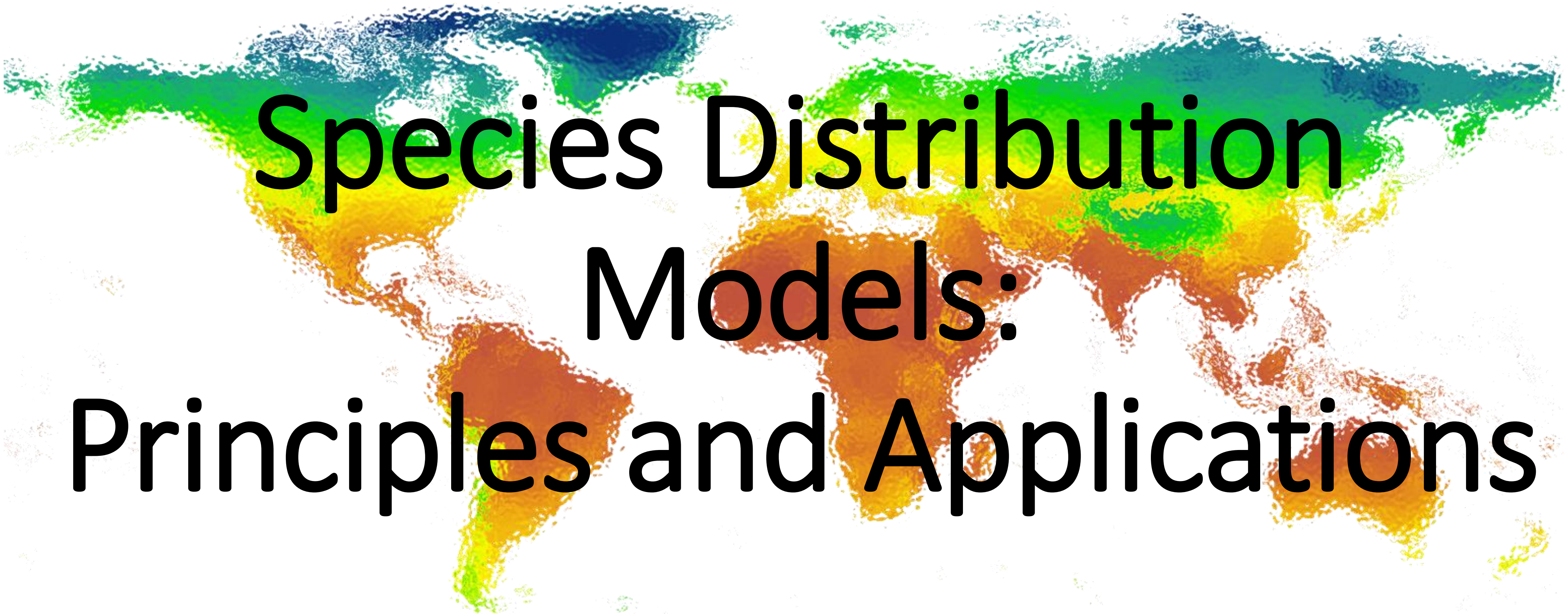


Species Distribution Models: Principles and Applications

Presenters: Nuno Cesar de Sá
Rosaleen March
Contributers: Niels Raes
Leon Marshall

CML
Institute of Environmental Sciences

Naturalis Biodiversity Center

Agenda: all about SDMs

- History
- Theory
- Principles
- Methodology
- Applications

Afternoon practical: Model your chosen species' habitat suitability under present and future climate conditions

The great debate: What is a niche???

Let's consider the concept of the niche—
If I knew what it meant I'd be rich.
It's dimensions are n
But a knowledge of Zen
Is required to fathem the bitch.
-- Grant Cottam and David Parkhurst

With your concept of niche I agree
But there's clearly one hitch I can see.
You blame the wrong sex
For the inherent hex,
For the niche is no she, but a he.
-- Joy Zedler

I'm amazed a smart woman like Joy
Would believe that a niche is a boy;
For a niche is elusive,
Deceitful, confusive –
It's quite clear it's a feminine ploy.
-- Grant Cottam

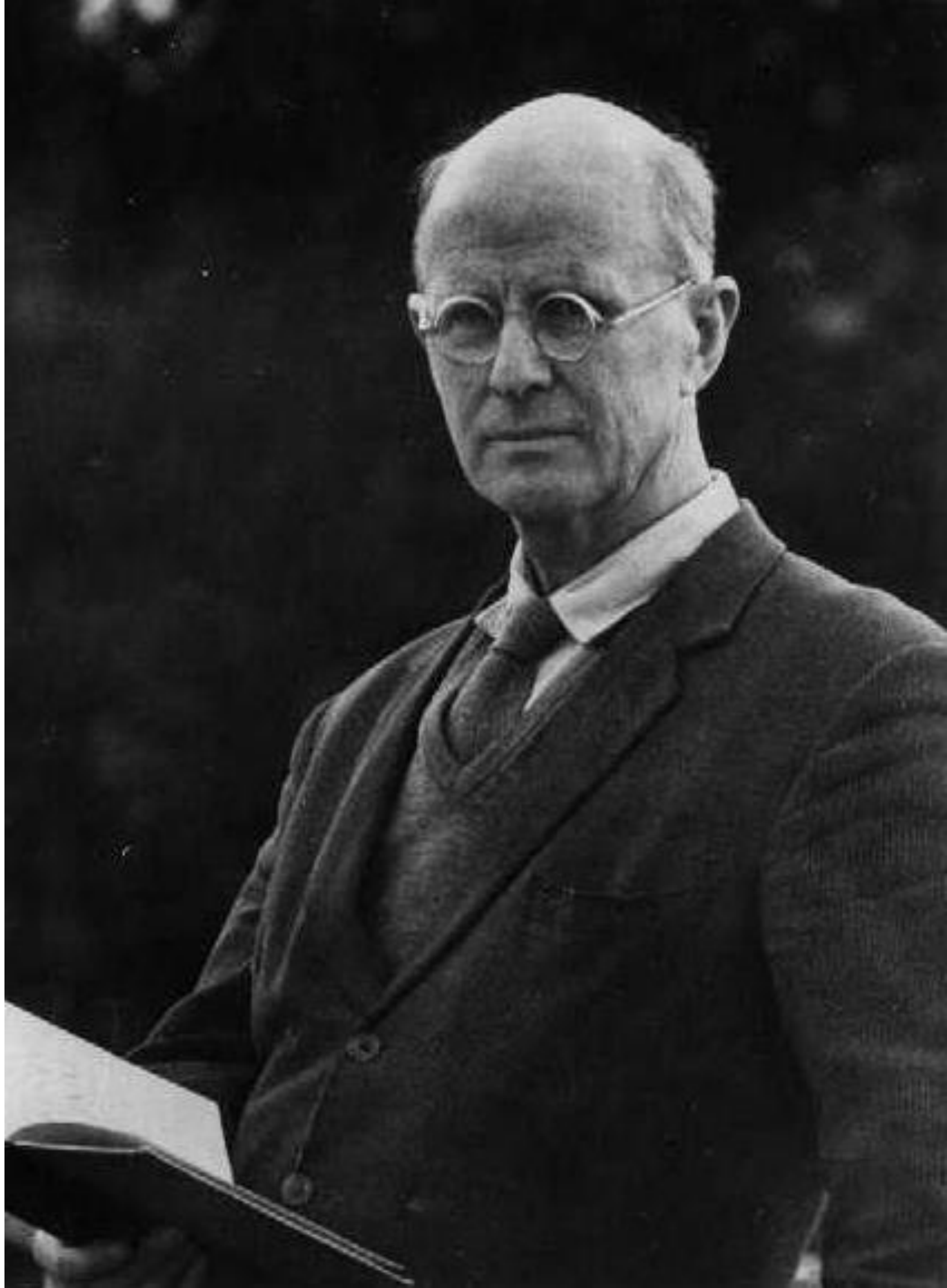
Hurlburt 1981

Joseph Grinnel (1877-1939)

- **Niche as habitat**
- First use of niche in published paper:
  - Grinnell, J. 1917. The niche relationships of the California Thrasher. The Auk 34:427-433
- First director of the Museum of Vertebrate Zoology Berkeley

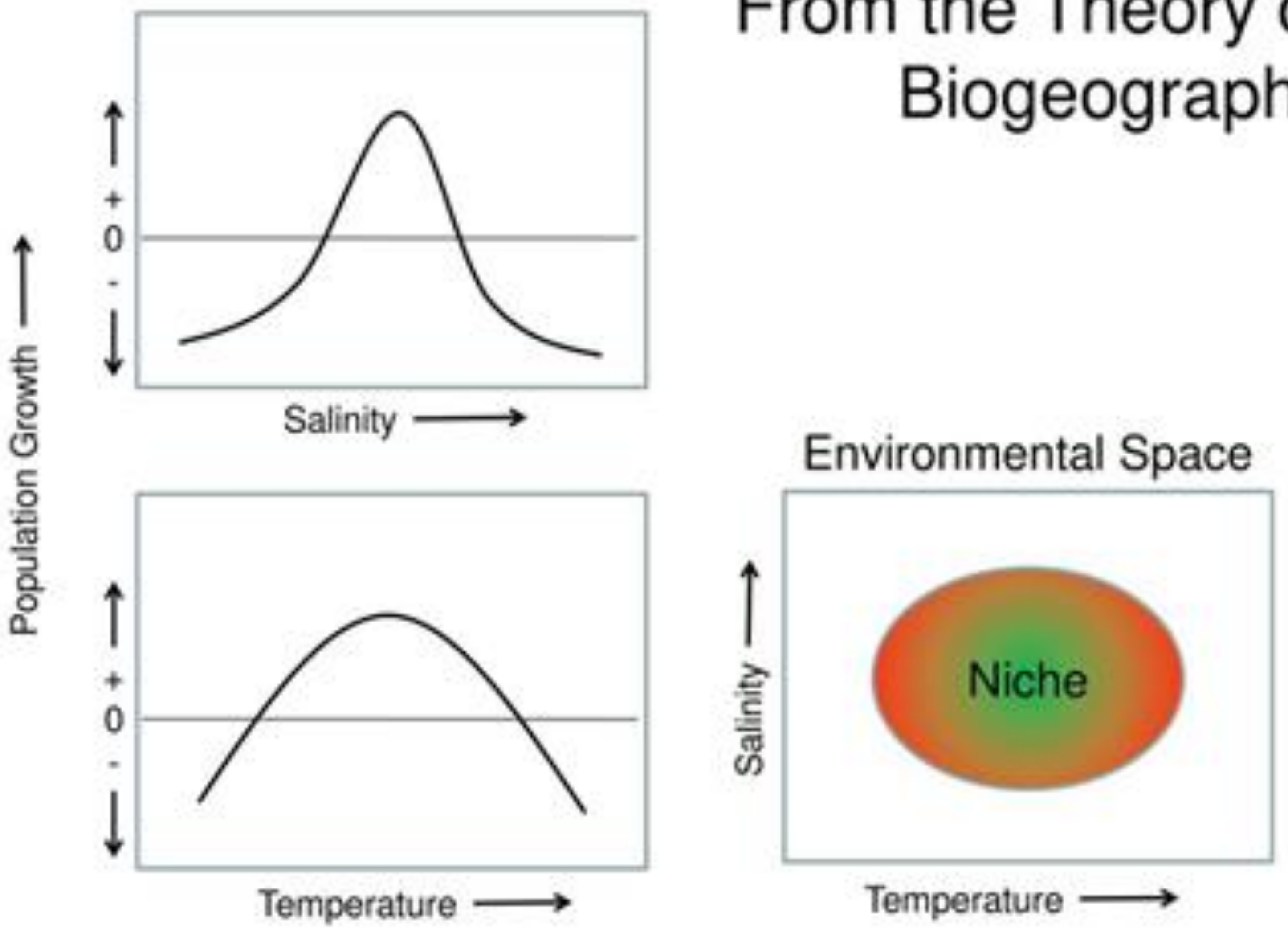
Charles Elton (1900-1991)

- **Niche as an occupation**
- Defined the niche a little differently – the role a species plays.
- Elton, C. 1927. Animal Ecology. Great Britain: William Clowes and sons Ltd.



Species are adapted to particular habitats.

From the Theory of Biogeography



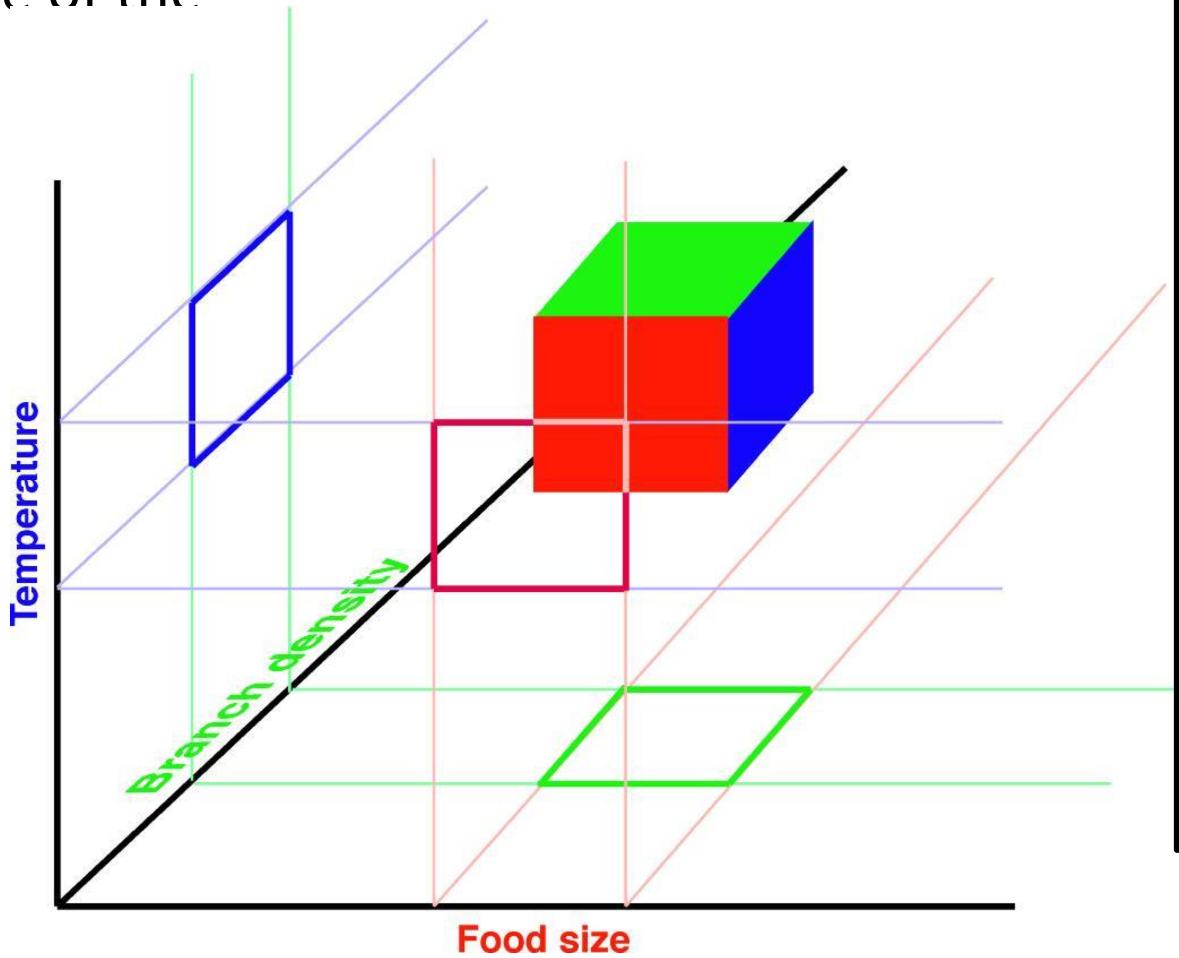
Brown, J.H., Lomolino, M.V. 1998, Biogeography: Second Edition. Sinauer Associates, Sinauer Massachusetts

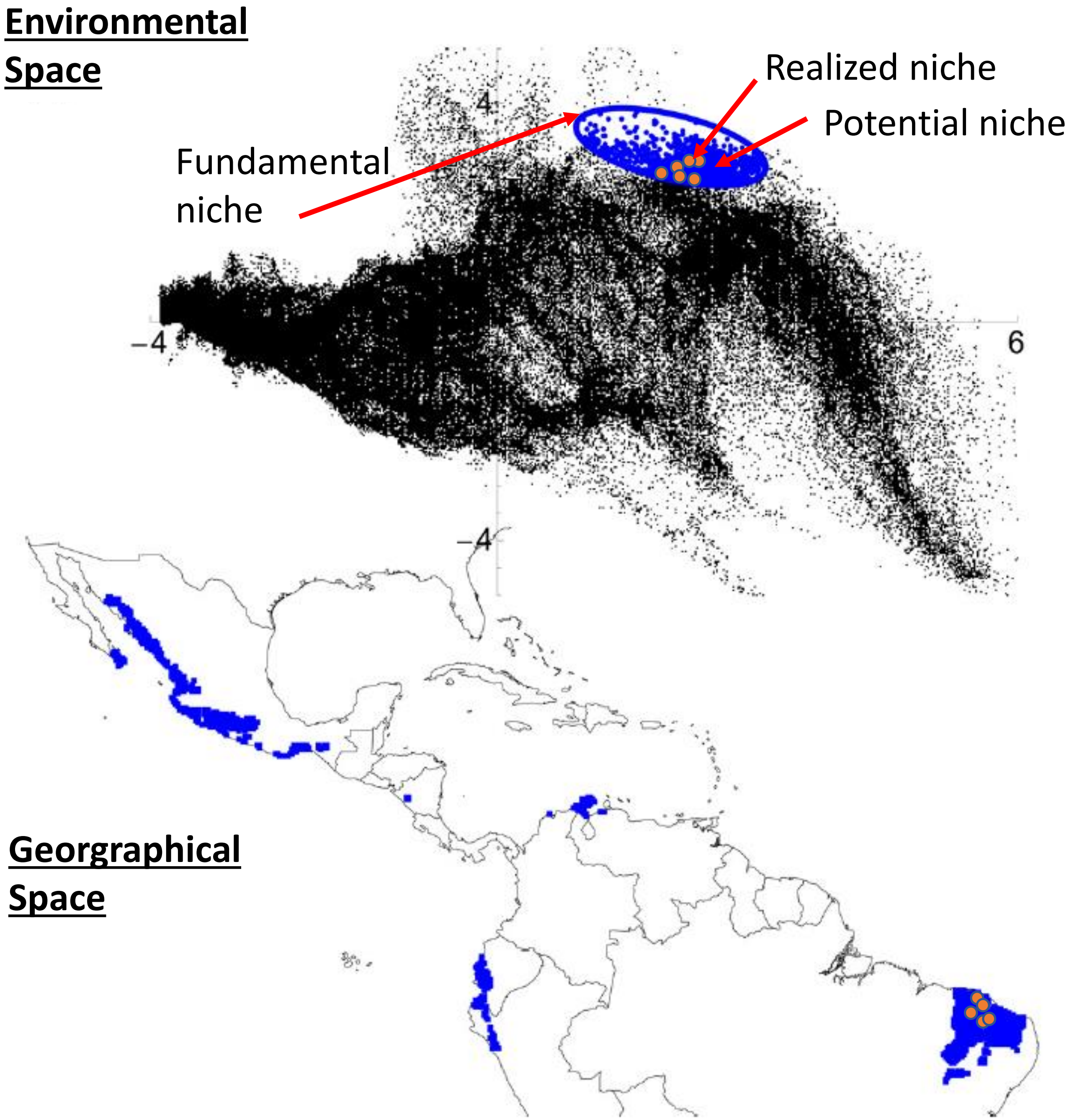
Evelyn Hutchinson (1903-1991)

- **Niche as an n-dimensional hypervolume**
    - Dimensions are environmental conditions and resources that allow a species to survive and reproduce
- Defined the concept of fundamental vs. realized niche
    - Hutchinson, G.E. 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology 22: 415-427.
- Father of modern ecology

Hutchinson's n-dimensional hypervolume

- Hutchinsonian Niche:
  - the sum total of an organism's use of the biotic and abiotic resoures in an environment.
- Generally includes:
  - Space utilization
  - Food consumption
  - Temperature range
  - Moisture requirements





Soberon & Nakamura 2009

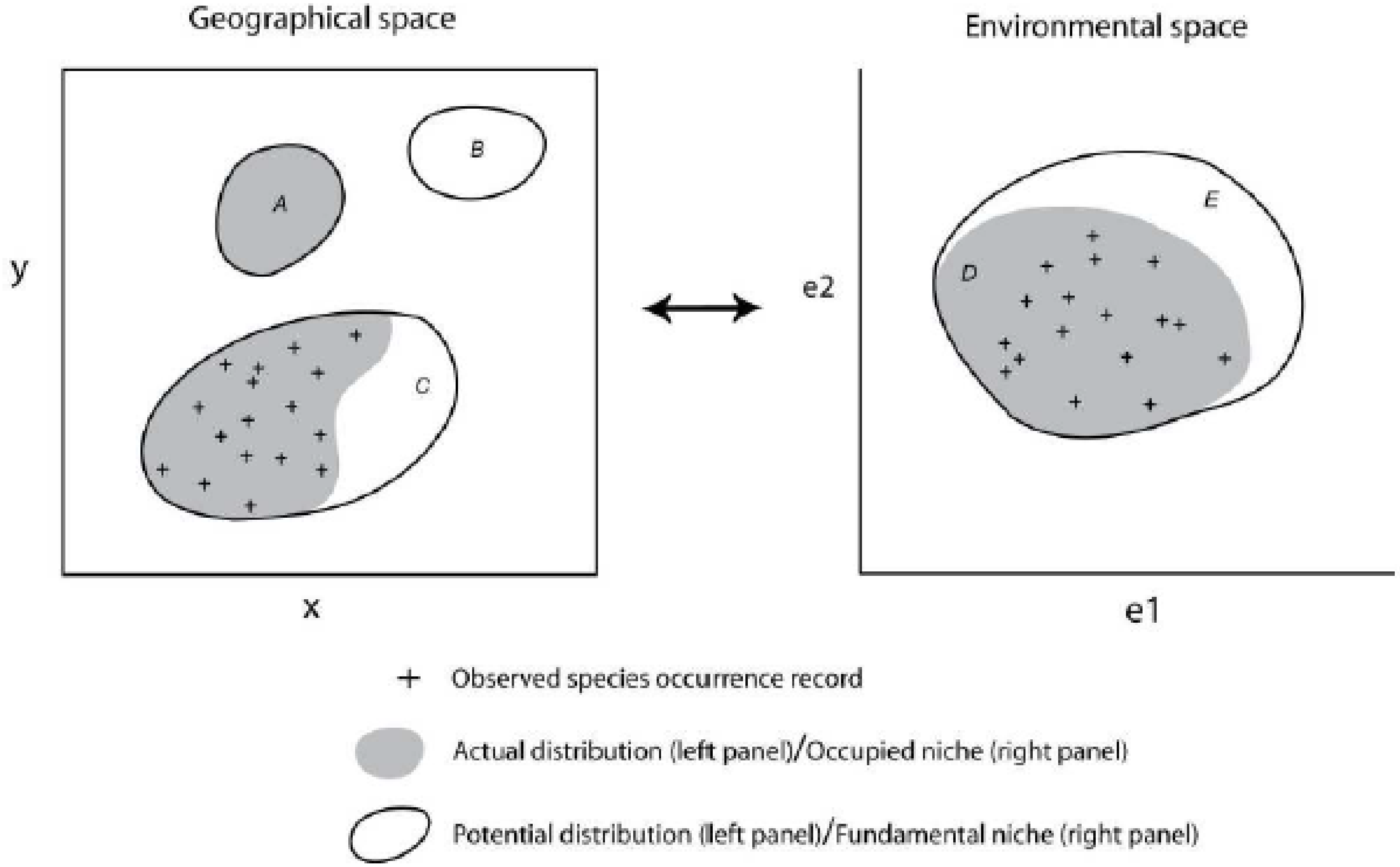
Fundamental Niche: portion of the environmental space (set of combinations of variables) capable of sustaining populations of a species

Potential Niche: part of the FN that actually exists in a given region and time

Realized Niche: part of the PN that the species actually uses, after effects of competitors and predators

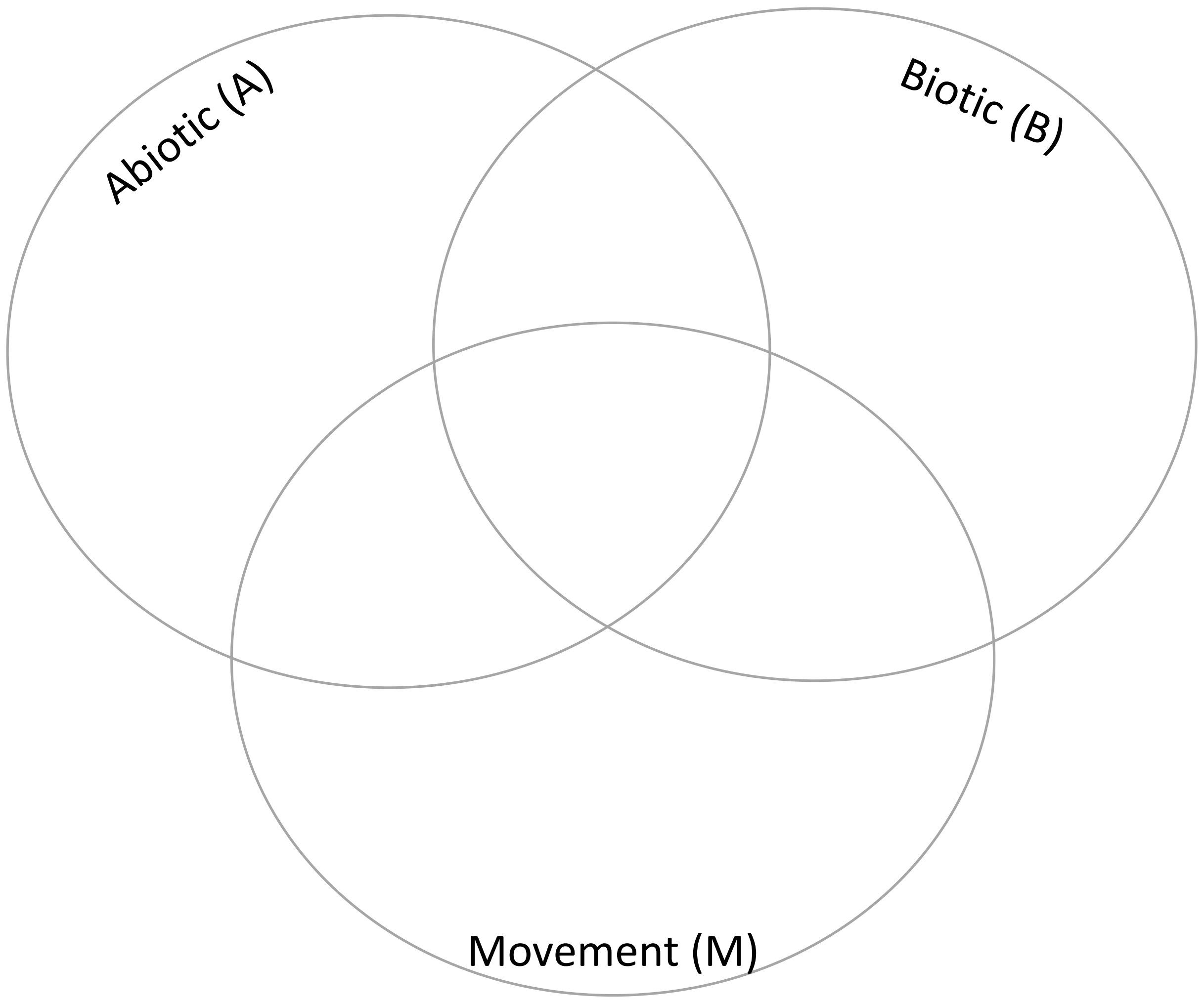
Tolerance Niche: the set of resources in which a population can survive, but not thrive (reproduce)

Environmental vs. Geographic space



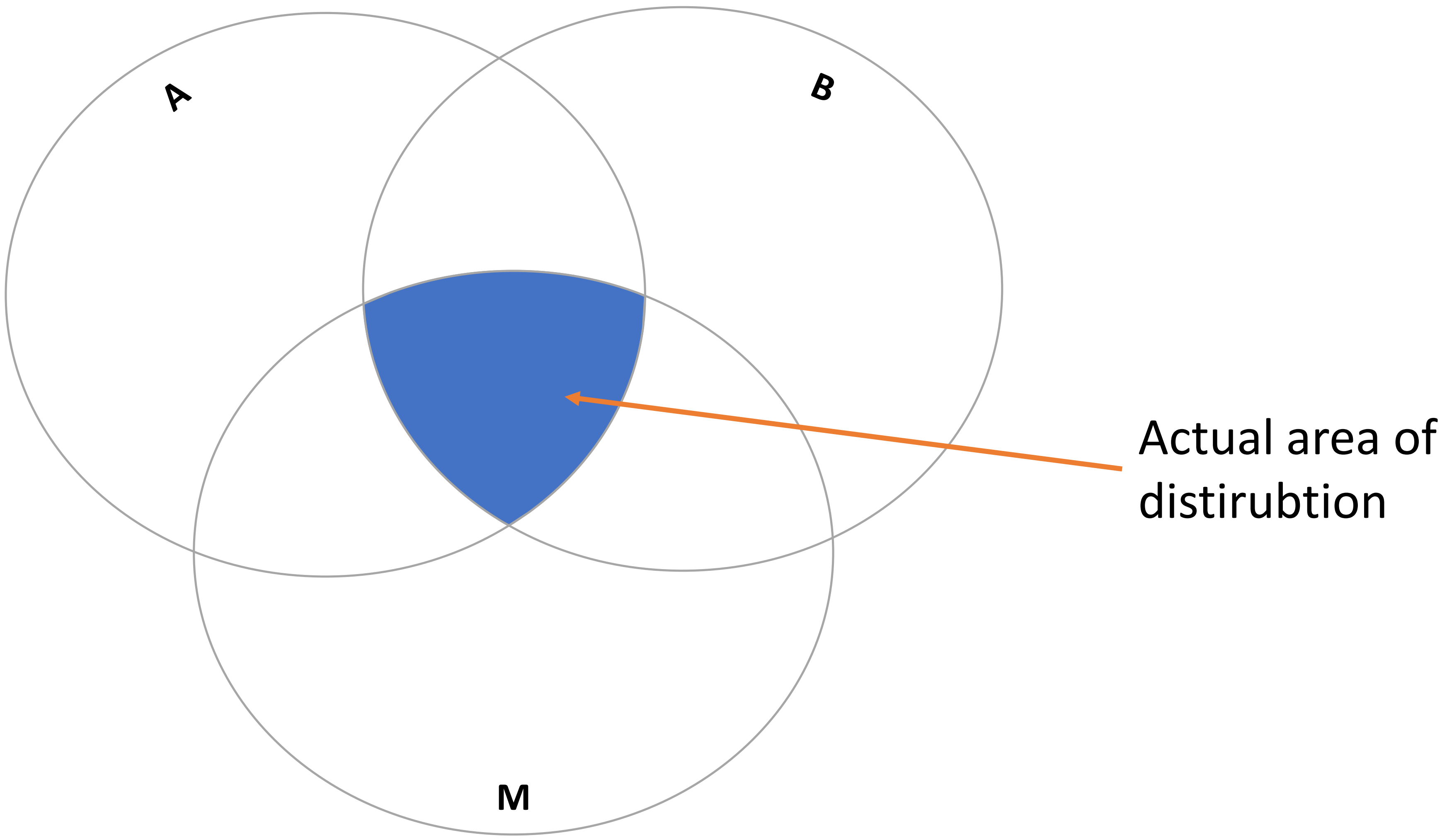
Pearson 2008. Species' Distribution Modeling for Conservation Educators and Practitioners

Niche Theory

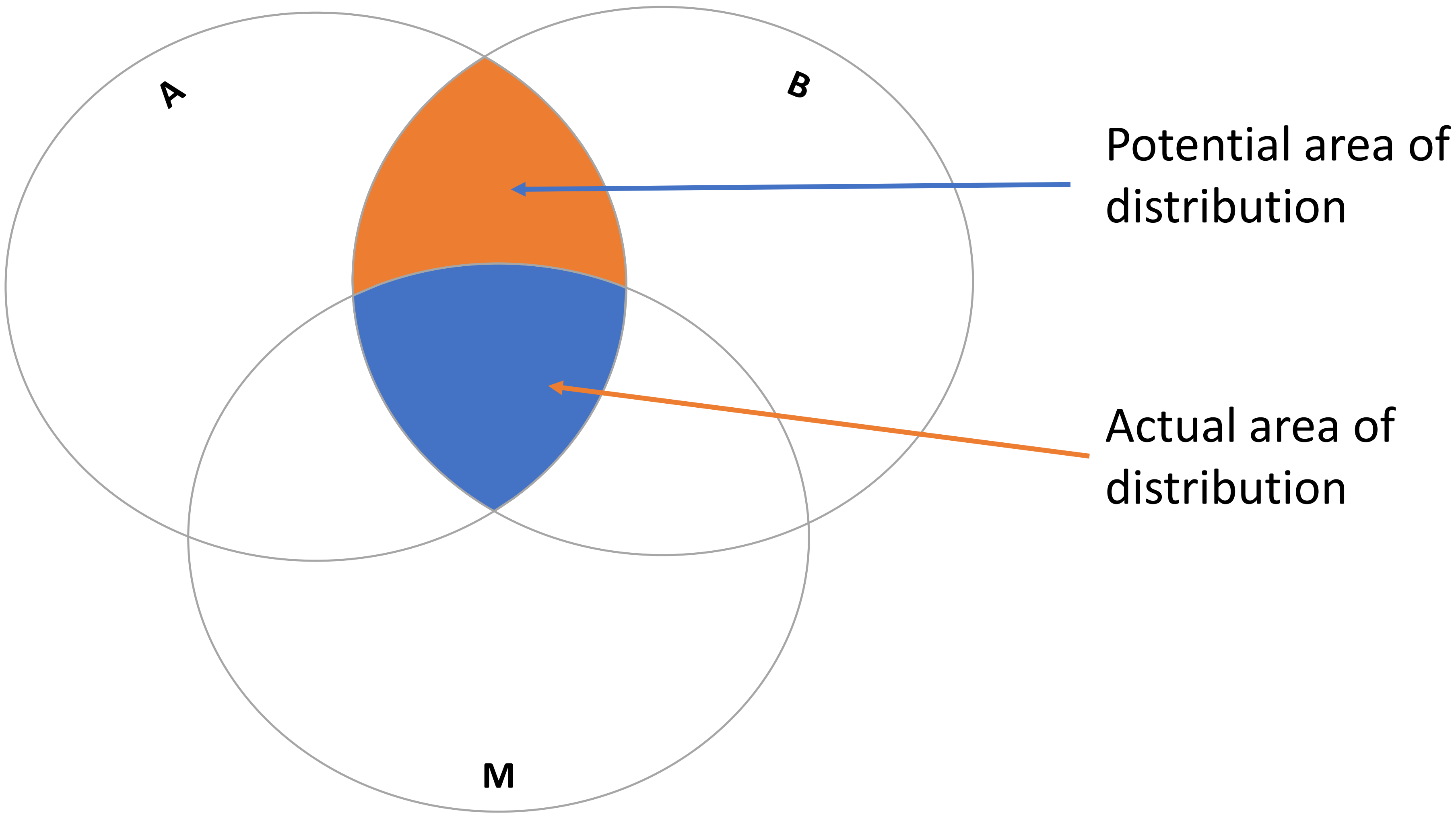


- **Abiotic**: region in the geographic space where scenopoetic conditions occur
- **Biotic**: region where biotic conditions would allow existence of viable populations
- **Movement**: region accessible to dispersal or colonization by the species over some relevant time interval

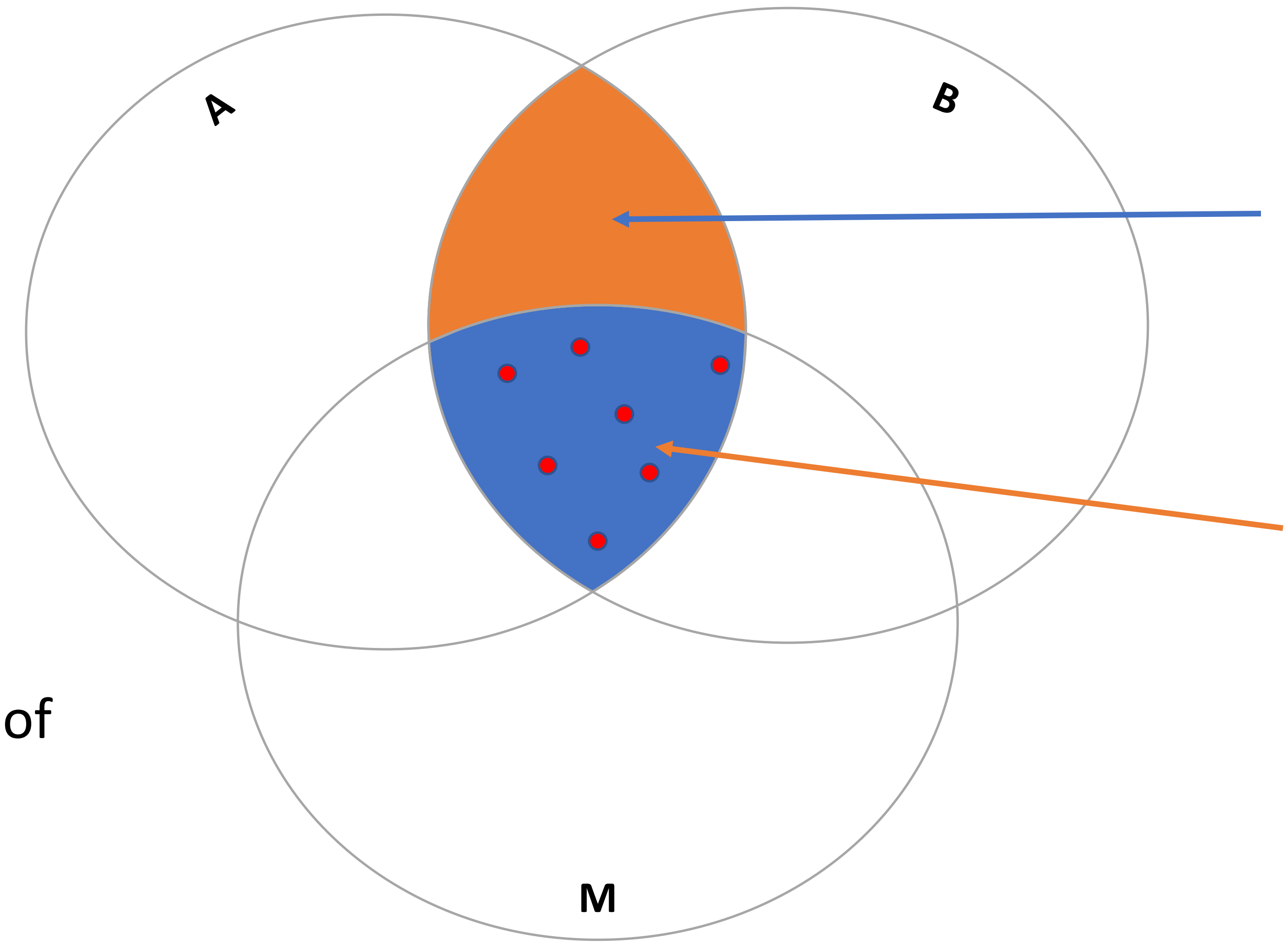
Niche Theory



Niche Theory

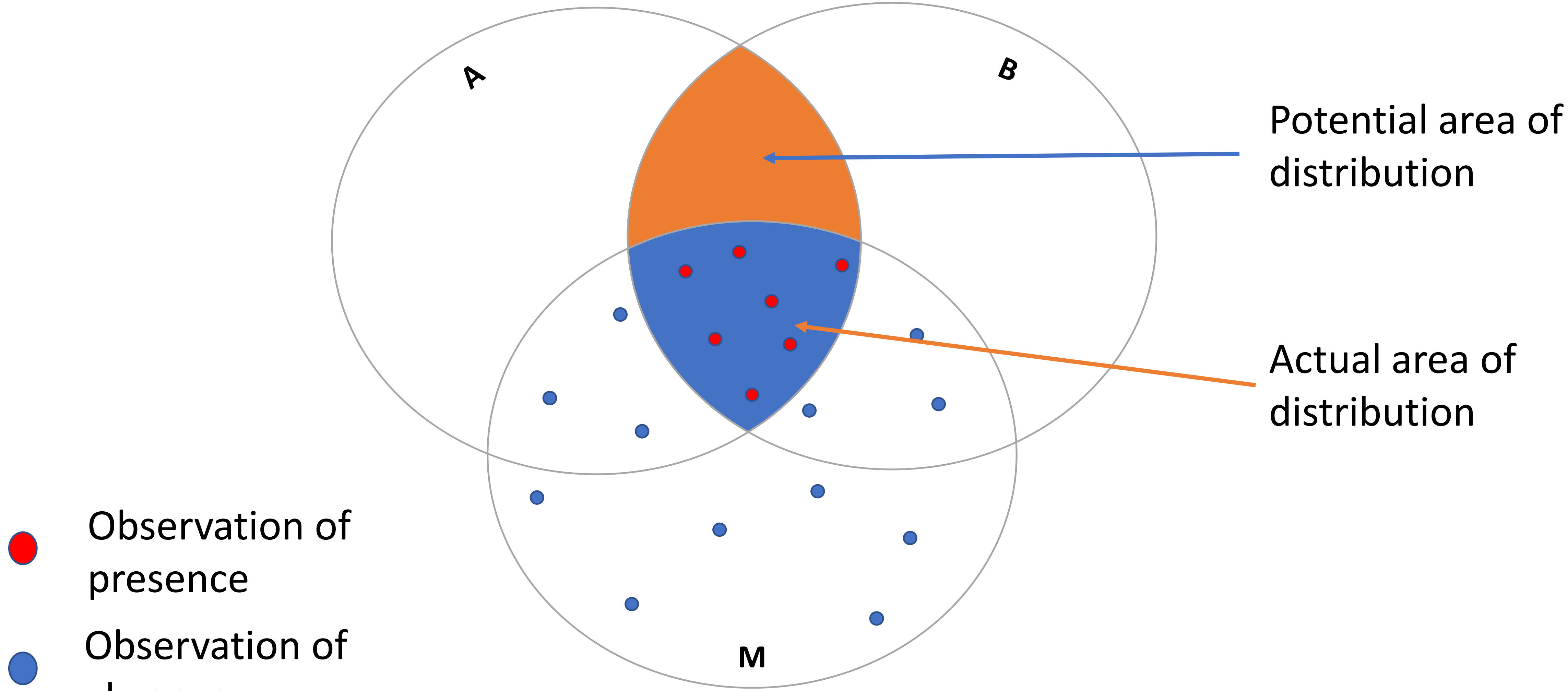


Niche Theory



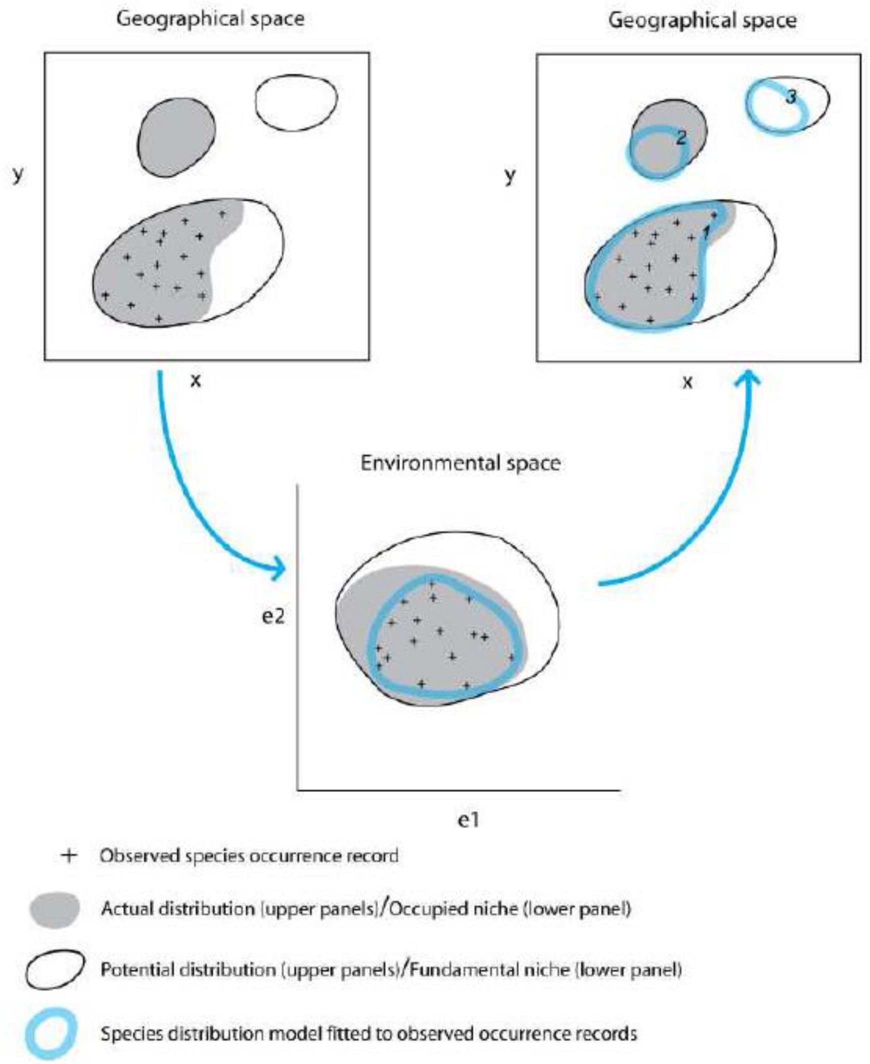
Observation of presence

Niche Theory



Niche theory → species distribution modelling

- SDMs attempt to predict the potential distribution of species by interpolating identified relationships between species occurrences and environmental predictors
  - Note that SDMs model the distribution of **suitable environments**, not the species' distribution.



Many names…

- Ecological niche modeling
- Species distribution modeling
- Habitat suitability modeling
- Habitat modeling
- Environmental niche modeling
- Climate niche modeling
- Climate envelope modeling
- Range mapping

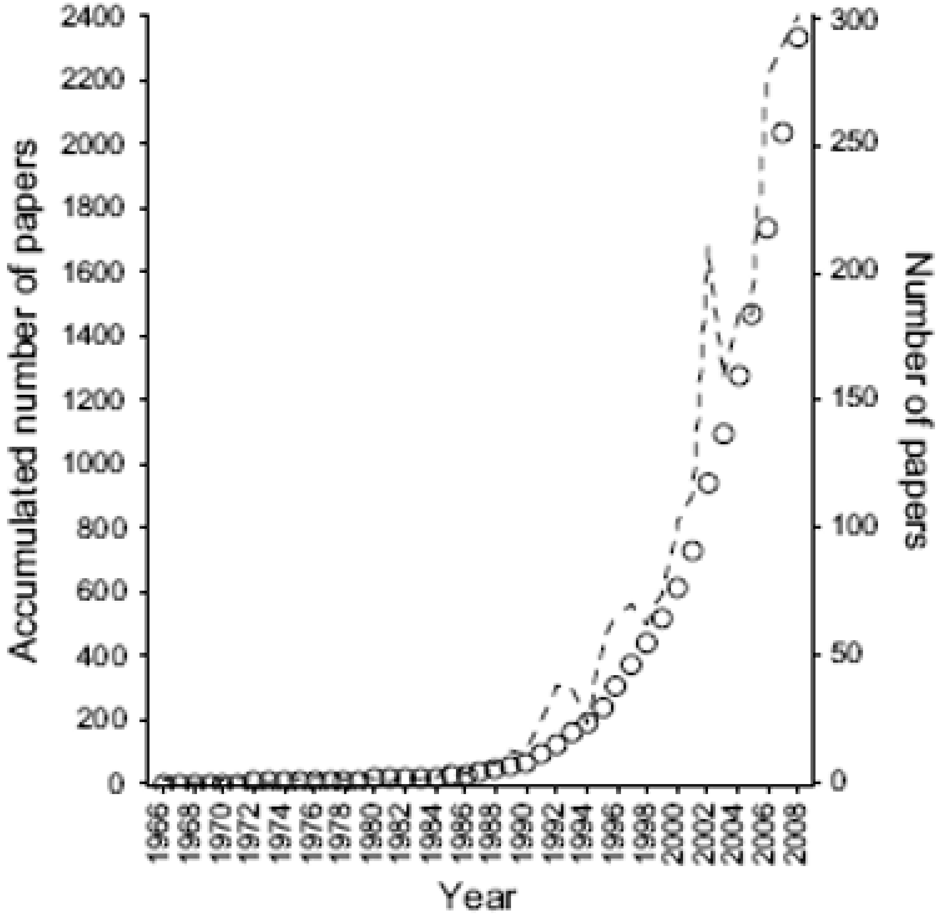


Figure 1. Number of published papers (line) and variation in the number of accumulated papers (points) on species distribution modelling found in ISI Web of Science after performing an exhaustive search by topics and authors.

Lobo *et al.* 2010 Ecography 33: 103-114

- ✓Species occurence records
- ✓Environmental data
- ✓Computing power

Why model??

- We need to know…
  - Where is species’ suitable habitat
  - Where was it in the past
  - Where will it be in the future
  - How fast will the habitat change
- It is difficult to get species’ environmental tolerances from experiments
- Makes use of vast resources – biodiversity collections

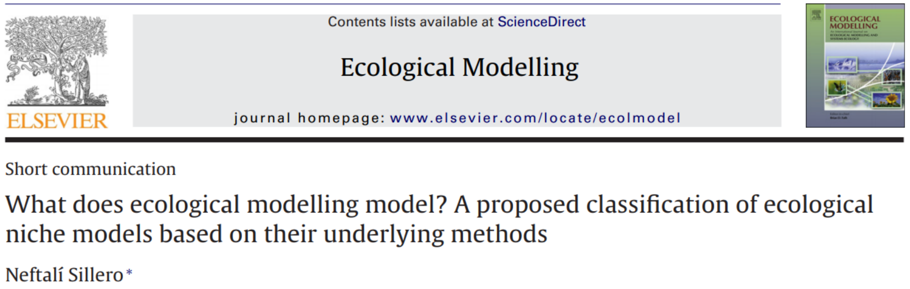
Species Distribution Modelling: Methodology

SDM –

- Group of algorithms used to estimate the spatial (+temporal) distribution of species

- Two main subgroups:
    - <u>Correlative/Empirical models</u>
    - Mechanistic model

- Correlative models: Hutchinson niche theory


Contents lists available at ScienceDirect

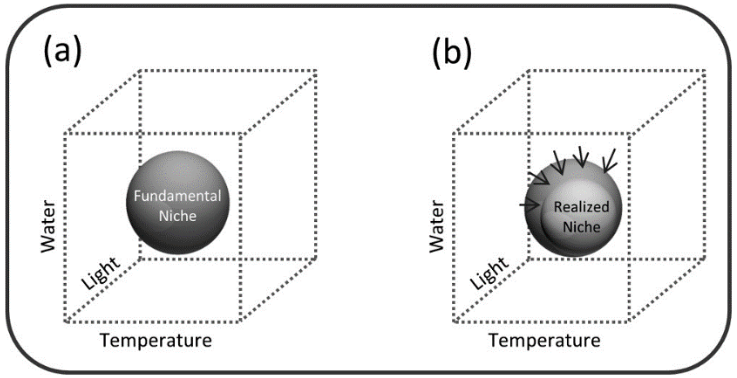
Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Short communication

What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods

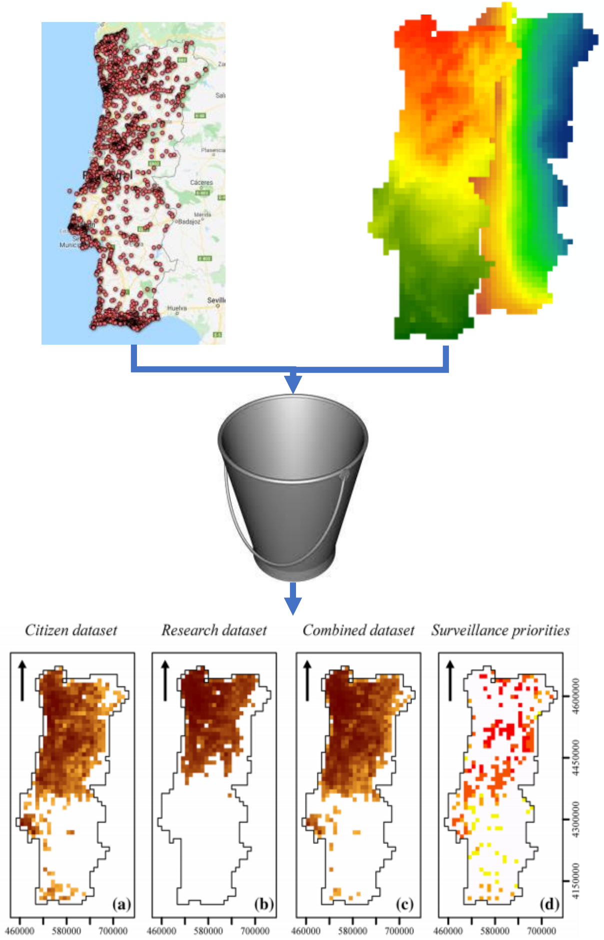
Neftalí Sillero*



Guisan, A., Thuiller, W., & Zimmermann, N. (2017). Overview, Principles, Theory, and Assumptions Behind Habitat Suitability Modeling. In *Habitat Suitability and Distribution Models: With Applications in R* (Ecology, Biodiversity and Conservation, pp. 9-58). Cambridge: Cambridge University Press. doi:10.1017/9781139028271.005

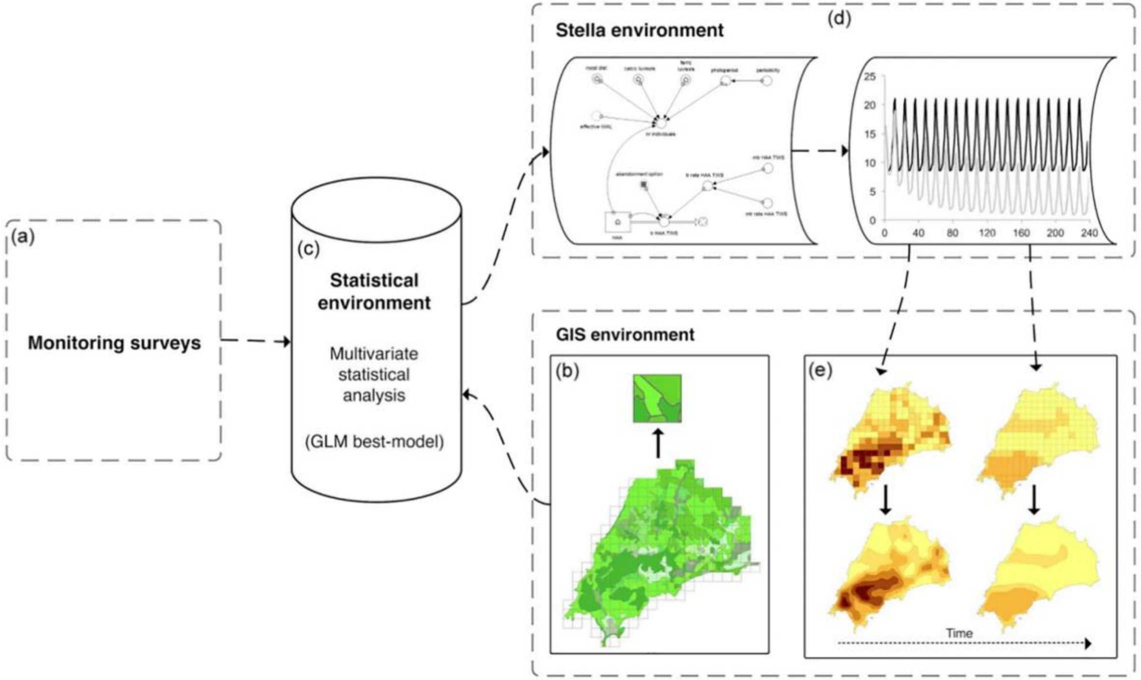
SDM – Correlative vs Mechanistic models

Correlative models



N César de Sá, et al - *Can citizen science data guide the surveillance of invasive plants? A model-based test with Acacia trees in Portugal*, Biological invasions, 2018

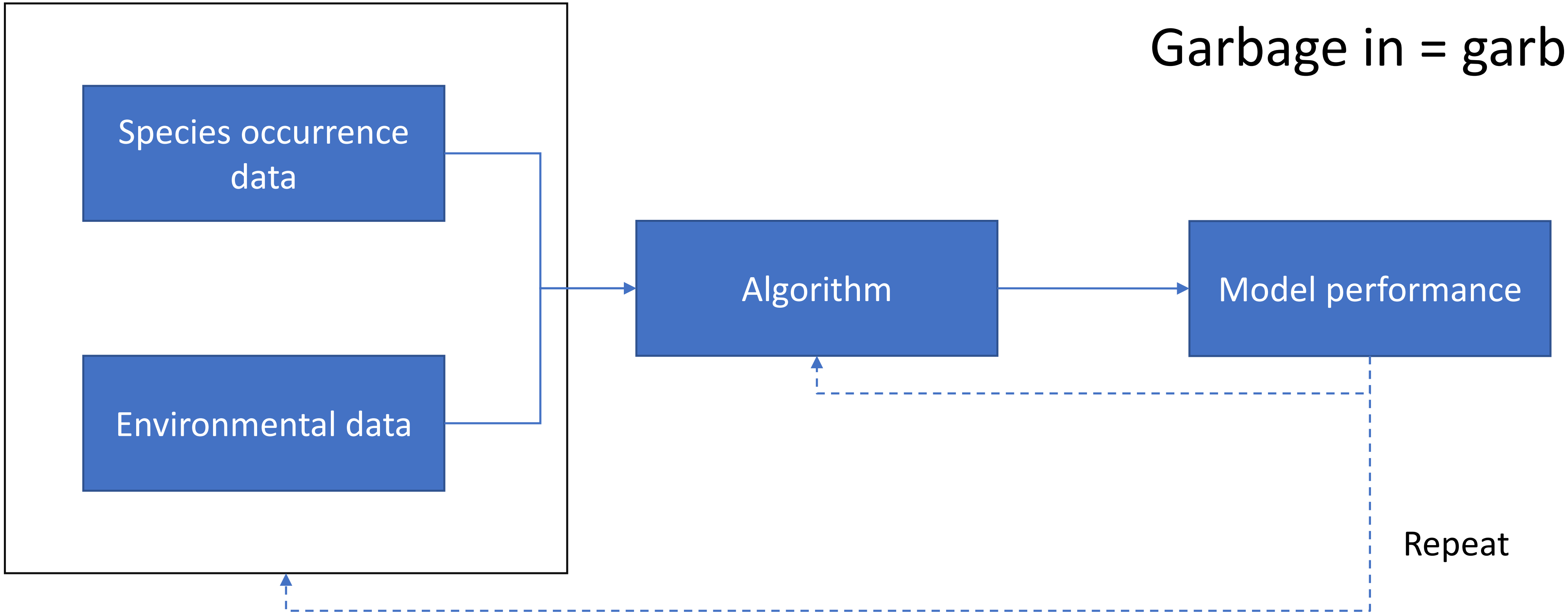
Mechanistic models



Francisco Morinha, Rita Bastos, Diogo Carvalho, Paulo Travassos, Mário Santos, Guillermo Blanco, Estela Bastos, João A. Cabral, A spatially-explicit dynamic modelling framework to assess habitat suitability for endangered species: *The case of Red-billed Chough under land use change scenarios in Portugal*,
Biological Conservation, Volume 210, Part A, 2017, Pages 96-106, ISSN 0006-3207, https://doi.org/10.1016/j.biocon.2017.04.013.

SDM – Overall Process

Garbage in = garbage out!



- Similar to any other regression/classification exercise
- Fundamental difference: <u>Niche theory</u>

SDM – Overall Process – common problems

Species occurrence data	Environmental data	Algorithm	Model performance
Species "BAM"	Ecological significance	Model selection	Accuracy
Sampling bias & efforts	Autocorrelation	Data needs	Overfit & underfit
	Spatial resolution	Model "logic" approach	Spatial auto-correlation

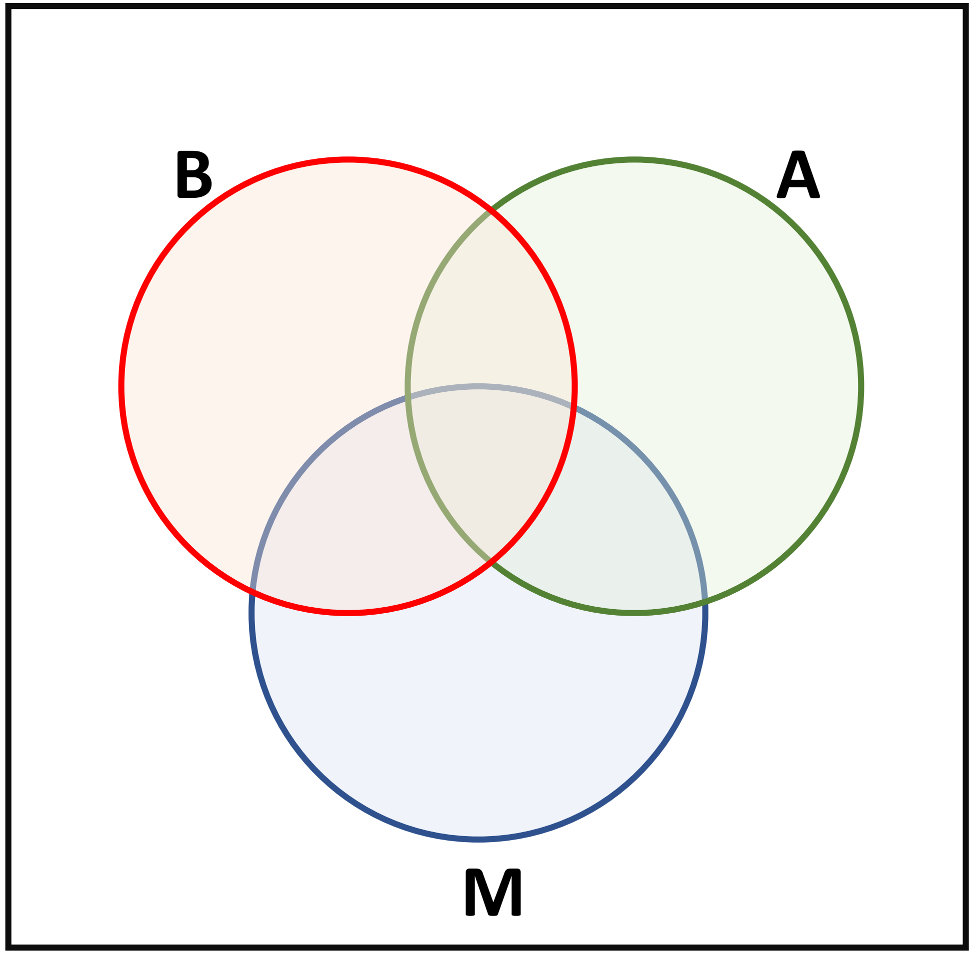
---

Ecological dimension: No "unified niche theory" yet

Different scales → Different dominant processes

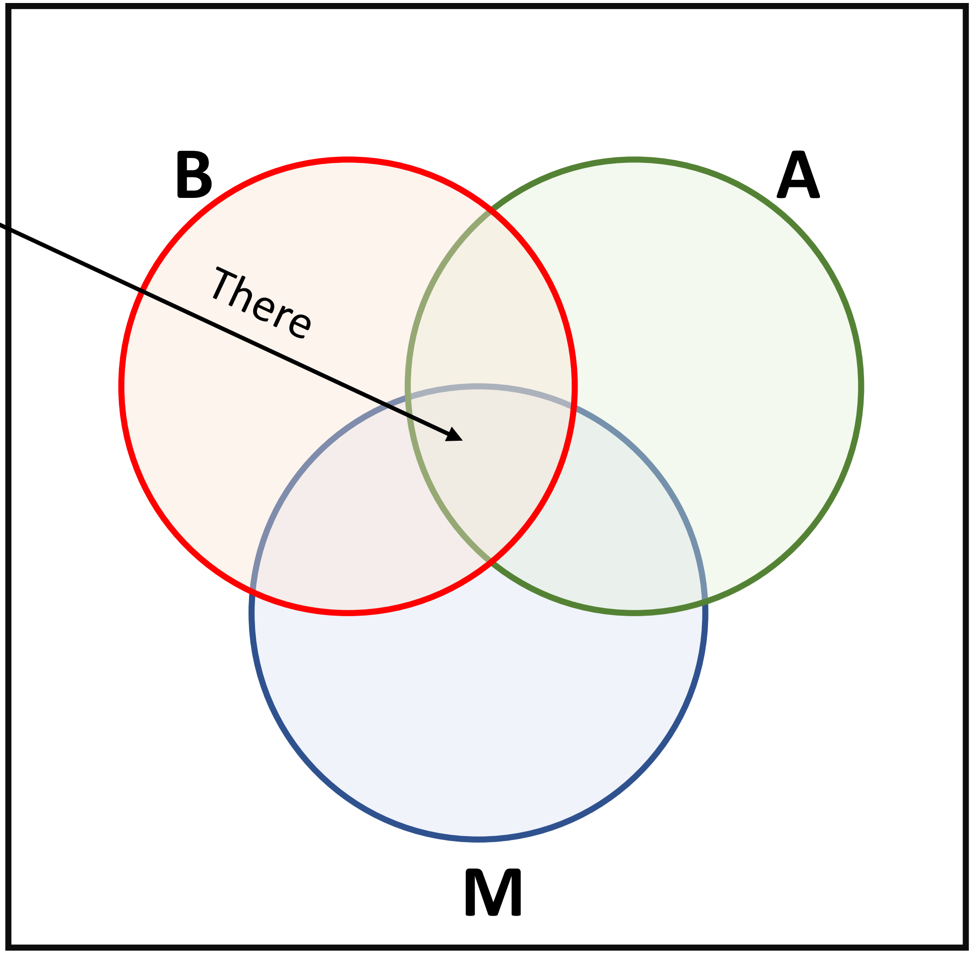
For the more curious: R.P. Anderson - *A framework for using niche models to estimate impacts of climate change on species distributions*, 2013

SDM – Occurrence data: Definitions



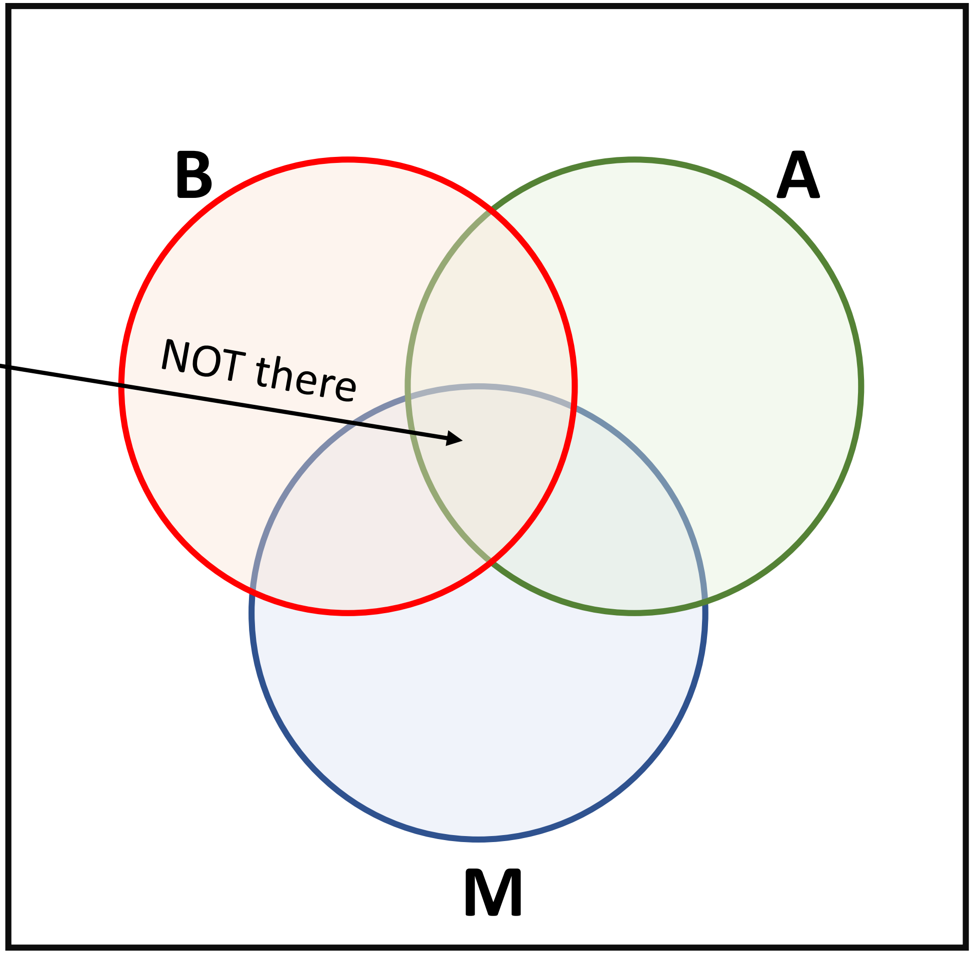
SDM – Occurrence data: Definitions

- **Presence**: Everywhere the species occurs



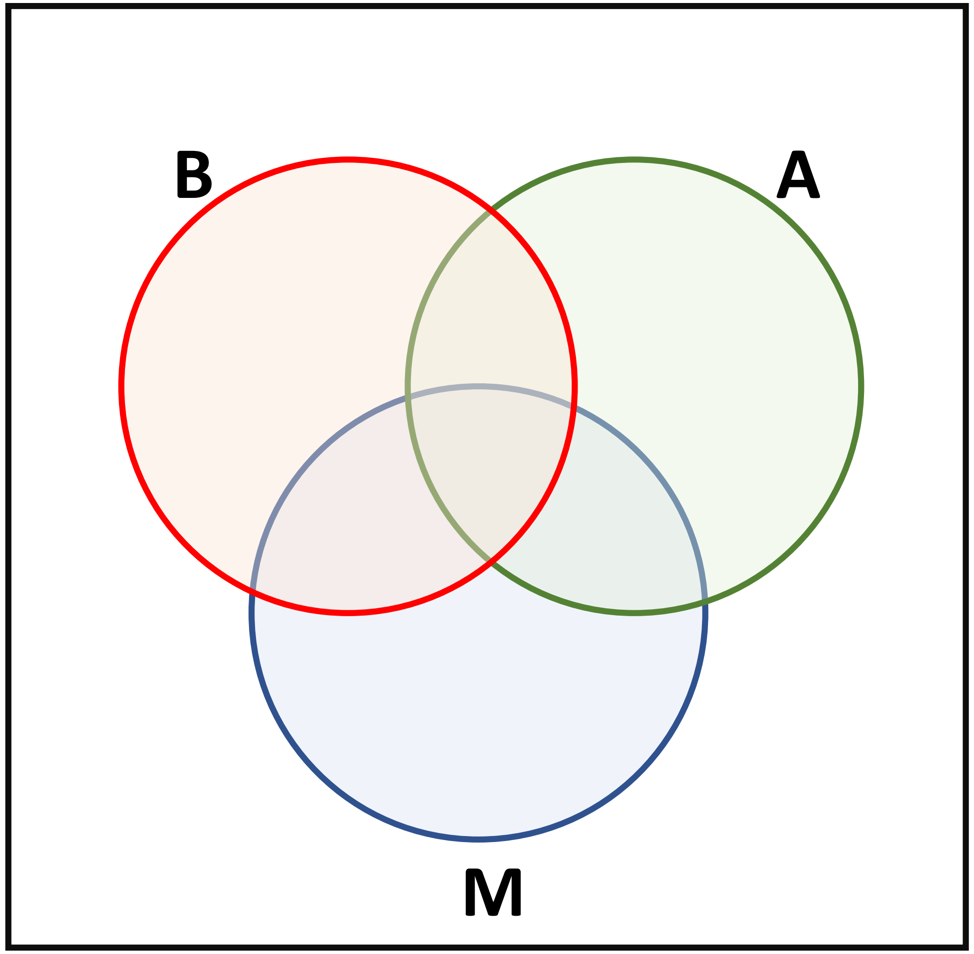
SDM – Occurrence data: Definitions

- **Presence**: Everywhere the species occurs
- **Absence**: Everywhere the species does not occur



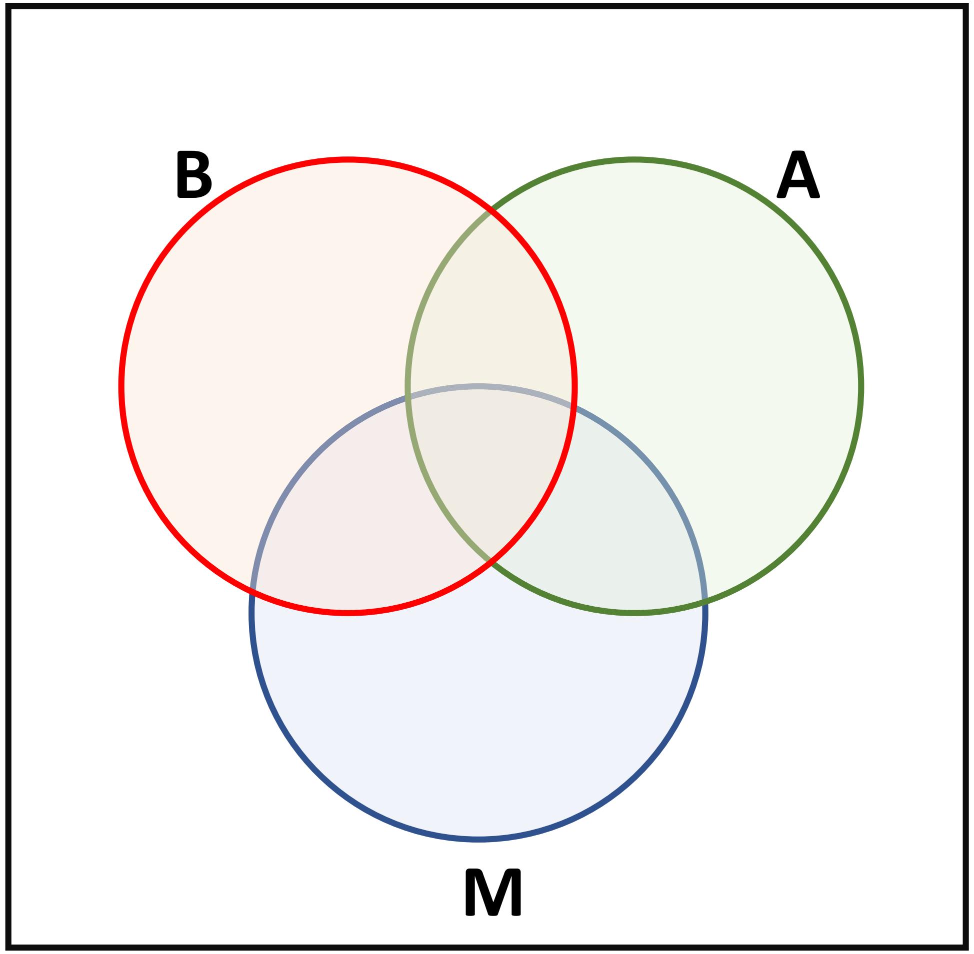
SDM – Occurrence data: Definitions

- **Presence**: Everywhere the species occurs
- **Absence**: Everywhere the species does not occur
- For modelling purposes:
  - **Pseudo-absence**: Places where the species probably does not occur
  - **Background data**: Biased or unbiased sample of the environment



SDM – Occurrence data: Definitions

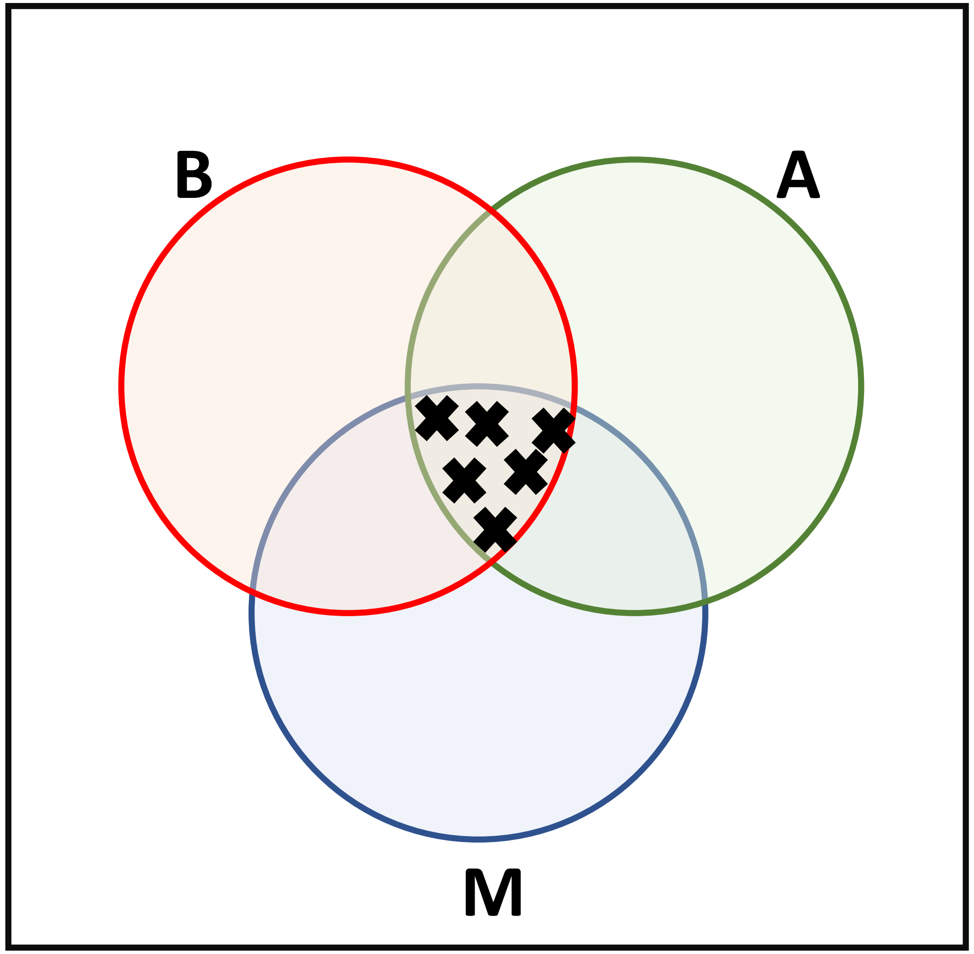
- What are good presences and good absences?



SDM – Occurrence data: Definitions

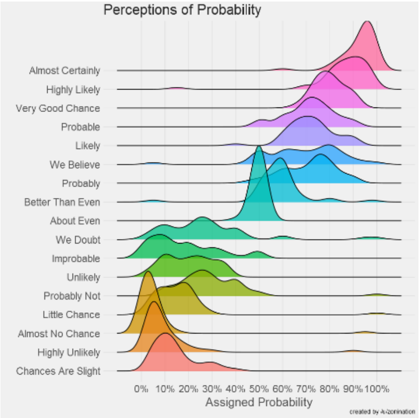
- What are good presences and good absences?
    - Good presences:
        - Very easy to define: <u>see it, it’s there</u>
        - Objectively:
            - Within the realized niche
            - Representative of the entire “niche space”
        - Sampling bias:
            - **DEPENDS:**
                - E.g. a rare species

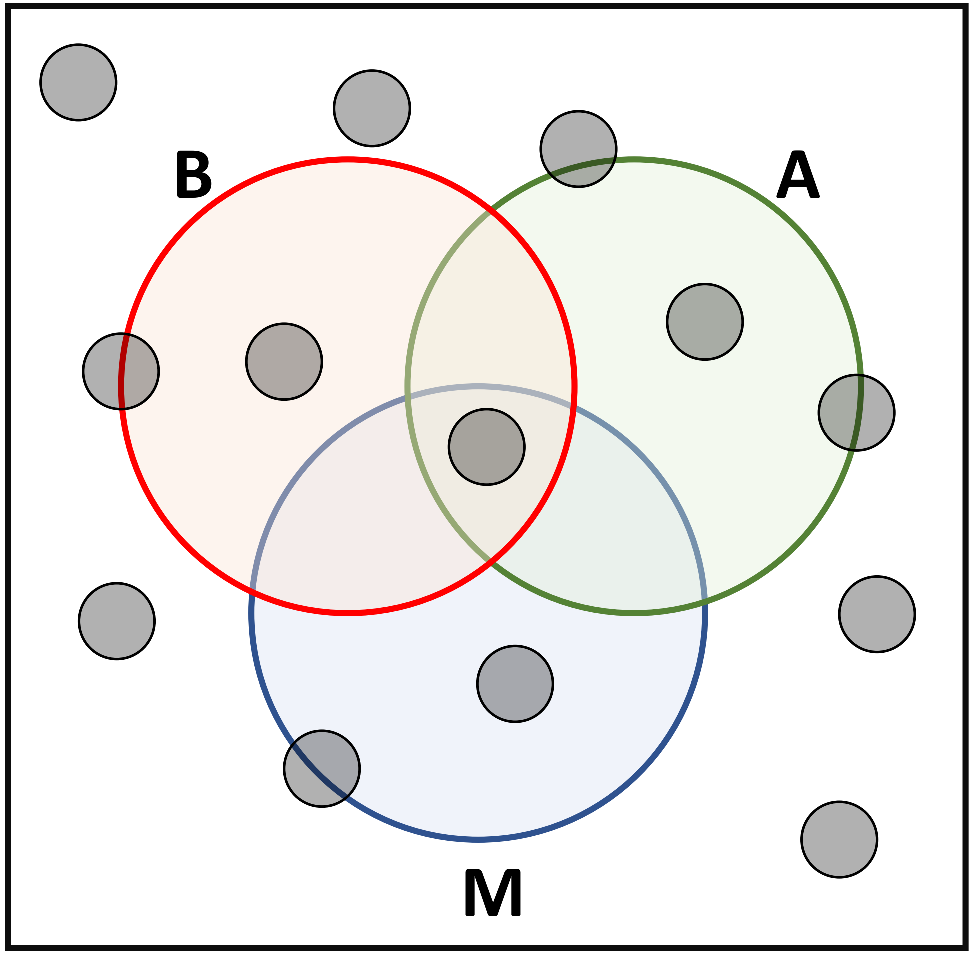
Exceptions.. Of course



SDM – Occurrence data: Definitions

- What are good presences and good absences?
  - Absences:
    - Harder to define: Didn’t see it, but might be there
    - Objectively:
      - Outside of the niche space



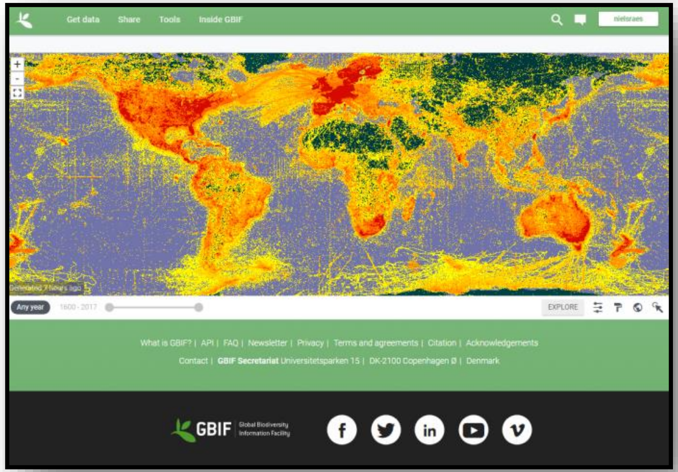


For the curious ones: Bayes theorem

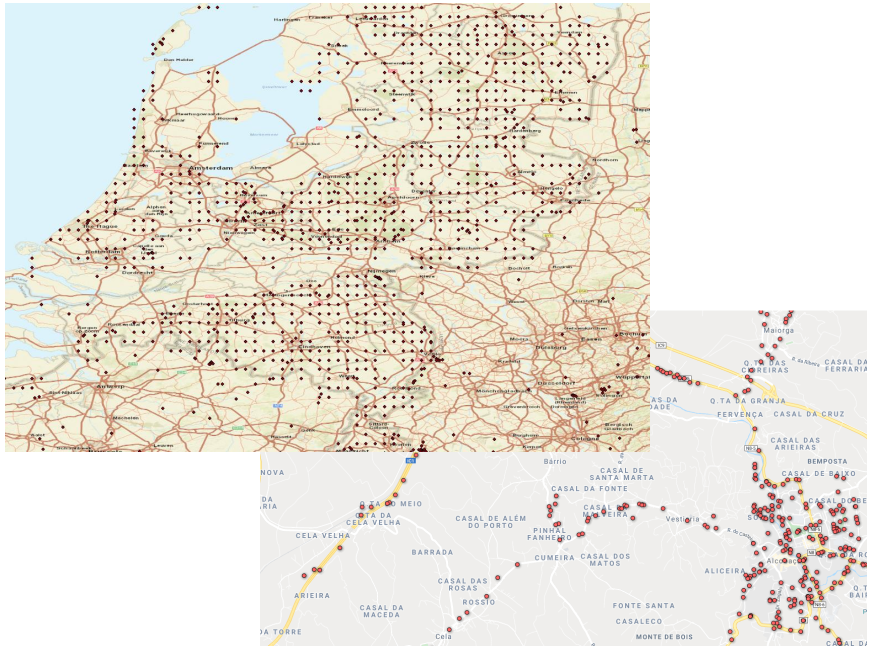
SDM – Occurrence data: Data quality

- Occurrence data quality is always on a gradient

Sampling collections



Sampling biases



Quality of data

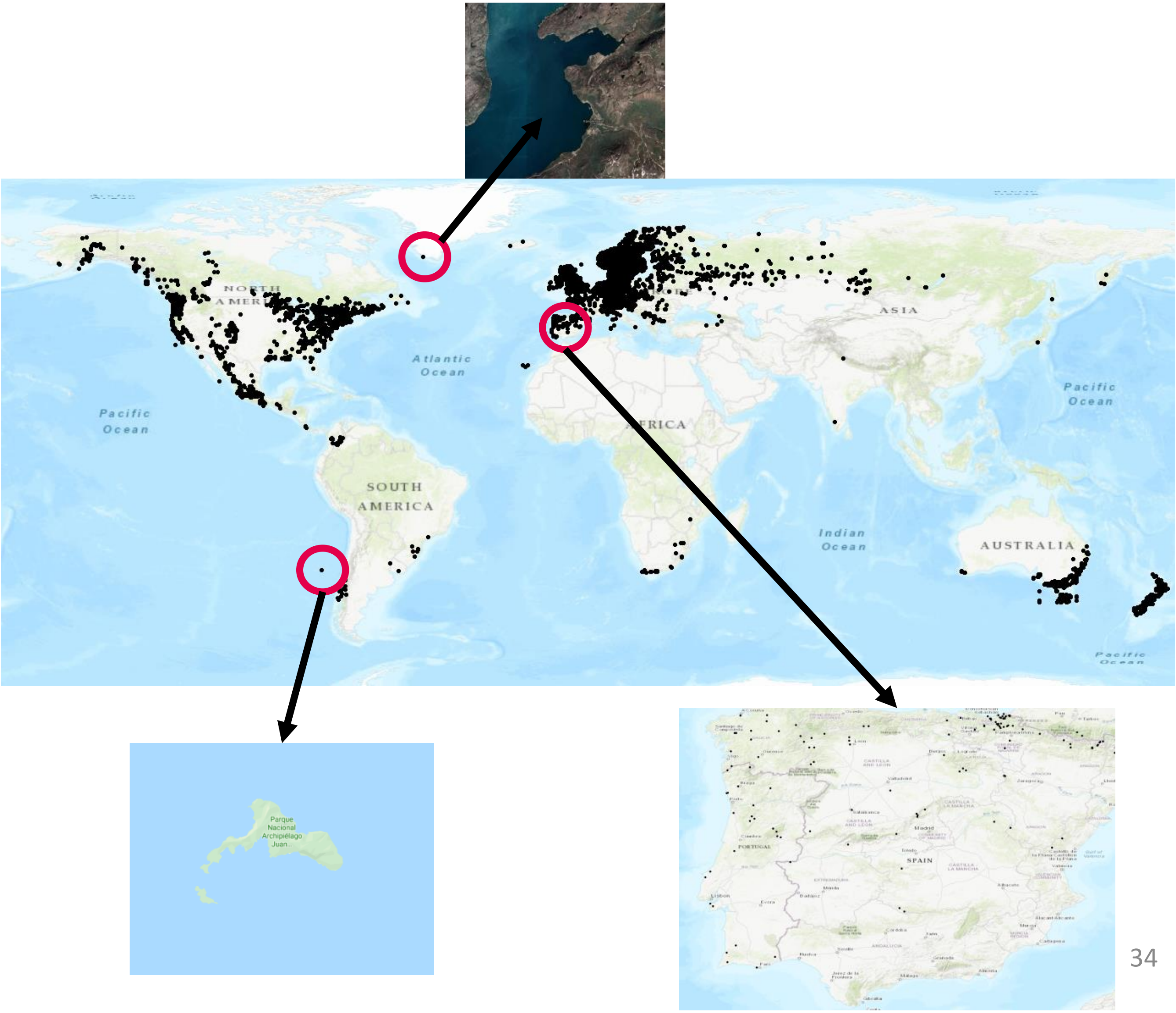
	Bryophytes	Mediterranean plants	Amazonian trees	Central American trees
Original names	1122	3047	1188	5113
Accepted	717 (63·9%)	2130 (69·9%)	471 (39·6%)	2595 (50·7%)
Synonym	287 (25·6%)	444 (14·6%)	37 (3·1%)	571 (11·2%)
Unresolved	79 (7·0%)	204 (6·7%)	11 (0·9%)	85 (1·7%)
Nonavailable	39 (3·5%)	269 (8·8%)	669 (56·3%)	1862 (36·4%)
Orthographic errors	17 (1·5%)	74 (2·4%)	8 (0·7%)	299 (5·8%)
Standardised names	955 (85·1%)	2940 (96·5%)	1040 (87·5%)	4720 (92·3%)

Verifying the quality of data is paramount..

SDM – Occurrence data: Data quality

- *Amanita muscaria*

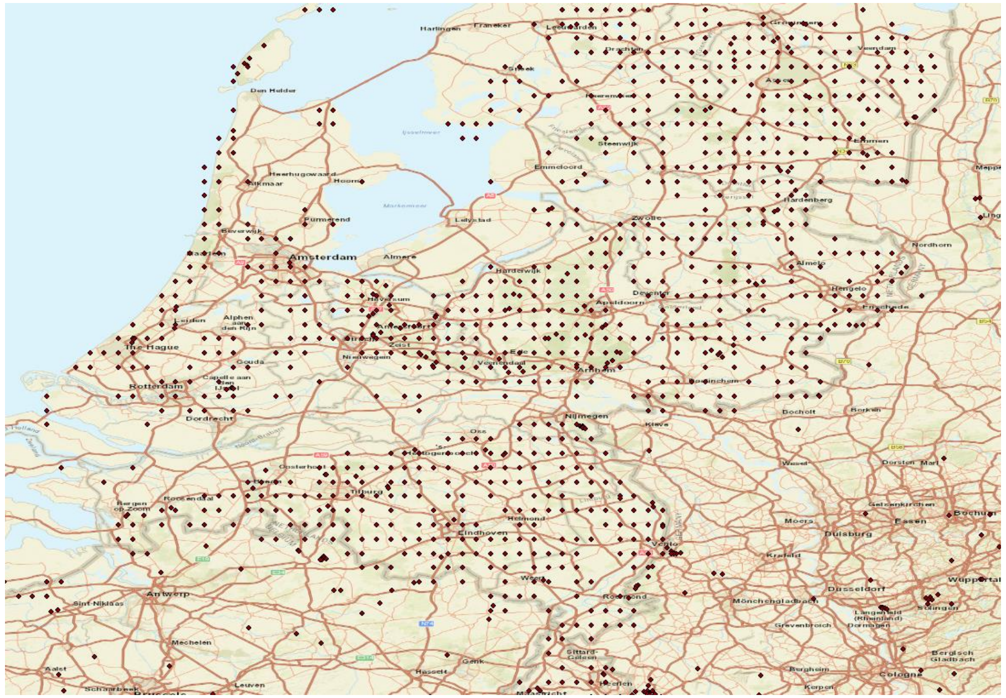
According to Wikipedia: “cosmopolitan” mushroom, native to conifer & deciduous forests throughout temperate and boreal regions.



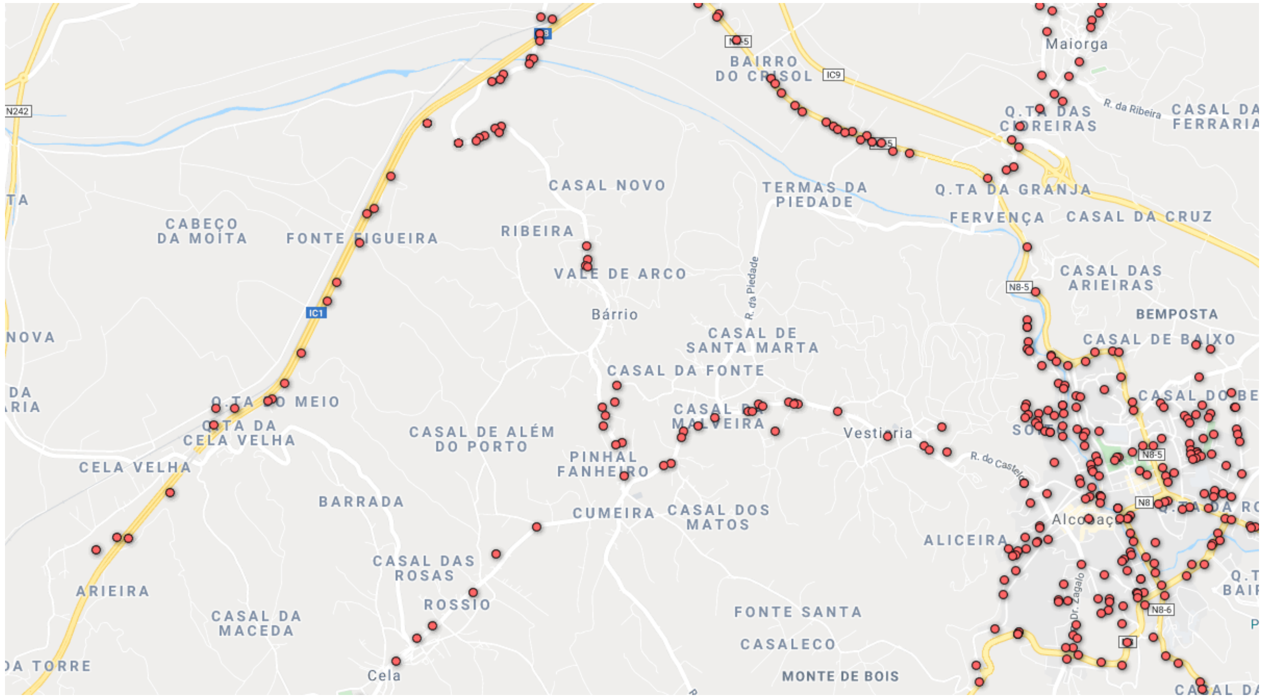
SDM – Occurrence data: Exploring

- Example – sampling strategy

Atlas data

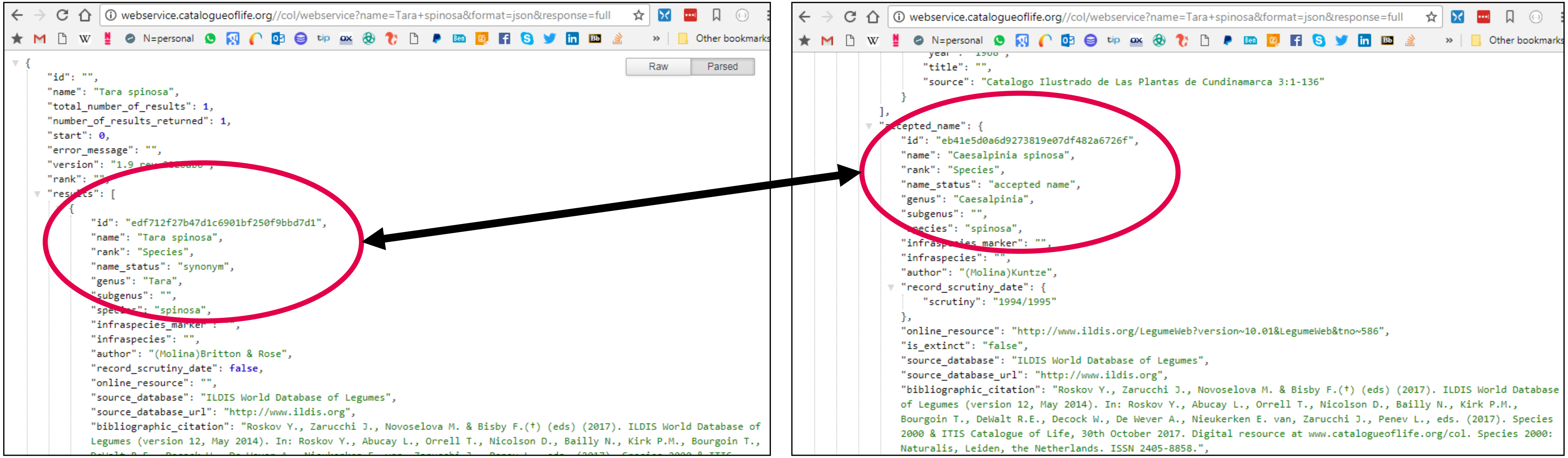


Citizen science data



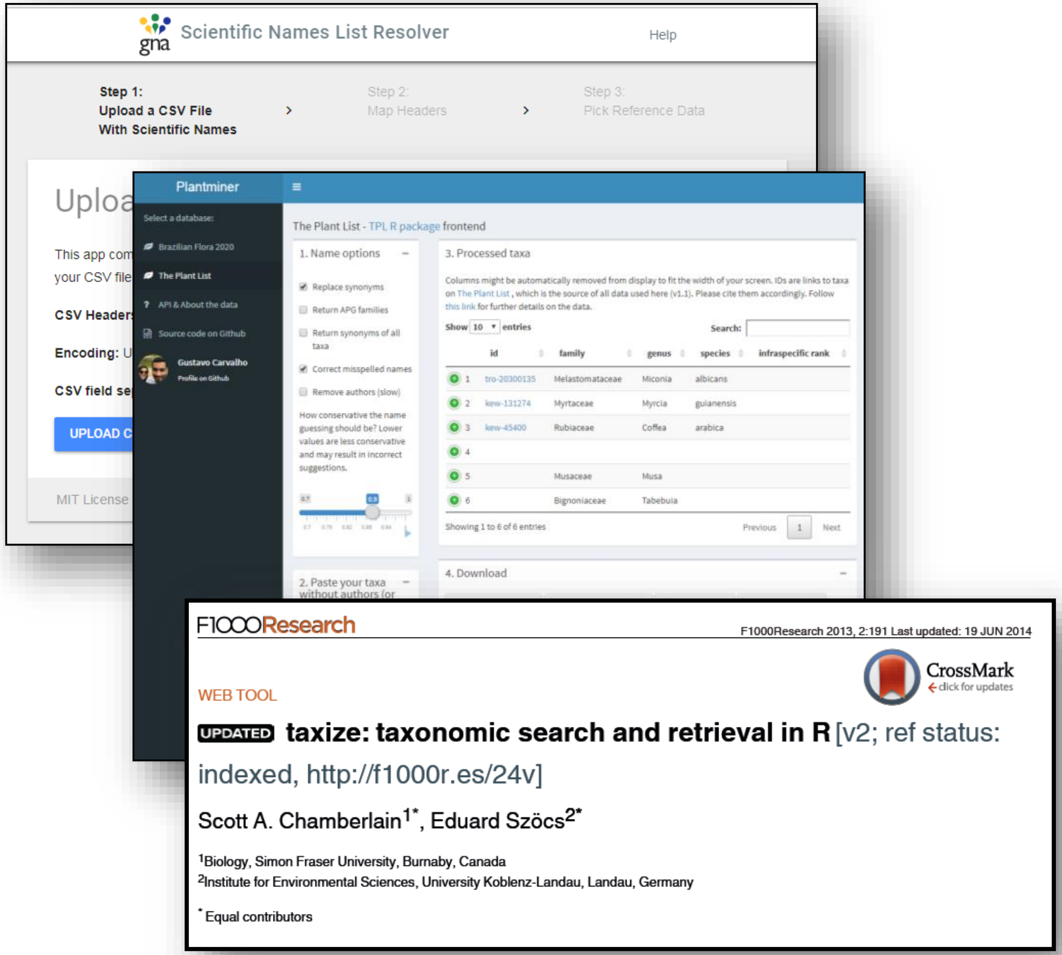
SDM – Occurrence data: Exploring

- Example: *Tara spinosa* also known as *Caesalpinia spinosa*



SDM – Occurrence data: Data quality

- How to compensate:
  - Spatial biases:
    - Many methods – not explored in this course (optional)
    - Suggestion: Use GIS
  - Data quality:
    - Some web services available
    - R packages also
    - <u>Use them to improve your species data</u>



Something to add?

Next: Environmental data

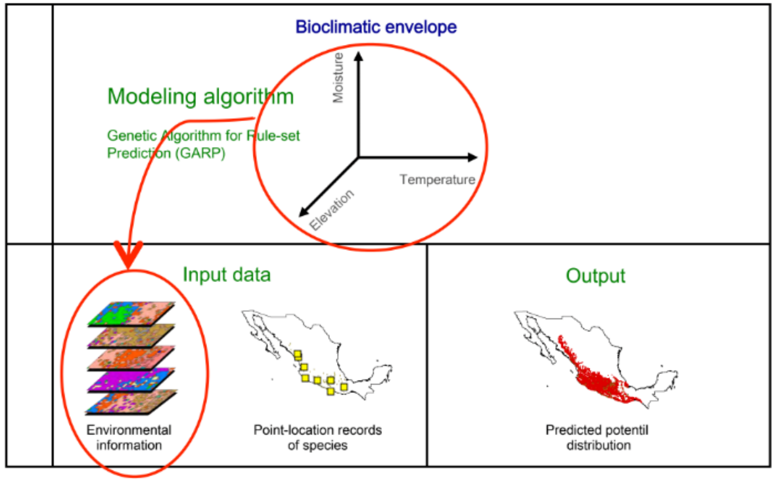
SDM – Environmental data

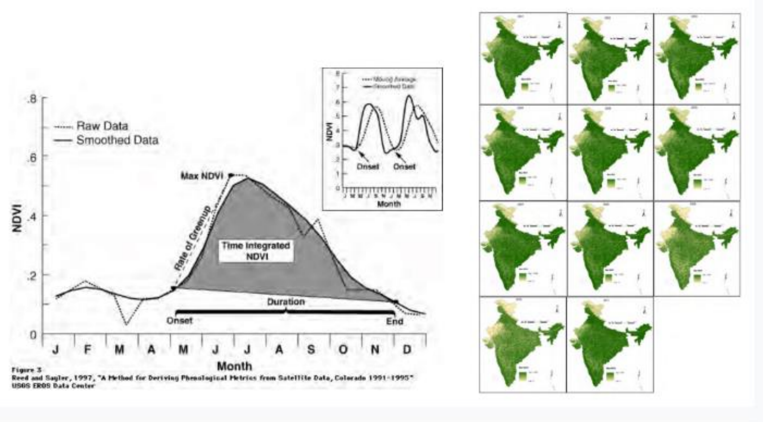
- It's the spatial representation of the "Environmental space"
  - Note: most often the abiotic space but can also represent the biotic space.

- Produced from:
  - Interpolation (e.g. Anuclim)
  - Climate models (e.g. GCM)
  - Remote Sensing
  - And many other sources

- Can be:
  - Direct measurements (e.g. Temperature)
  - Proxy variables (e.g. NDVI)

- Dimensions:
  - Spatial resolution
  - Temporal resolution





Recommended resources: http://biodiversity-informatics-training.org/bi-curriculum/enm-sdm/

SDM – Environmental data

A. Climatic data:
    - WorldClim (worldclim.org)
    - CliMond (climond.org)
    - GCM Downscaled (ccafs-climate.org)

SDM – Environmental data

WORLDCLIM

BIOCLIM

Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g., BIOCLIM, GARP). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

They are coded as follows:

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (P2/P7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (P5-P6)
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

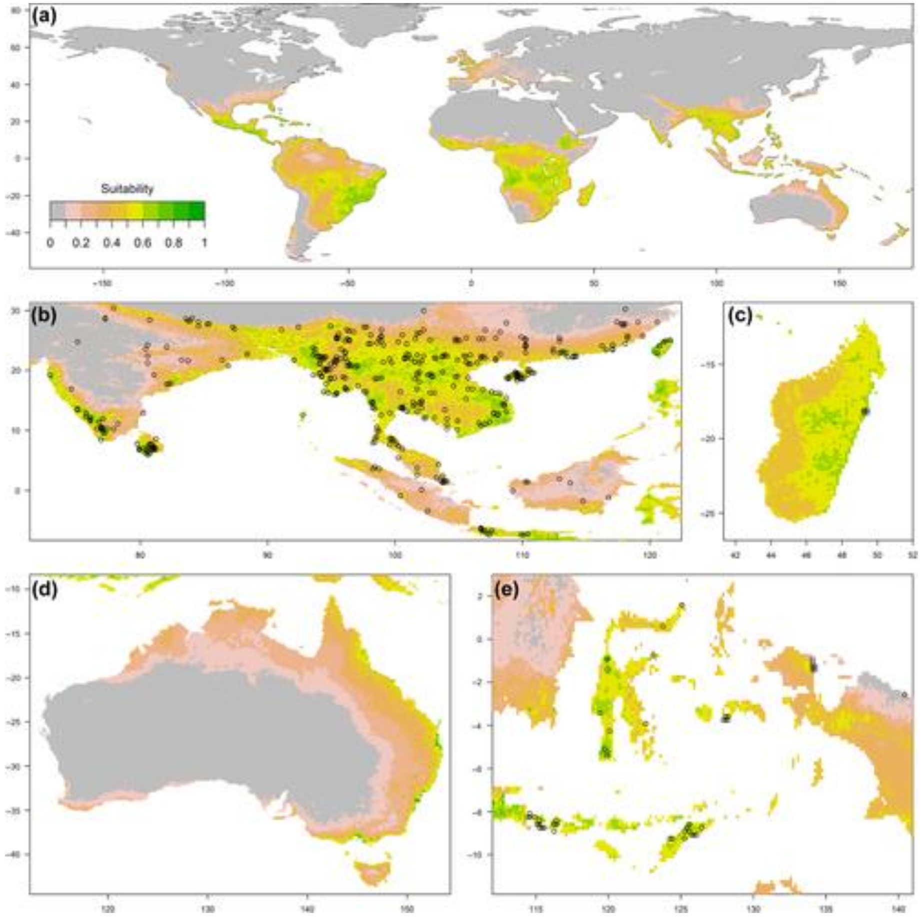
This scheme follows that of ANUCLIM, except that for temperature seasonality the standard deviation was used because a coefficient of variation does not make sense with temperatures between -1 and 1).

This AML (Arc-Info script) was used to generate these layers.

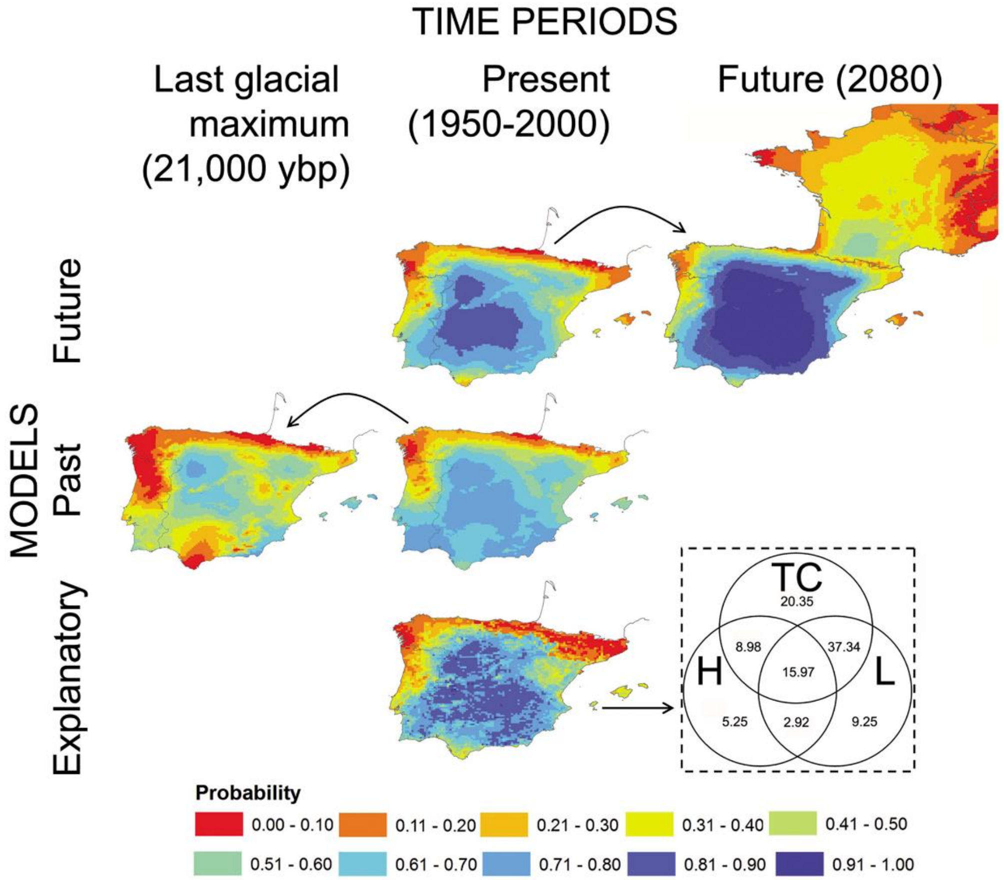
- Is it the only data source? NO!
- But it's the most used.
- Represents the ~ climatic conditions between 1950 and 2000
- Yes there are newer versions.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

SDM – Environmental data



Tingley, R. , García‐Díaz, P. , Arantes, C. R. and Cassey, P. (2018), *Integrating transport pressure data and species distribution models to estimate invasion risk for alien stowaways*. Ecography, 41: 635-646. doi:10.1111/ecog.02841



Acevedo P, Melo-Ferreira J, Real R, Alves PC (2012) Past, Present and Future Distributions of an Iberian Endemic, *Lepus granatensis*: Ecological and Evolutionary Clues from Species Distribution Models. PLOS ONE 7(12): e51529. https://doi.org/10.1371/journal.pone.0051529

SDM – Environmental data

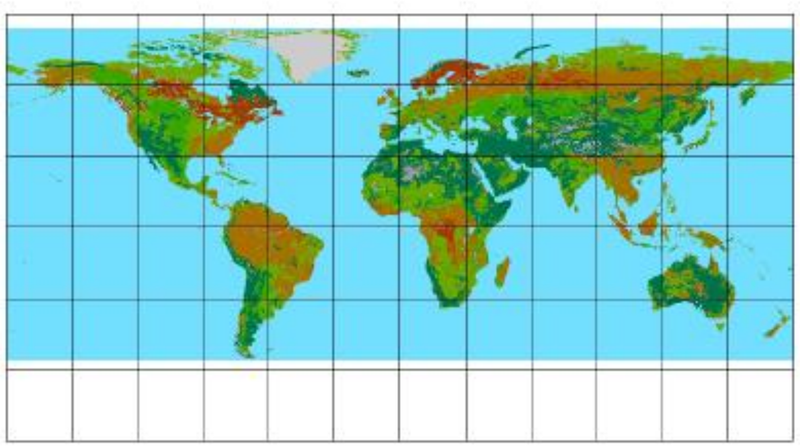
A. Climatic data:
   - WorldClim (worldclim.org)
   - CliMond (climond.org)
   - GCM Downscaled (ccafs-climate.org)

B. Soil/edaphic data
   - FAO (fao.org/geonetwork/)
   - ISRIC (isric.org) – International Soil Reference and Information Centre
   - Harmonised World Soil Database (iiasa.ac.at)

SDM – Environmental data

ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.0)

Niels H Batjes
(June 2006)



Structure of table *WISEparameterEstimates*

Name	Type	Description
CLAF	Text	FAO-Unesco (1974) Legend code
PRID	Text	profile ID (as documented in table DSMWComposition)
Drain	Text	FAO soil drainage class
Layer	Text	code for depth layer (from D1 to D5; e.g. D1 is from 0 to 20 cm)
TopDep	Integer	depth of top of layer (cm)
BotDep	Integer	depth of bottom of (cm)
CFRAG	Integer	coarse fragments (> 2mm)
SDTO	Integer	sand (mass %)
STPC	Integer	silt (mass %)
CLPC	Integer	clay (mass %)
PSCL	Text	FAO texture class
BULK	Single	bulk density (kg dm^{-3})
TAWC	Integer	available water capacity (cm m^{-1}, -33 to -1500 kPa conform to USDA standards)
CECs	Single	cation exchange capacity ($cmol_c$ kg^{-1}) for fine earth fraction
BSAT	Integer	base saturation as percentage of CECsoil
CECc	Single	CECclay, corrected for contribution of organic matter ($cmol_c$ kg^{-1})
PHAQ	Single	pH measured in water
TCEQ	Single	total carbonate equivalent (g C kg^{-1})
GYPS	Single	gypsum content (g kg^{-1})
ELCO	Single	electrical conductivity (dS m^{-1})
TOTC	Single	organic carbon content (g C kg^{-1})
TOTN	Single	total nitrogen (g kg^{-1})
CNrt	Single	C/N ratio
ECEC	Single	effective CEC ($cmol_c$ kg^{-1})

SDM – Environmental data

A. Climatic data:
   - WorldClim (worldclim.org)
   - CliMond (climond.org)
   - GCM Downscaled (ccafs-climate.org)

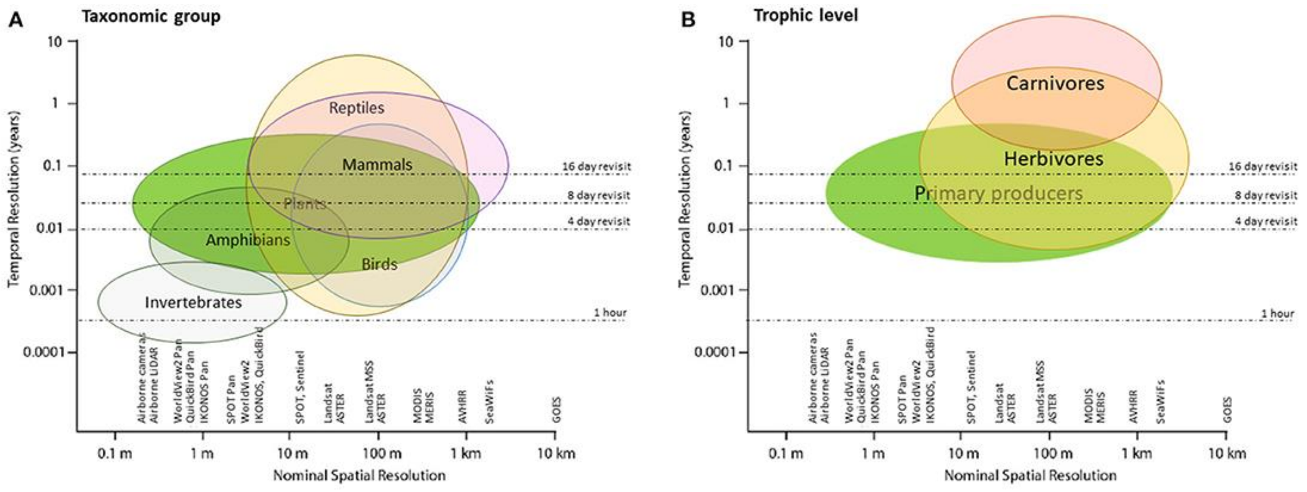
B. Soil/edaphic data
   - FAO (fao.org/geonetwork/)
   - ISRIC (isric.org) – International Soil Reference and Information Centre
   - Harmonised World Soil Database (iiasa.ac.at)

C. Other variables: E.g. Remotely sensed data

SDM – Environmental data

Environmental drivers	RS predictors	Habitat quality	Nutritional value	Seasonality/life cycle	Selected references
SOIL					
Soil type	Spectral features*, such as reflectance in the absorption region of specific constituent minerals, etc.	✓	✓		Guanter et al., 2015
Soil moisture	Spectral indices* (e.g., NDWI) or transformations (e.g., wetness); data from the SMOS Earth Explorer	✓	✓		Papes et al., 2012
CLIMATE					
Temperature	Thermal data* (LST)	✓		✓	Cord and Rödder, 2011
Precipitation	Cloud cover*; precipitation data derived from CHIRPS	✓	✓	✓	Wilson and Jetz, 2016
VEGETATION					
Vegetation structure	Laser scanning metrics* (e.g., tree height, canopy height, canopy vertical structure, etc.); parameters derived from RTM	✓	✓		Bradbury et al., 2005
Vegetation condition	Spectral indices* (e.g., NDVI, EVI) or transformations (greenness and brightness); parameters derived from RTM	✓	✓		Santos et al., 2016
Productivity	Biophysical parameters* (e.g., fPAR, LAI); parameters derived from RTM	✓	✓	✓	Coops et al., 2009
Plant stress	Spectral indices* (e.g., PRI, EWT); fluorescence data	✓	✓	✓	Saatchi et al., 2008
Land surface phenology	Phenological metrics from time series* (e.g., start/length of the growing season, senescence, etc.)	✓	✓	✓	Leitão et al., 2010
Nutrients	Spectral features*, such as reflectance in specific absorption features of nitrogen, etc.	✓	✓		Sheppard et al., 2007
Landscape configuration	Landscape and surface metrics relating to fragmentation, connectivity, heterogeneity, texture*, etc.	✓		✓	Bellis et al., 2008
Habitat information	Habitat type (LCC), fractional cover* of functional types (trees, grass, etc.)	✓	✓	✓	Wessels et al., 2004
DISTURBANCES					
Disturbances	Distance metrics* (e.g., to nearest road or settlement); Change products from LCC or fractional cover; Indices derived from time series (e.g., DI)	✓			Devictor et al., 2008
Human Impact	Stable nighttime lights* derived from the DMSP, land use intensity	✓			Escobar et al., 2015

*NDWI, Normalized Difference Water Index; SMOS, Soil Moisture Ocean Salinity; LST, Land Surface Temperature; CHIRPS, Climate Hazards group InfraRed Precipitation with Station data; RTM, Radiative Transfer Models; NDVI, Normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; fPAR, fraction of Photosynthetically Active Radiation; LAI, Leaf Area Index; PRI, Photochemical Reflectance Index; EWT, Equivalent Water Thickness; LCC, Land Cover Classification; DI, Disturbance Index; DMSP, Defense Meteorological Satellite Program. *Denotes which variable is used in the selected reference.*

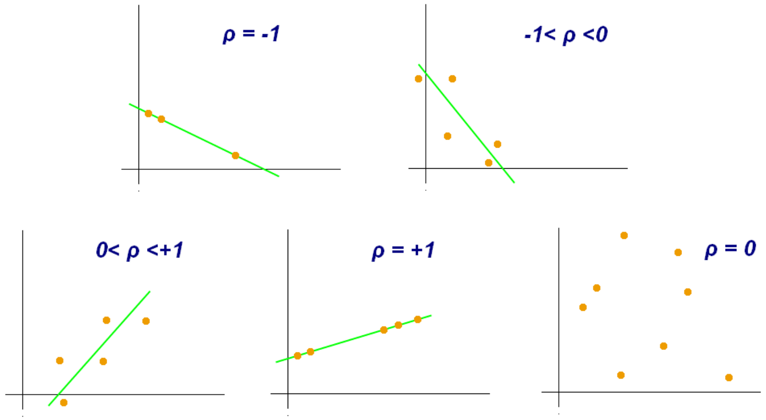


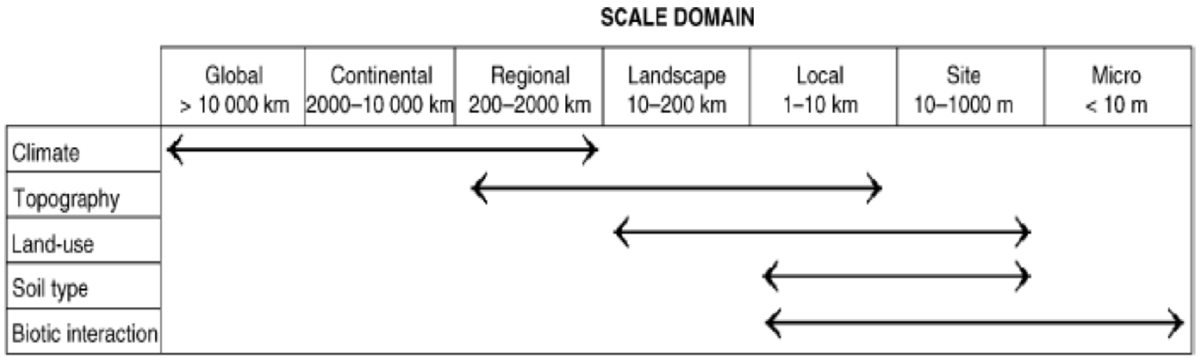
Leitão, PJ and Santos MJ, *Improving Models of Species Ecological Niches: A Remote Sensing Overview*, Frontiers in Ecology and Evolution, 2019

<u>Use at your own risk!</u>

SDM – Environmental data

- Problems:
  - Autocorrelation:
    - Pairwise correlation
      - Model multicolinearity
    - Spatial autocorrelation
  - Scale:
    - Ecological processes happen at different scales

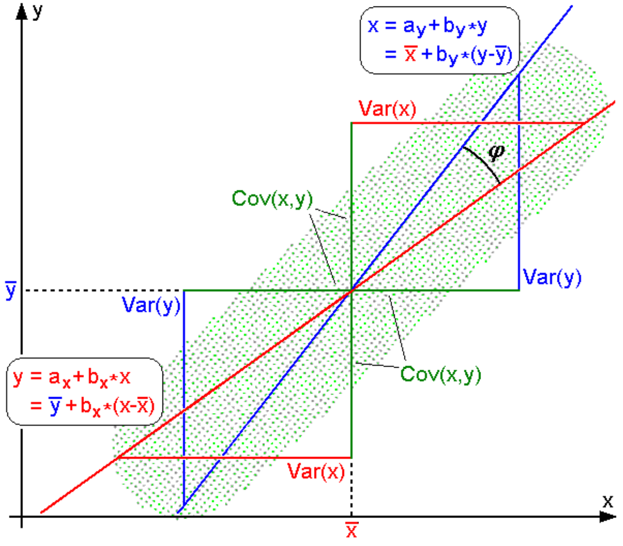




Pearson & Dawson (2003) Global Ecology & Biogeography **12:** 361–371

SDM – Environmental data

- What is correlation:
  - A statistical association/dependence between two variables
  - Common measured by the Pearson coefficient



https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \qquad \text{(Eq.1)}$$

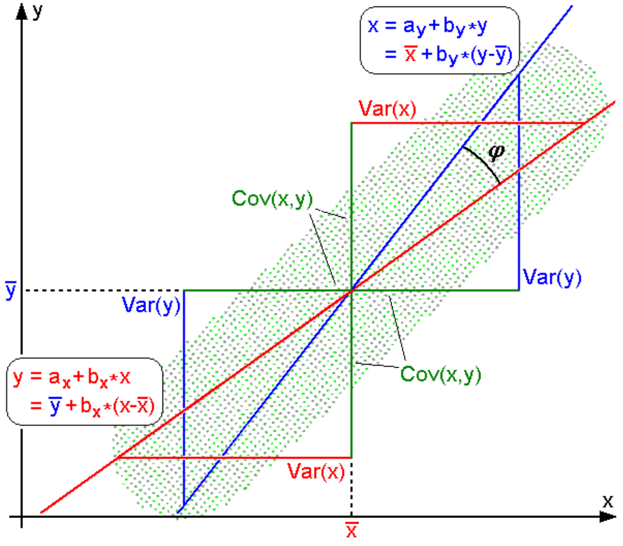
Testing your data for autocorrelation:

1. Pairwise comparison of the Pearson coefficient between each variable
2. If > 0.7, consider removing a variable
   1. Implication is that any variance is already well explained by one of the variables

SDM – Environmental data

- What is multicollinearity:
  - A statistical association/dependence of one variables towards N other variables

  - Common measure: Variance Inflation factor



https://en.wikipedia.org/wiki/Multicollinearity

$$VIF_i = \frac{1}{1 - R_i^2}$$

Testing your data for multicolinearity:

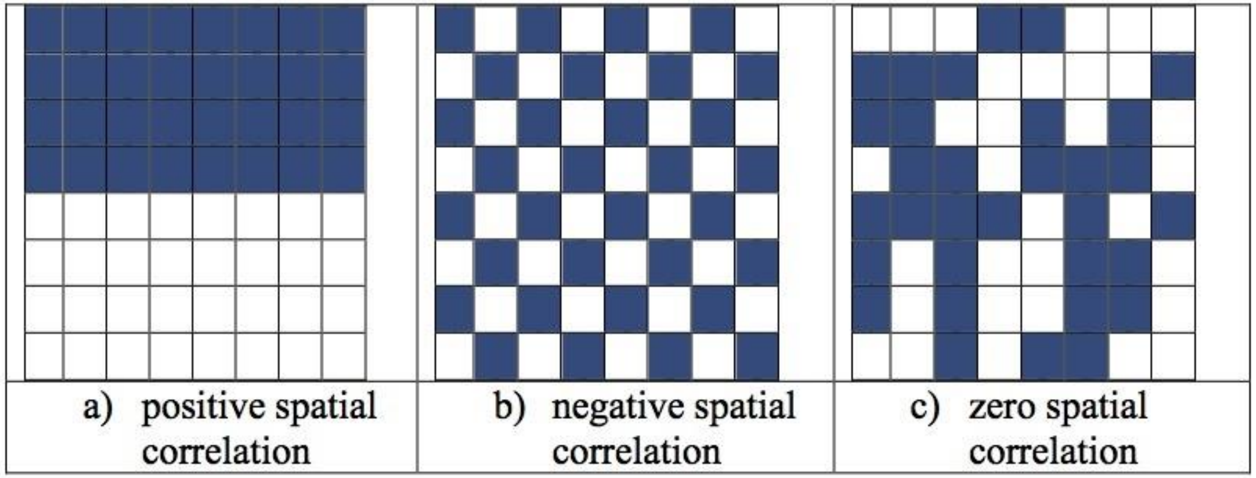
1. Simple linear regression of the target variable against all others

2. If VIF > 10, consider removing the chosen variable

   1. The implication is that the test variable is linearly dependent to the combination of the others

SDM – Environmental data

- What is spatial autocorrelation
  - It’s a autocorrelation, in space. Meaning the degree at which correlations are observed in space
  - Commonly measured using Moran’s I or Geary’s C



Ways to correct this not explored in this course.

- ArcGIS has some tools to evaluate this

- Important notion:
  - <u>Spatial autocorrelation of species occurrences is NOT model autocorrelation</u>

- Species can be inherently spatially autocorrelated
- <u>Models should NOT have spatial biases</u>

Something to add?

Next: the fun part – model algorithms

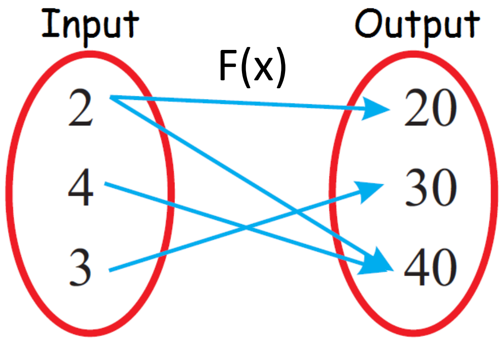
SDM – Algorithms

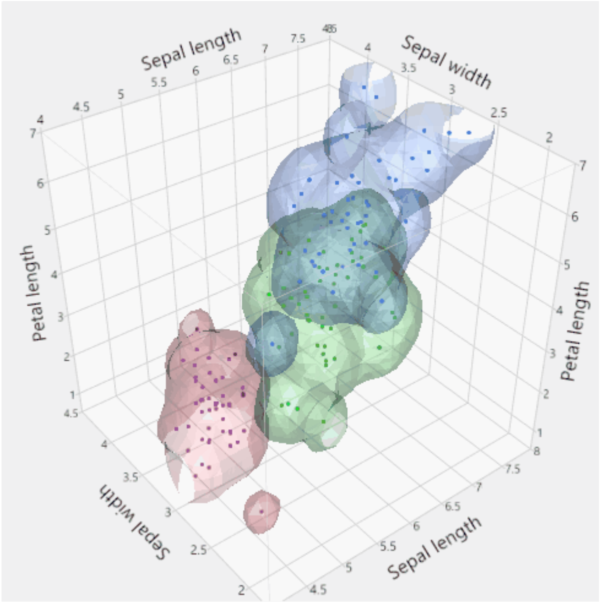
- At it's core they are "mapping functions"

- Mapping functions:
  - A function that maps the value of one domain to another

- Most algorithms are <u>classical machine learning</u>
  - ML is to teach a computer to perform a task without giving direct instructions

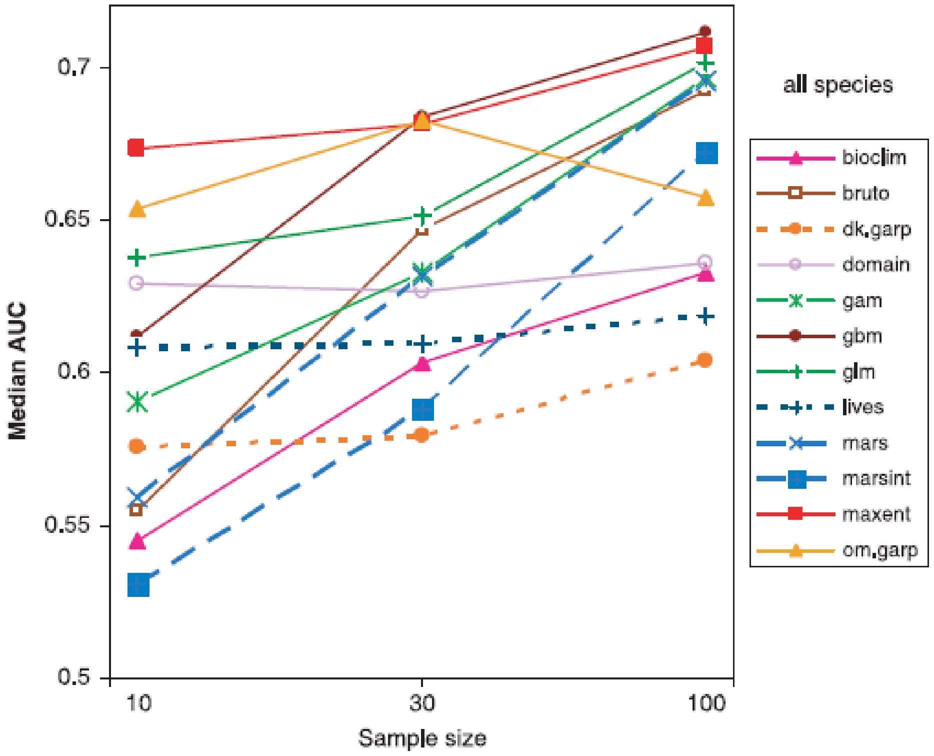
- Objective is to "teach" the computer the n-hyperdimensional niche
  - Remember Hutchinson niche definition





SDM – Algorithms

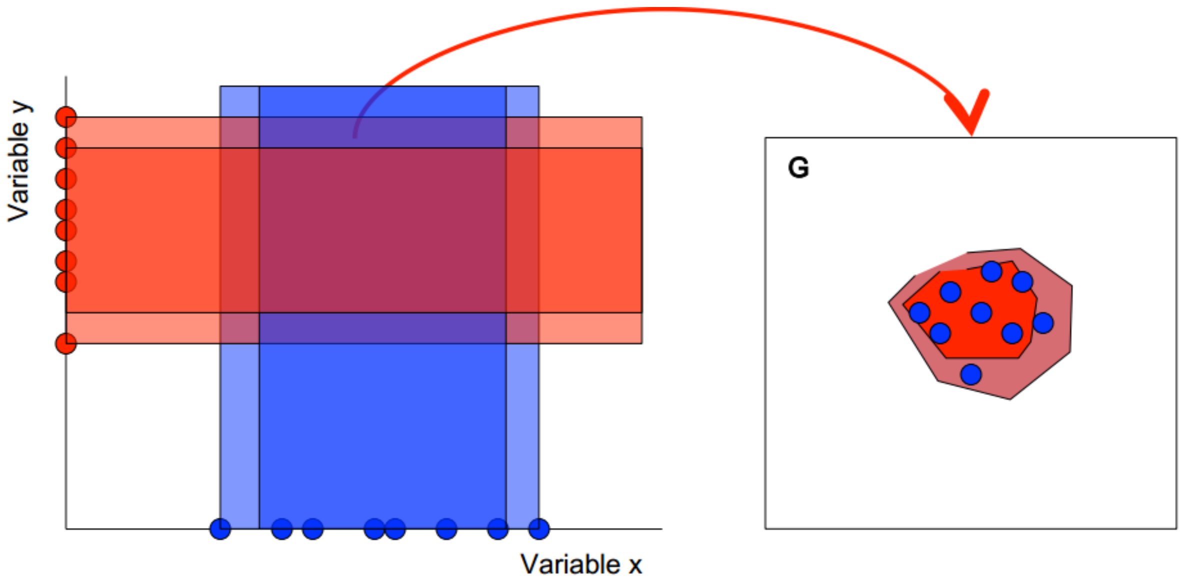
Method(s)[1]	Model/software name[2]	Species data type	Key reference/URL
Gower Metric	DOMAIN*	presence-only	Carpenter et al. 1993 http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/ http://diva-gis.org
Ecological Niche Factor Analysis (ENFA)	BIOMAPPER*	presence and background	Hirzel et al. 2002 http://www2.unil.ch/biomapper/
Maximum Entropy	MAXENT*	presence and background	Phillips et al. 2006 http://www.cs.princeton.edu/~scl
Genetic algorithm (GA)	GARP[3]*	pseudo-absence[4]	Stockwell and Peters 1999 http://www.lifemapper.org/deskt
Artificial Neural Network (ANN)	SPECIES	presence and absence (or pseudo-absence)	Pearson et al. 2002
Regression: generalized linear model (GLM), generalized additive model (GAM), boosted regression trees (BRT), multivariate adaptive regression splines (MARS)	Implemented in R[5]	presence and absence (or pseudo-absence)	Lehman et al. 2002 Elith et al. 2006 Leathwick et al. 2006 Elith et al. 2007
Multiple methods	BIOMOD	presence and absence (or pseudo-absence)	Thuiller 2003
Multiple methods	OpenModeller	depends on method implemented	http://openmodeller.sourceforge.



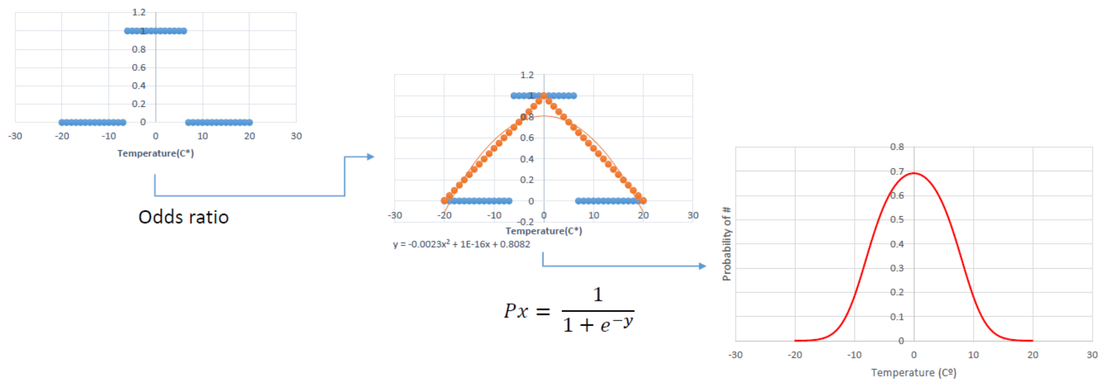
Some examples..

SDM – Algorithms

- **Bioclim:**
  - Climatic envelope
  - Statistically infers “best thresholds”
  - Very intuitive

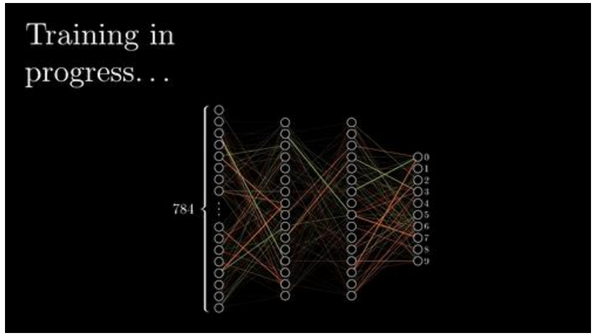


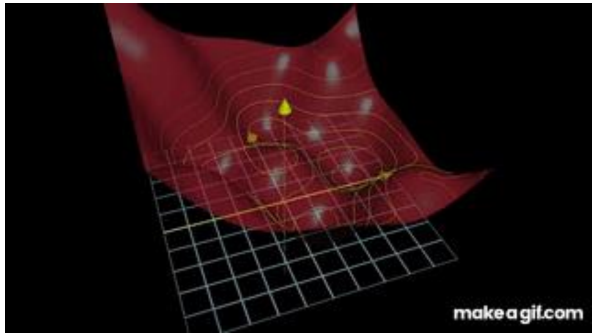
- **Generalized Linear methods:**
  - Uses “odds” at each step to fit a non-linear method.
  - Plugs in “logit function”



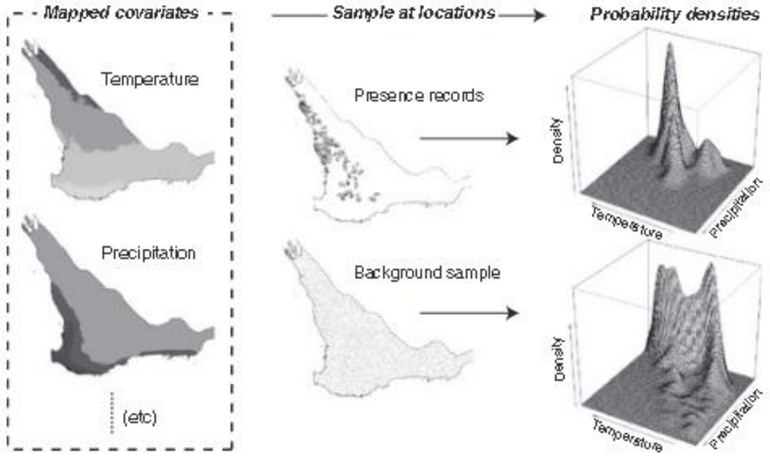
SDM – Algorithms

- **Artificial Neural Network**
  - A complex network of interacting mapping functions
    - At each step, tries new sets of parameters until it's good enough





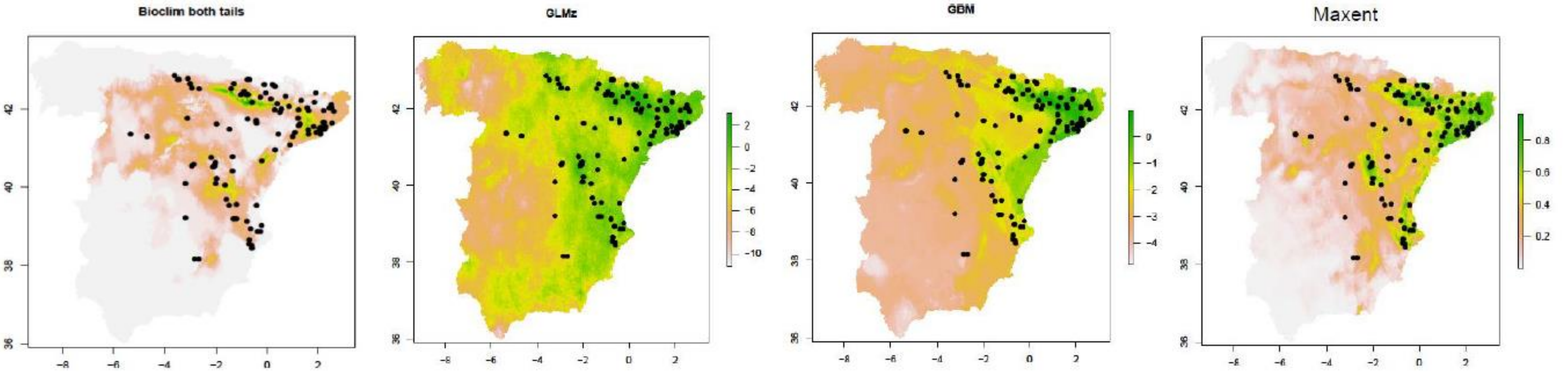
- <u>**Maxent – Maximum Entropy**</u>
  - Probabilística approach
  - Maximizes the "entropy" – or the variance of the data.
  - Similar structure to GLM's



$$\max_{\alpha,\beta} \frac{1}{m}\sum_{i=1}^{m} \ln\left(f(\mathbf{z}_i)e^{\eta(\mathbf{z}_i)}\right) - \sum_{j=1}^{n} \lambda_j|\beta_j|$$

Highly recommended reading:
Elith, J. , Phillips, S. J., Hastie, T. , Dudík, M. , Chee, Y. E. and Yates, C. J. (2011), A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17: 43-57. doi:[10.1111/j.1472-4642.2010.00725.x](10.1111/j.1472-4642.2010.00725.x)

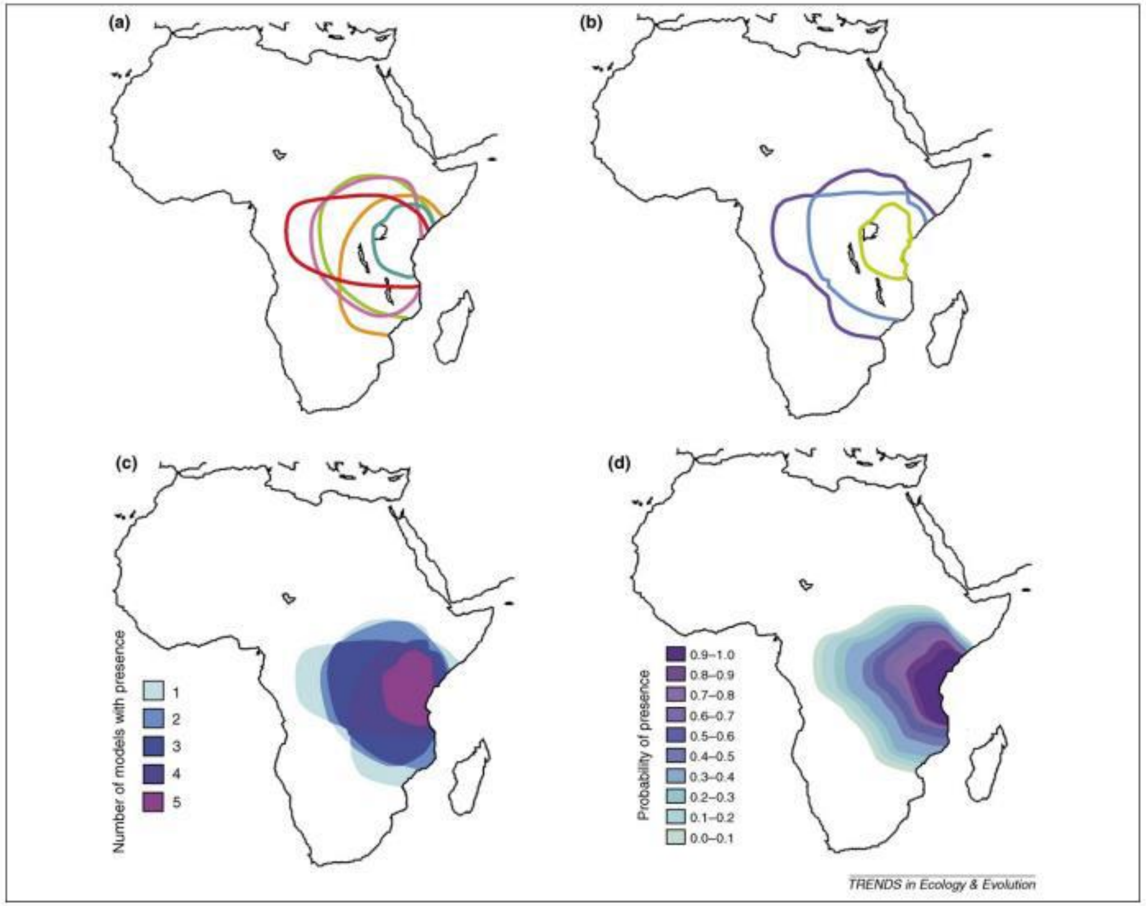
SDM – Algorithms



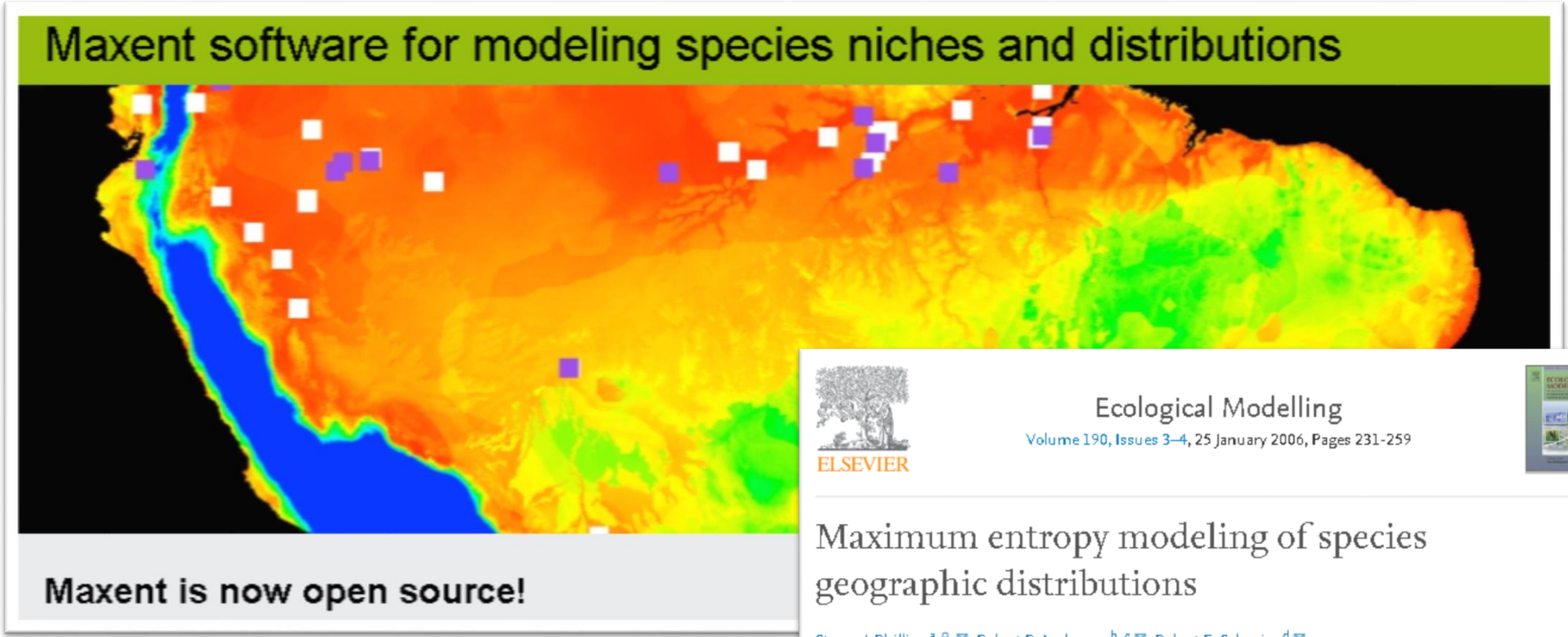
1. Why are the outputs so different?
2. Which one is best?
3. Is it possible to just combine all of them?

SDM – Algorithms

- The outputs are different because the models are different
  - *Each can capture specific "details"*
- The best? *No silver bullet.*
- Can we combine them? *Yes, it's what is most commonly used now: Biomod2 R package*



SDM – Algorithms: Maxent



Ecological Modelling
Volume 190, Issues 3–4, 25 January 2006, Pages 231-259

Maximum entropy modeling of species geographic distributions

Steven J. Phillips [a], Robert P. Anderson [b, c], Robert E. Schapire [d]

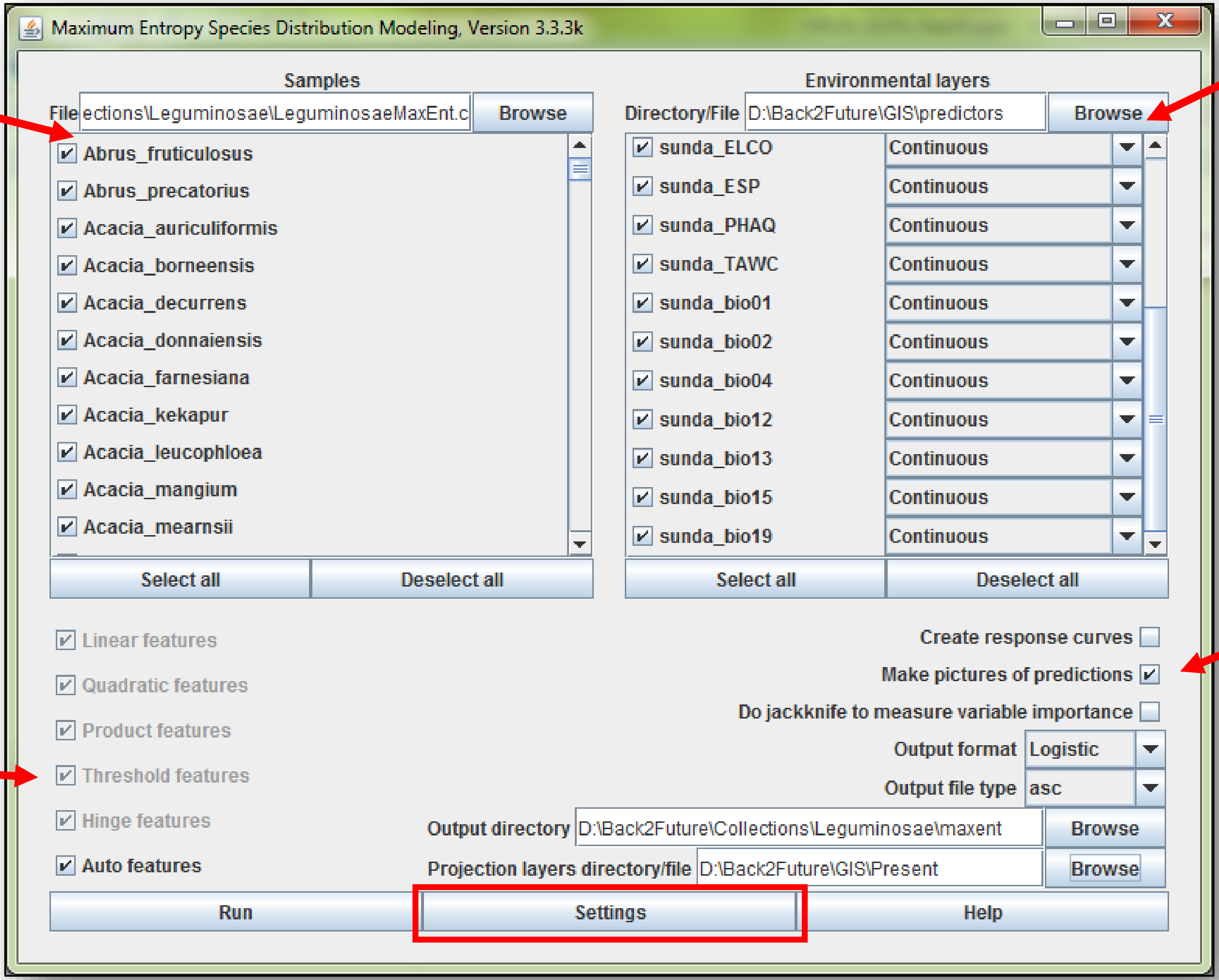
https://biodiversityinformatics.amnh.org/open_source/maxent/

https://www.andersonlab.ccny.cuny.edu/resources

SDM – Algorithms: Maxent

List of species

List of env. vars



Extra outputs

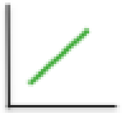
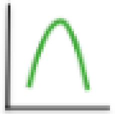
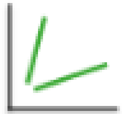
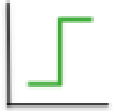
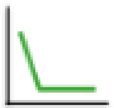
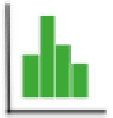
List of feature fitting methods

SDM – Algorithms: Maxent

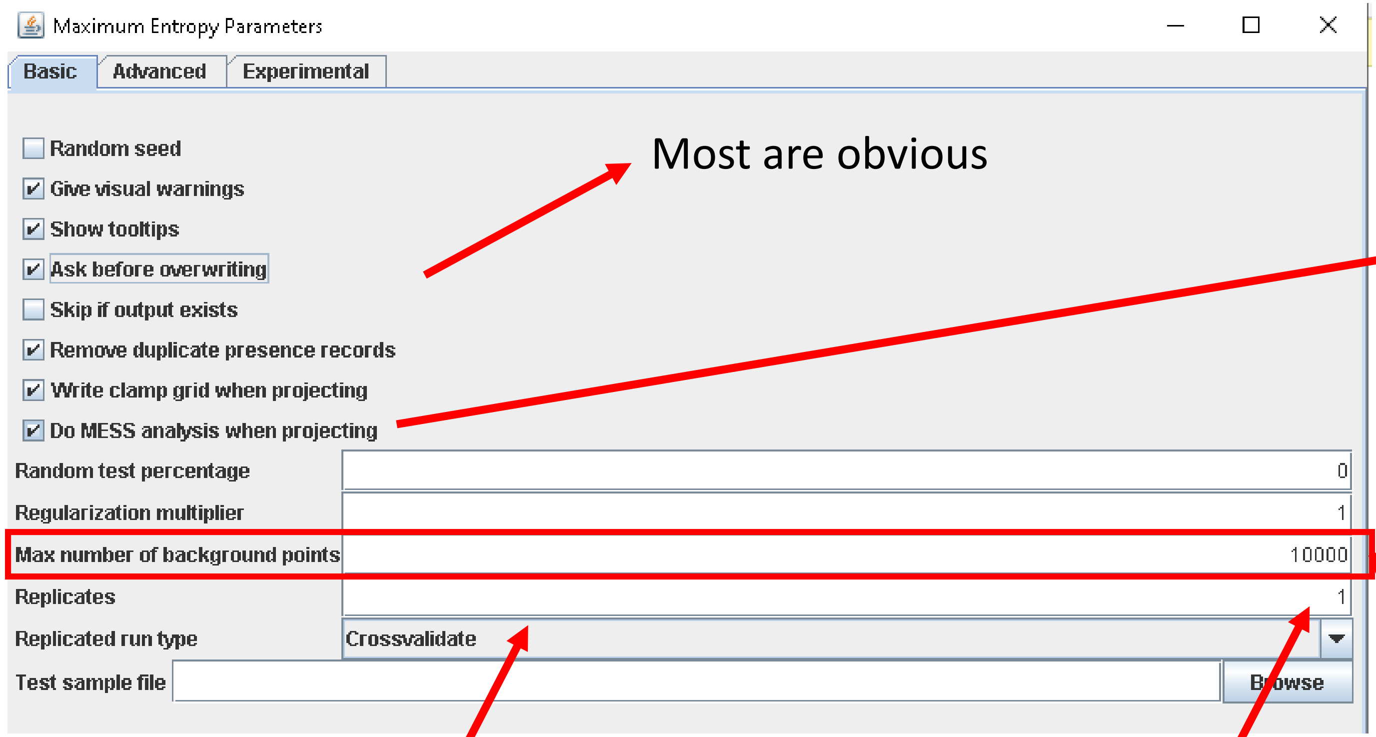
What are "feature":

ML lingo for covariates

Here used as a transformation function

Feature type	Interpretation	Constraint	Shape
Linear	Continuous variable	The *mean* of each environmental variable at an unknown location should be close to the mean of that variable in known occurrence locations.	
Quadratic	Square of the variable	The *variance* of each environmental variable at an unknown location should be close to the variance of that variable in known occurrence locations.	
Product	Pairs of continuous variables – allows for interactions	The *co-variance* of two environmental variables at an unknown location should be close to the co-variance of those variables in known occurrence locations.	
Threshold	Conversion into binary response based on a threshold	The proportion of predicted occurrences with values above the threshold (binary response = 1) should be close to the proportion of known occurrences.	
Hinge	As threshold type, but response after the threshold (knot) is linear	The mean above the knot of each environmental variable at an unknown location should be close to the mean above the knot of that variable in known occurrence locations.	
Categorical	Categorical variable	The proportion of predicted occurrences in each category should be close to the proportion of observed occurrences in each category.	

SDM – Algorithms: Maxent



Most are obvious

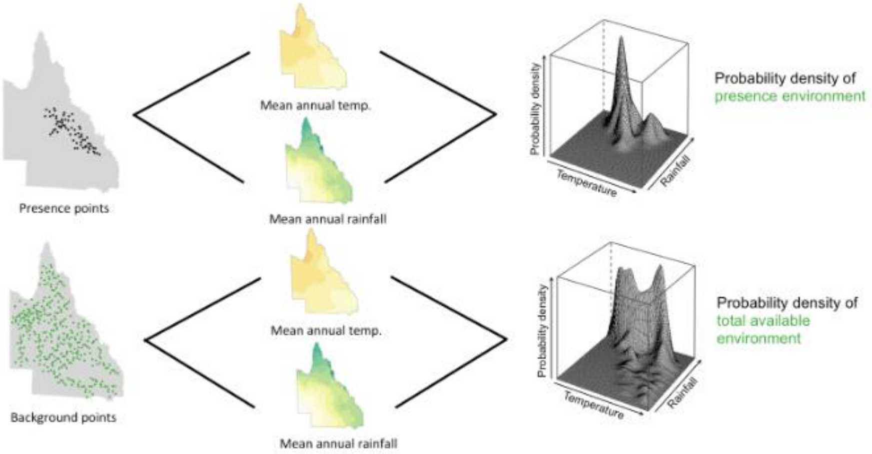
Type of acc. evaluation

Nº of repetitions

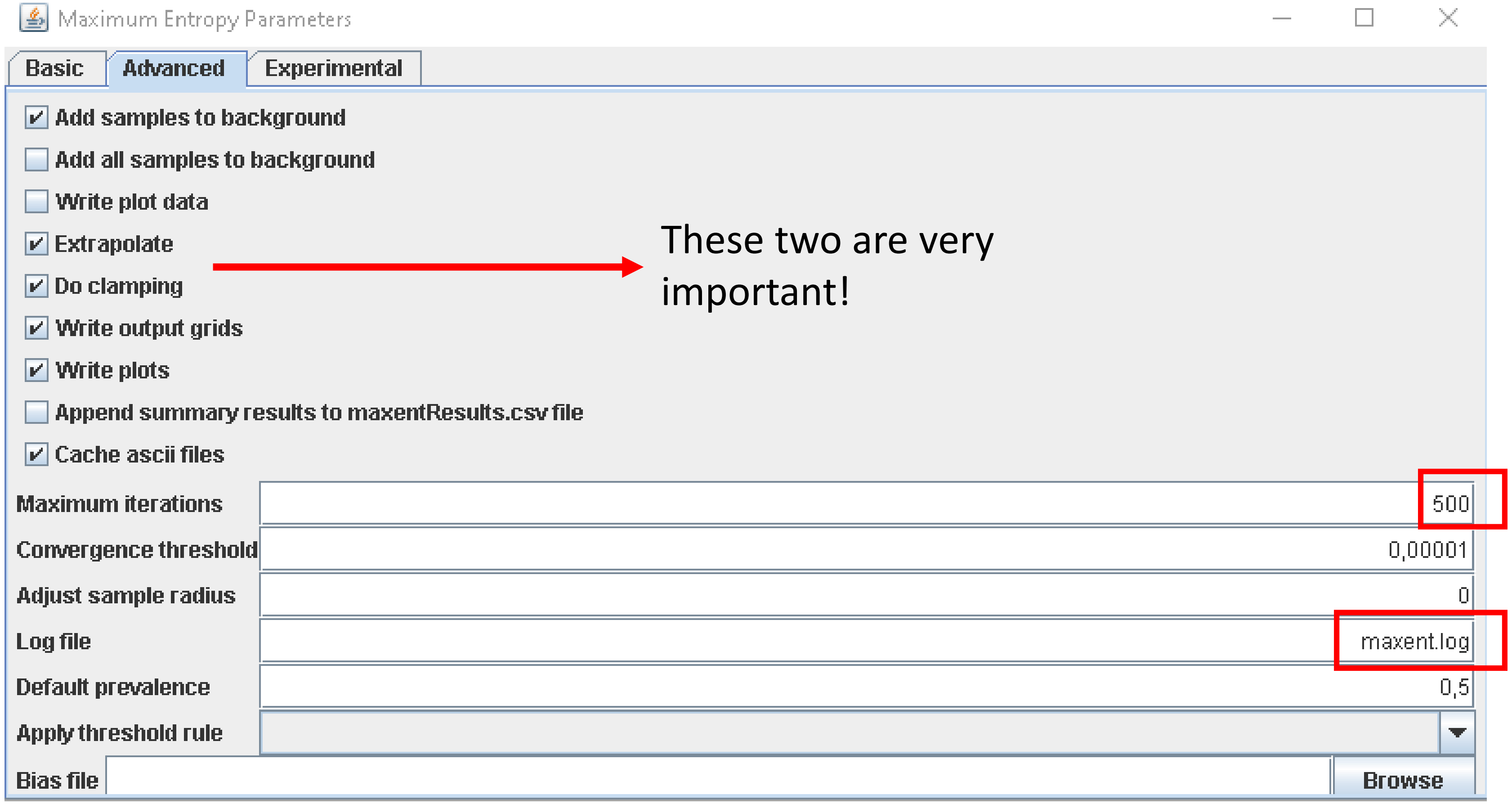
Multivariate Environmental Similarity Surfaces

-A measure of the similarity of the environmental variables in training vs the prediction environment

Elith, J. , Kearney, M. and Phillips, S. (2010), The art of modelling range-shifting species.
Methods in Ecology and Evolution, 1: 330-342. doi:10.1111/j.2041-210X.2010.00036.x



SDM – Algorithms: Maxent

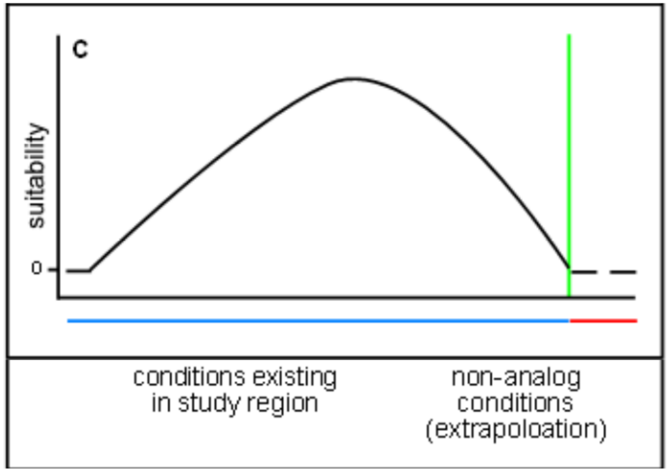


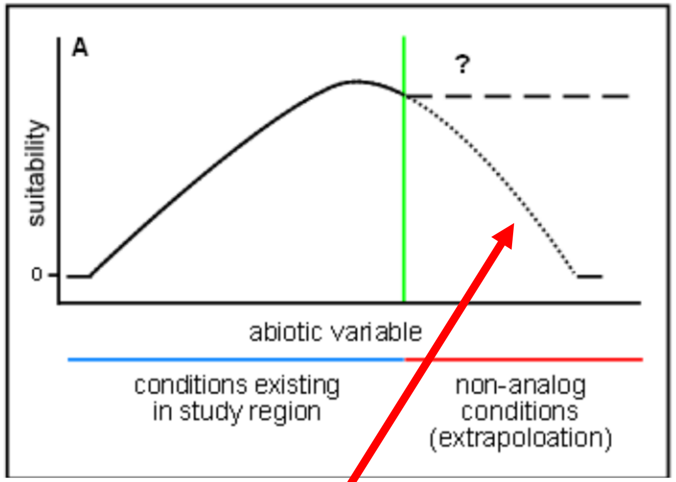
These two are very important!

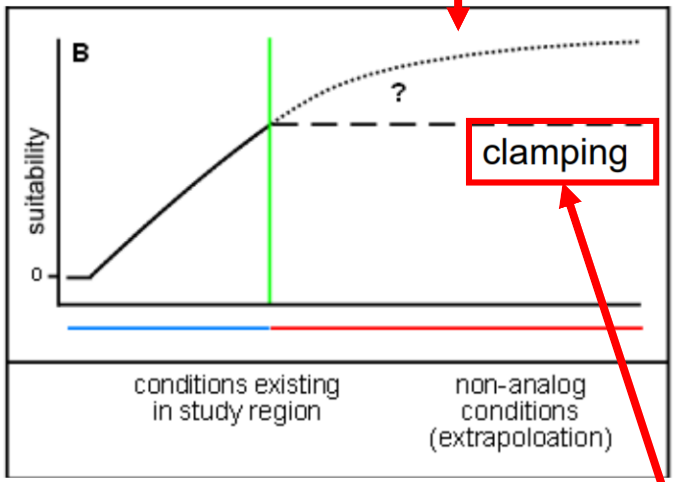
SDM – Algorithms: Maxent

“Extrapolate”

This is good news!







This is bad news

Clamping

“fade by clamping”

SDM – Algorithms: Maxent

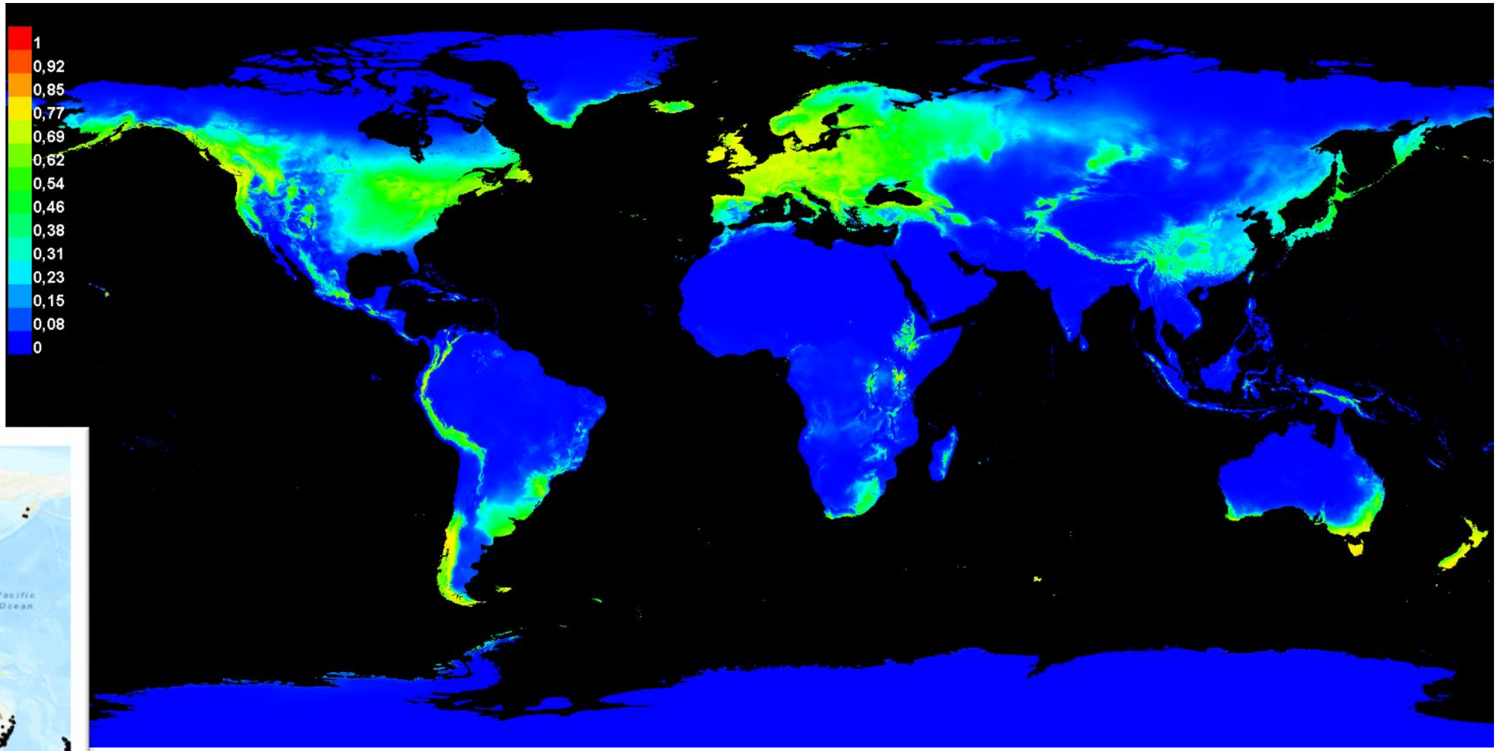
WorldClim 1.4: Current conditions (~1960-1990)

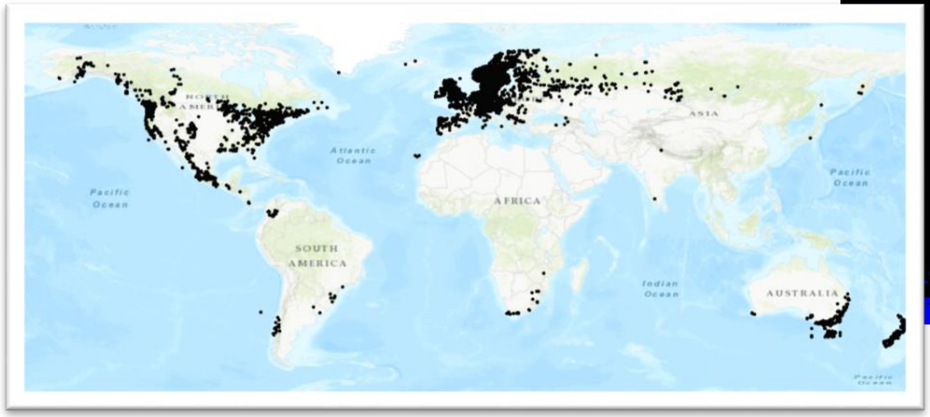
If you need the highest resolution (**30 arc-seconds (~1 km)**) then you can **download by tile**. See the Methods page for more info on how these data were generated, and this page for info on details about the data (such as units).

Generic grid format

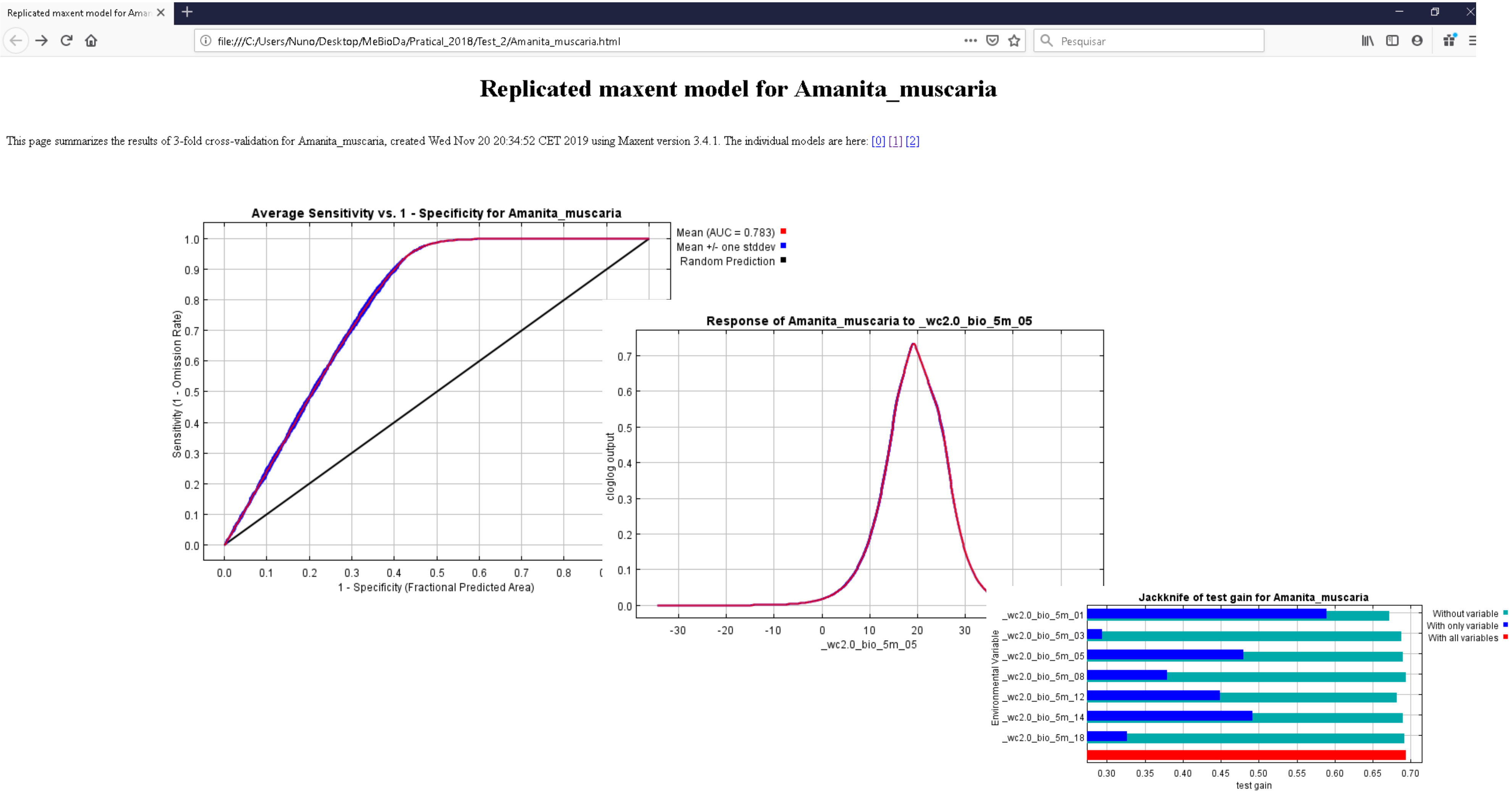
variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C * 10)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C * 10)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C * 10)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
bioclimatic variables	bio 10m	bio 5m	bio 2.5m	bio1-9, 10-19

SDM – Algorithms: Maxent





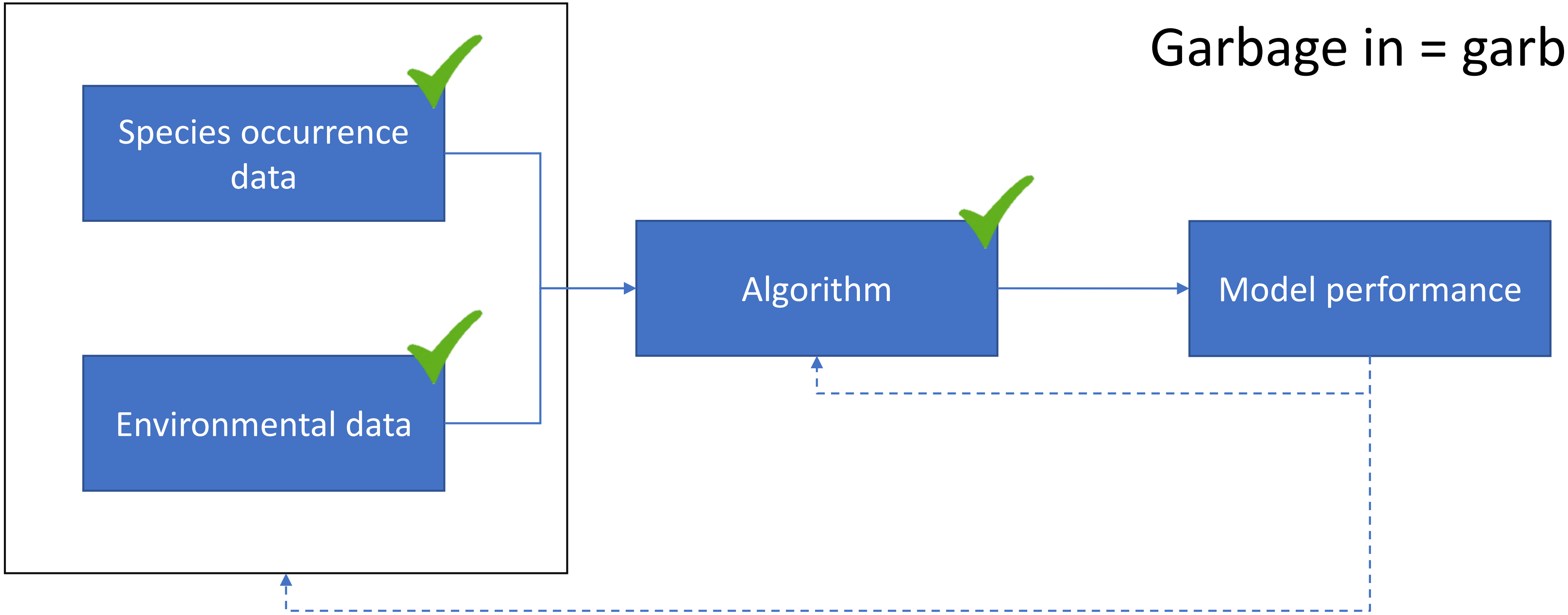
SDM – Algorithms: Maxent diagnostics



Something to add?

Next: The frustrating part - Model performance

SDM: Model evaluation



Garbage in = garbage out!

SDM is evaluated like “any other” classification exercise

SDM: Model evaluation

Confusion matrix		**Reference data**	
		Presence	**Absence**
Model predictions	**Presence**	**True presences**	**False presences**
	Absence	**False absences**	**True absences**

Type I error

Type II error

…But for most cases in SDM we do not have true absences….

SDM: Model evaluation

<table>
<tr><td></td><td></td><td colspan="2">True condition</td><td></td><td colspan="2"></td></tr>
<tr><td></td><td>Total population</td><td>Condition positive</td><td>Condition negative</td><td>Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$</td><td colspan="2">Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$</td></tr>
<tr><td rowspan="2">Predicted condition</td><td>Predicted condition positive</td><td>True positive</td><td>False positive, Type I error</td><td>Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$</td><td colspan="2">False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$</td></tr>
<tr><td>Predicted condition negative</td><td>False negative, Type II error</td><td>True negative</td><td>False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$</td><td colspan="2">Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$</td></tr>
<tr><td></td><td></td><td>True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$</td><td>False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$</td><td>Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$</td><td rowspan="2">Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR–}}$</td><td rowspan="2">$F_1$ score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$</td></tr>
<tr><td></td><td></td><td>False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$</td><td>Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$</td><td>Negative likelihood ratio (LR–) $= \frac{\text{FNR}}{\text{TNR}}$</td></tr>
</table>

All of these measures provide different insights onto our model quality

But unfortunately, presence-only SDM to not have absence data by definition..

SDM: Model evaluation

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

All of these measures provide different insights onto our model quality

But unfortunately, presence-only SDM to not have absence data by definition..

These provide insights into how good it detects positives or negatives

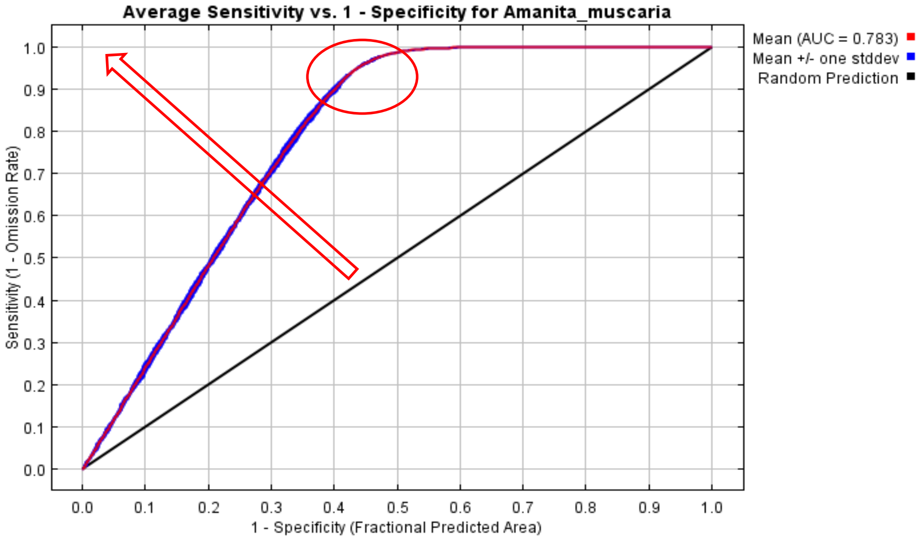
$$Sensitivity = \frac{TP}{TP + FN}$$

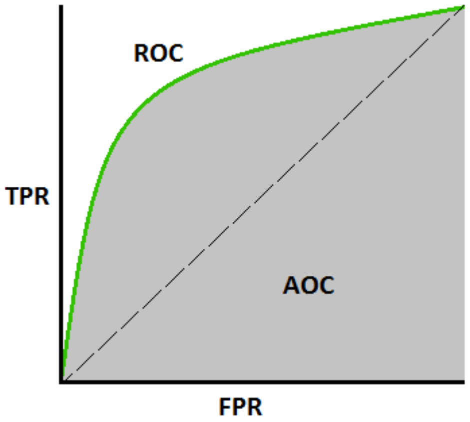
$$Specificity = \frac{TN}{TN + FP}$$

Some estimate of the “discriminatory power”

SDM: Model evaluation

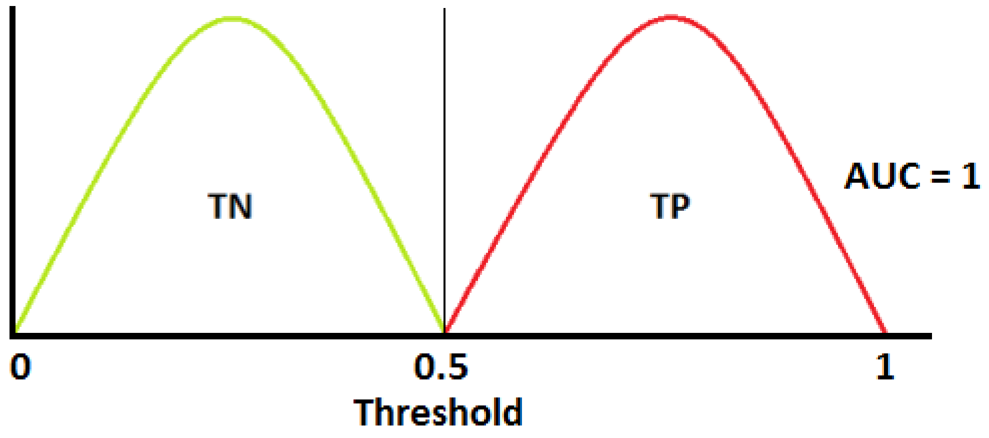
- AUCROC
  - *A measure of the discrimination power of the model*
- ROC curve:
  - Plot the values of TPR and FPR on a 2 dimensional axis with varying thresholds
- AUC:
  - The integral of the area under the curve
  - It’s a unitary square .: Max area = 1





Check this quick explanation: https://www.youtube.com/watch?v=4jRBRDbJemM

SDM: Model evaluation

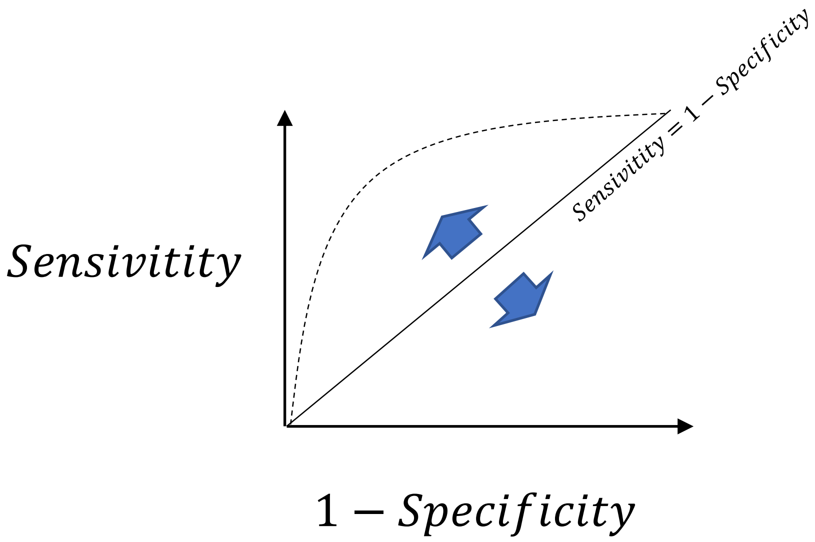


If the classification is perfect, then its possible to find a "threshold" that perfectly separates the two classes

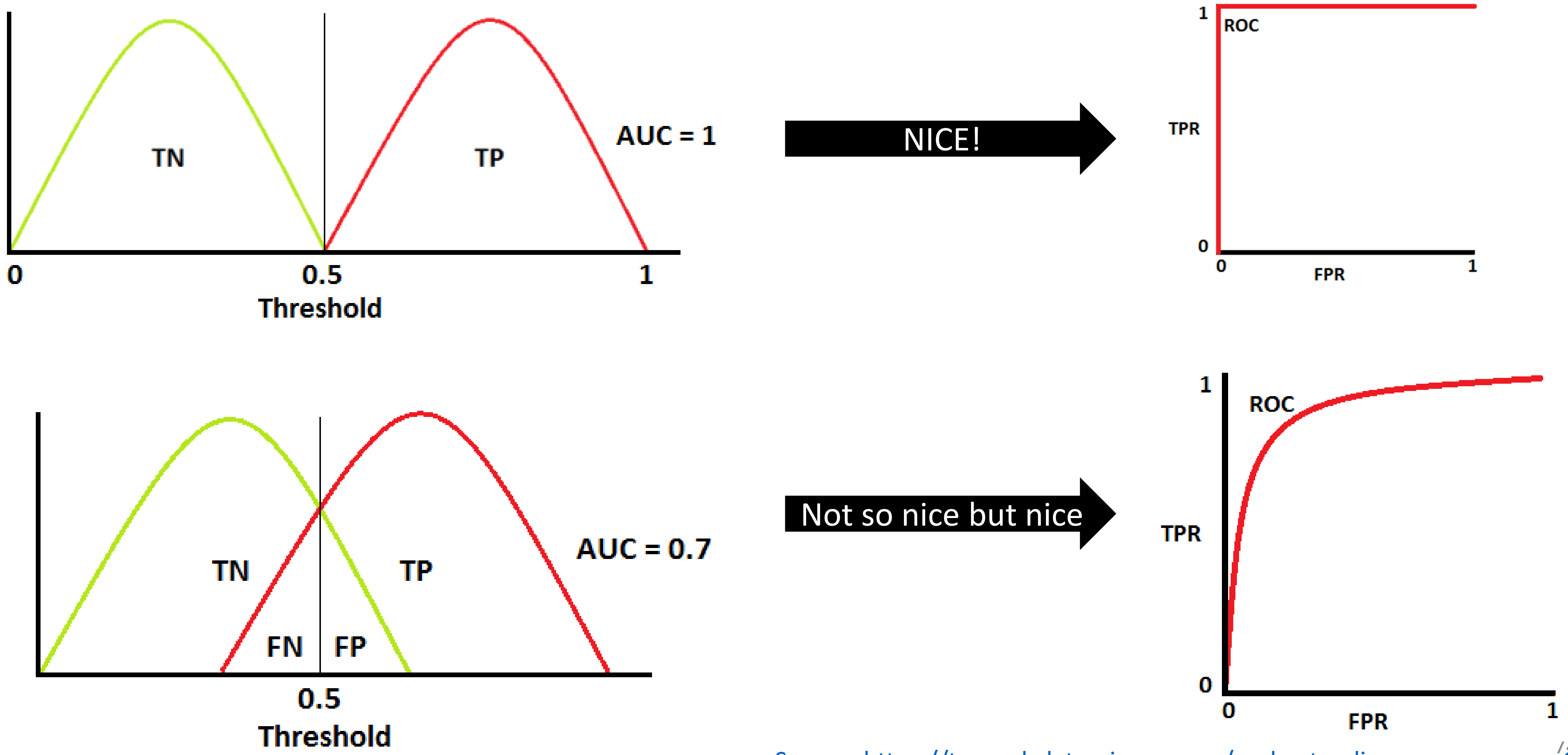
An alternative is then to find this threshold to provide an inference of the best ability of the model

Here is where AUCROC comes in play:

Prob.	Th: 0.25	Th: 0.5	Th: 0.75
0.1	0	0	0
0.4	1	0	0
0.75	1	1	1
0.8	1	1	1
0.56	1	1	0
0.2	0	0	0



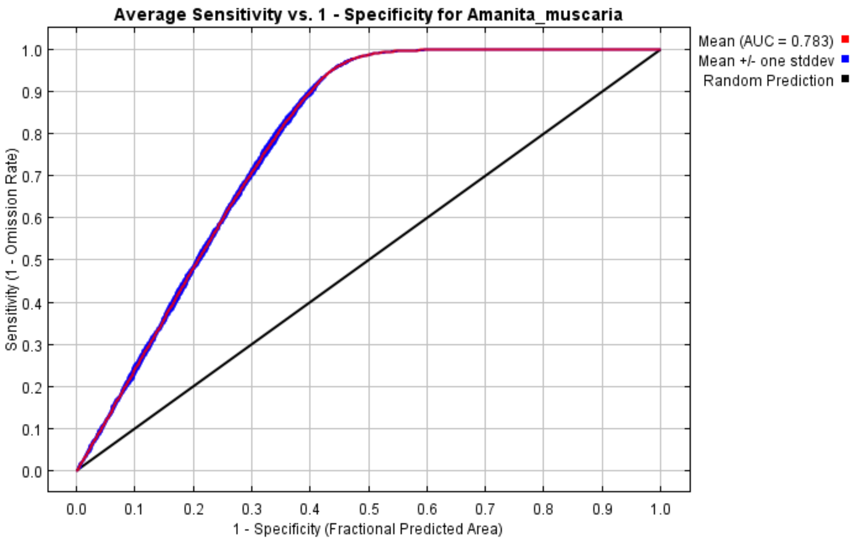
SDM: Model evaluation



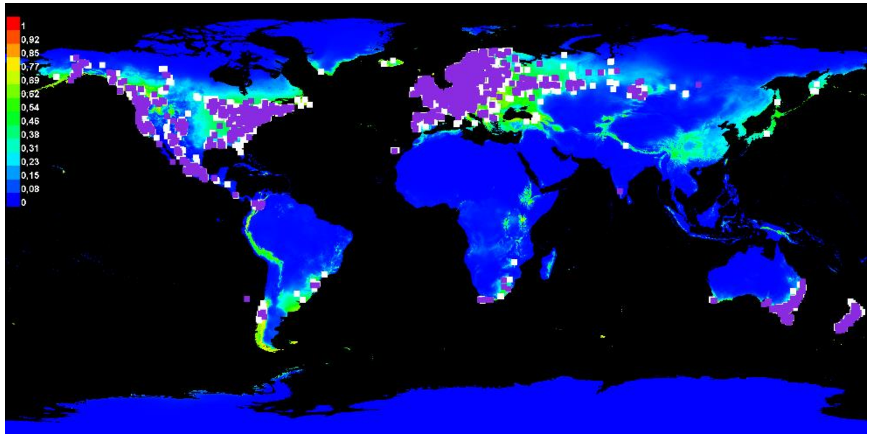
[Source: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5](https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5)

SDM: Model evaluation

- But Maxent only takes in presence data still so:
  - Presences – input data
  - "Absences" – Background data
    - Thus, pseudo-absences



  - Limitations:
    - Samples "background" as pseudo-absences
    - <u>AUC is now dependant on "prevalence"</u>



SDM: Model evaluation

PROBLEM !!!

"When AUC statistics are applied to presence-only data and pseudo-absences, the maximum achievable AUC value is **no longer 1, BUT 1- *a*/2**; where *a* stands for the true species' distribution, which we typically do not know" (Phillips et al. 2006)

Implications:

- Since Maxent uses background as PA .: your AUC will be dependant on the total area
  - <u>Be aware of this!</u>
- Use other methods e.g. Boyce index or Null model

x			
		x	
x			
	x		

$$1 - \frac{\frac{4}{16}}{2} = 1 - 0.125 = 0.875$$

x			x
		x	
x	x		
	x		

$$1 - \frac{\frac{6}{16}}{2} = 0.8125$$

x		x	
	x	x	x
x		x	x
	x	x	

$$1 - \frac{\frac{10}{16}}{2} = 0.6875$$

SDM: Model evaluation

- Null model:
  - Pattern-generating model based on random sampling from a known or imagined distribution, or randomizations of ecological data (Gotelli and McGill 2006).

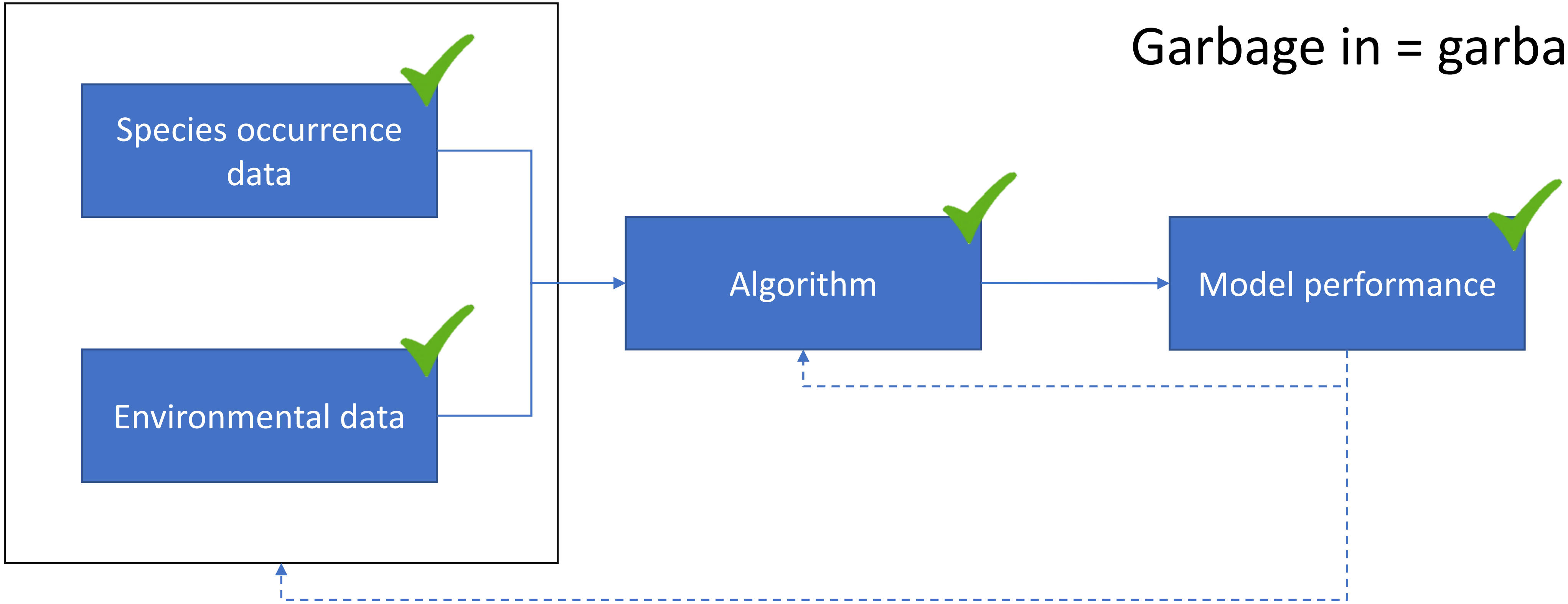
- **Objective:**
  - Tests whether chance alone is enough to explain the observed patterns.
  - Straightforward and closely resembles hypothesis testing in conventional statistical analyses.

A null-model for significance testing of presence-only species distribution models

Niels Raes and Hans ter Steege

N. Raes (raes@nhn.leidenuniv.nl), National Herbarium of the Netherlands, Leiden Univ. branch, Einsteinweg 2, NL-2300 RA, Leiden, the Netherlands H. ter Steege, Inst. of Environmental Biology, Section Plant Ecology and Biodiversity and the National Herbarium of the Netherlands, Utrecht Univ. branch, Utrecht Univ., Sorbonnelaan 14, NL-3584 CA, Utrecht, the Netherlands.

SDM: Model evaluation



Garbage in = garbage out!

Let’s recap!

SDM Recap

- Occurrence data:
  - Should be representative of the ecological niche of the species
    - For this class, you should assess that by GIS + Auxiliary information
  - There will be biases!
    - And errors – check your data

- Environmental data:
  - Should be selected according to ecological motifs
    - Area of training should be clipped to species distribution – remember AUC
  - Check for statistical correlation:
    - Pairwise pearson correlation & multicollinearity

SDM Recap

- **Modelling algorithm:**
  - There are many options but we will focus on MAXENT
  - You can find a lot of extra resources on it:
    - https://biodiversityinformatics.amnh.org/open_source/maxent/
    - https://www.andersonlab.ccny.cuny.edu/resources
  - Check all the options, in particular:
    - Clampin; Number of background points; MESS; Jacknife test of variable importance
  - Interpret the diagonostics!
- **Model evaluation:**
  - Similar to any other classification exercise
  - Maxent AUCROC is dependant on the prevelance
    - Both the nr of points as well as the overall AOI used

Assumptions

SDM assumptions	Violations of SDM assumptions	Consequences of violating SDM assumptions on model performance	Solutions to improve SDM performance
Assumption 1: Species is at equilibrium with environmental conditions in its native range	Native range is restricted by biotic interactions (e.g., competition, predation, human disturbance, etc.)	Underprediction of potential regions of suitability	Use historical range information for model training
	Native range is restricted by dispersal limitation	Underprediction of potential regions of suitability	
Assumption 2: Niche stability	Evolutionary or behavioral adaptation to environmental conditions in introduced area	Underprediction of potential regions of suitability	Shorten timescale of analysis
	New ecological relationships in introduced range	Overprediction or underprediction of potential regions of suitability	
Assumption 3: Training samples are representative of environmental conditions in native range	Training samples are biased	Underprediction of potential regions of suitability	Use design- or model-based environmental stratifications to target underrepresented areas for additional field data collection
			Generate random pseudo-presence points across native range
	Few training samples are available	Underprediction of potential regions of suitability	Generate adequate number of random pseudo-presence points from native range
Assumption 4: Climatic conditions between training and introduced areas are analogous	Novel climatic conditions occur in introduced area; modeled responses extrapolate beyond range of values for environmental predictors found in native range	Overprediction or underprediction of potential regions of suitability	Use a clamping procedure to limit predictions in regions with novel climatic conditions

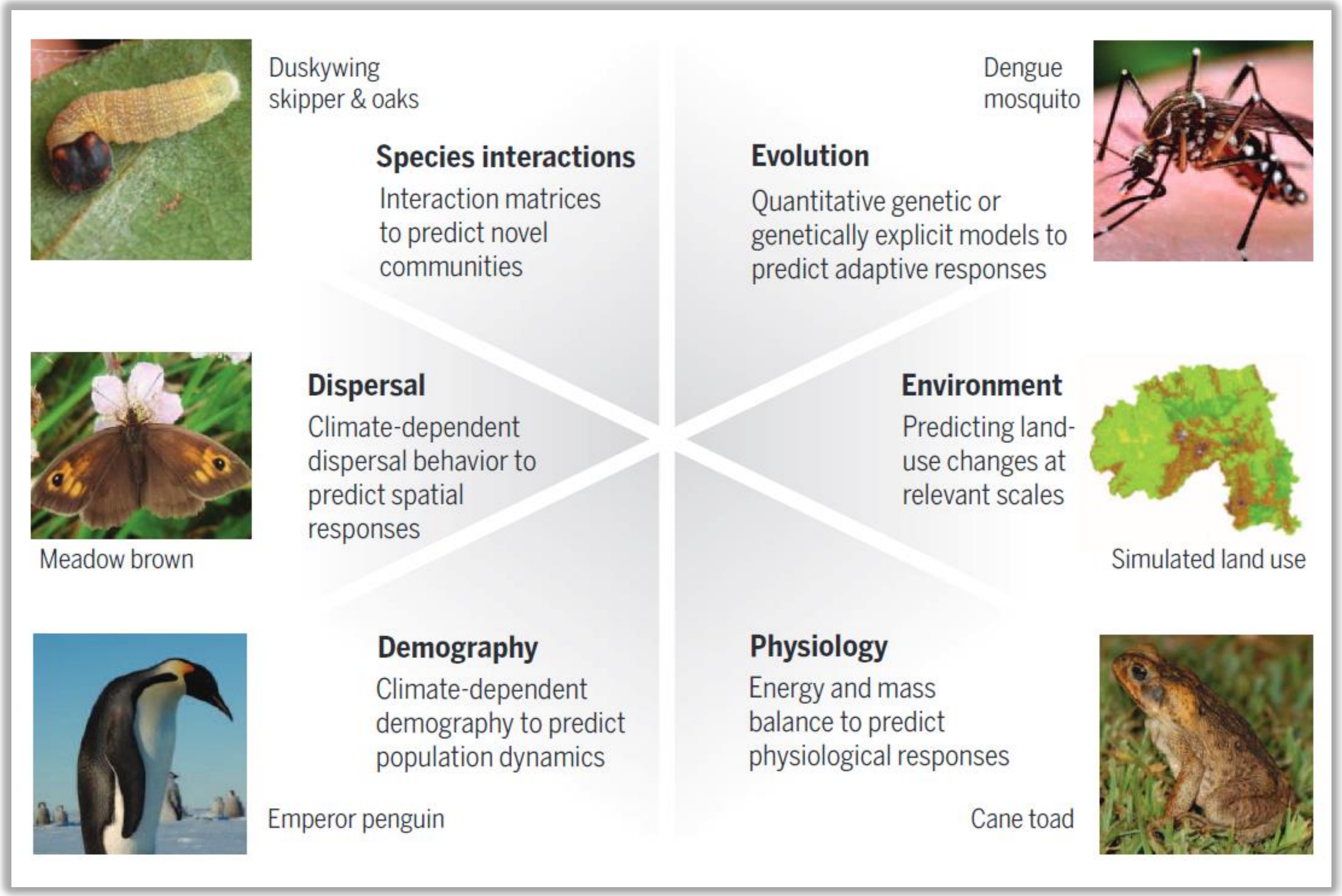
Richmond et al 2010

Assumptions

- Correlations. that variables used reflect the niche requirements of a species
- Equilibrium and habitat saturation (suitable habitat is fully occupied)
- Dispersal and landscapes. That individuals have ability to disperse to suitable locations within their niche space.
- Biotic interactions. That species respond independently to the environmental factors that determine its niche space and thurs habitat and distribution.
- Adaptation and evolution. Niche conservatism

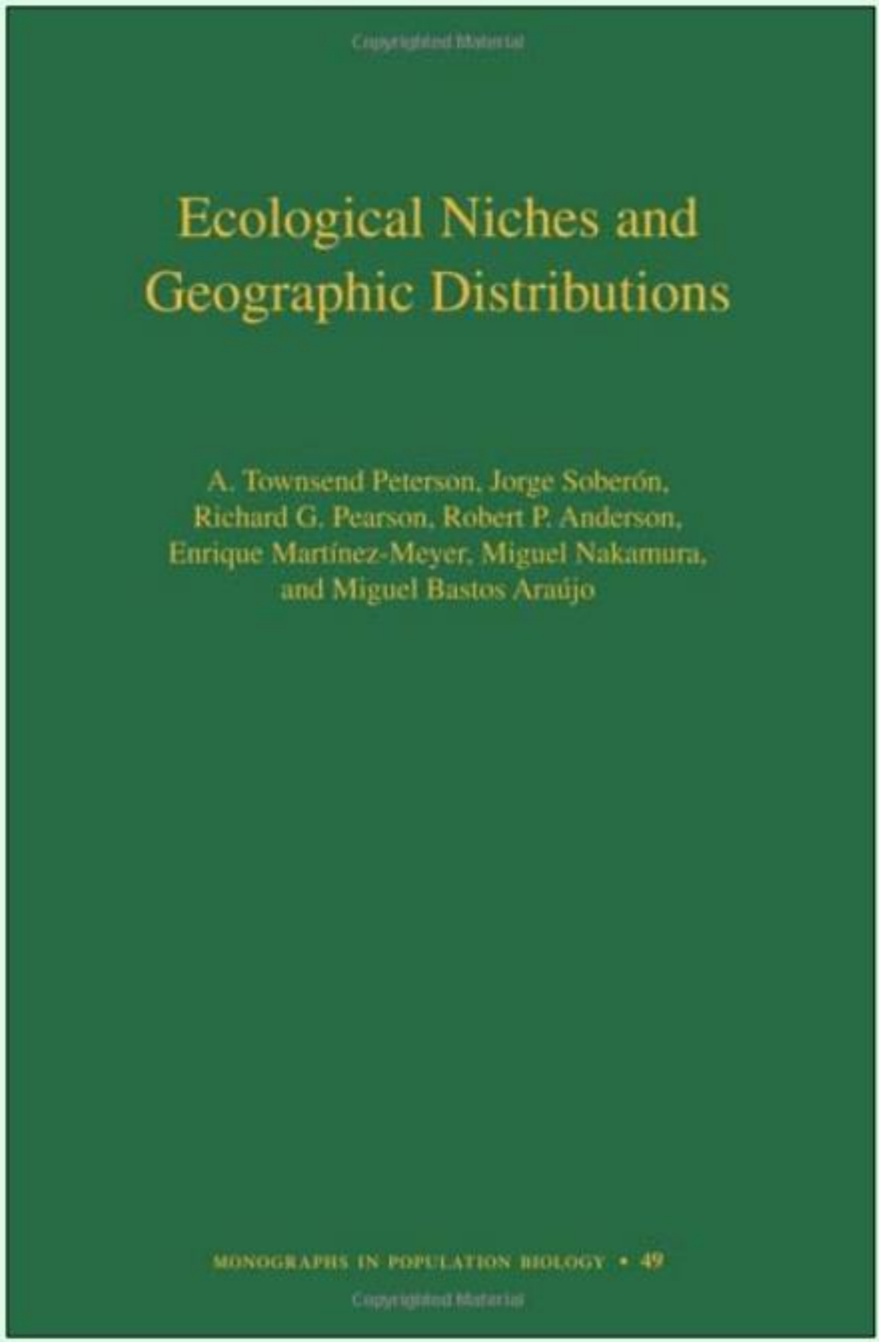
Wiens et al 2009

What is missing?



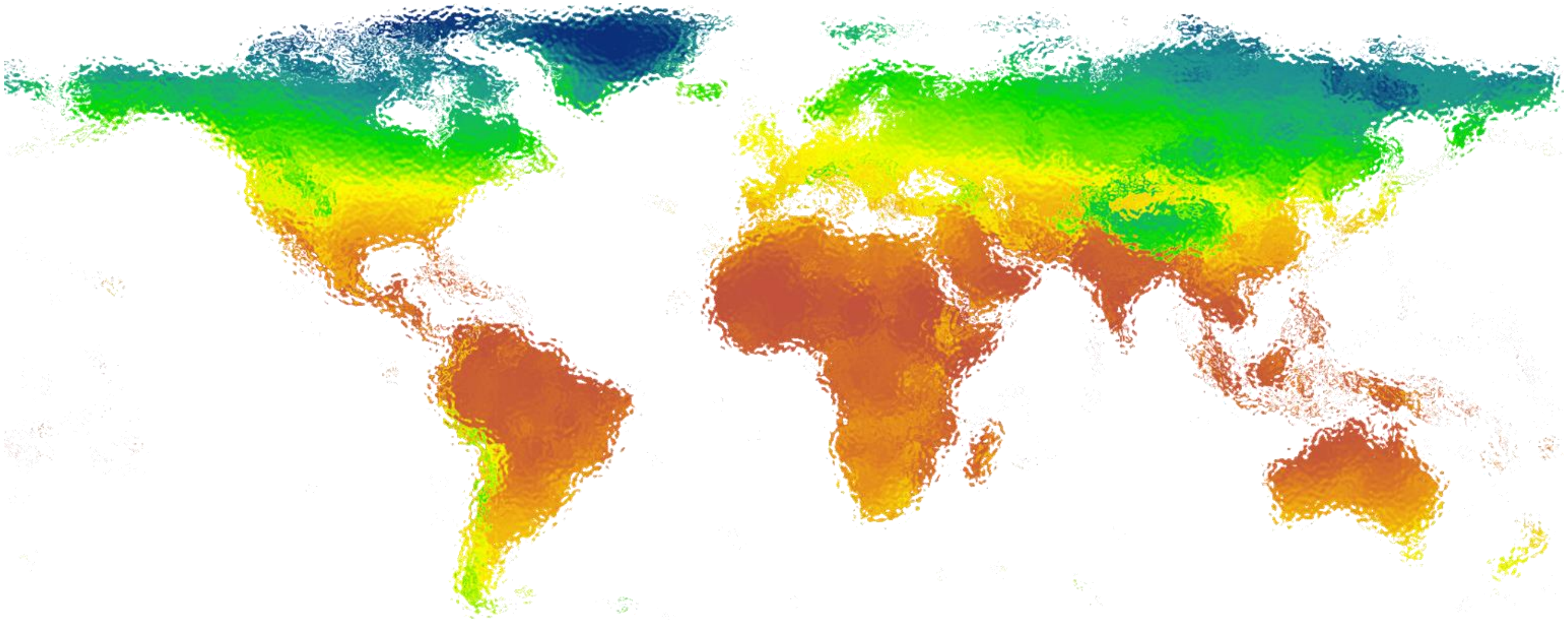
Urban *et al.,* - Improving the forecast for biodiversity under climate change - 2016 - Science

Naturalis Biodiversity Center

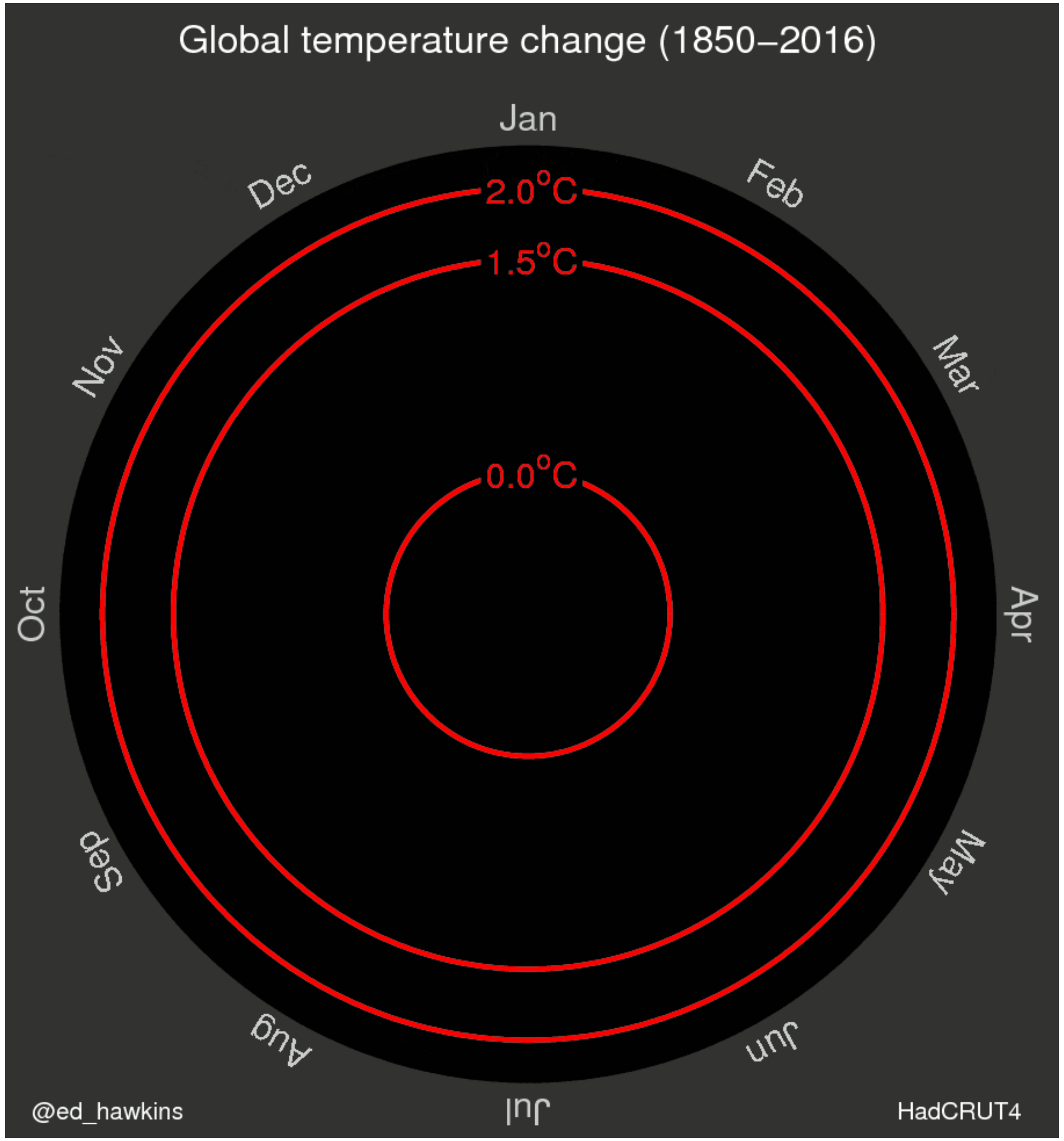

Copyrighted Material
Ecological Niches and
Geographic Distributions
A. Townsend Peterson, Jorge Soberón,
Richard G. Pearson, Robert P. Anderson,
Enrique Martínez-Meyer, Miguel Nakamura,
and Miguel Bastos Araújo
MONOGRAPHS IN POPULATION BIOLOGY • 49
Copyrighted Material

Practical Uses of Species Distribution Models:

Forecasting



Global Change

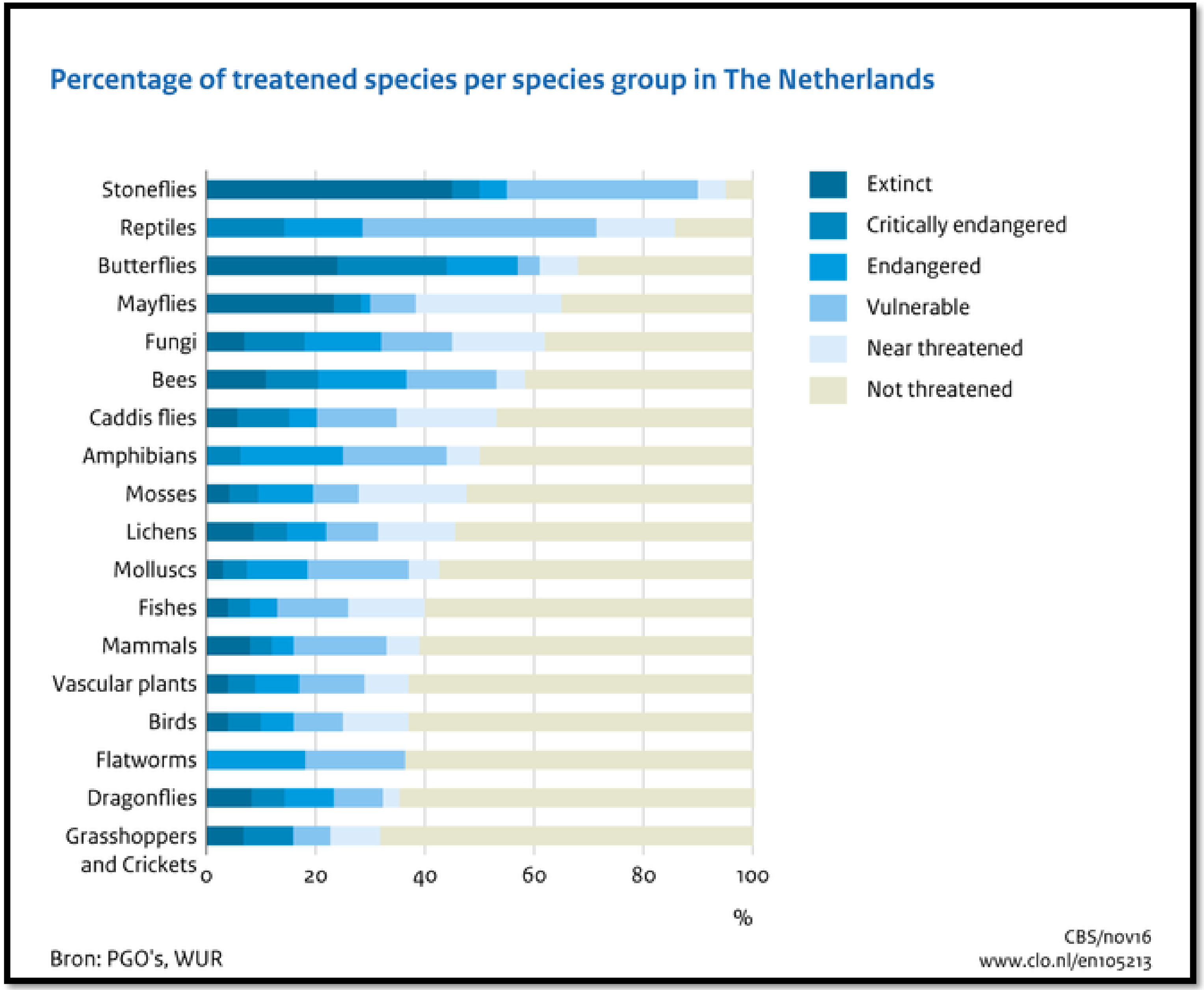


Invasive

Disease vectors

Endangered

Crop pollinators


Percentage of treatened species per species group in The Netherlands
Stoneflies
Reptiles
Butterflies
Mayflies
Fungi
Bees
Caddis flies
Amphibians
Mosses
Lichens
Molluscs
Fishes
Mammals
Vascular plants
Birds
Flatworms
Dragonflies
Grasshoppers and Crickets
0
20
40
60
80
100
%
Extinct
Critically endangered
Endangered
Vulnerable
Near threatened
Not threatened
Bron: PGO's, WUR
CBS/nov16
www.clo.nl/en105213

Climate models

- Model the interactions between the atmosphere, oceans, land surface, ice – and the sun.

- Attempt to reproduce the past and predict the future

2050

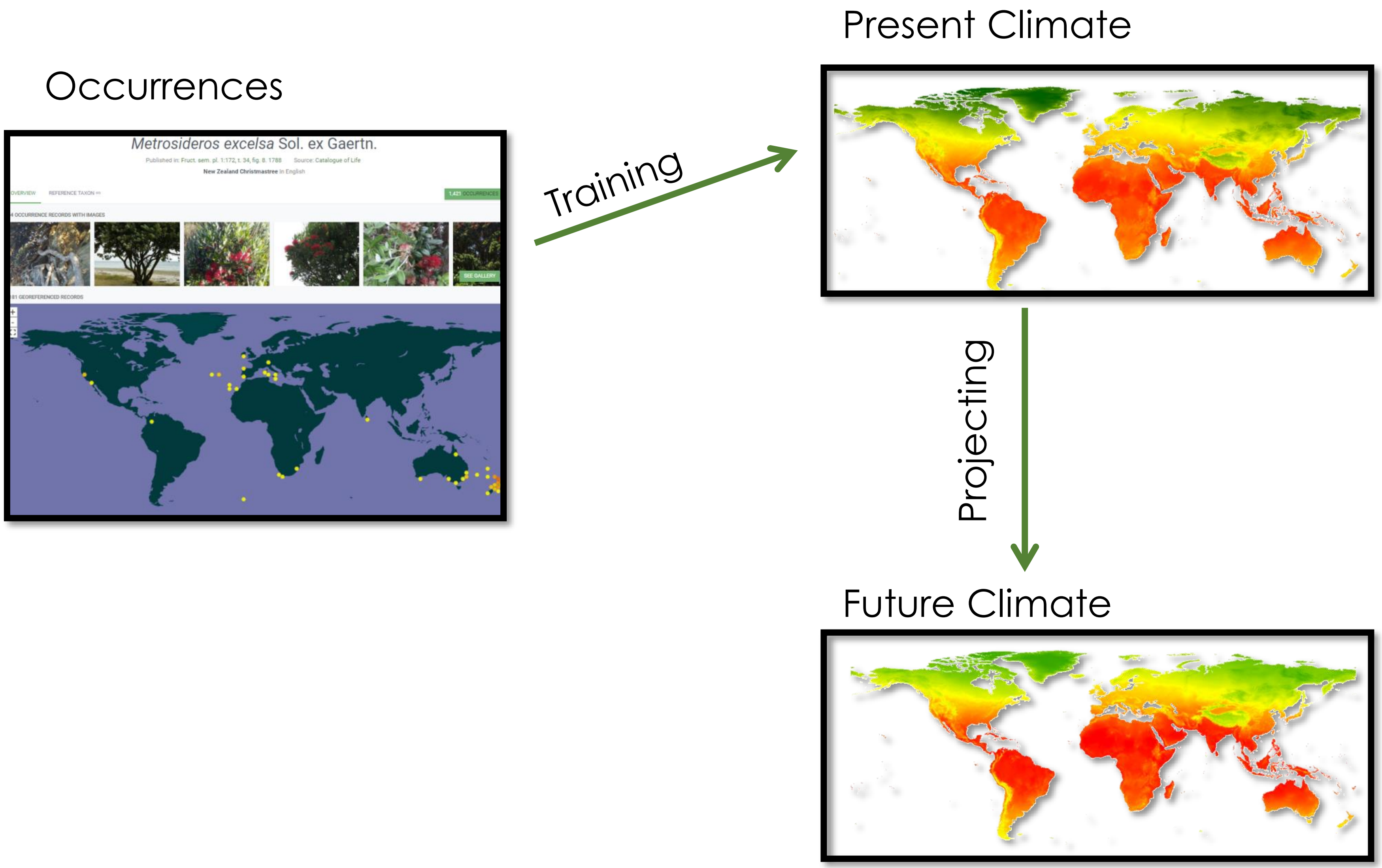
GCM	code	rcp26	rcp45	rcp60	rcp85
ACCESS1-0 (#)	AC		tn, tx, pr, bi		tn, tx, pr, bi
BCC-CSM1-1	BC	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
CCSM4	CC	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
CESM1-CAM5-1-FV2	CE		tn, tx, pr, bi		
CNRM-CM5 (#)	CN	tn, tx, pr, bi	tn, tx, pr, bi		tn, tx, pr, bi
GFDL-CM3	GF	tn, tx, pr, bi	tn, tx, pr, bi		tn, tx, pr, bi
GFDL-ESM2G	GD	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	
GISS-E2-R	GS	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
HadGEM2-AO	HD	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi
HadGEM2-CC	HG		tn, tx, pr, bi		tn, tx, pr, bi
HadGEM2-ES	HE	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi	tn, tx, pr, bi

Naturalis Biodiversity Center

Why so many?

- Portray interactions with respect to a multitude of processes operating on many different space and time scales.
- Different choices about which elements of the physics to emphasize
- Ensembles of models map out range of possible futures and help us understand their uncertainties
- Predicting socioeconomic development is another challenge

Bioclimatic envelope modeling



Naturalis Biodiversity Center

Temperature	
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

Moisture	
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

http://www.worldclim.org/bioclim

Storylines

<u>Scenarios</u>

Since we can't know future conditions for sure we need to develop multiple scenarios based on our most educated guesses and models.

Scenario	GRAS	BAMBU	SEDG
Climate projection	Corresponds to the IPCC SRES A1FI storyline and its assumptions	SRES A2 (the best fitting SRES scenario available at the time of calculation, though SRES A1B would have fitted better to past emission trajectories)	SRES B1 (SRES scenario with the lowest emissions, but not as low as 450 p.p.m. CO_2 stabilization assumed so the SEDG storyline differs significantly from B1)
EU Common Agricultural Policy	Dismantling payments for production (1st pillar) and for rural development and environment (2nd pillar)	Shift from 1st to 2nd pillar results in polarisation: intensification of high yielding locations, neglect of low yielding ones	Spatially explicit support structure to maintain (organic) agriculture throughout the landscape (only the 2nd pillar transfers remain)
EU funds	Phasing out, considered as subsidies	Focused on infrastructure development and growth in poor regions	Focused on local green development and opportunities, education and employment
Energy policy	Efficiency, some renewable energies based on cost calculations	Efficiency, aiming at 20% reduction of greenhouse gas emissions by 2020, 80% by 2080. Increase in nuclear and renewable energy	Aiming at 75% reduction of CO_2 emissions by 2050 through savings, changing consumption patterns and renewable energies
Transport policy	Increased efficiency due to market pressure, no policy to shift the modal split or even reduce transport	Technological improvements and change in share of different modes of mobility (walking, cycling, trains, cars, boats, planes) – modal split	Transport reduction priority, plus modal split change (through pricing and infrastructure supply), technical improvements
EU chemicals policy: REACH	Focus on innovation and competitiveness. REACH not rigorously implemented	REACH implemented	REACH plus; filling gaps, e.g. for nano-materials, endocrine disruptors, metals.
Trade policy	Strong support for World Trade Organization and free trade	Promoting free trade except in 'strategic areas'	Global sourcing reduced for cost reasons; phytosanitarian controls

IPCC, Intergovernmental Panel on Climate Change; SRES, Special Report on Emissions Scenarios; REACH, registration, evaluation, authorization and restriction of chemical substances; SEDG, sustainable European development goal.

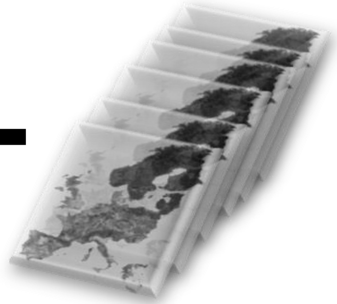
Spangenberg et al., 2012 - Scenarios for investigating risks to biodiversity - Global Ecology and Biogeography

Naturalis Biodiversity Center

SDM - Training

Bumblebee Collection Europe (48 Species)

\+

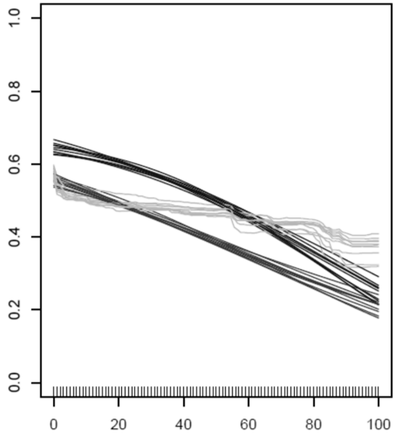


European LU + Climate

\+

Algorithms:
- GLM
- MAXENT
- GBM

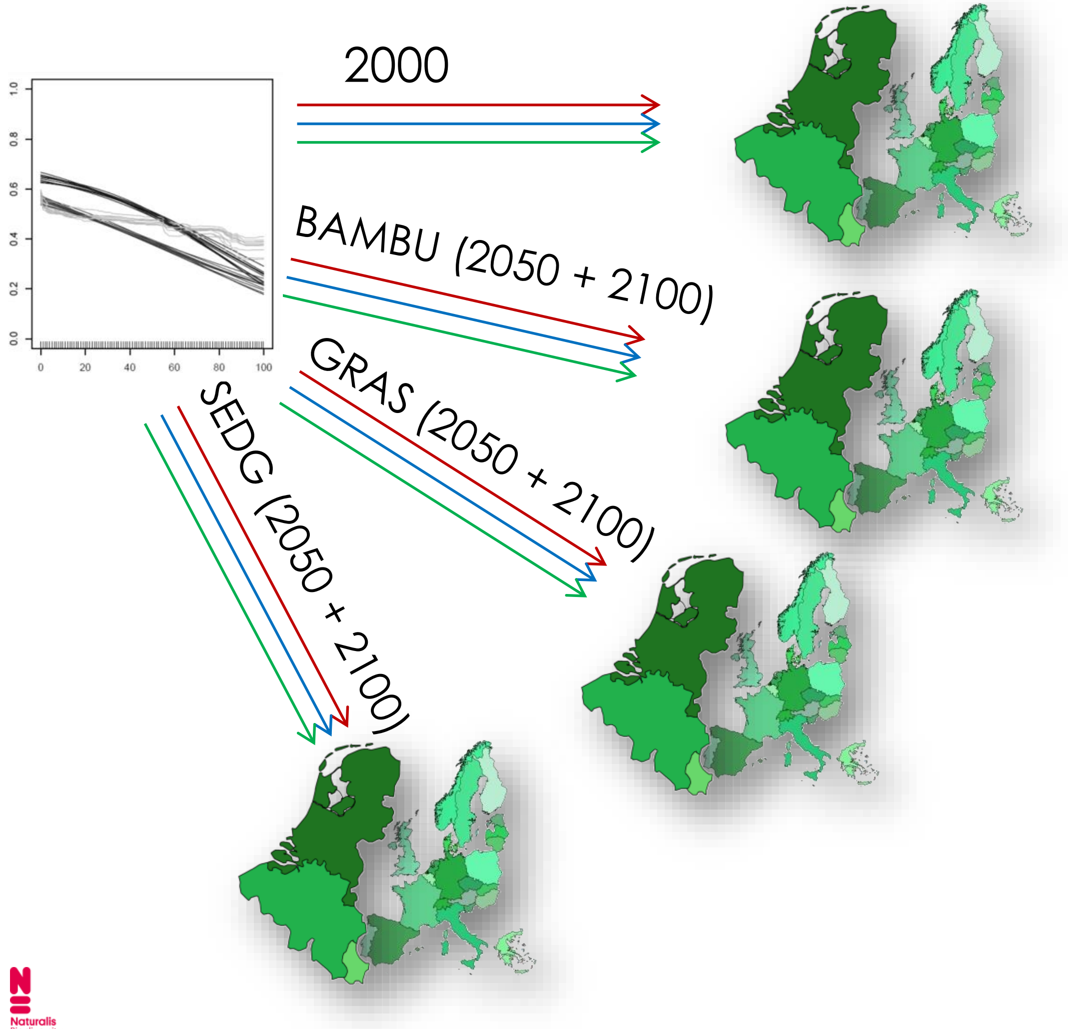
=



ENSEMBLE

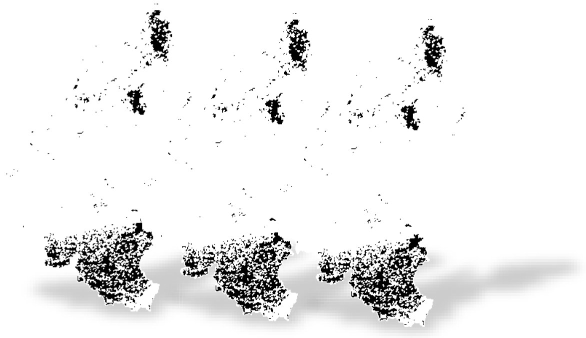
SDM - Projecting

Model Projections



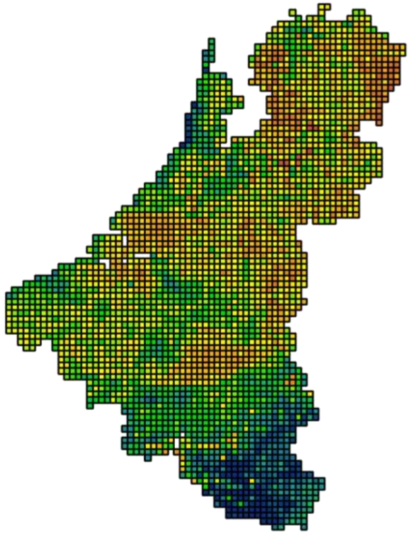
Output

- Variable Contribution
- Model Performance
- Distribution Maps
    - Binary
    - Habitat Suitability
- Results
    - Range Shifts
    - Range Change



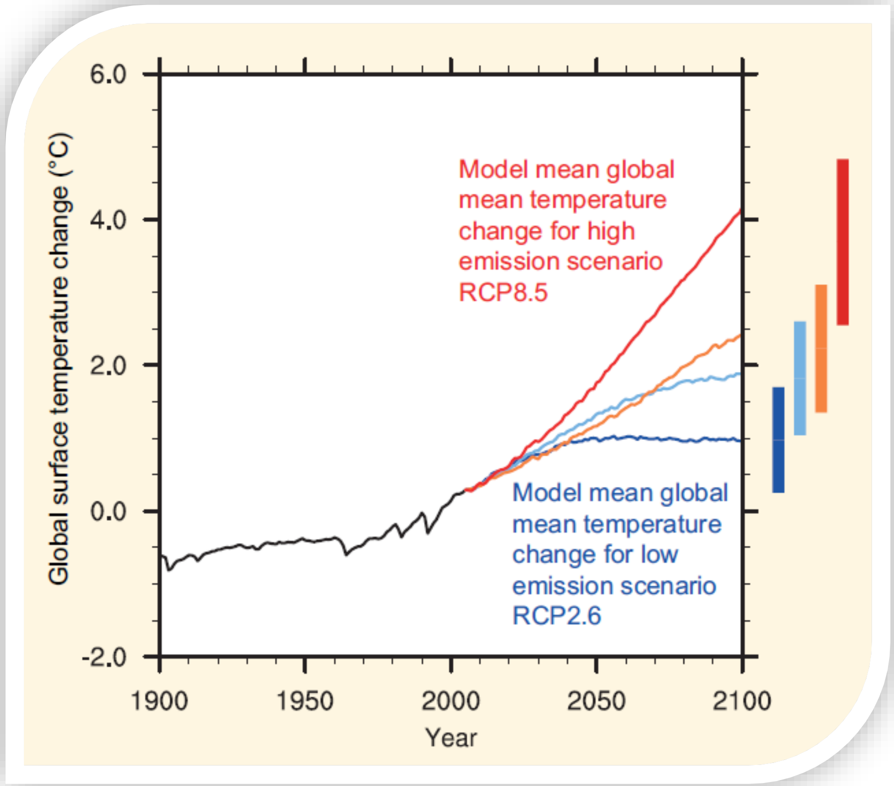
Measurements

- Range change
- Extinction
- Shift in Suitable Habitat
- Latitudinal and Longitudinal Shifts
- Species richness change
- Community composition
- Conservation status (IUCN)



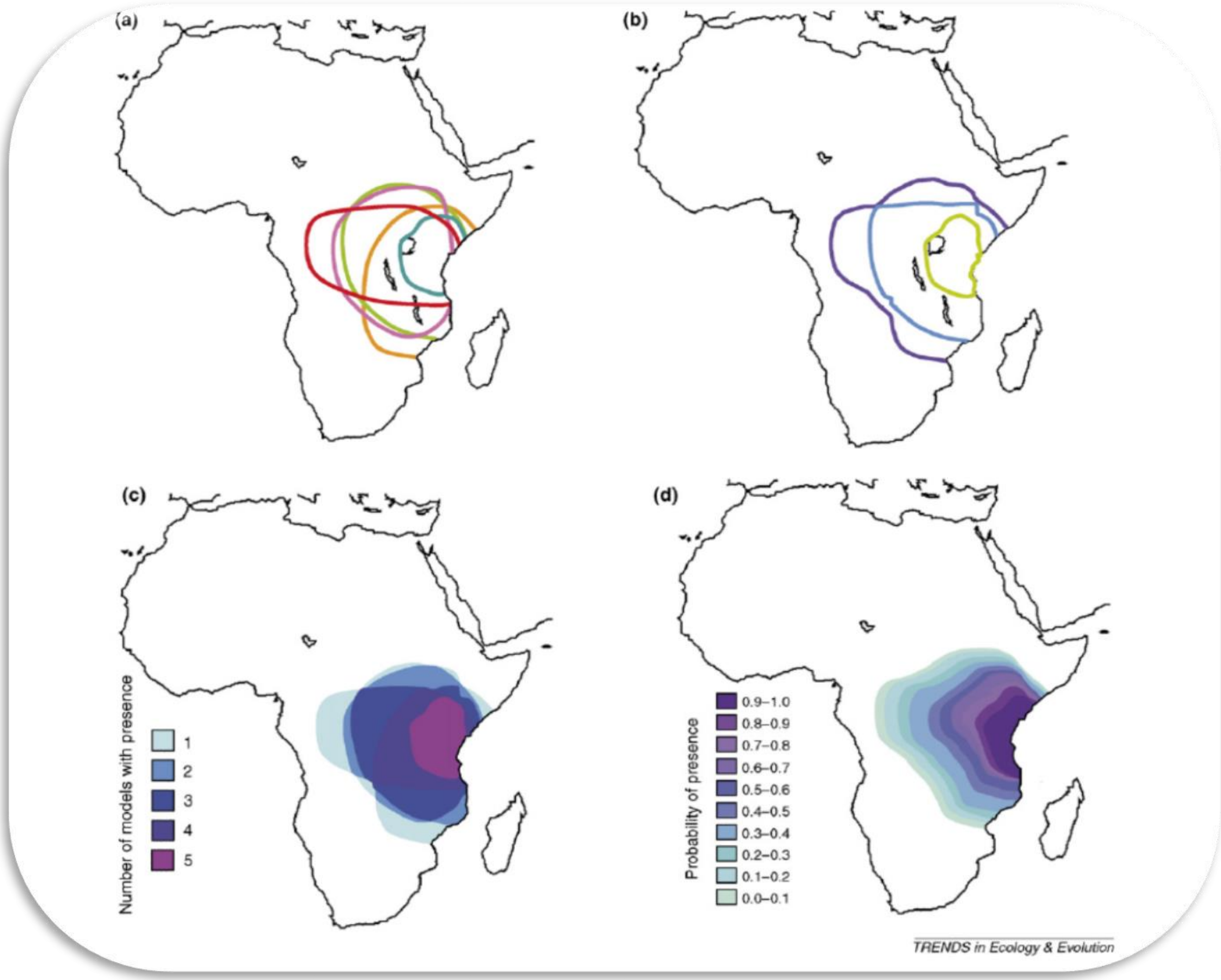
Uncertainty

- Collection records
  - Sample size
  - Range size
- Covariates
  - Different climate models
  - Covariate selection
- Model
  - Algorithm
  - Parameter selection



IPCC, 2013

Uncertainty



Araujo *et al.*, 2006 - Reducing uncertainty in projections of extinction risk from climate change - TREE

Uncertainty in SDMs

- Do not model actual species distributions
- Do not account for all abiotic and biotic variables (interspecific relationships, physical barriers, etc.)
- Poor sampling in environmental space can inhibit proper modeling
- Source-sink problem (individuals can be found in unsuitable habitats)
- Type 4 Errors
  - Can predict distributions that are neither part of the actual or potential distribution
- Less certainty when extrapolating data
  - Predict distribution for environments with variables outside the range of what was put into the model

“We think that this is the most extreme version and it’s not based on facts. It’s not data driven. We’d like to see something that is more data driven. It’s based on modeling, which is extremely hard to do when you’re talking about the climate.”

“all models are wrong,
but some are useful”

- George E. P. Box

Examples

Leon Marshall

How does change in major land use classes affect the projected distribution of European Bumblebees predicted under climate change?

Naturalis Biodiversity Center

Variable Selection

- Arable
- Forest
- Grassland
- Permanent crops
- Urban

- Growing degree days
- Water balance
- Temperature range
- μ rainfall wettest month
- μ diurnal range

1) Dynamic Climate Only

2) Static Land Use and Dynamic Climate

3) Dynamic Land Use and Dynamic Climate

Naturalis Biodiversity Center

SDM - Training

Bumblebee Collection Europe (48 Species)

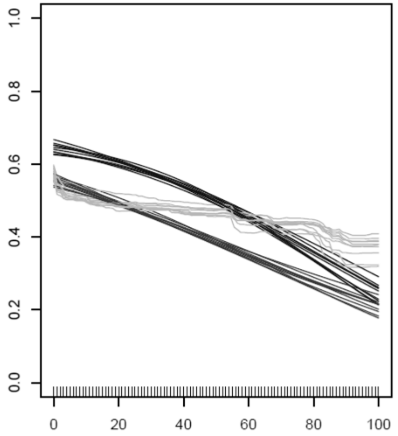
+

European LU + Climate

+

Algorithms:
- GLM
- MAXENT
- GBM

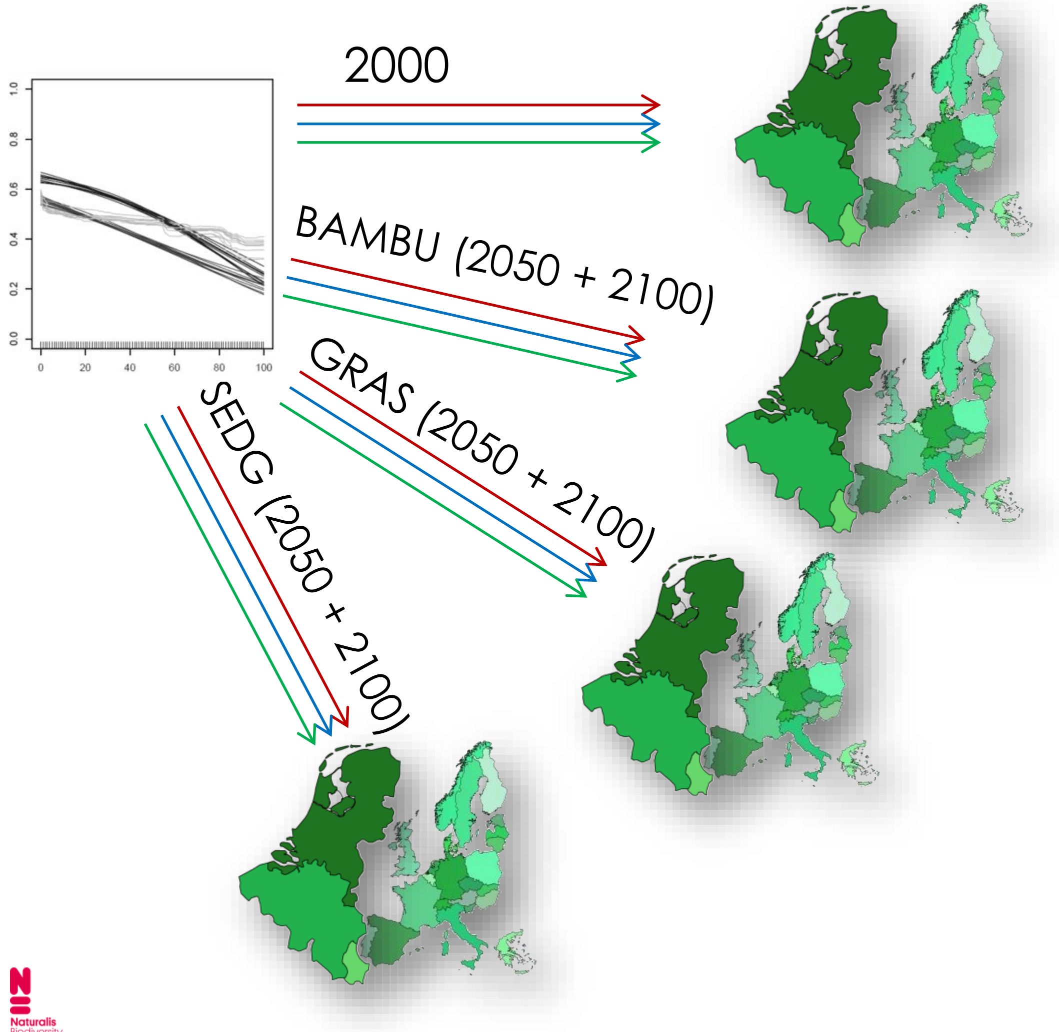
=



ENSEMBLE

SDM - Projecting

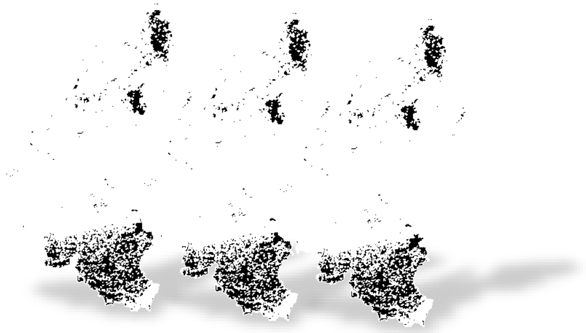
Model Projections



=

Output

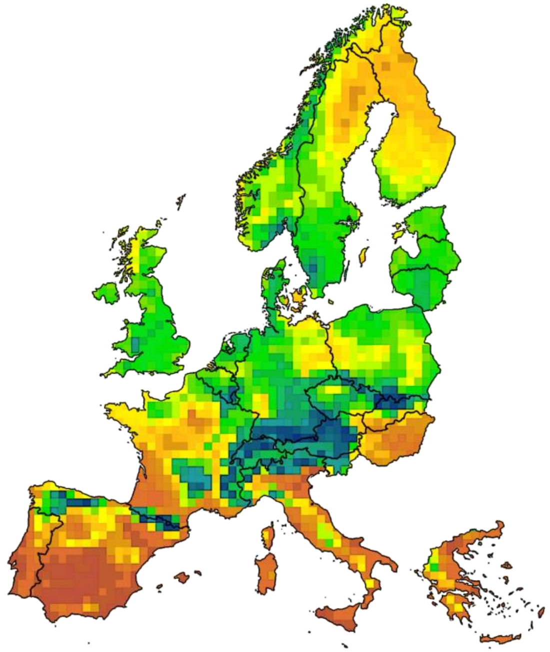
- Variable Contribution
- Model Performance
- Distribution Maps
  - Binary
  - Habitat Suitability
- Results
  - Range Shifts
  - Range Change



Naturalis Biodiversity Center

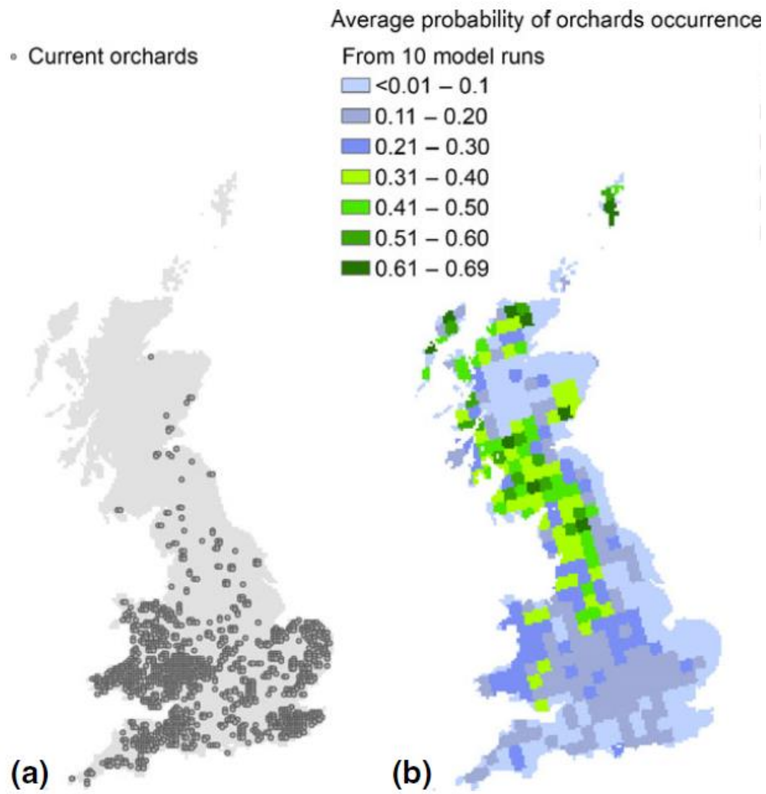
Conclusions

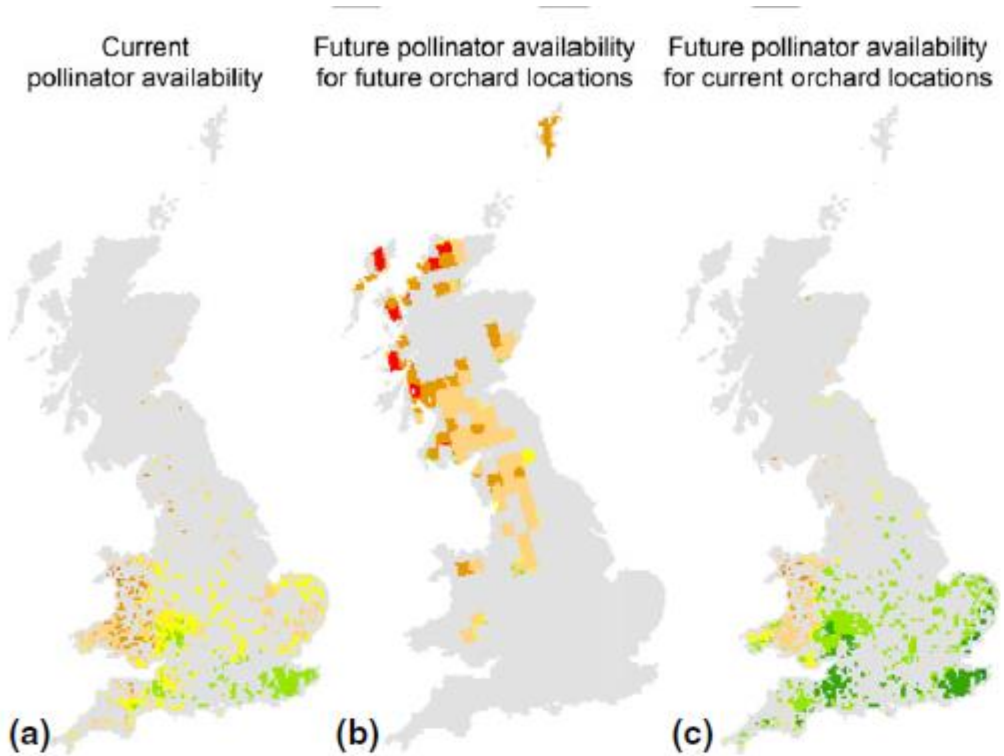
- Range loss
  - misrepresented by climate only models?
  - under-predicted with static land use models?
- Large variability in species responses
- Dynamic land use models capture different habitat suitability envelopes. Not simply level up or down.



Naturalis Biodiversity Center

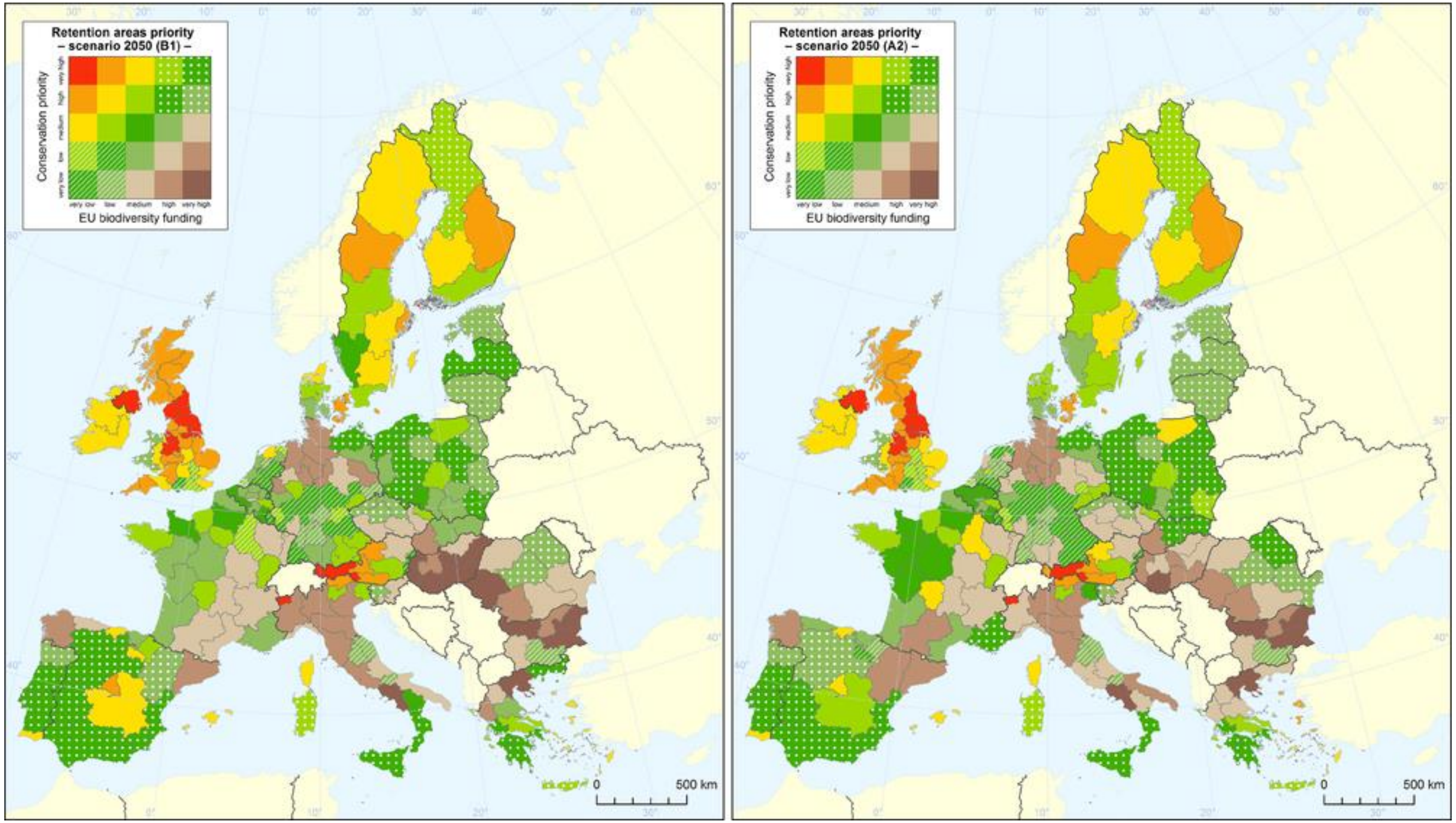
Agriculture – crop management





Polce *et al.*, 2014 - Climate-driven spatial mismatches between British orchards and their pollinators: Increased risks of pollination deficits - Global Change Biology

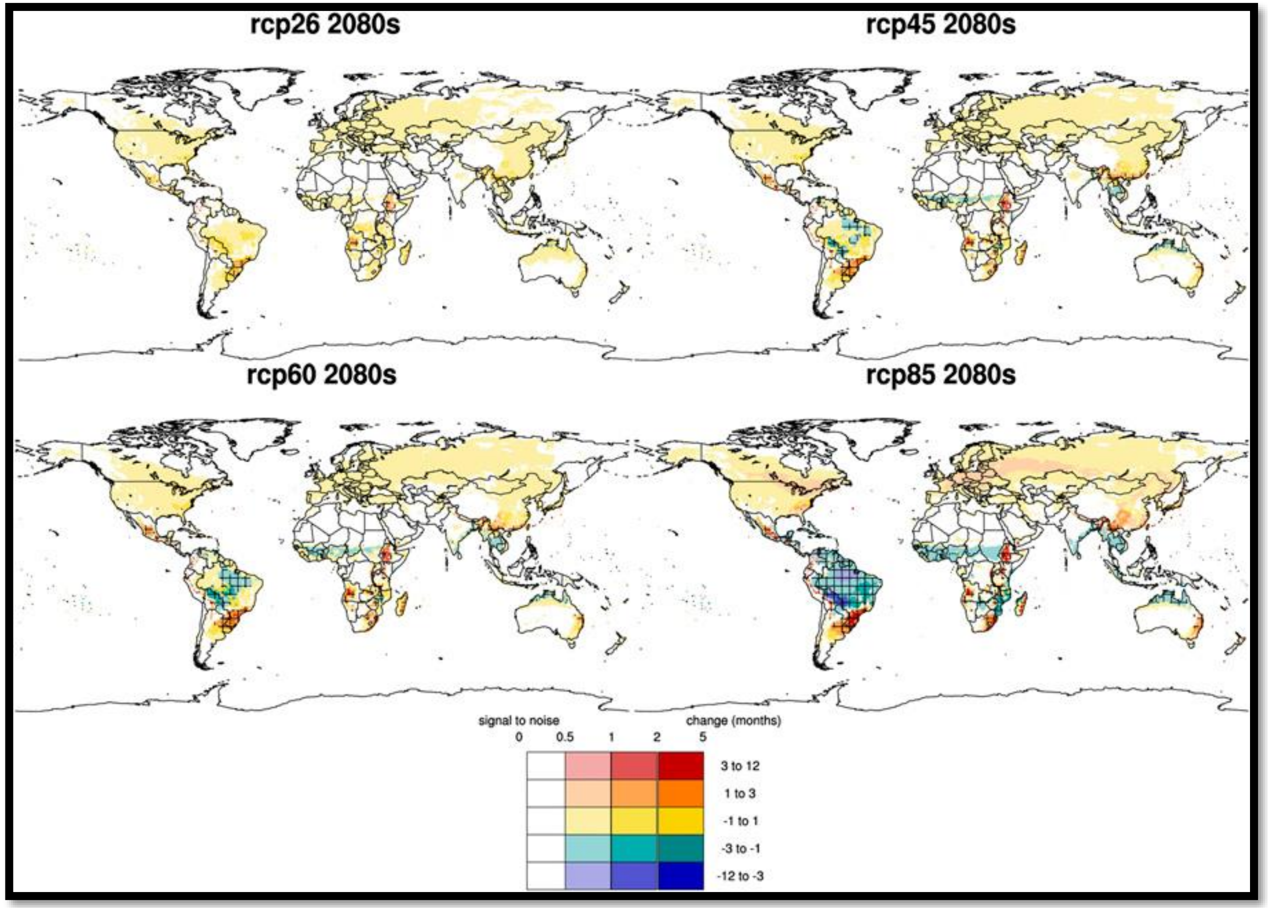
Conservation Funding



Lung *et al.*, 2014 - Biodiversity Funds and Conservation Needs in the EU Under Climate Change – Conservation Letters

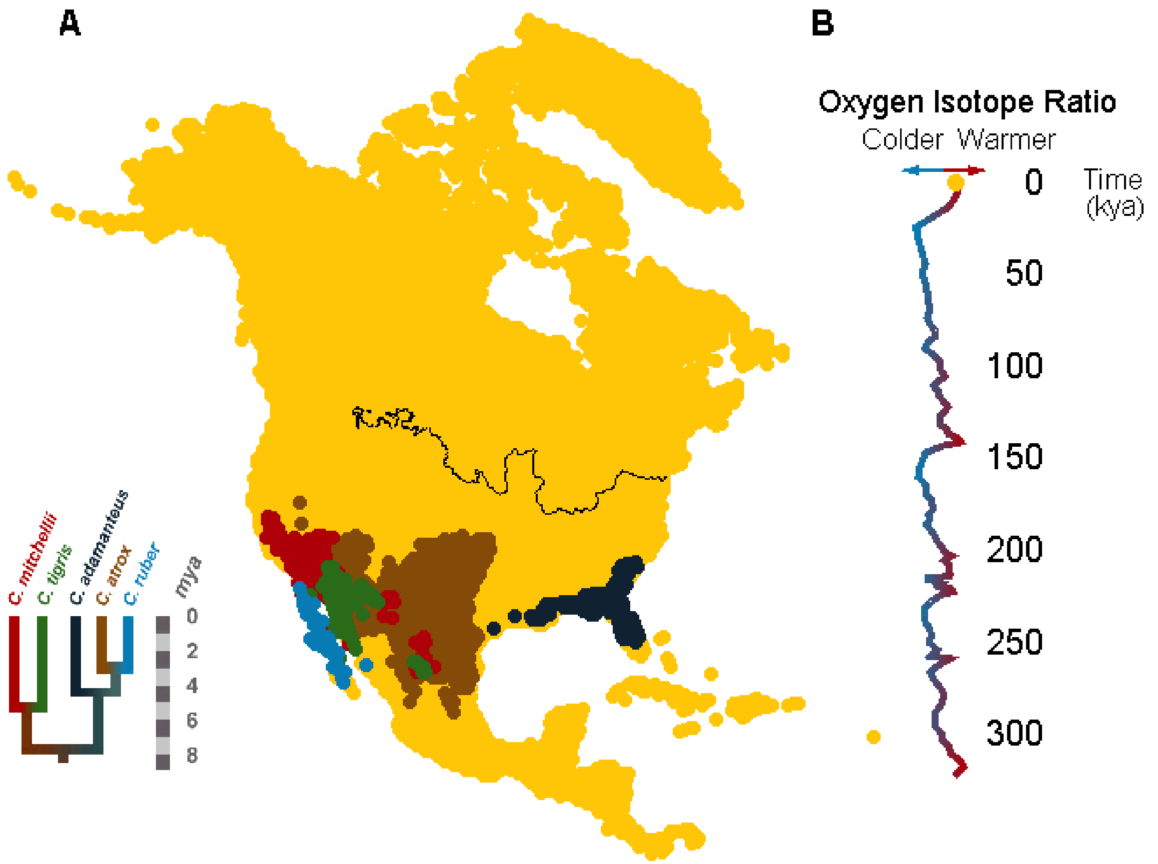
Human health

Disease (E.g. Malaria)



Caminade *et al.*, 2014 - Impact of climate change on global malaria distribution- PNAS

Paleophylogeographic models for five rattlesnake species



Lawing, A. M., & Polly, P. D. (2010). Geometric morphometrics: Recent applications to the study of evolution and development. *Journal of Zoology*

“2011 the worst single year drought in Texas history”

Jun/Jul/Aug
Departures from average
1895-2011

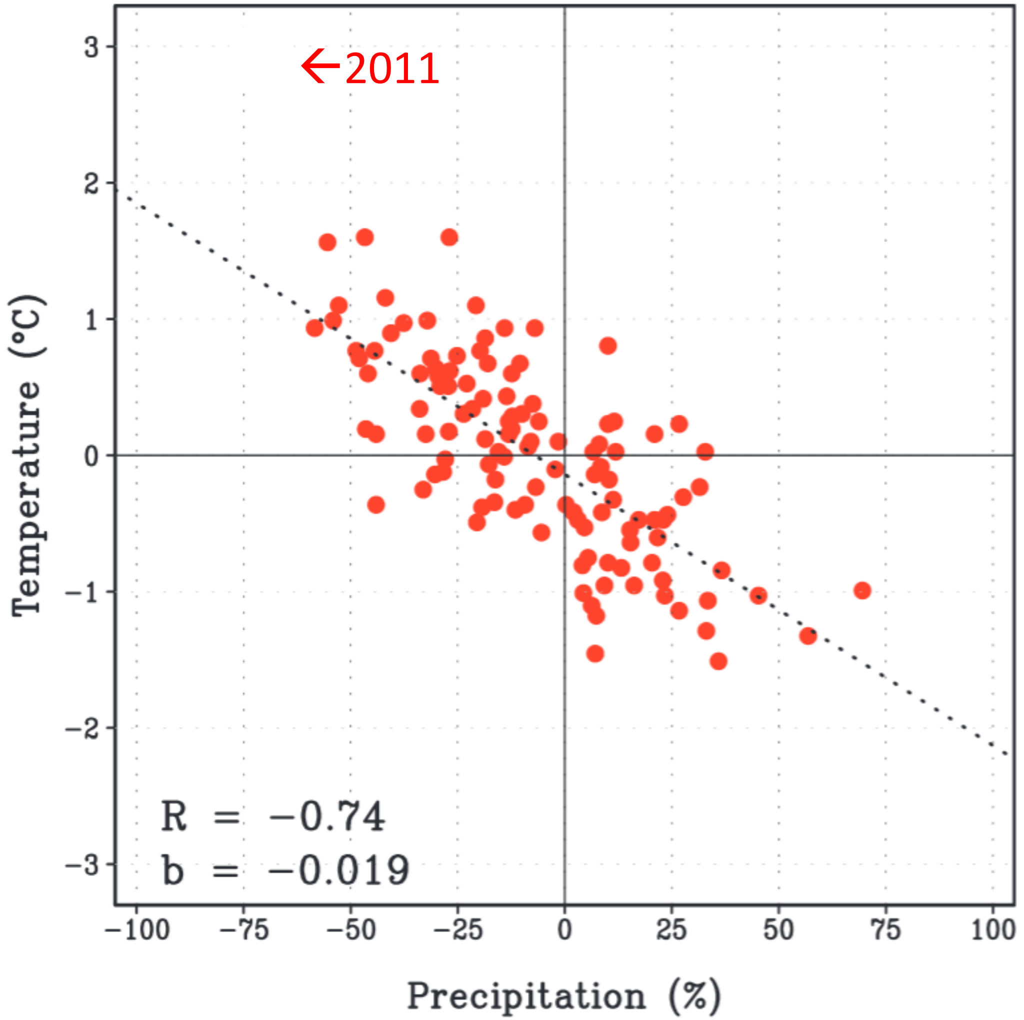


FIG. 5. The historical relationship between JJA Texas averaged rainfall departures (% of climatology) and surface temperature departures (°C). Each dot corresponds to a summer during 1895–2010, and the 2011 value is indicated by the blue wagon wheel. Inset values are for the correlation *R* and the slope of the linear fit expressed as degree Celsius per percent precipitation departure.

Source: Hoerling, M. et al., 2013: Anatomy of an Extreme Event. *J. Climate*, **26**, 2811–2832.

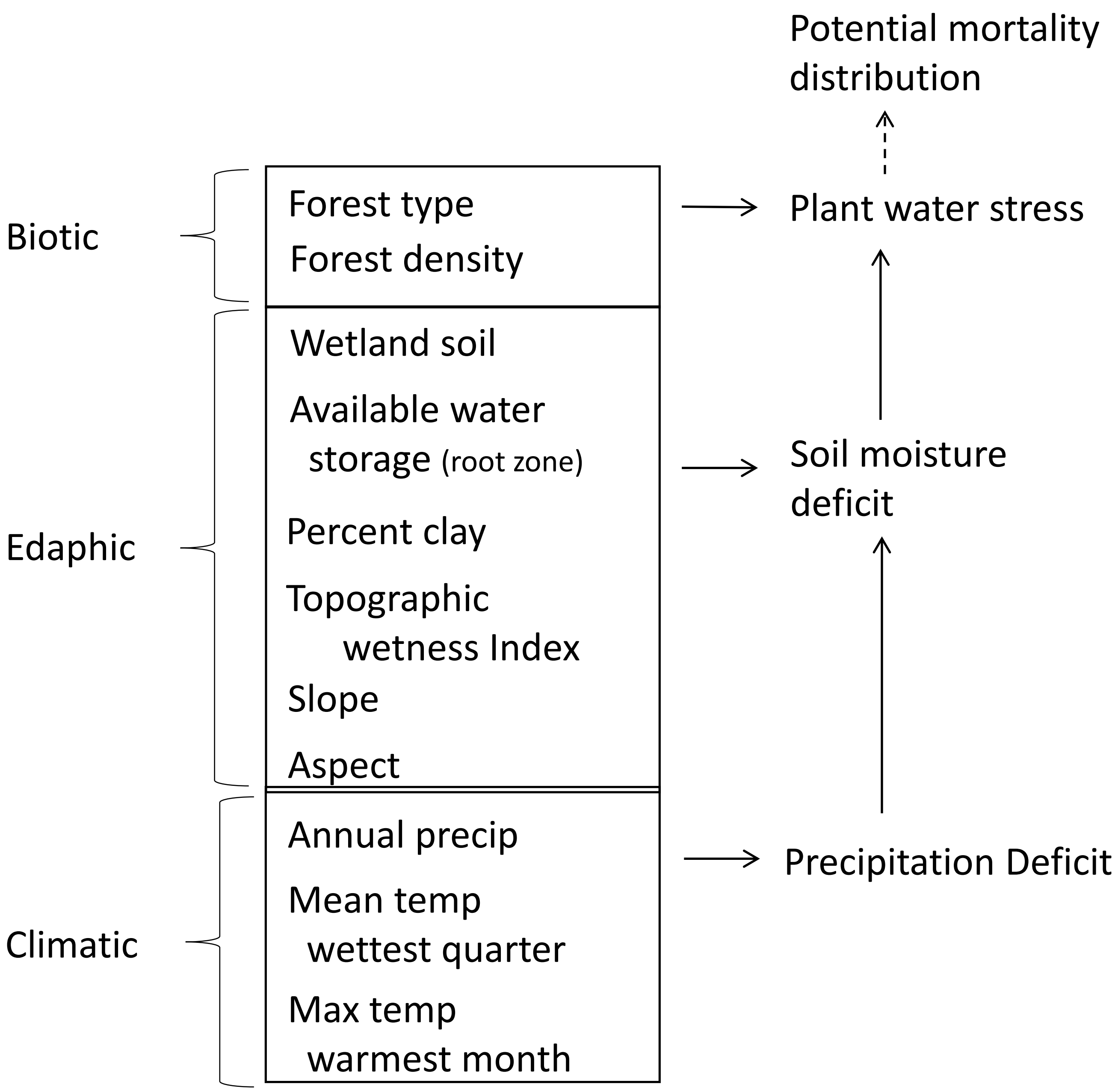
Central Region
April 2012

What drives drought-related tree mortality?

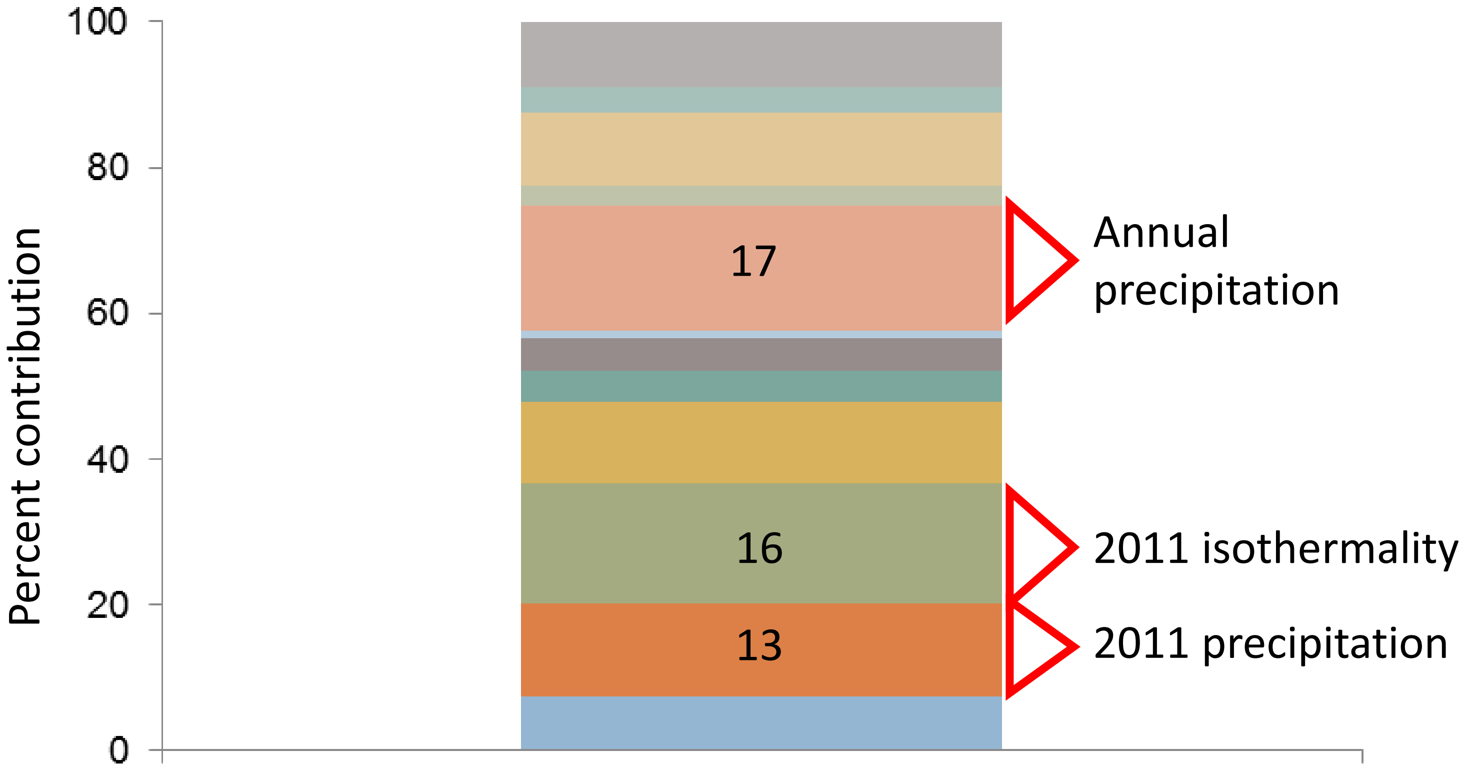
- What is the contribution of long-term climate vs. time-of-drought conditions?

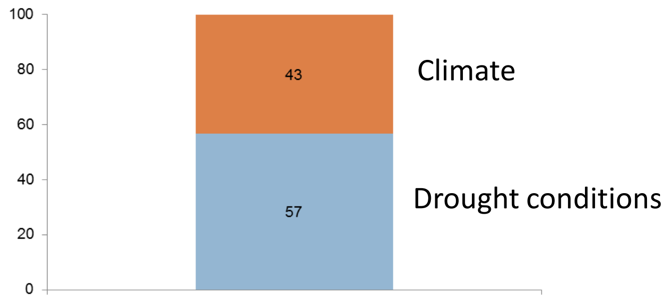
- What is the relative contribution of climate, biotic, and edaphic variables?
  - How do contributions differ with scale?
  - How do contributions differ by region?

Texas Forest Service documents trees that died in the 2011 drought.


Potential mortality distribution
Biotic
Forest type
Forest density
Plant water stress
Edaphic
Wetland soil
Available water storage (root zone)
Percent clay
Topographic wetness Index
Slope
Aspect
Soil moisture deficit
Climatic
Annual precip
Mean temp wettest quarter
Max temp warmest month
Precipitation Deficit

Climate & drought conditions

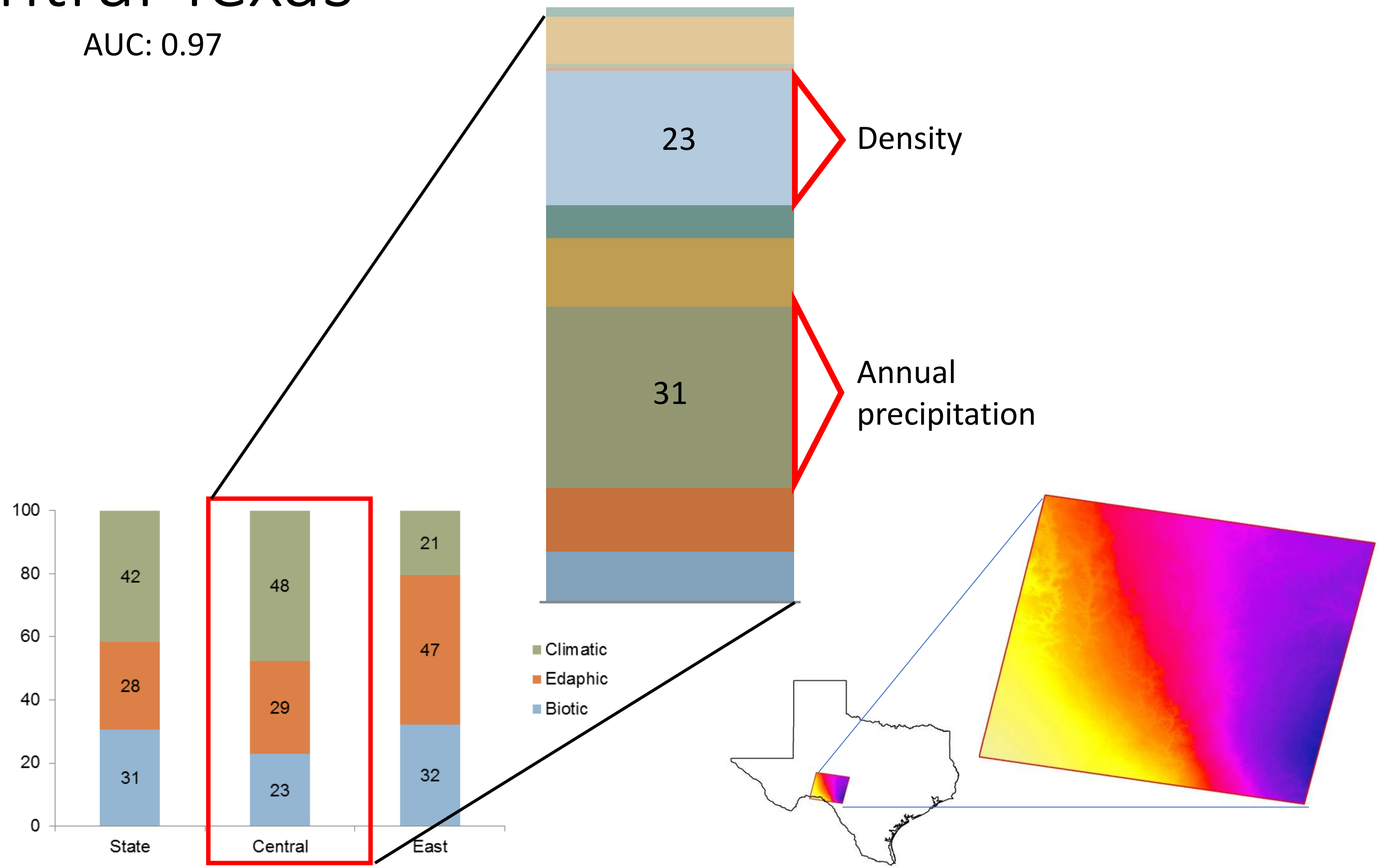




AUC: 0.84

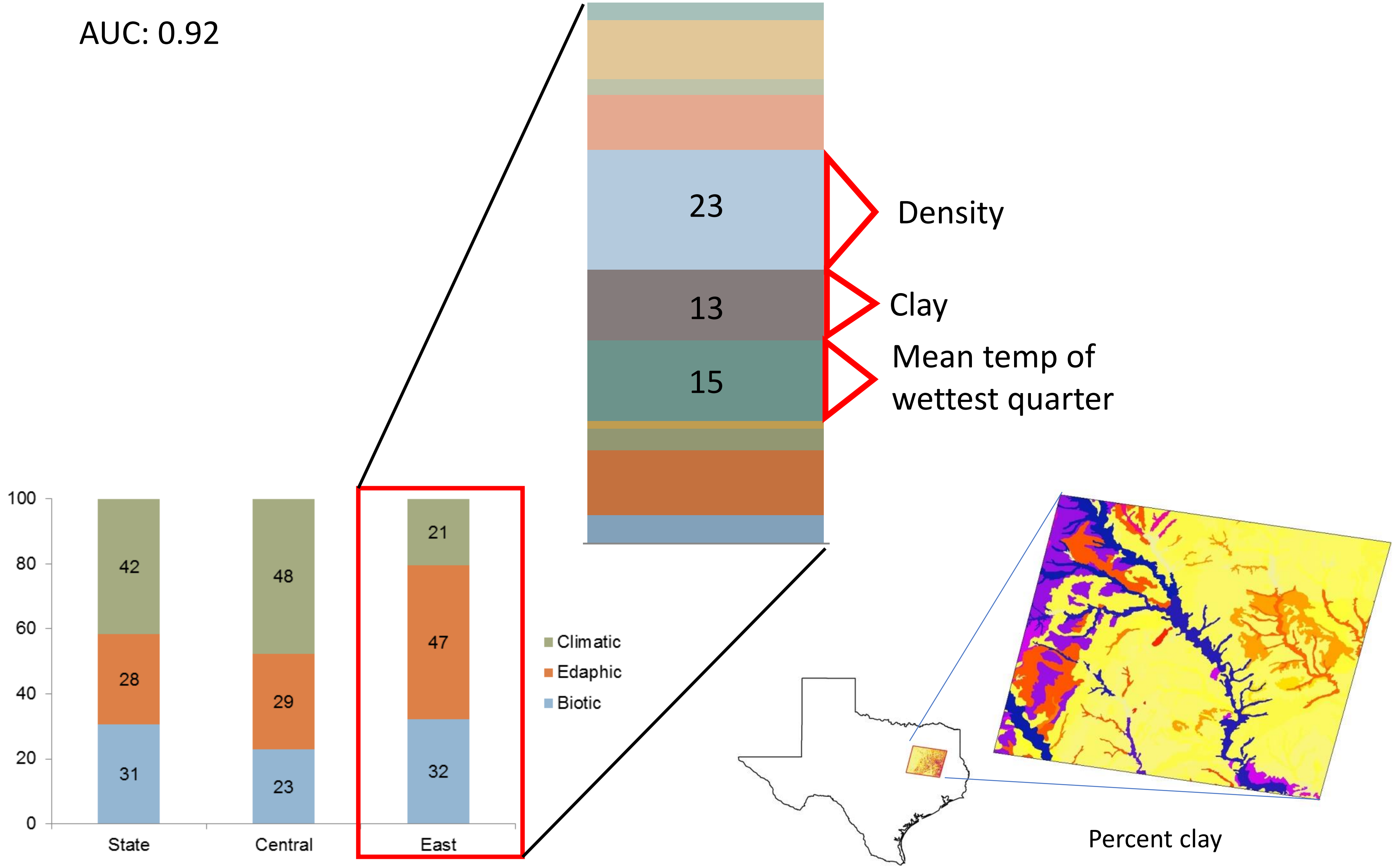
Central Texas

AUC: 0.97



East Texas

AUC: 0.92



Percent clay

What are Invasive Alien Species?

- Non-native species which:
  - Reproduces too much and too fast
  - Harms the ecosystem
  - Causes economic damage
  - Negative impact on human health

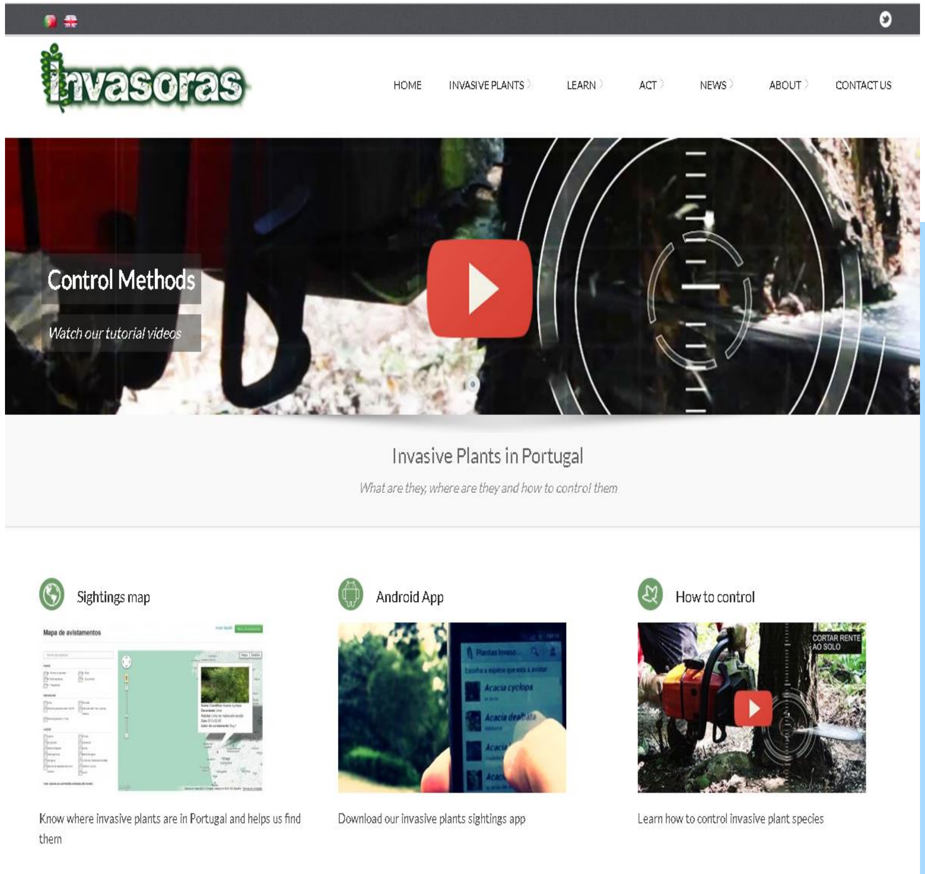
- One of the most important threats to biodiversity and ecosystem services

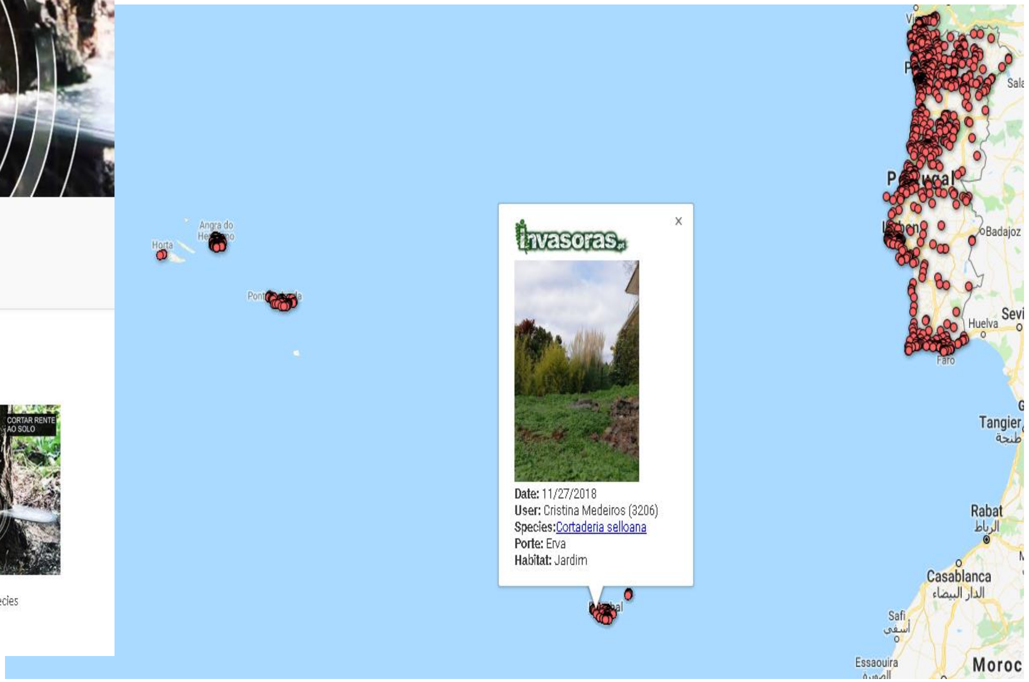
- Threat increased by globalization and climate change

Flowering Acacia (IAS) and pine – no space for other species

Louisiana crawfish – both in PT and NL

Humanly impossible for small team to monitor their distribution/spread



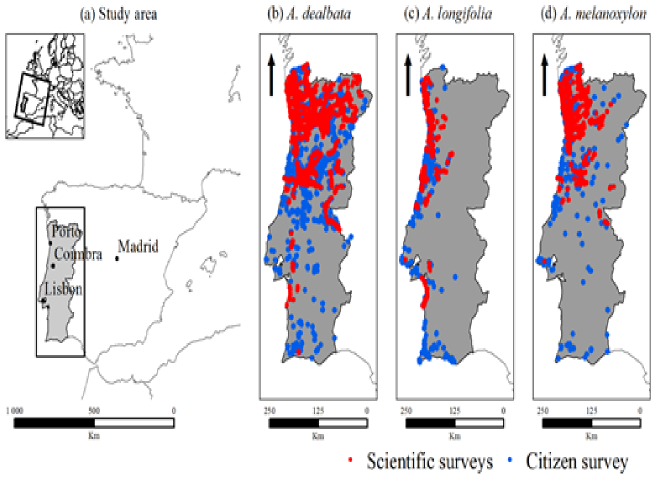


Online platform to increase awareness on the topic of Invasive Alien Species - http://invasoras.pt/en/

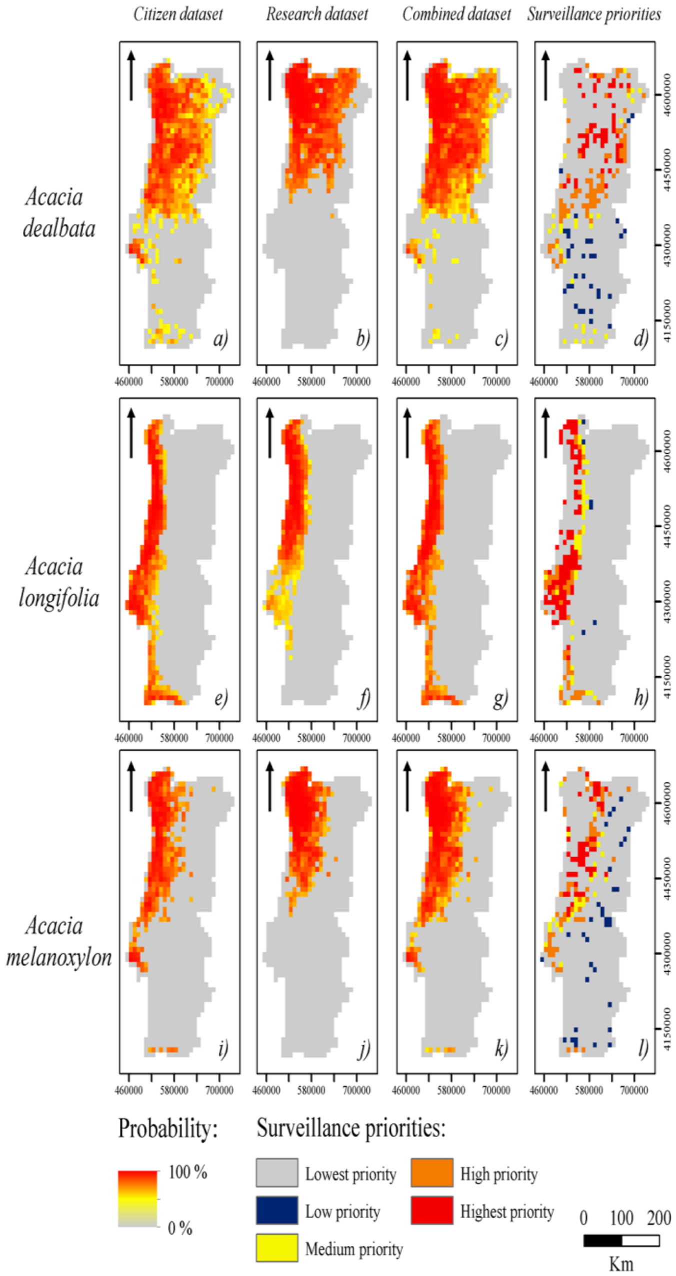
Allows online submission of sightings or using an Android app – data available through a google fusion table

- > 16 000 sightings recorded
- > Each occurrence validated by the team (Very high accuracy)

Can citizen science help improving the surveillance efforts of IAP?



Ensemble modelling (biomod2)



- Citizen data models showed much higher geographical dispersion of invasive alien plants
  - But also spatial biases (towards roads, cities, population centers)
  - Combining citizen and scientist data improved the models (scientists sampling strategy is biased towards the known ecology characteristics of the species)
- The disagreement between the final models is indicative of data deficiency – thus indentifying áreas with need for improved surveillance

Some published uses of SDMs in conservation biology

From Pearson 2008

Type of use	Example reference(s)
Guiding field surveys to find populations of known species	Bourg et al. 2005, Guisan et al. 2006
Guiding field surveys to accelerate the discovery of unknown species	Raxworthy et al. 2003
Projecting potential impacts of climate change	Iverson and Prasad 1998, Berry et al. 2002, Hannah et al. 2005; for review see Pearson and Dawson 2003
Predicting species' invasion	Higgins et al. 1999, Thuiller et al. 2005; for review see Peterson 2003
Exploring speciation mechanisms	Kozak and Wiens 2006, Graham et al. 2004b
Supporting conservation prioritization and reserve selection	Araújo and Williams 2000, Ferrier et al. 2002, Leathwick et al. 2005
Species delimitation	Raxworthy et al. 2007
Assessing the impacts of land cover change on species' distributions	Pearson et al. 2004
Testing ecological theory	Graham et al. 2006, Anderson et al. 2002b
Comparing paleodistributions and phylogeography	Hugall et al. 2002
Guiding reintroduction of endangered species	Pearce and Lindenmayer 1998
Assessing disease risk	Peterson et al. 2006, 2007