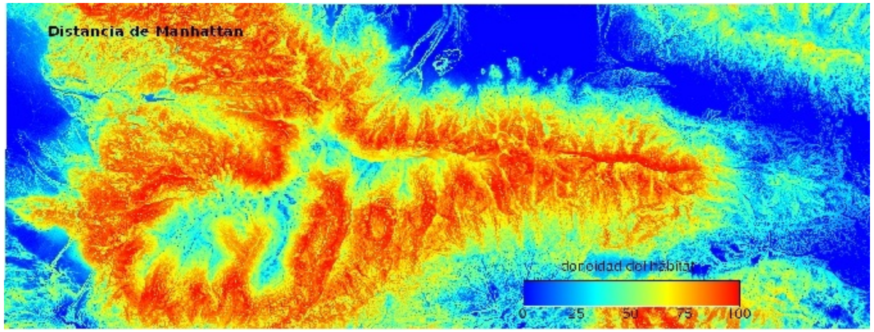
specificity → $E = D/(B+D)$

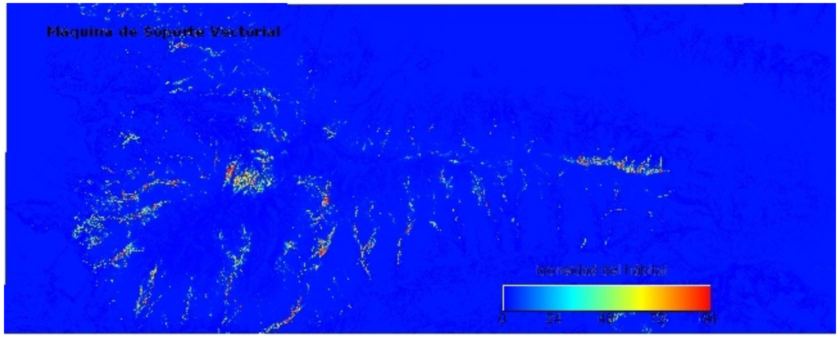
true skill statistic → $TTS = S + E - 1$

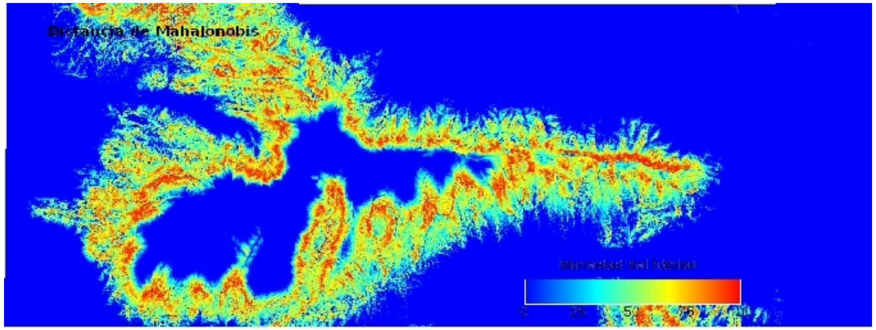
ENSEMBLE MODEL FORECASTING

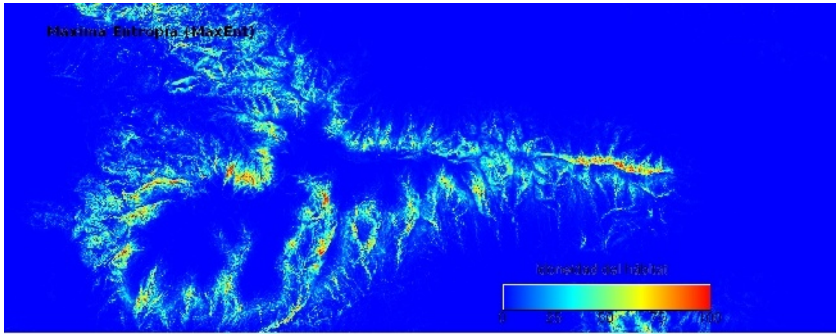






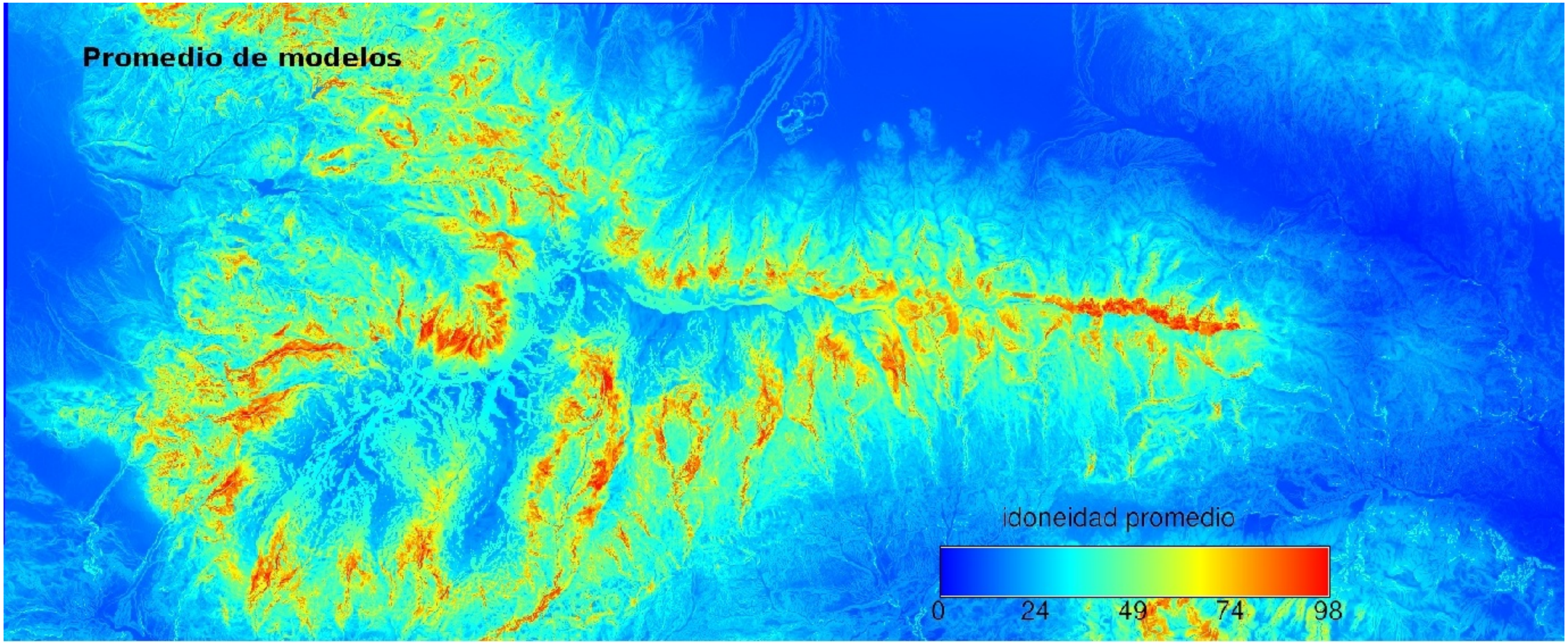




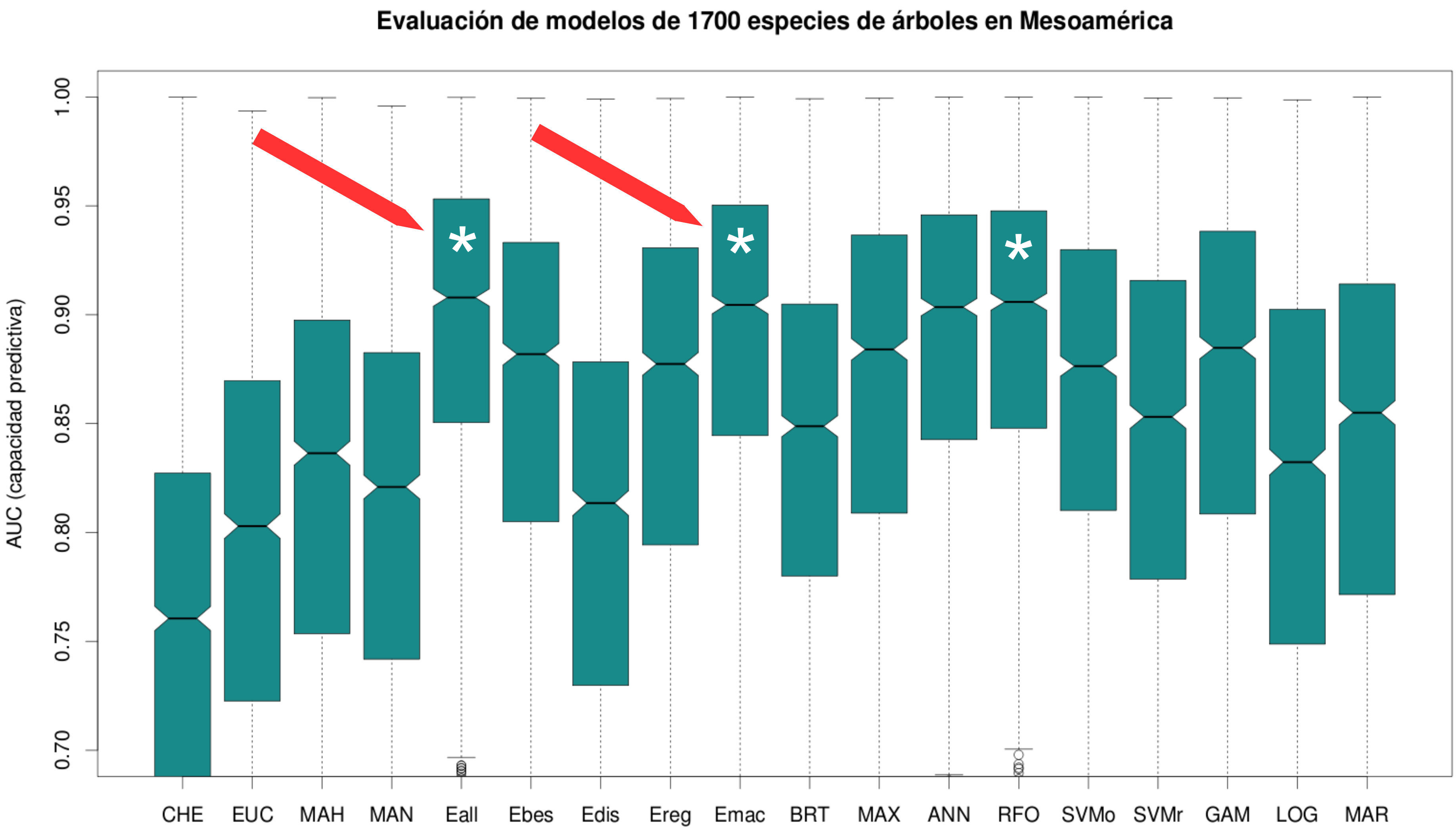


ENSEMBLE MODEL FORECASTING

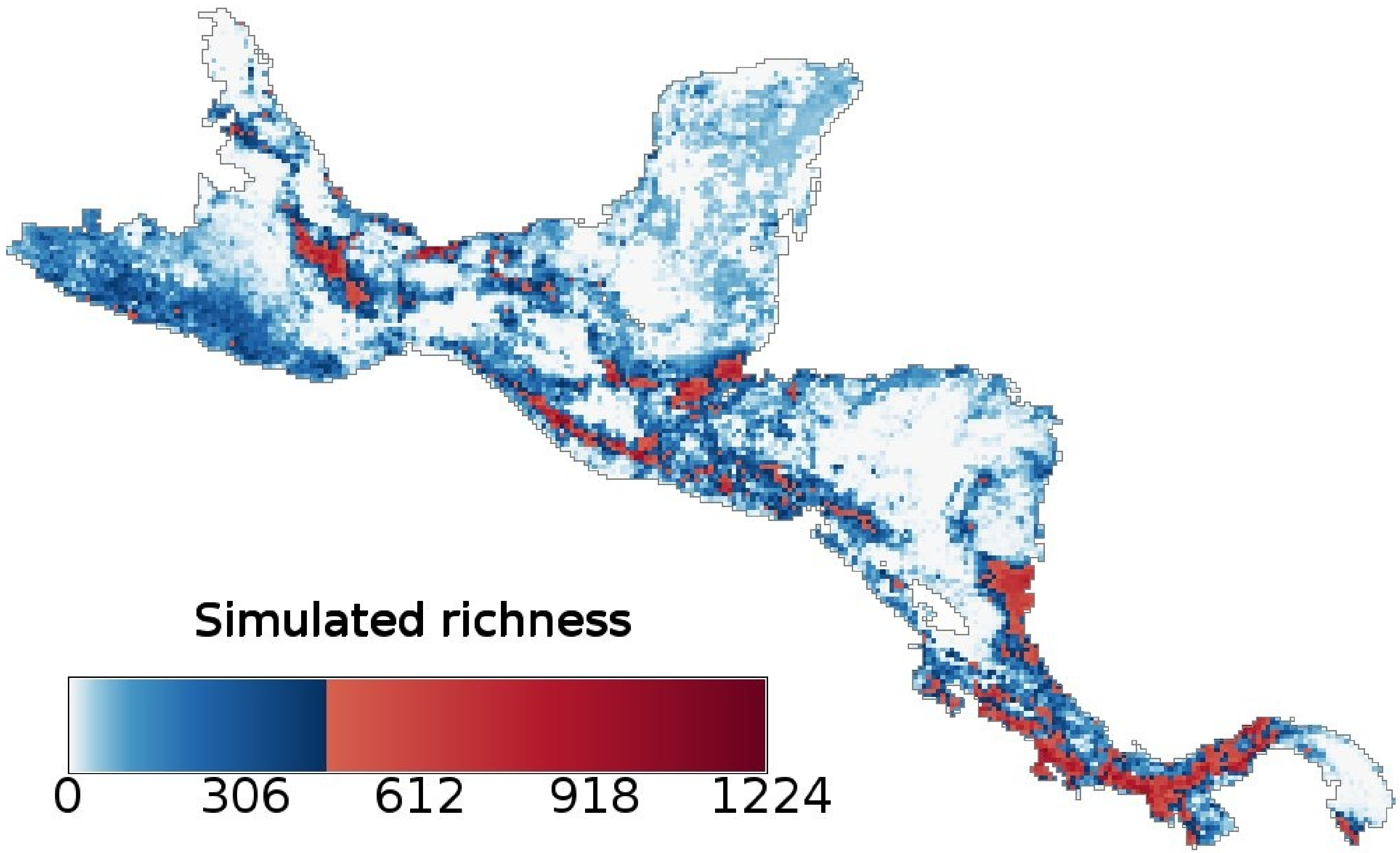
AVERAGE OF THE PREVIOUS MODELS



ENSEMBLES ACTUALLY WORK!

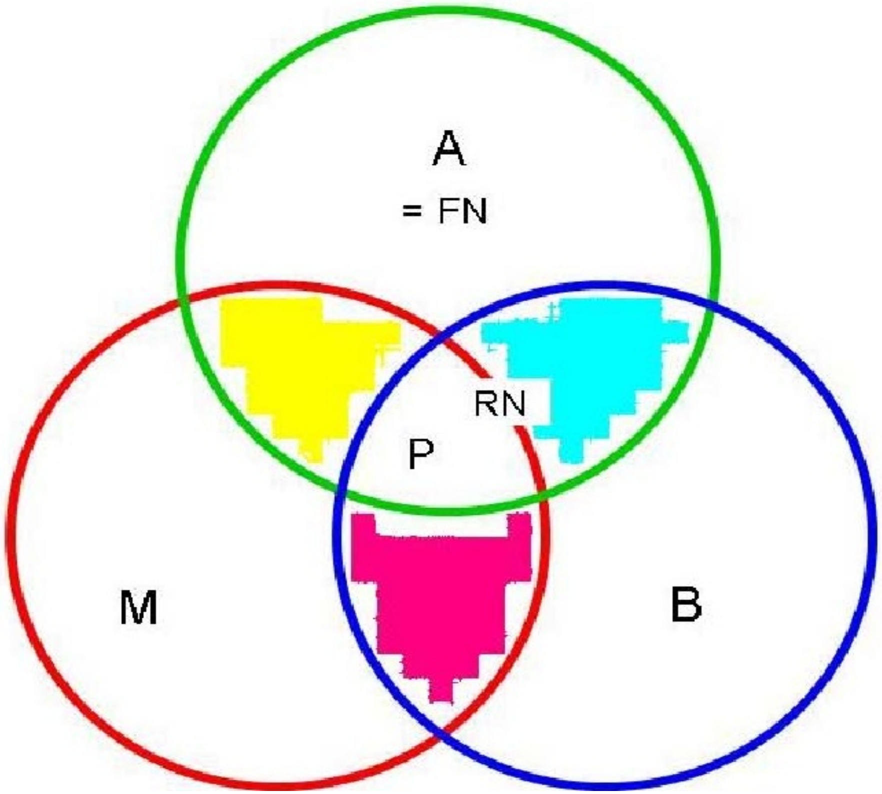


RICHNESS MODELS



Benito et al. 2013 MEE

INTERPRETATION



BAM diagram

- **A** – abiotic environment = **FN** (fundamental niche)
- **B** – biotic environment
- **M** – accessible habitat
- **RN** – ***realized niche***
- **P** - ***presence***

Soberón 2005 Biodiversity Informatics

INTERPRETATION

Soberón (2005):

"*Whether the result is interpreted as the species' distribution, the spatial extent of its fundamental niche, or some other phraseology, these algorithms* ***only find regions that 'resemble' those where occurrence points are located****.*"

APPLICATIONS
OF
SPECIES DISTRIBUTION MODELS

VEGETATION MAPPING

Box *et al.* (1981) Predicting physiognomic vegetation types with climate variables. *Vegetatio* 45, 127-139

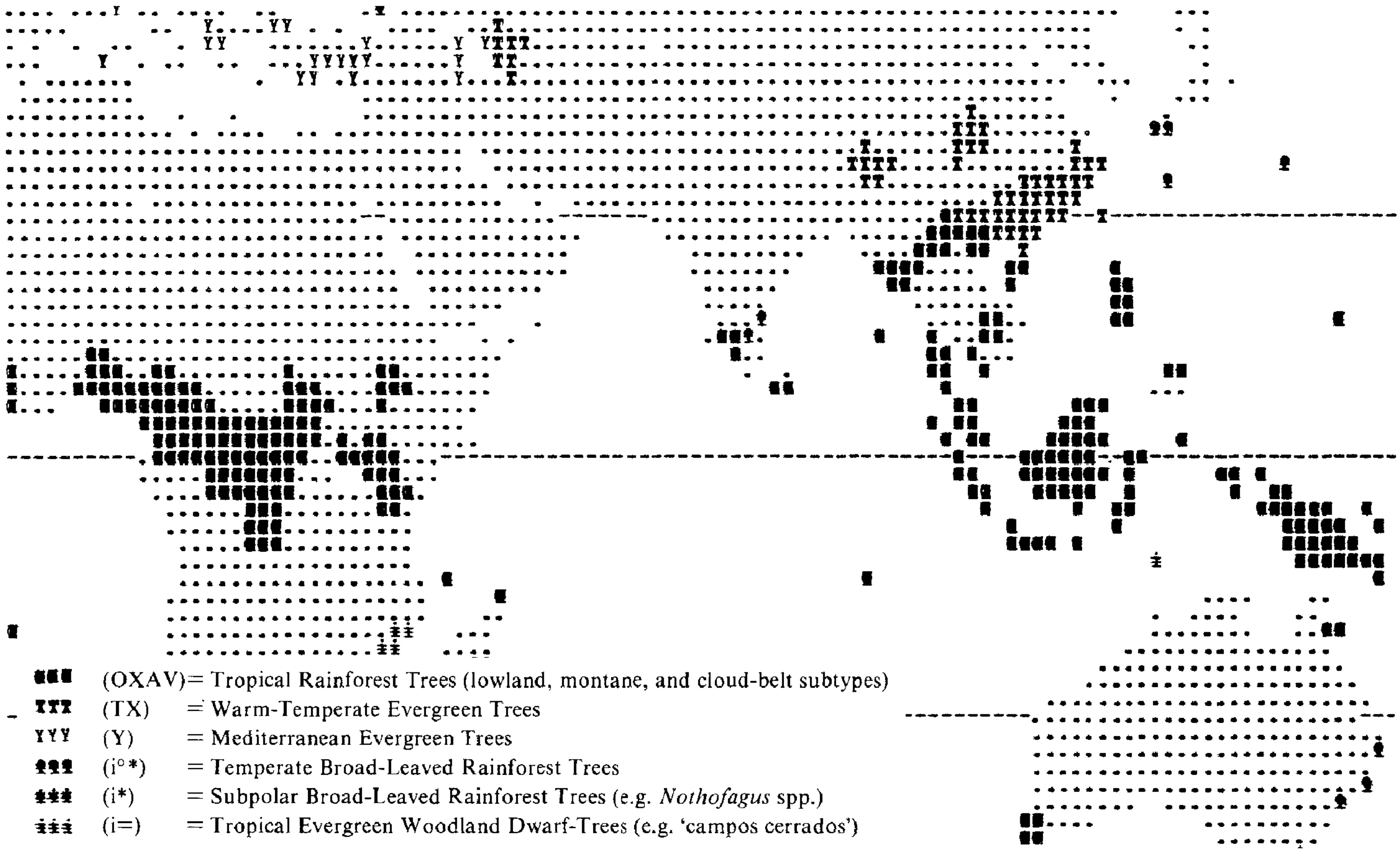
- **Objective**: mapping of broad vegetation types
- **Methods**: bioclimatic envelopes
- **Results**: 85 % success in predicting plant types and 50 % in predicting vegetation structure.
- **Conclusion**: Simple models can predict vegetation patterns at continental to global scales with remarckable accuracy.

Box *et al.* (1981)

Bioclimatic envelopes

	TMAX	TMIN	DTY	PRCP	MI	PMAX	PMIN	PMTMAX
Tropical rainforest								
Max.	30.0	28.0	12.0	*****	****	*****	****	*****
Min.	20.0	18.0	0.0	1 400.	1.00	150.	5.	10.
Raingreen monsoon forest								
Max.	35.0	28.0	15.0	*****	3.00	*****	30.	*****
Min.	23.0	14.0	0.0	800.	0.90	100.	0.	20.
Summergreen broad-leaved trees								
Max.	32.0	10.0	50.0	*****	****	*****	****	*****
Min.	15.0	–20.0	8.0	300.	0.80	40.	2.	30.
Short-needled boreal/ montane trees								
Max.	22.0	3.0	60.0	*****	****	*****	****	*****
Min.	10.0	–25.0	10.0	100.	0.60	25.	0.	10.

Box *et al.* (1981)

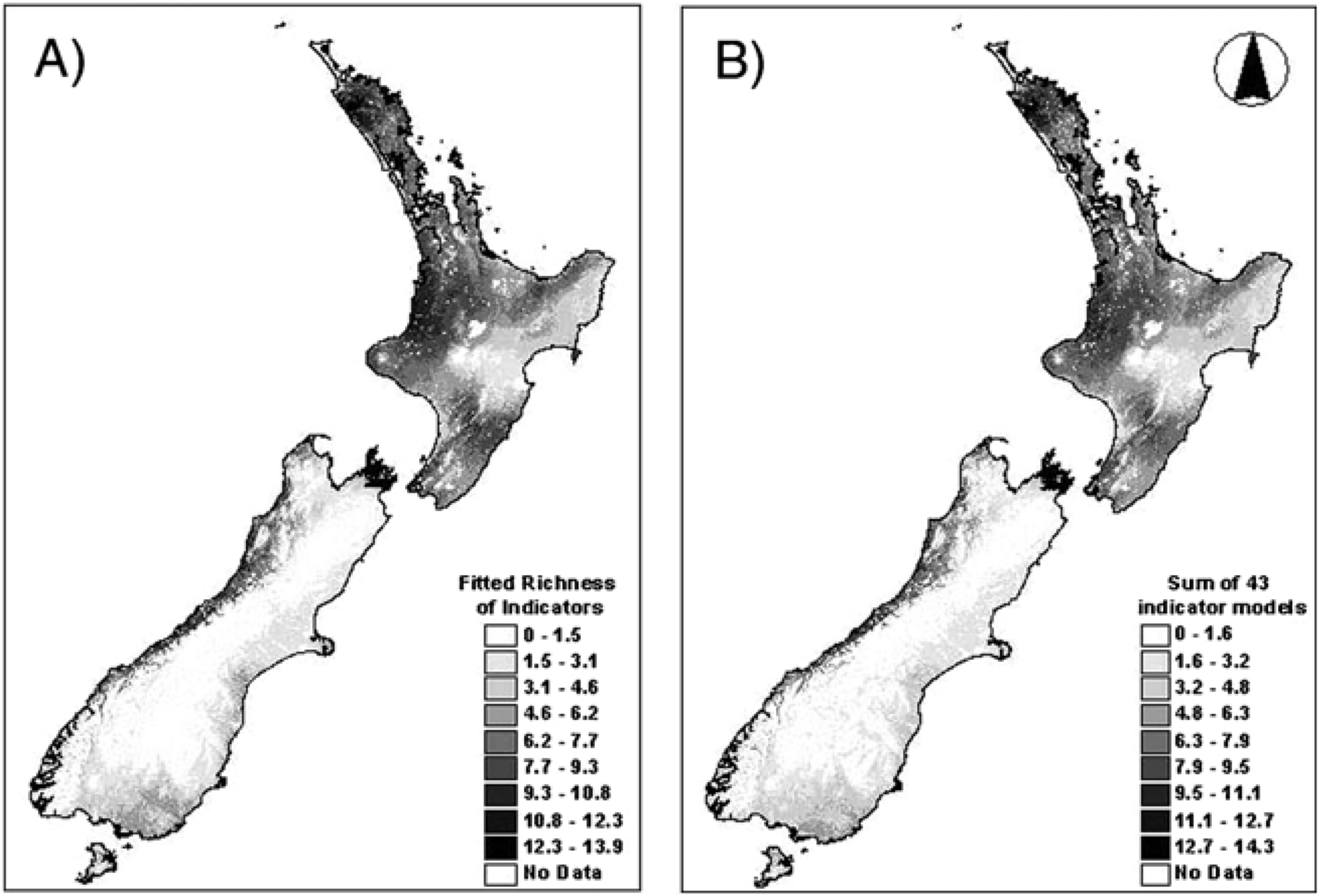


RICHNESS MODELLING

Lehmann *et al.* 2002. Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation* 11, 2217-2238

- **Objective**: Biodiversity modelling.
- **Methods**: Stacked SDMs compared with a classic richness-modelling method.
- **Results**: High agreement between models.
- **Conclusion**: Stacked SDMs allow to identify biodiversity hotspost, useful in conservation planning.

Lehmann *et al.* 2002

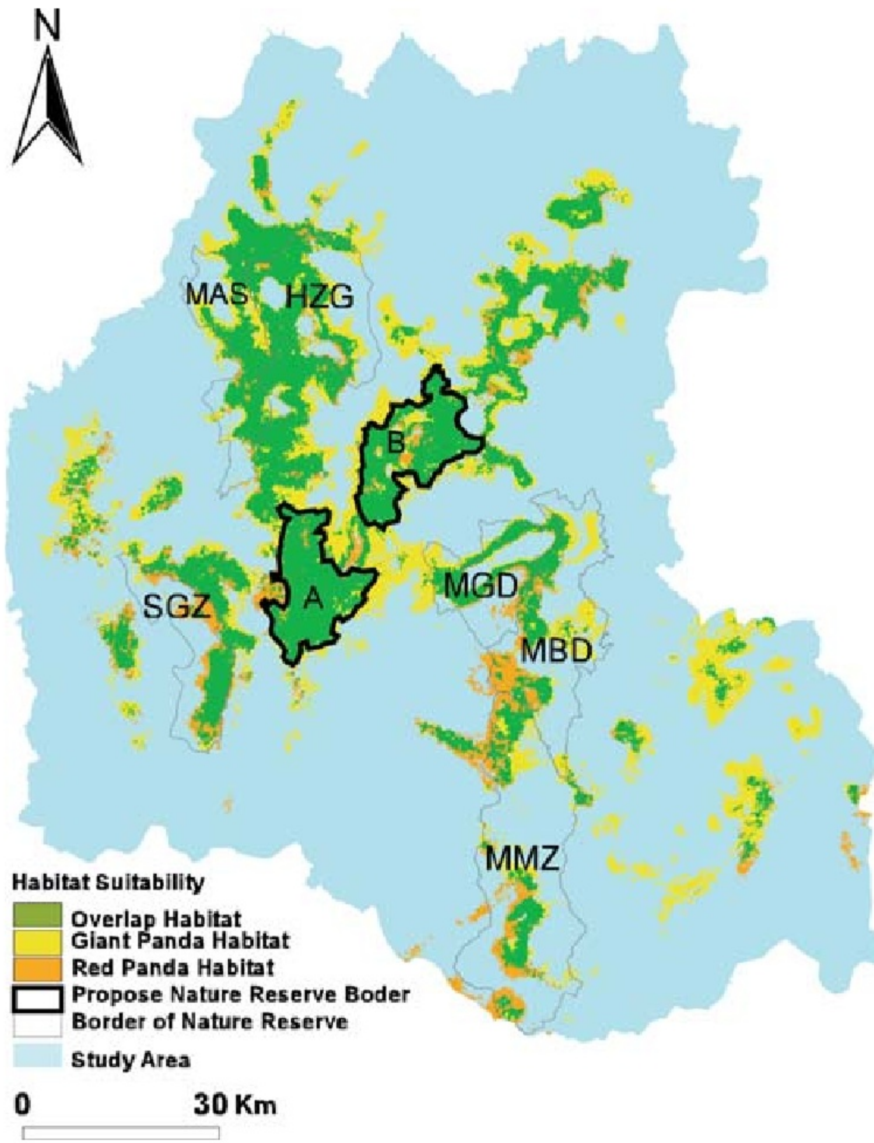


CONSERVATION PLANNING

Dunwu *et al.* 2002. Ecological niche modeling of the sympatric giant and red pandas on a mountain-range scale. *Biodiversity and Conservation* 18, 2127-2141

- **Objective**: Find suitable conservation areas for Red and Giant pandas.
- **Methods**: Overlap of SDMs of each species.
- **Results**: Current reserves have non-optimal designs, SDMs can help to make them more efficient.
- **Conclusions**: SDMs can be useful for conservation planning!

Dunwu et al. 2002

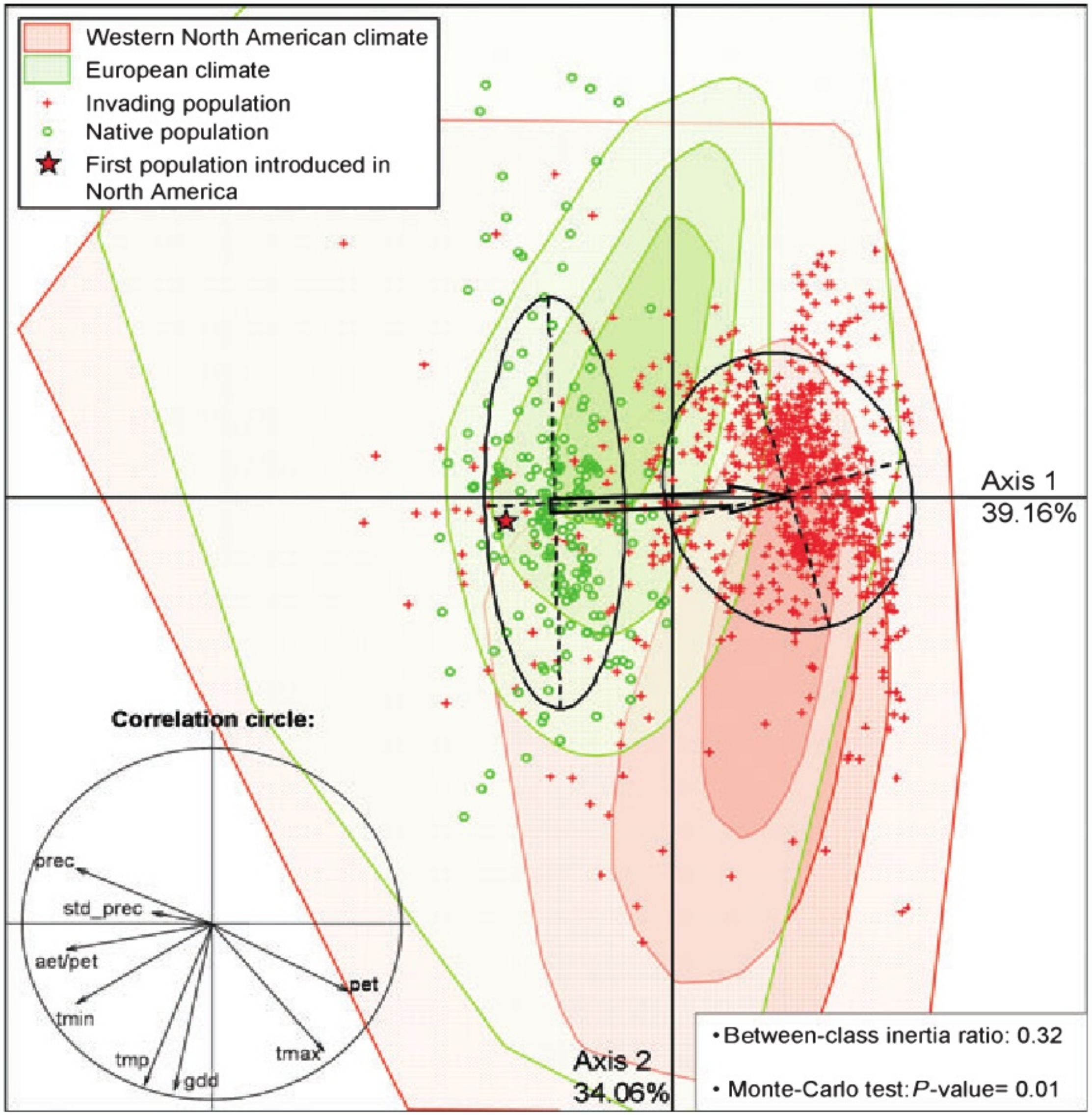


UNDERSTANDING BIOLOGICAL INVASIONS

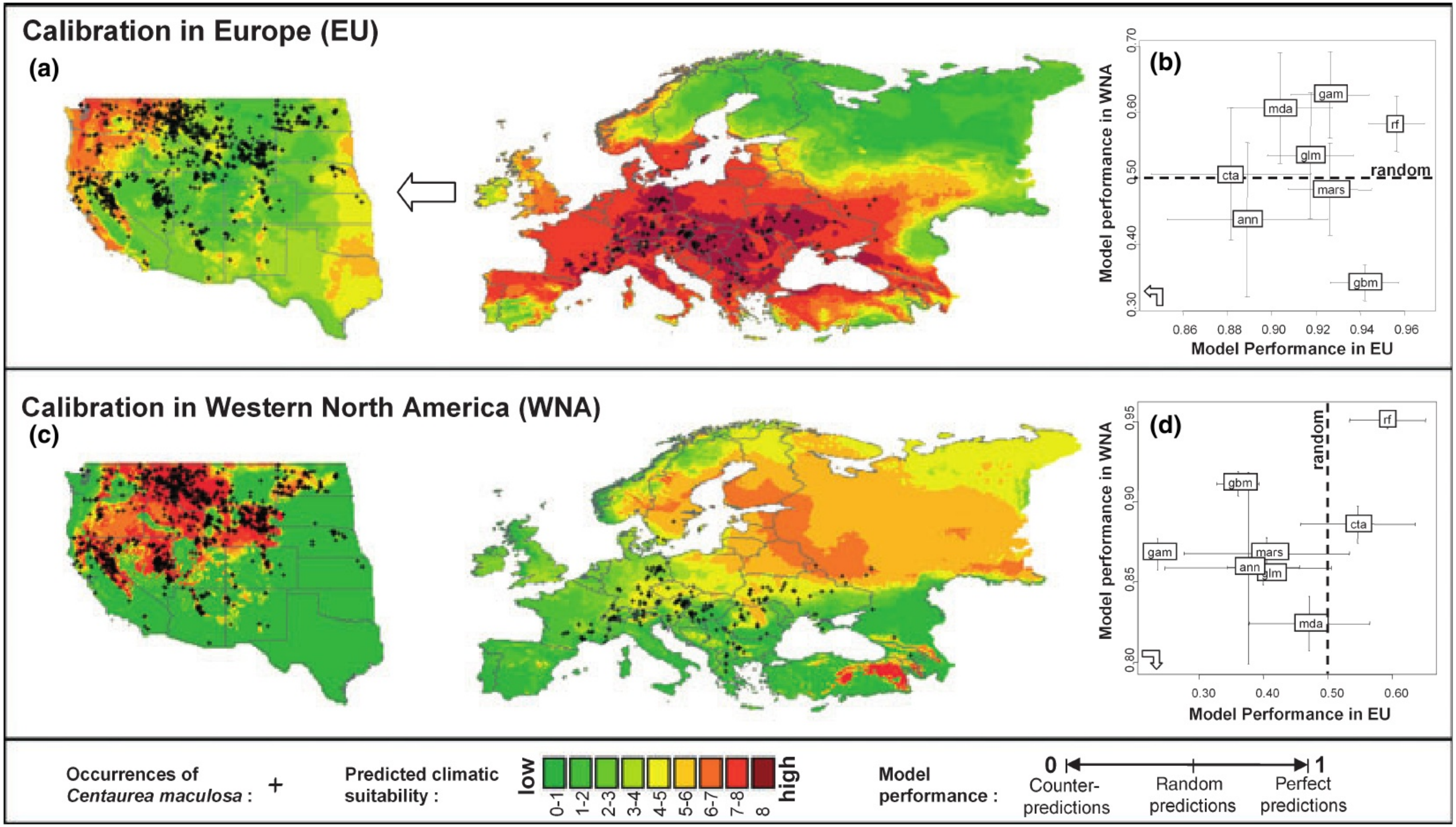
Broenninmann *et al.* 2007. Evidence of climatic niche shift during biological invasions. *Ecology Letters* 10, 701-709

- **Objective**: Evaluate how niche dynamics promotes invasions.
- **Methods**: SDMs of an invasive species in native and invaded ranges.
- **Results**: The target species expanded its niche after invasion.
- **Conclusion**: SDMs can predict introduction areas, but can have issues to forecast invaded ranges due to niche expansion.

Broenninmann *et al.* 2007



Broenninmann *et al.* 2007

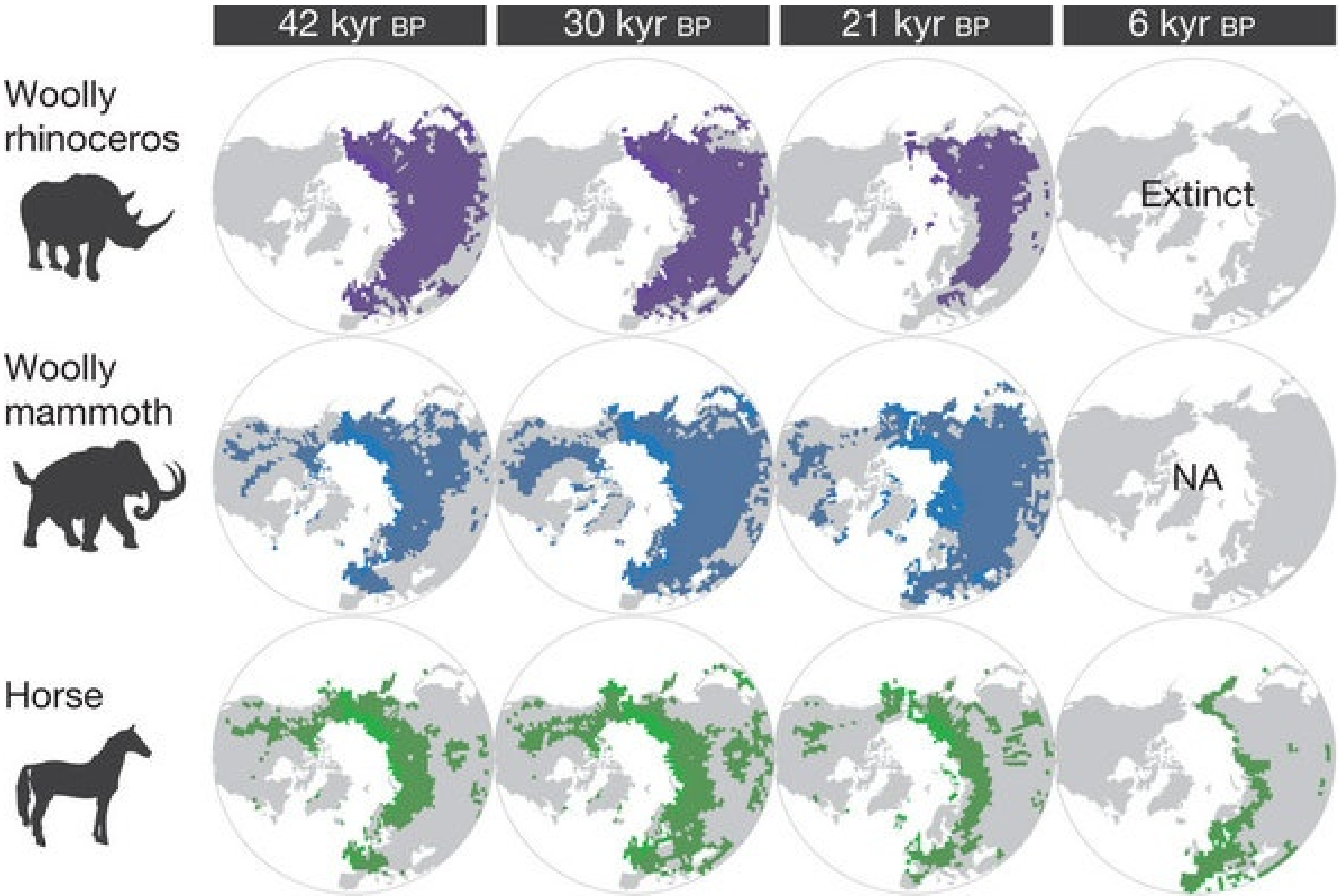


PALAEODISTRIBUTION ANALYSIS

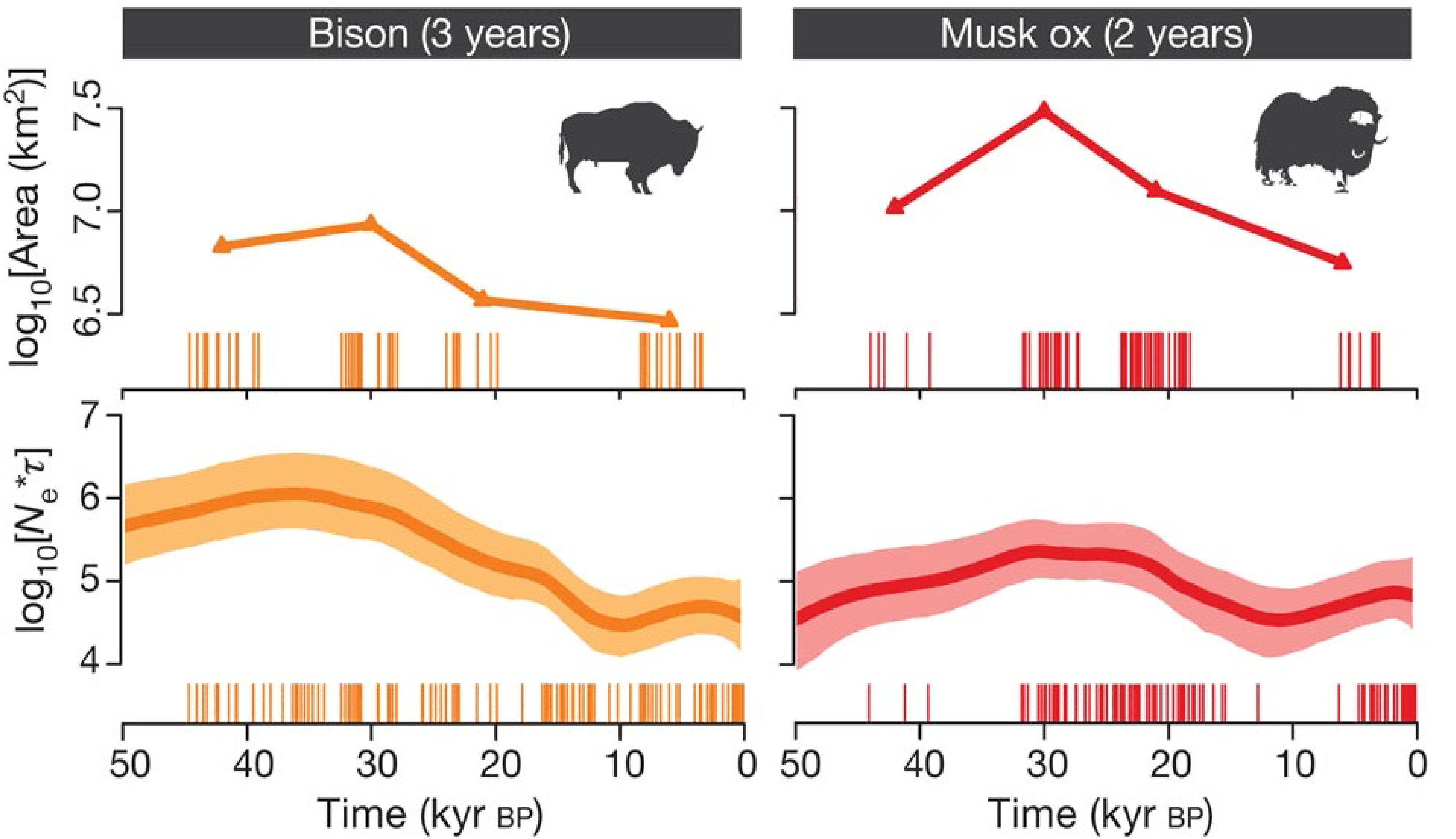
Lorenzen *et al.* 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479, 359-364

- **Objective**: Test the effect of climate and humans on the distribution of Quaternary megafauna.
- **Methods**: SDMs, ancient DNA, fossil record.
- **Results**: Different species showed different responses to climate change and huma presence.
- **Conclusiones**: SDMs coupled with other methods can provide new insights on palaeoecological dynamics.

Lorenzen *et al.* 2011



Lorenzen *et al.* 2011

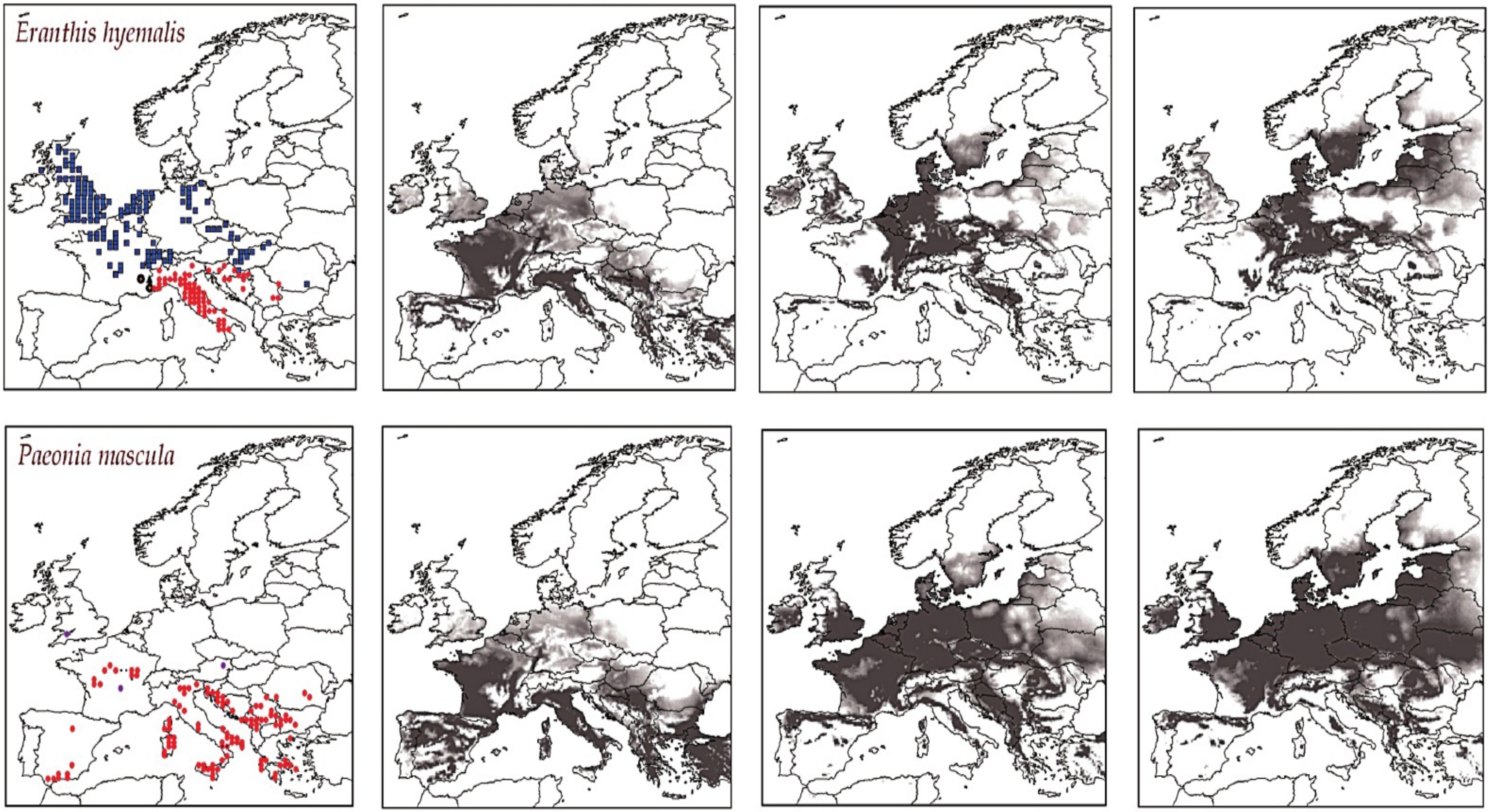


RANGE SHIFT UNDER CLIMATE CHANGE

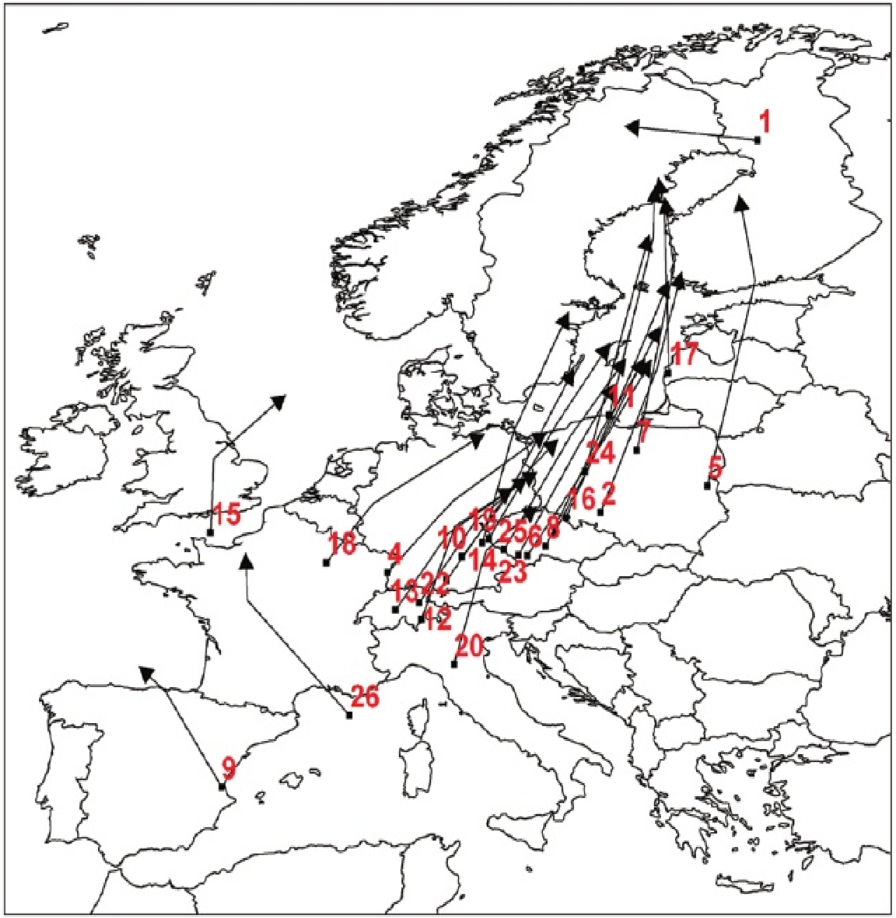
Skov y Svenning 2004. Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography* 27, 366-380

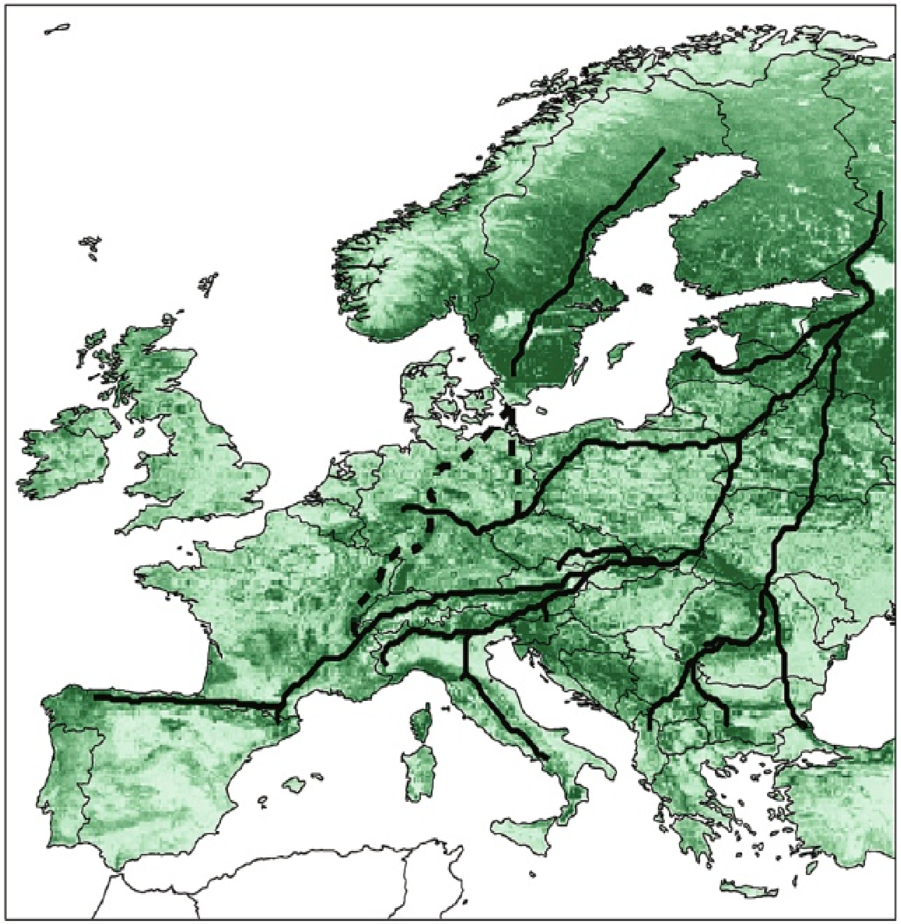
- **Objective**: Forecast range shift under climate change in Europe.
- **Methods**: Fuzzy bioclimatic envelopes.
- **Results**: Range reduction for all studied species.
- **Conclusions**: Low dispersal rates can hinder the ability of species to track their bioclimatic niche.

Skov y Svenning 2004



Skov y Svenning 2004





R

PRIMEROS PASOS CON R y Rstudio

- Tools>Global Options> General> untick “Restore .Rdata
- Tools>Global options>Pane Layout, tcik “Files” (upper right panel)
- Tools> Global options> Packages> CRAN mirror> Tick a local mirror.
- Menú Session> Set Working Directory> c:/taller1/intro_R/datos
- Open .../**taller1/intro_R/introduccion.R**

PREPARING ENVIRONMENTAL PREDICTORS AND PRESENCE RECORDS

MAIN R SCRIPT

1_prepara_variables_y_presence.R

COLLINEARITY ANALYSIS

- Two steps process:
  - Correlation matrix visualized as dendrogram to select meaningful predictors.
  - VIF (variance inflation analysis) to make sure that remaining predictors are not a linear combination of the other remaining predictors.

PREPARING PRESENCES

- Adjusting presences to the study area
- Checking taxonomic coherence
- Removing fossil record
- Remove duplicated records
- Check accuracy of coordinates
- Reduce spatial autocorrelation

PREPARING BACKGROUND, PSEUDO-ABSENCE AND ABSENCE DATA

- **Background**: comprehensive sampling of available ecological conditions (overlaps presences).
- **Pseudo-absence**: random points not overlapping presences.
- **Absences**: we don’t have real absences, but we will create them following a realistic criteria (just for pedagogic purposes!)

PRESENCE TABLES

species	x	y	var_1	var_2	var_n	presence
Ursus	26.48	63.01	456	856	...	1
Ursus	23.60	50.20	546	452	...	1
Ursus	30.20	57.34	569	145	...	1
Ursus	32.10	62.35	412	456	...	0
Ursus	28.45	27.41	158	658	...	0
Ursus	30.24	31.22	123	456	...	0
...	...	...	...	...	...	...


That's all Folks!