european social fund in the czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION, YOUTH AND SPORTS

OP Education for Competitivness

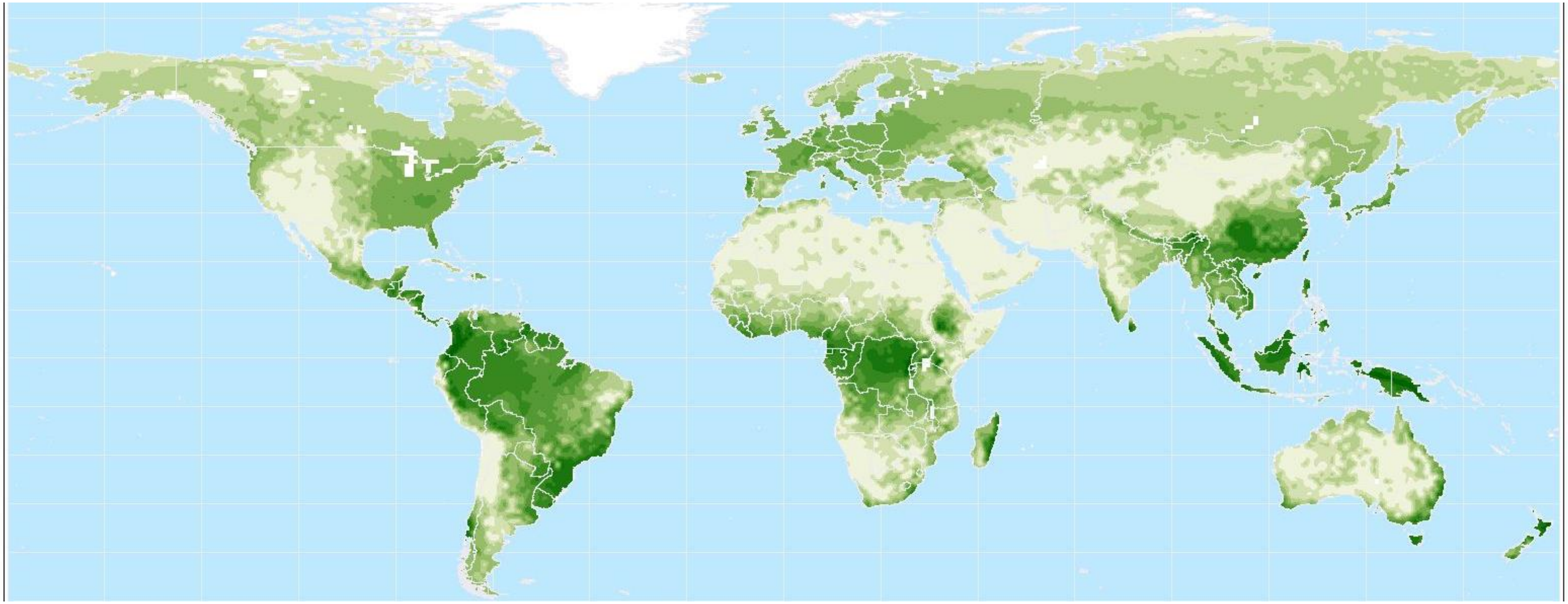
INVESTMENTS IN EDUCATION DEVELOPMENT

Mapping and modeling species distributions

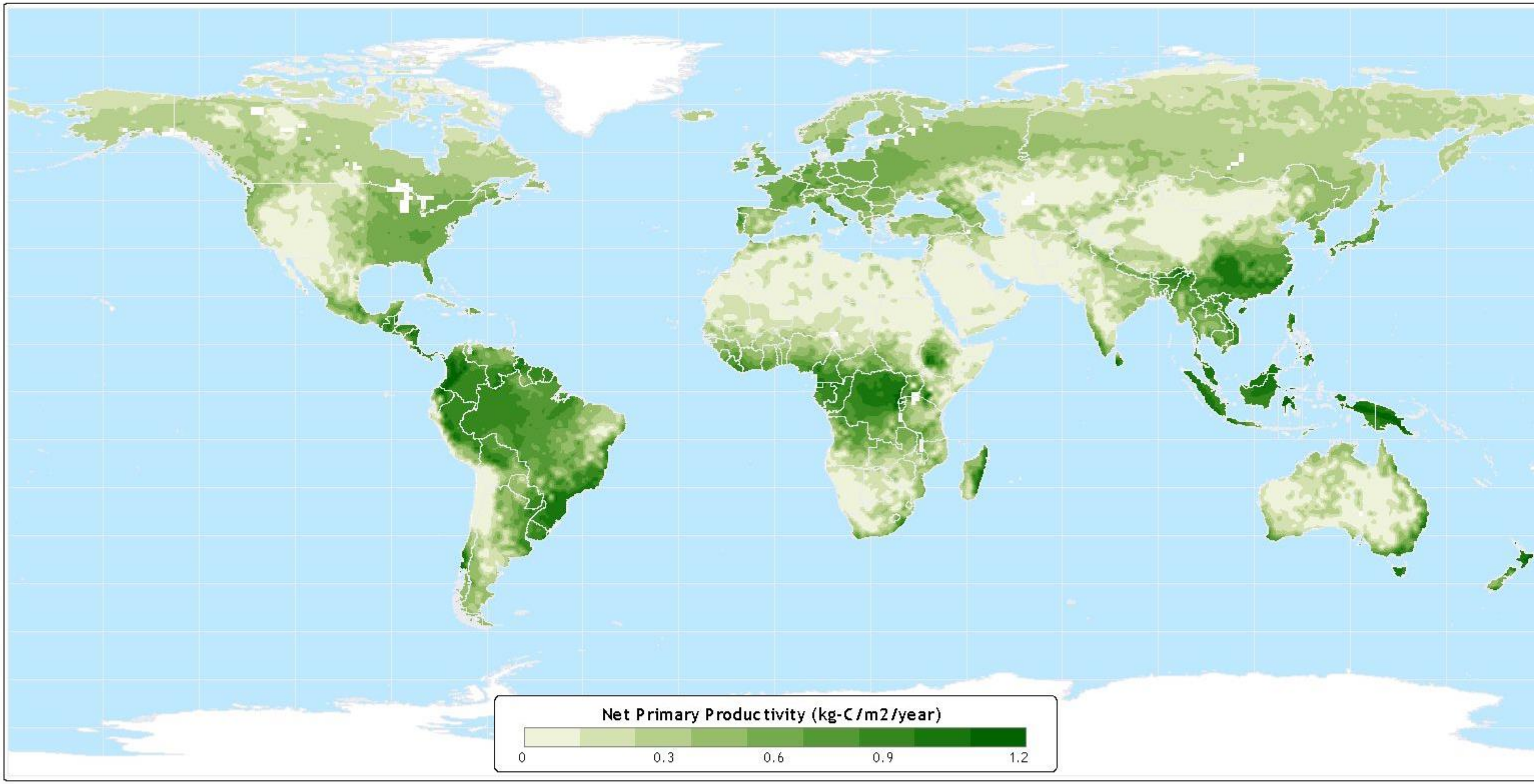
Department of Botany and Zoology, Masaryk University

Bi9661 Selected issues in Ecology, Autumn 2013

Borja Jiménez-Alfaro, PhD



Net Primary Productivity



Data taken from: IBIS Simulation
(Kucharik, et al. 2000)
(Foley, et al. 1996)

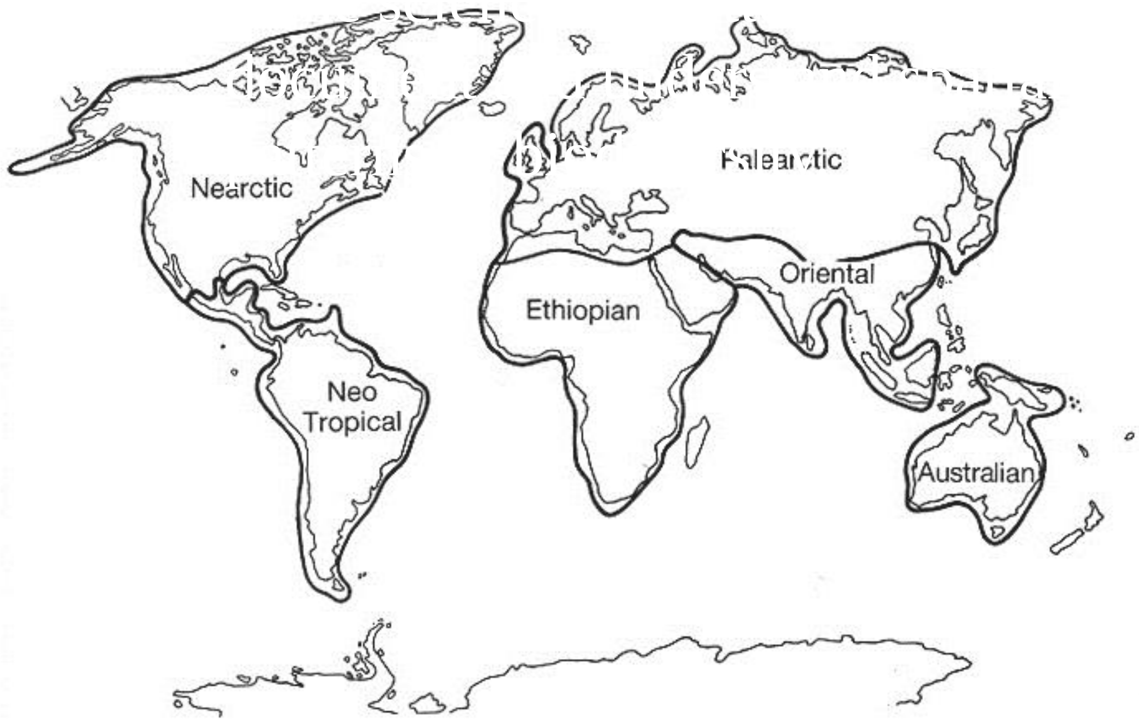
Atlas of the Biosphere
Center for Sustainability and the Global Environment
University of Wisconsin - Madison

Introduction:

ASSESSING SPECIES DISTRIBUTIONS

Biogeography

“attempts to explain why species and higher taxa are distributed as they are, and why the diversity and taxonomic composition of the biota vary from one region to another”



Philip Sclater
(1829-1913)

Historical biogeography

Reconstruct the origins, dispersal, and extinctions of taxa
Primarily focused on evolution, dispersal and vicariance

Ecological biogeography

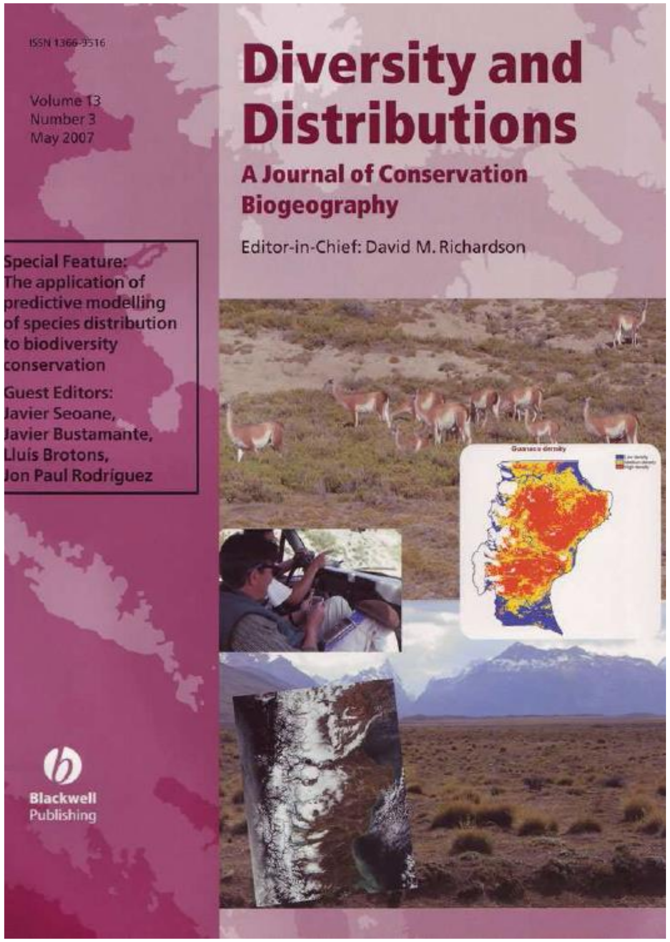
Primarily focused on present distributions, species responses to biotic environment and interactions with other organisms

Paleoecology

Combines historical and ecological biogeography, investigating the relationships between communities (abundance, distribution, and diversity of species) and abiotic conditions

Conservation biogeography

Work on the protection and restoration of natural environments



Special feature:
The application of predictive modeling of species distribution to biodiversity conservation

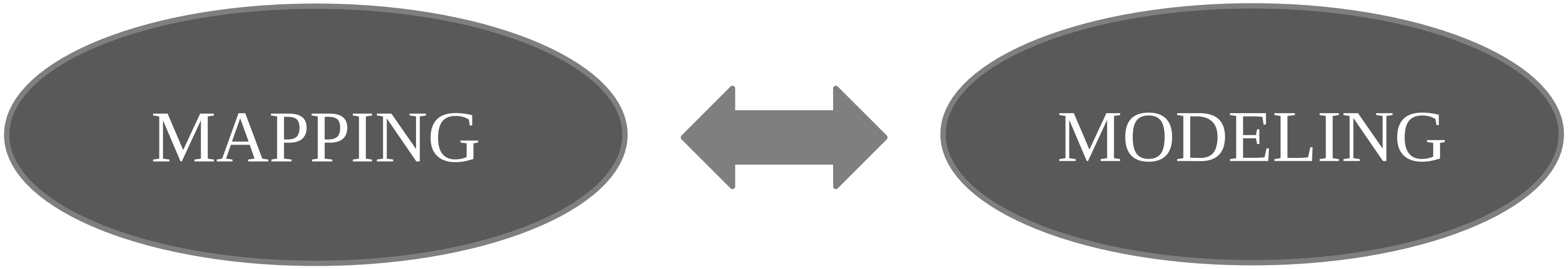
Special issue
Diversity and distributions
13 (3) 2006

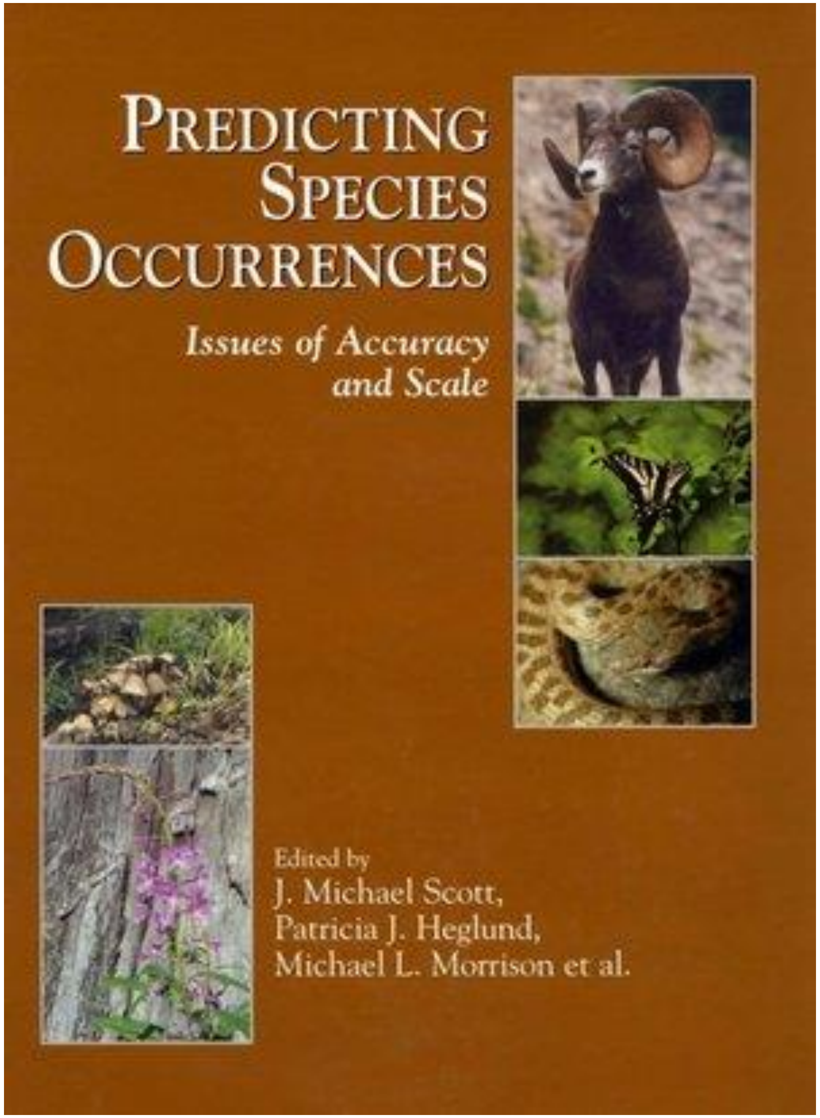
Recent tools in biogeography

Computational power (computers)

Geographic Information Systems

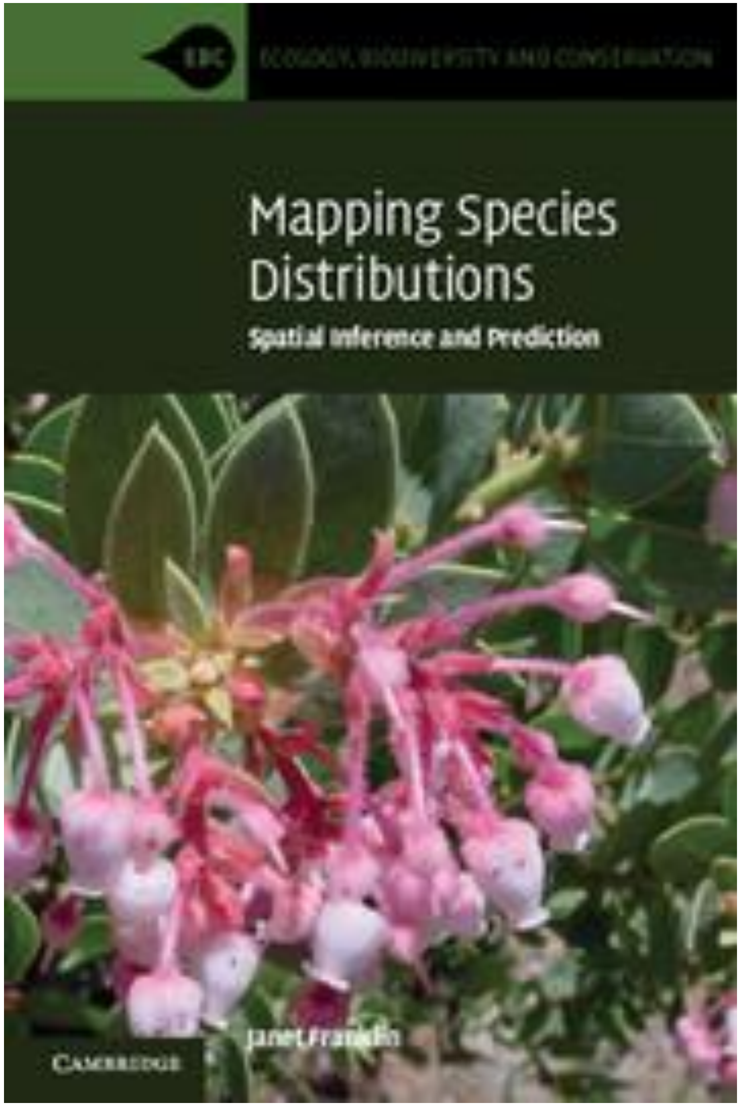
Geostatistics




PREDICTING SPECIES OCCURRENCES
Issues of Accuracy and Scale
Edited by
J. Michael Scott,
Patricia J. Heglund,
Michael L. Morrison et al.

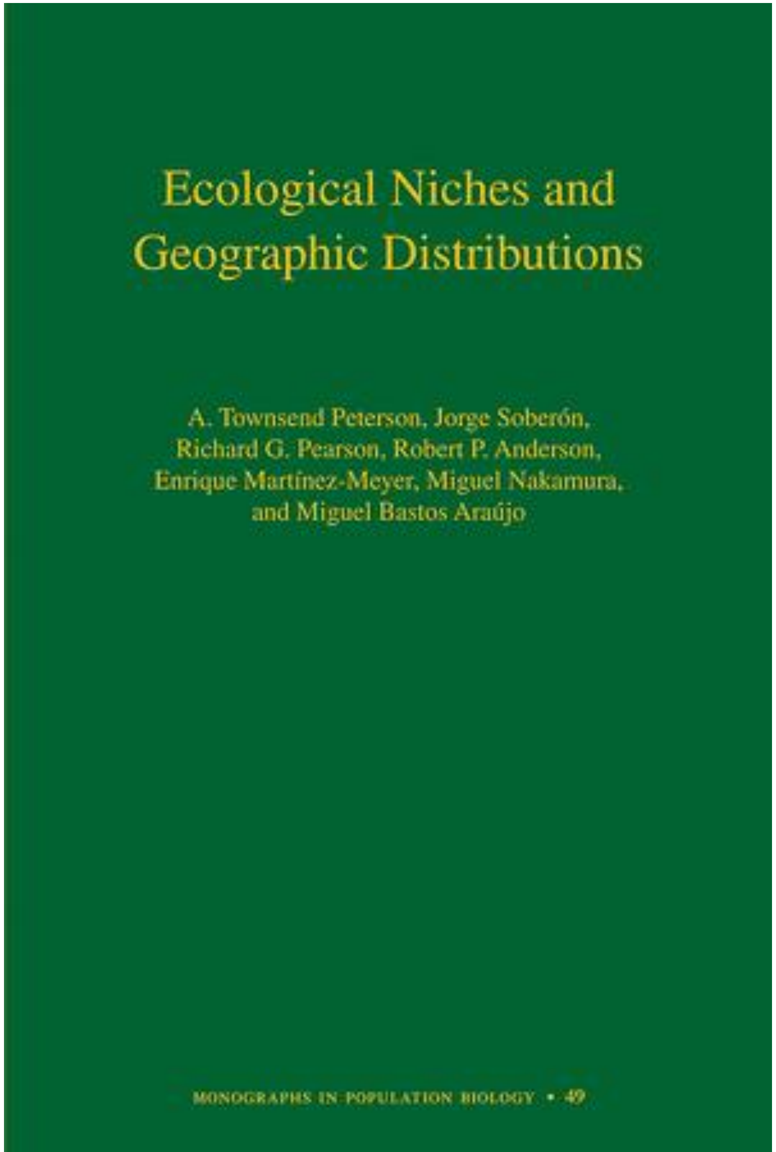
2002

ASSESSING SPECIES DISTRIBUTIONS



2010

ASSESSING SPECIES DISTRIBUTIONS

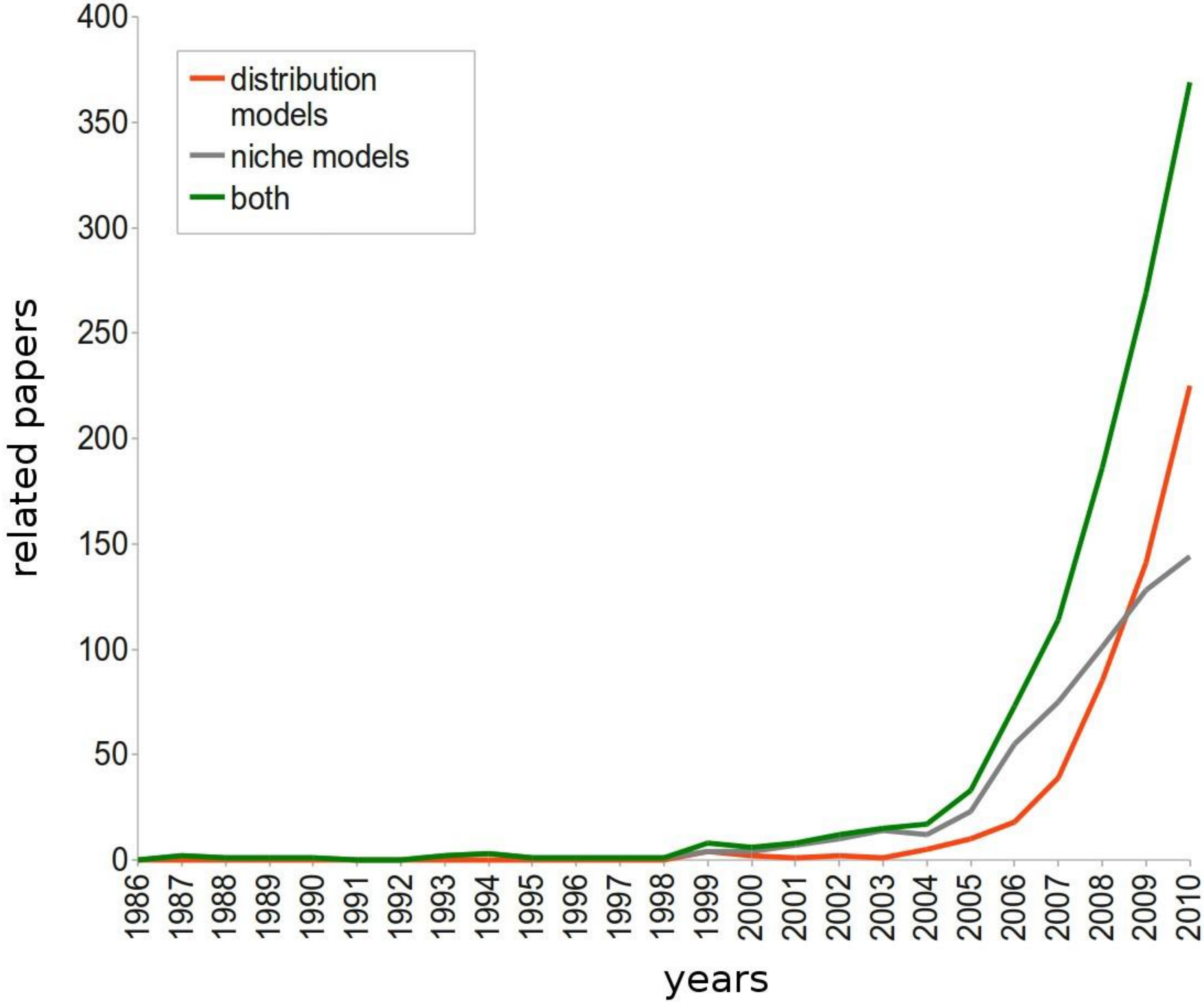


2011

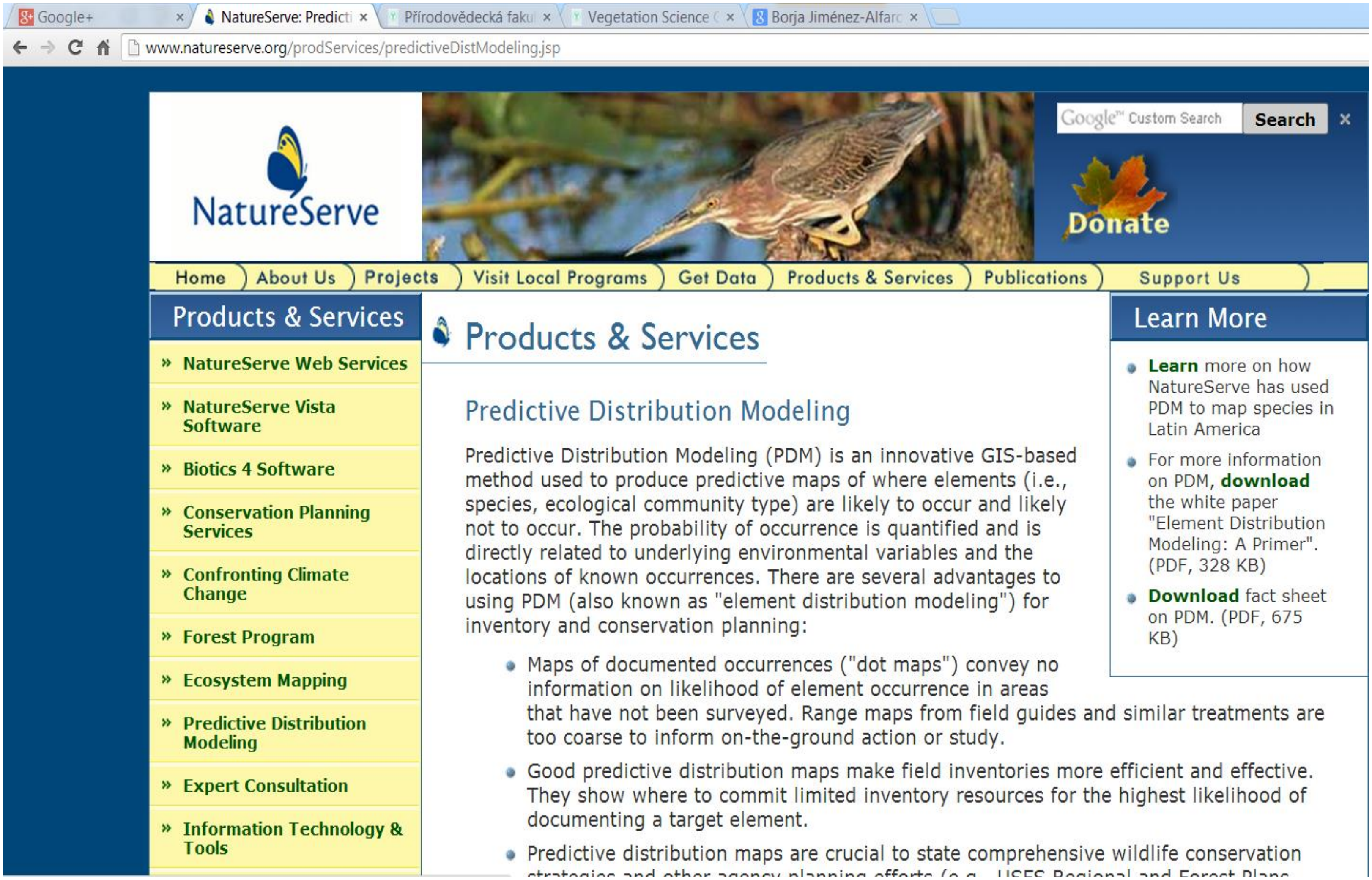
Species distribution models (ecological niche models) are used for:

Predicting species occurrences and estimating ranges
Modeling ecological spatial responses
Reconstructing past distributions
Biogeography of genetic and physiological data
Assessing responses to climate changes
Establishing diversity patterns (endemicity, richness)

and much more…



400
350
300
250
200
150
100
50
0
related papers
distribution models
niche models
both
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
years



Google+ | NatureServe: Predicti | Přírodovědecká faku | Vegetation Science | Borja Jiménez-Alfarc

www.natureserve.org/prodServices/predictiveDistModeling.jsp

NatureServe

Google™ Custom Search | Search

Donate

Home | About Us | Projects | Visit Local Programs | Get Data | Products & Services | Publications | Support Us

Products & Services

- » NatureServe Web Services
- » NatureServe Vista Software
- » Biotics 4 Software
- » Conservation Planning Services
- » Confronting Climate Change
- » Forest Program
- » Ecosystem Mapping
- » Predictive Distribution Modeling
- » Expert Consultation
- » Information Technology & Tools

Products & Services

Predictive Distribution Modeling

Predictive Distribution Modeling (PDM) is an innovative GIS-based method used to produce predictive maps of where elements (i.e., species, ecological community type) are likely to occur and likely not to occur. The probability of occurrence is quantified and is directly related to underlying environmental variables and the locations of known occurrences. There are several advantages to using PDM (also known as "element distribution modeling") for inventory and conservation planning:

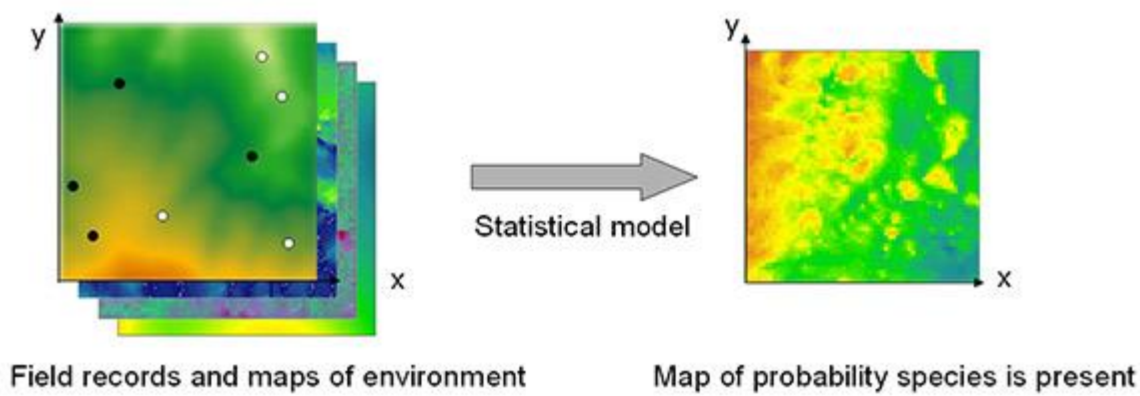
- Maps of documented occurrences ("dot maps") convey no information on likelihood of element occurrence in areas that have not been surveyed. Range maps from field guides and similar treatments are too coarse to inform on-the-ground action or study.
- Good predictive distribution maps make field inventories more efficient and effective. They show where to commit limited inventory resources for the highest likelihood of documenting a target element.
- Predictive distribution maps are crucial to state comprehensive wildlife conservation

Learn More

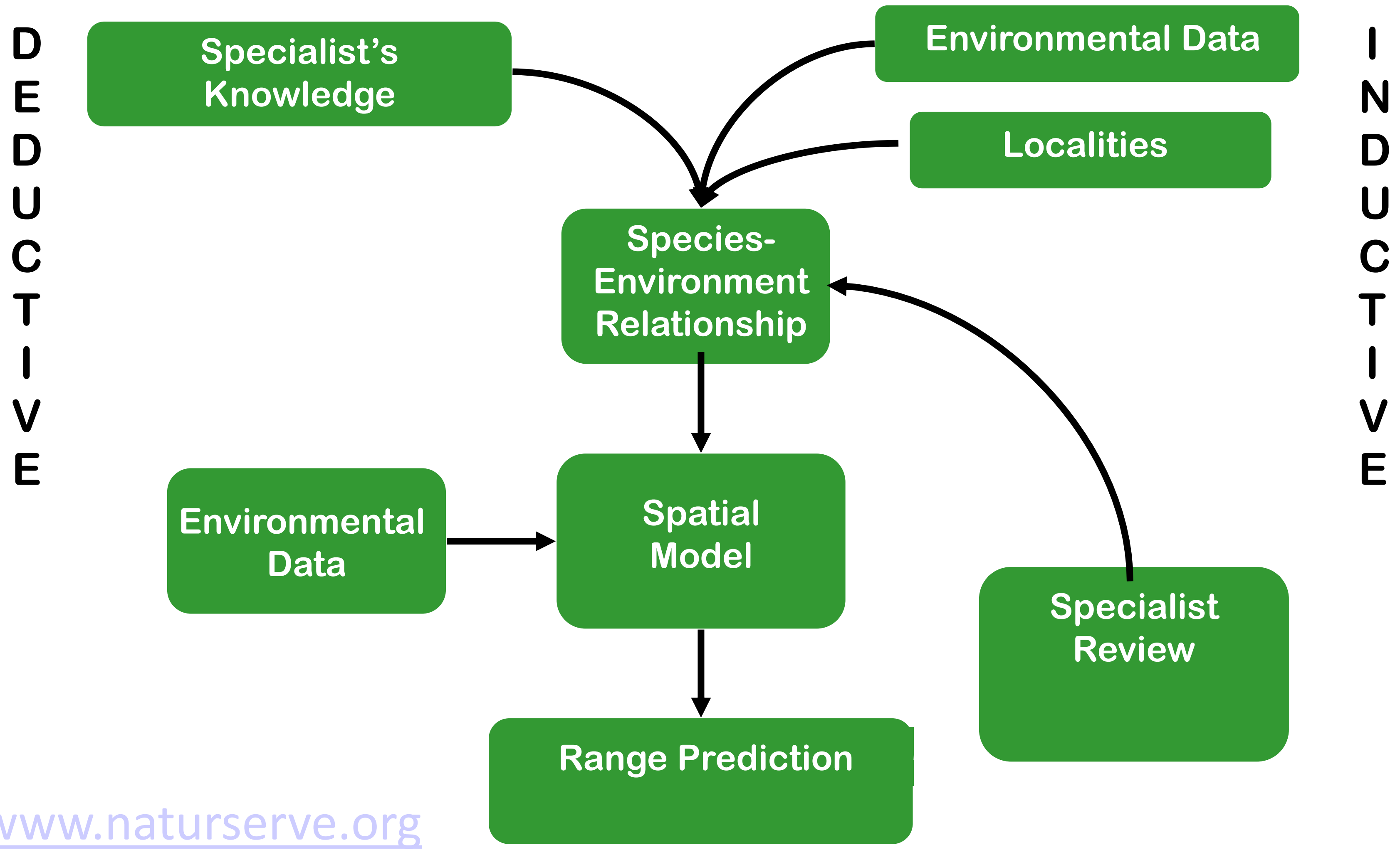
- **Learn** more on how NatureServe has used PDM to map species in Latin America
- For more information on PDM, **download** the white paper "Element Distribution Modeling: A Primer". (PDF, 328 KB)
- **Download** fact sheet on PDM. (PDF, 675 KB)

Three main steps:

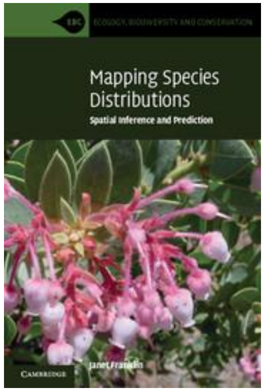
1. compile **spatial data** associated with the target element and **environmental data** for the area of interest
2. build a **statistical model** based on the association of the element to environmental variables at sites of known occurrence
3. Map the **model via GIS** across the area of interest



Distribution Modeling



www.naturserve.org



Franklin
2009

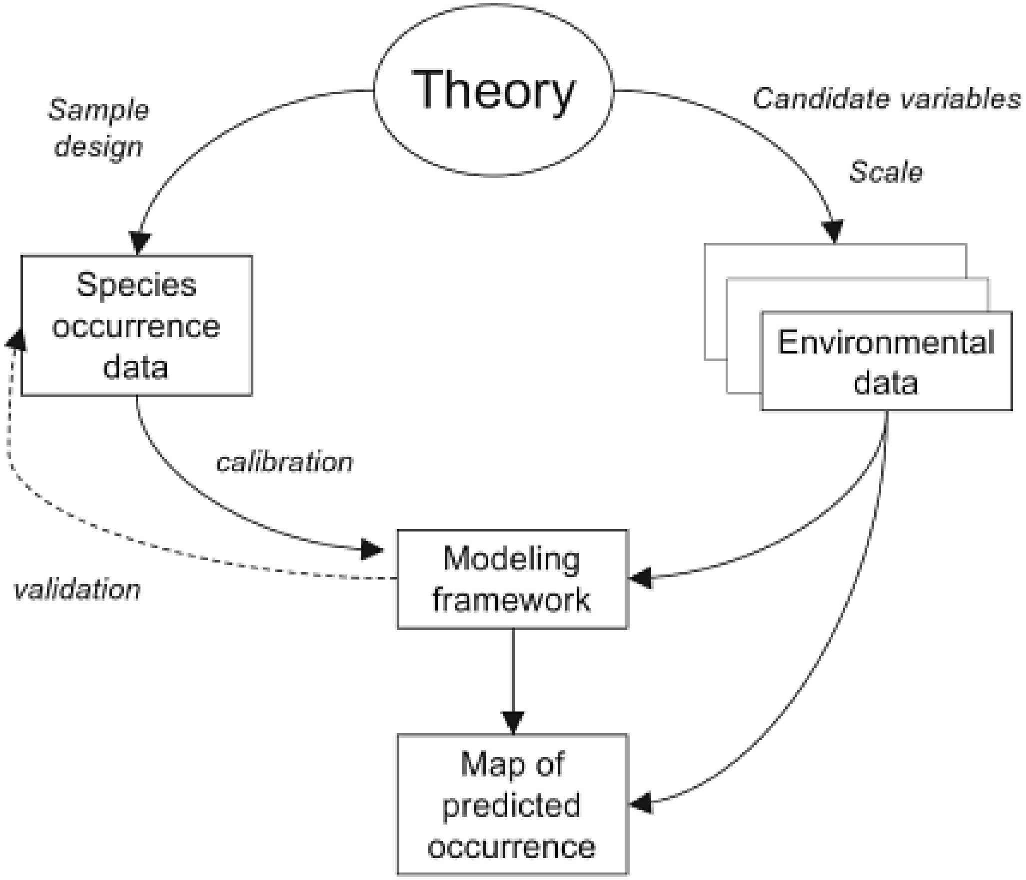
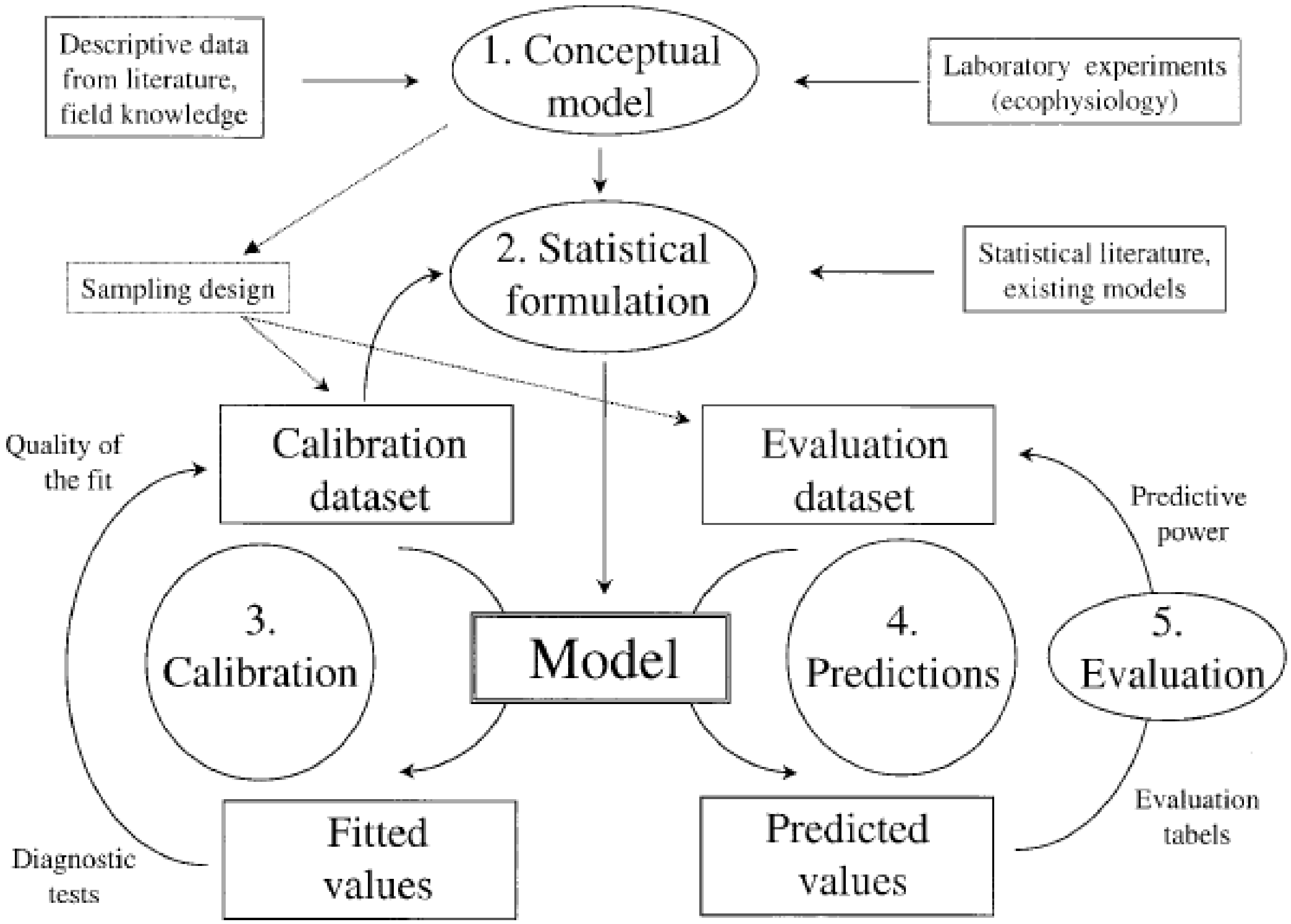


Fig. 1.2. Diagram showing the components of species distribution modeling. Biogeographical and ecological theory and concepts frame the problem, and identify the characteristics of the species and environmental data required to calibrate an appropriate empirical SDM and apply it to produce a map of predicted species occurrence or suitable habitat.

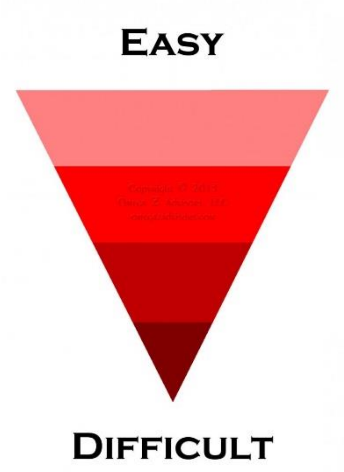
Model building process (from Guisan and Zimmerman 2000)



Distribution modeling (per se) is EASY

Just some technical skills are required

Anyone can compute it with user-friendly software



Applying distribution modeling is more TRICKY

You need a good purpose to do it

(research question, conservation goal)

You must know how to do it properly

About this course:

Part 1 – Mapping

Dealing with occurrence data
Environmental variables
Spatial terms and PRACTICE with GIS

Part 2 – Modeling

Background theory (niche concept)
Modeling methods
Maximum Entropy and PRACTICE with MaxEnt

Part 3 – Mapping and Modeling

Model implementation and evaluation
Applications and future challenges
Using your OWN DATA and GROUP PRACTICE

Part 1:

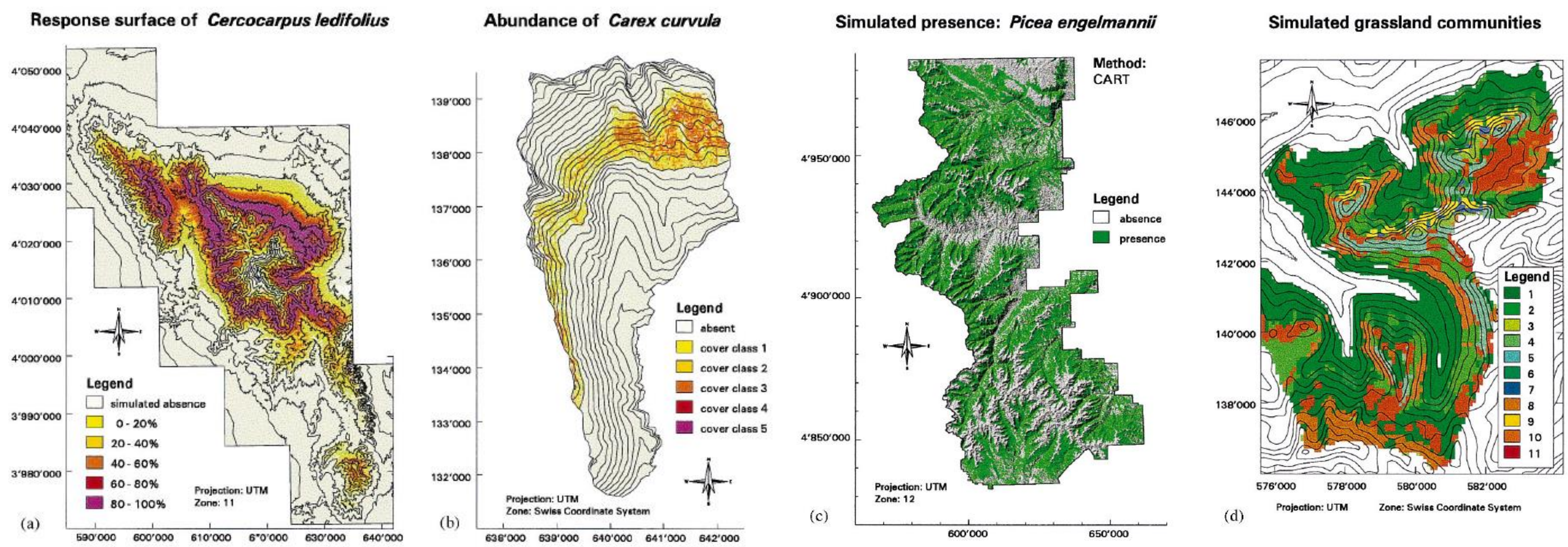
MAPPING

OCCURRENCE DATA

What is occurrence data?

A **record** for one species/organism/community in one **locality**

Presences (and **absences**) are the MAIN dependent variable for Species Distribution Modeling (but there are more)



Guisan & Zimmerman (2000)

Where to obtain occurrence data?

Two main options:

1. Desing your own field sampling

Pro: you have control on your data

Con: many times you have not time or money

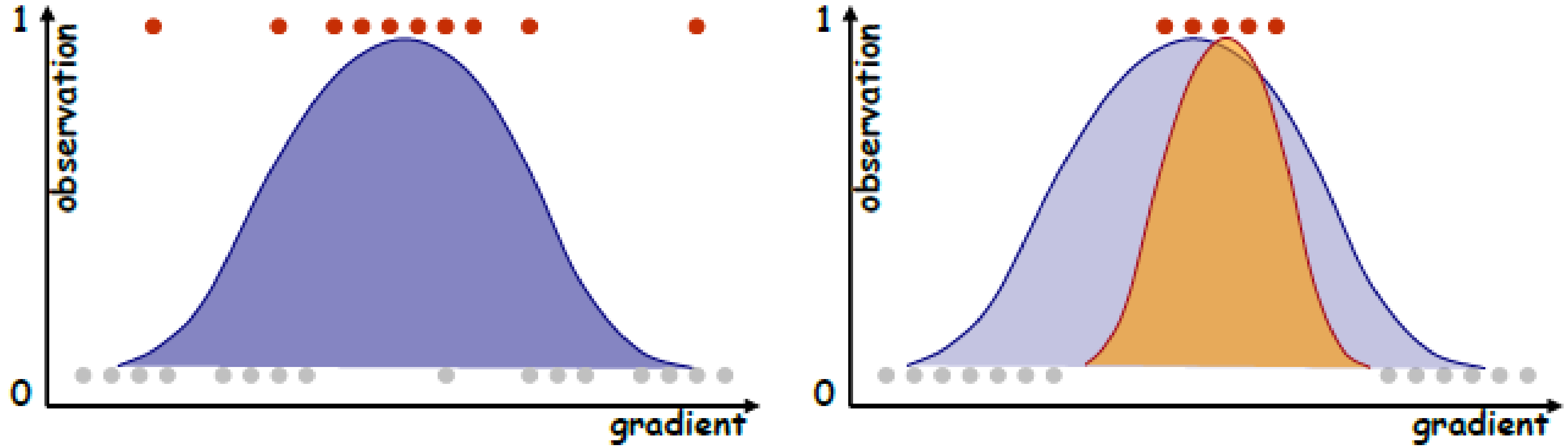
2. Using existing data (e.g. biodiversity databases)

Pro: huge amount of data around the world

Con: uncertainties on sampling, accuracy, etc.

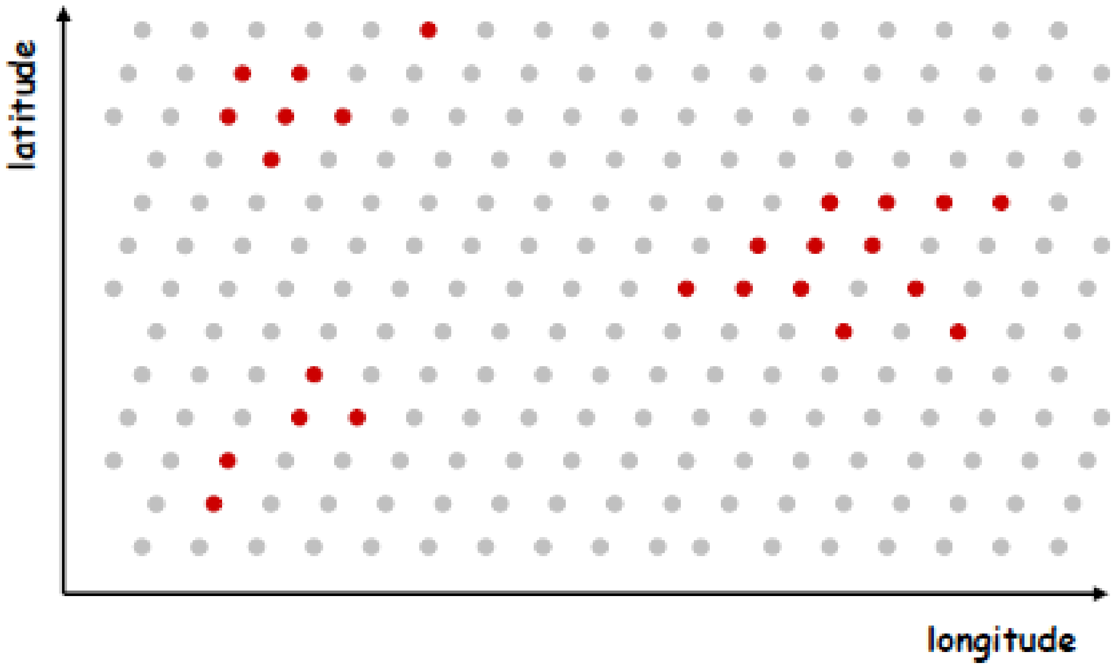
1. Sampling design

Probabilistic design is required to quantify the species responses along gradients, in order to consider the edges of environmental distribution



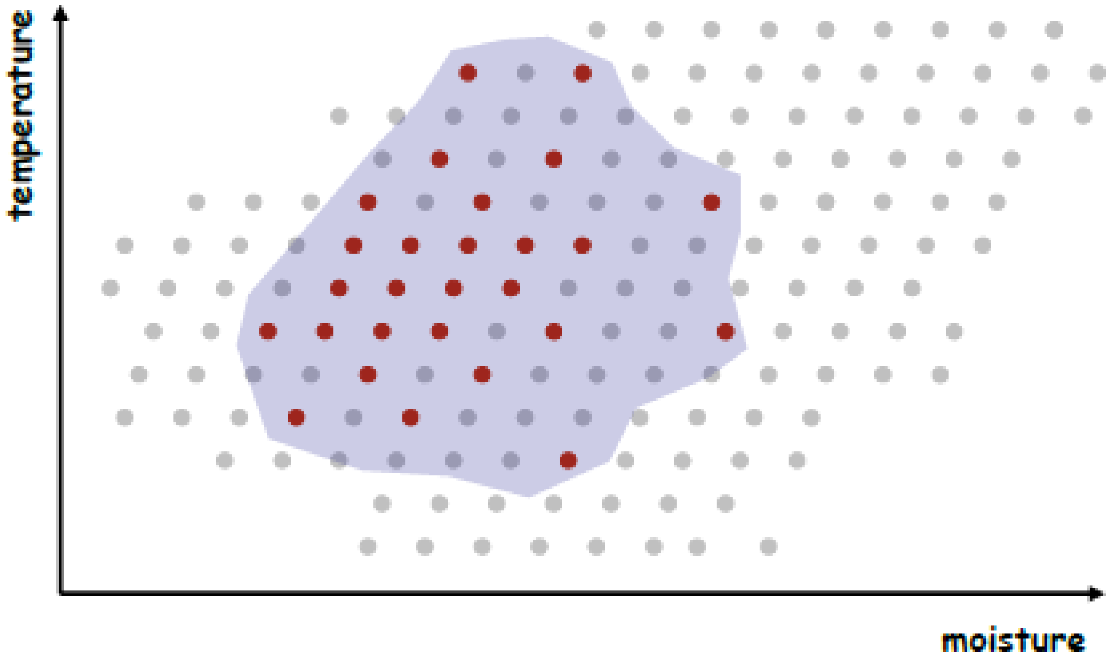
Geographic distribution of a species

Species are often distributed in a patchy manner at the scale of the landscape (non-random structure of L).



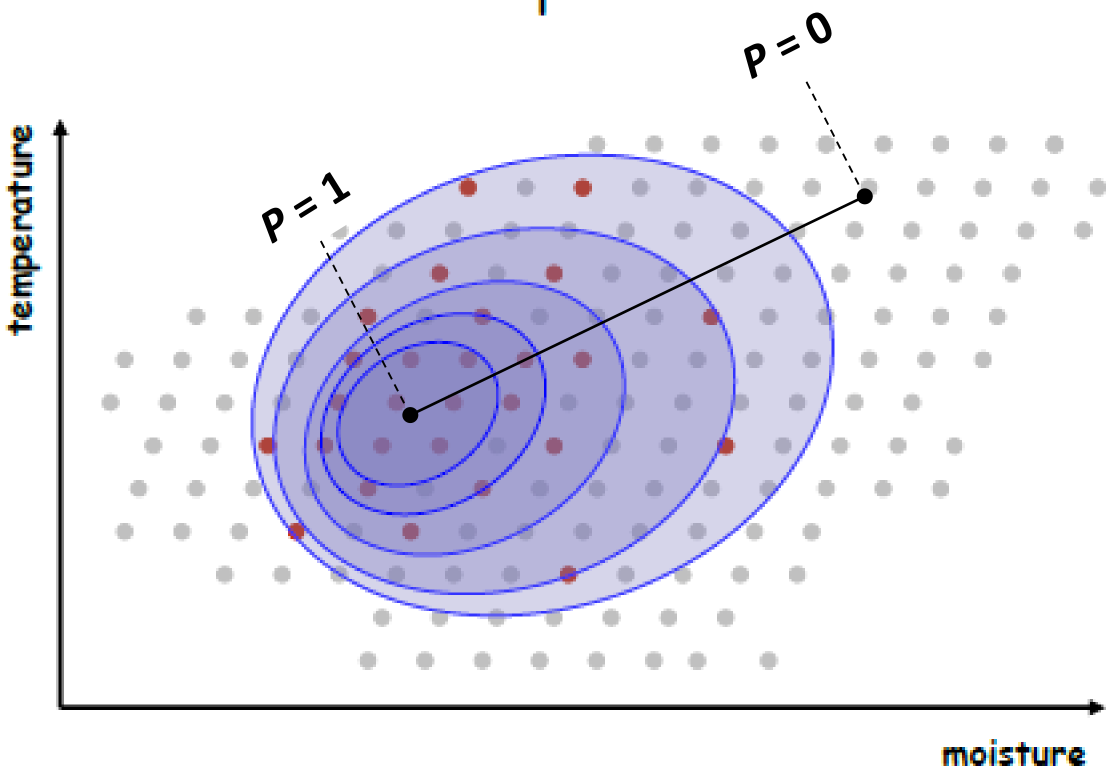
Distribution in the environmental space

If we translate this distribution into env. space, then the distribution often looks quite different



Distribution in the environmental space is different!

Our goal is then to quantify the density distribution or the likelihood to find a species in the environm. space



There is no universal design for all questions, but…

Simple random design is used for relatively homogeneous spaces (when the probabilities of occurrences are equal) but it is not a good option if you have to sample organisms which are rare or disjuntly distributed

Regular, **systematic**, **clustered** or **stratified** designs are prefered to sample occurrence data if the organism is clearly influenced by geographial, environmental or topotraphical gradients

Even better can be mixed designs, e.g **random stratified sampling**

You “should” avoid **bias** survey sampling

-> Geographic bias: along roads, near the cities,…

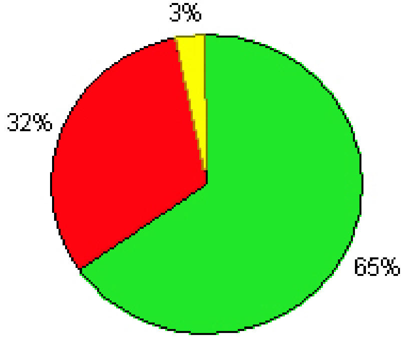
-> Taxonomic bias: wrong identification of species

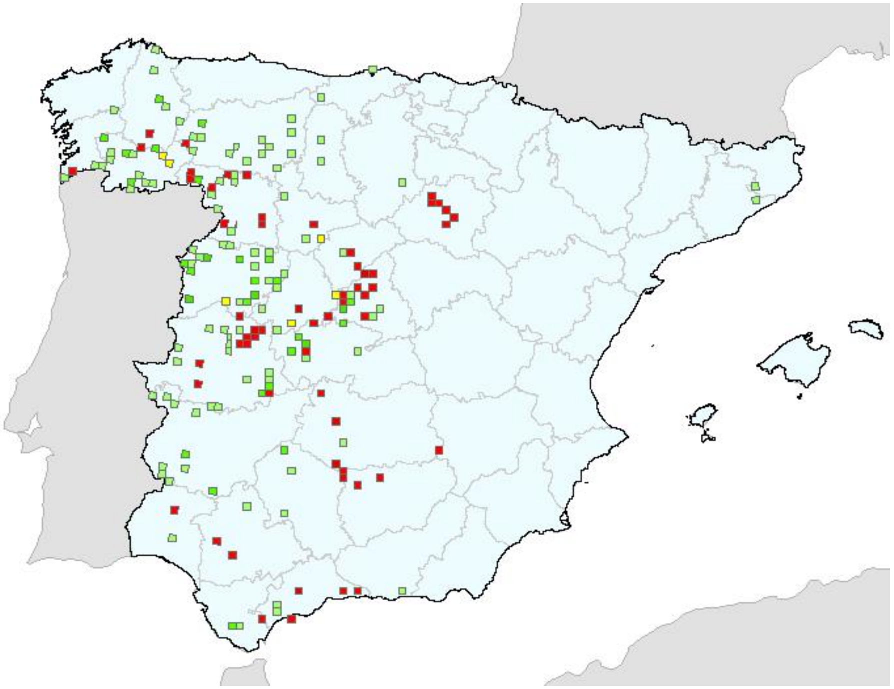
E.g.: identification of bias in biological collections of *Lupinus hspanicus* (Parra-Quijano et al.): diferent geographic cover

Genetic bank

Both

Herbarium / literature



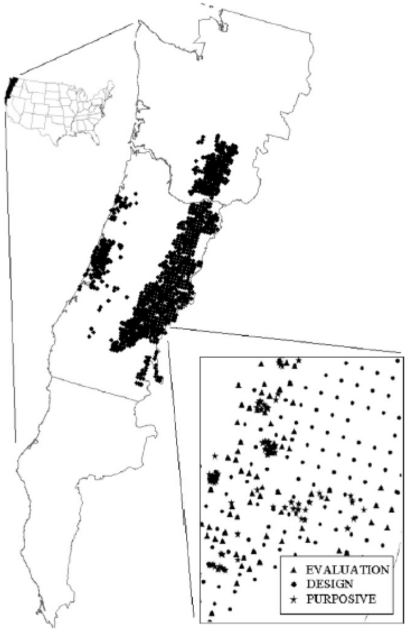


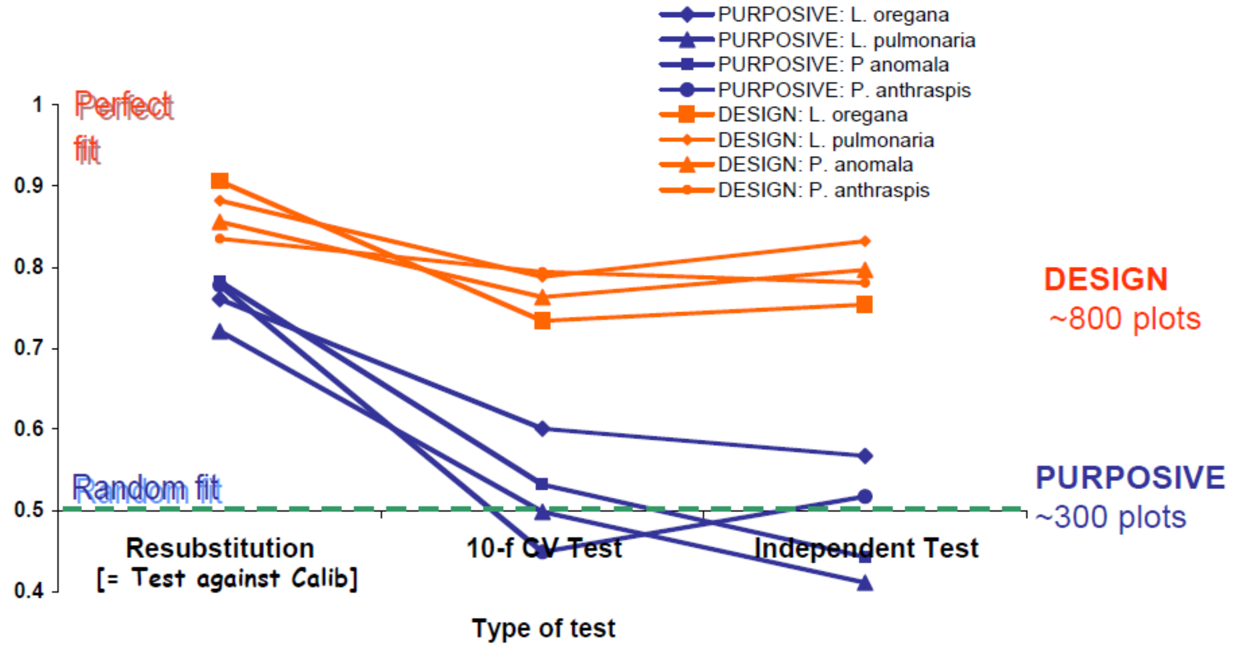
You “should” avoid **purpose** sampling

-> non-probabilistic, based on aprioristic knowledge

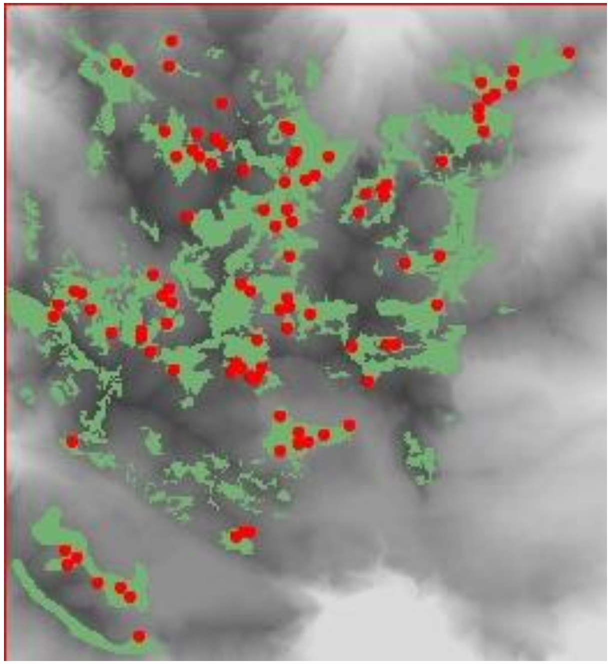
-> usually produces undersampling of the study subject

E.g.: Comparison of sampling survey desings for predicting lichen species in USA (Edwards et al. 2006)

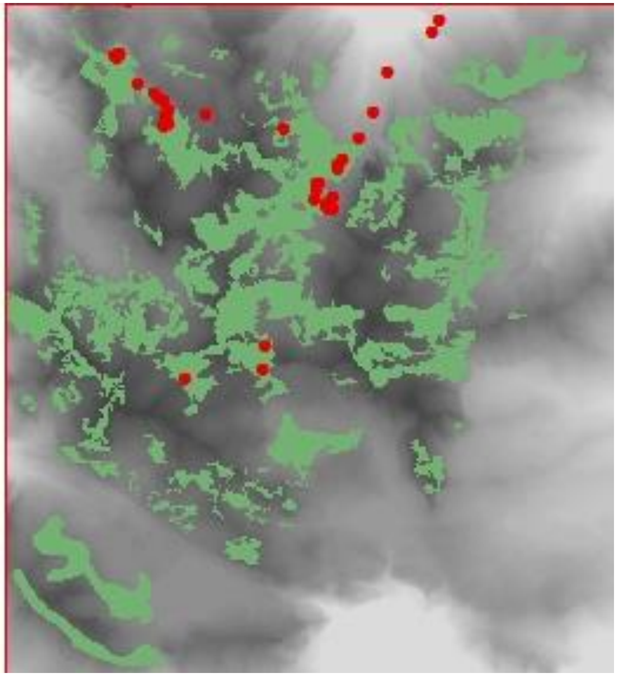




A visual example of **designed** versus **purposive** sampling (vegetation plots in Picos de Europa, Spain)

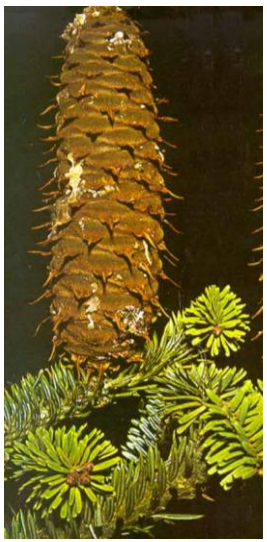


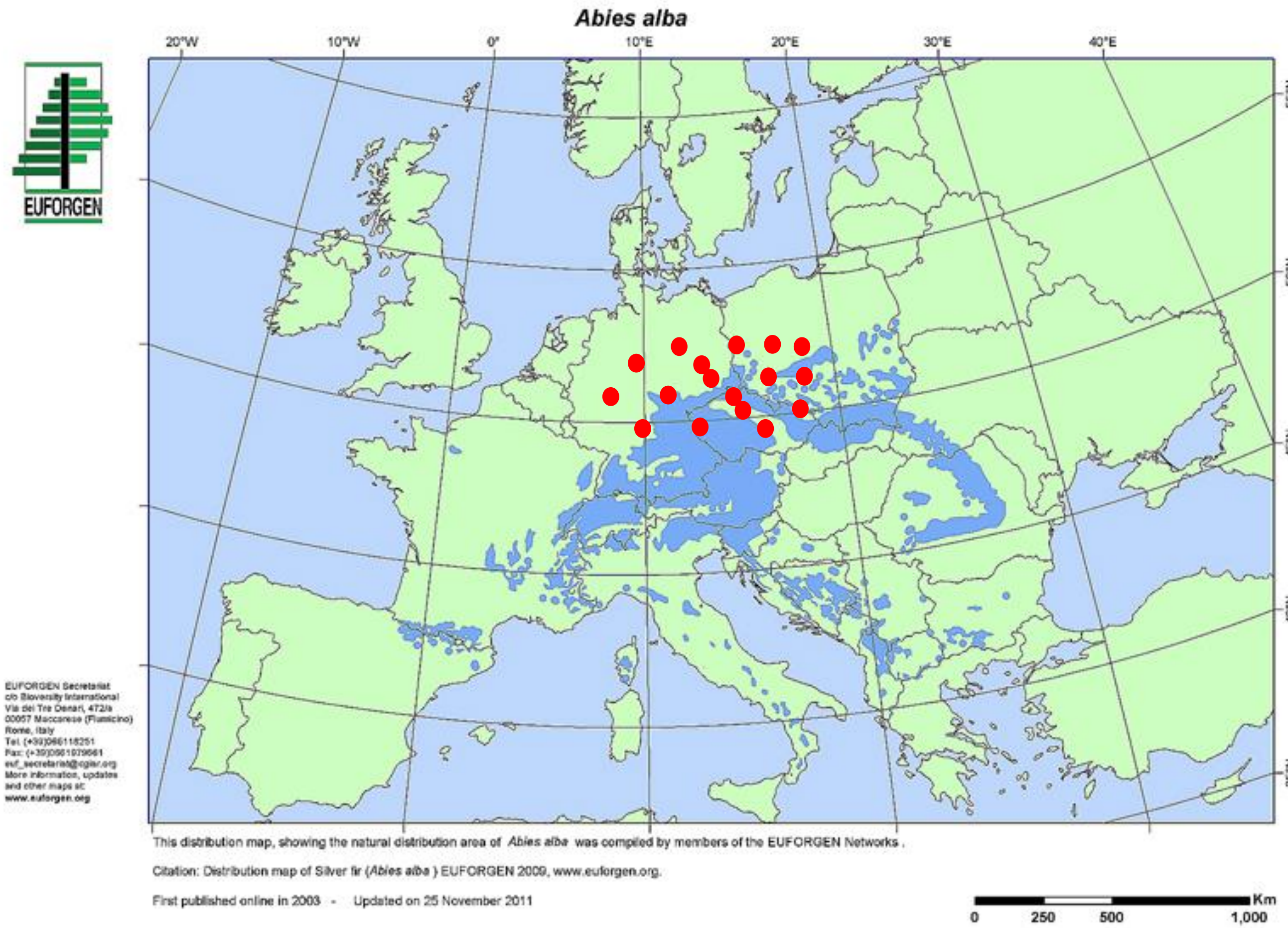
Designed (systematic)
(N = 80)



Purposive (biased)
(N = 100)

More important that the number of observations is the degree to which the range of the environmental space occupied by the species are covered in the sample (= **COMPLETENESS**) and the frequency of events (records of species presences) from the sample (= **PREVALENCE)**





How many samples?
Or, a better question,
What is the **mínimum sample** size for my study**?**

For SDMs, there are some rules:

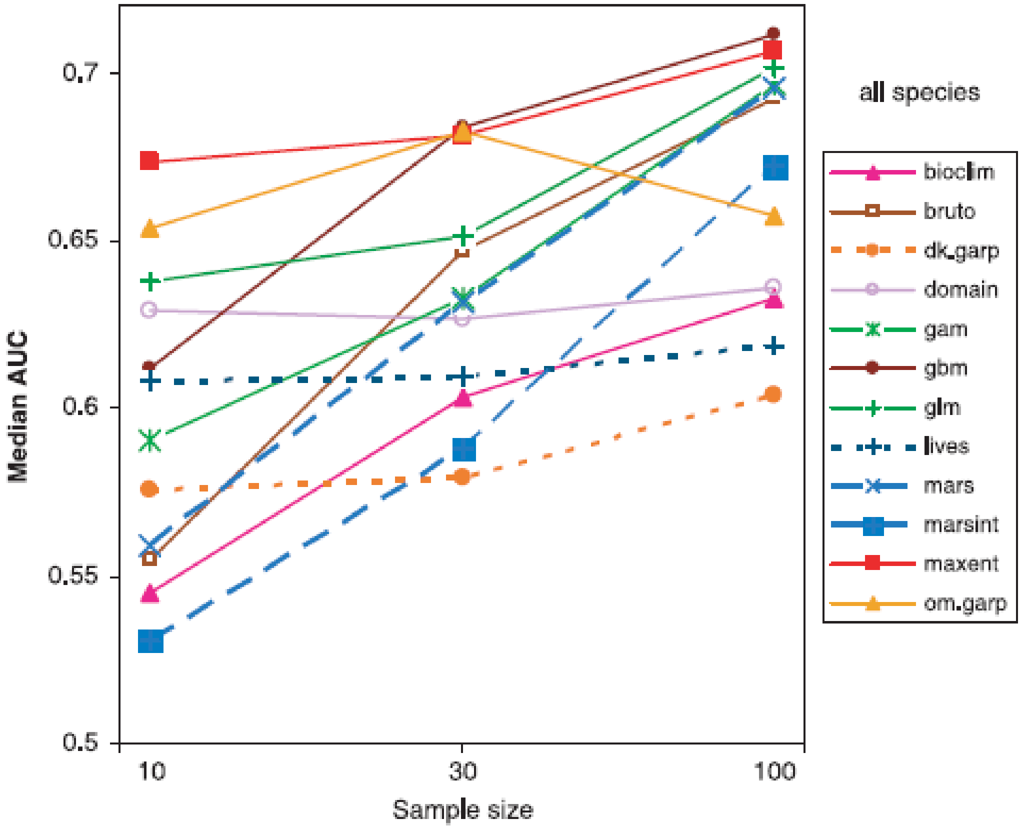
→ A mínimum of 50 observations can be fine

→ 20-40 times as many observations as predictors

→ For rare species and some algorithms, 20 occurrences can be enough…!

Very few samples can be valid for rare organisms

... but it depends on the method



Wisz et al. 2008. Effect of sample size on the performance of species distribution models.
Diversity and Distributions 14: 763

In summary, the quality of our data for modeling distributons will depend on many factors:

-> **EXTENT** of the study area and **ACCURACY** of occurences

-> The **ECOLOGY** of the species

-> How we sample the **ENVIRONMENTAL SPACE**

-> How many **PRESENCES** and **ABSENCES** are sampled

-> The **PREDICTORS** and the modeling **METHOD**

2. Using existing data (e.g. biodiversity databases)

-> SDMs are mainly used to map unkwon species distributions
-> Species mapping has however a long history using known distributions from many different sources

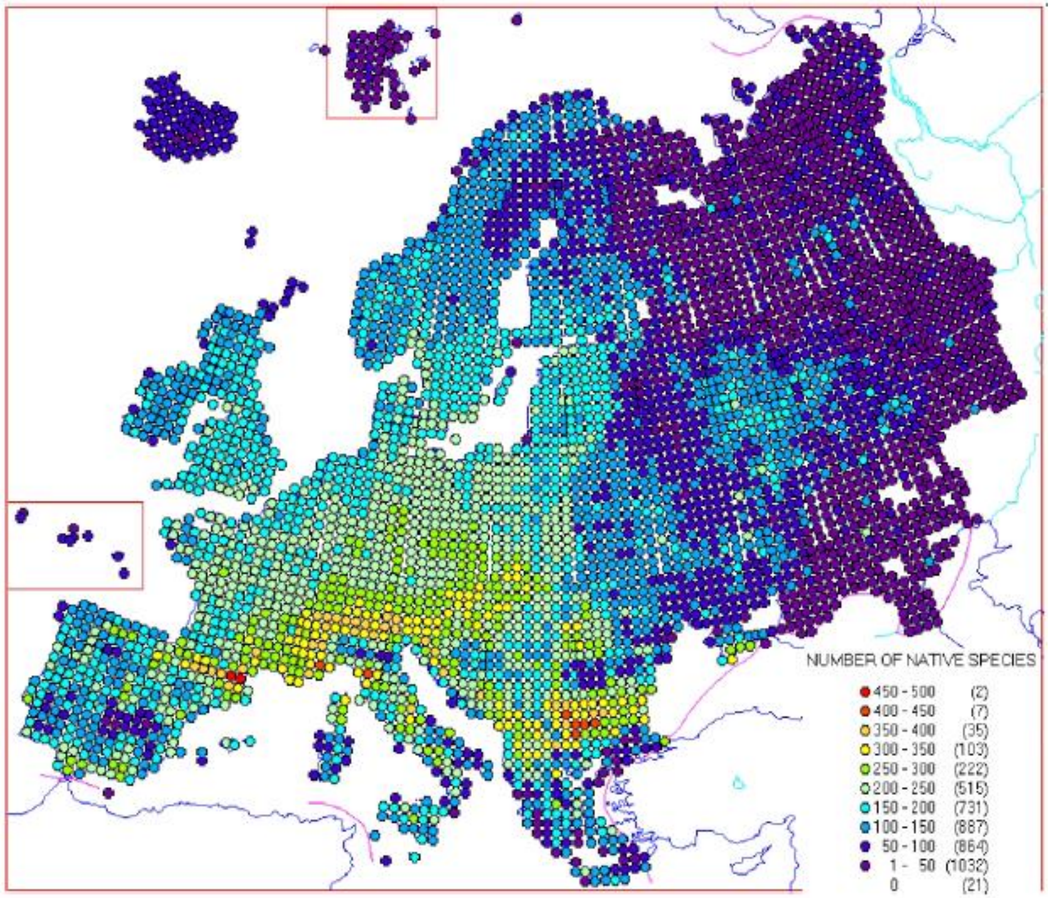
Main types of sources:

- **Grid-based atlases** (compilation of information)
- **Natural history collections** (museums, botanic gardens)
- **Surveys** (conservation, vegetation or faunistic surveys)

Grid-based Atlases

Pro: Cover large territories and represent distribution ranges well

Con: Coarse grain (10 km, 50 km) and small spatial acuracy



ATLAS FLORAE EUROPAEAE

50km x 50 km

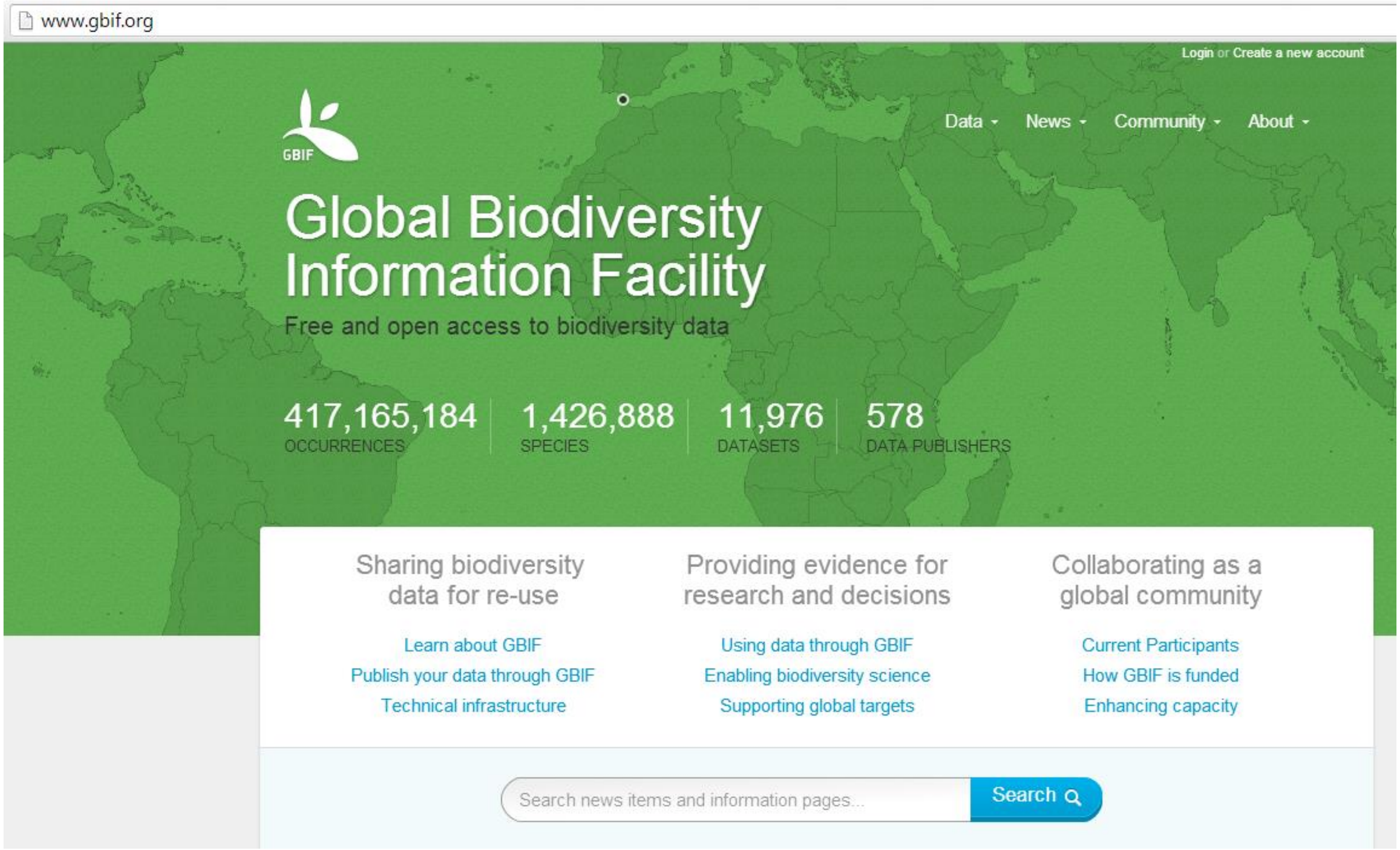
2559 species

(20% of European flora)

Natural history collections

Pro: Large amount of data for all the world

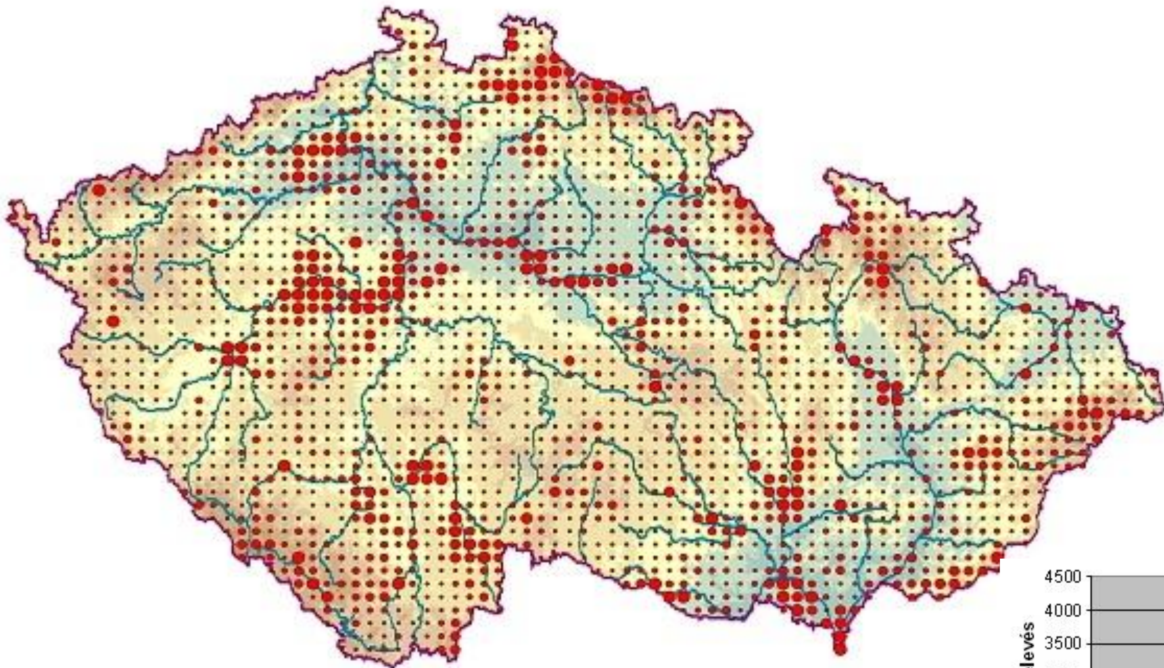
Con: Low spatial resolution and high uncertainty



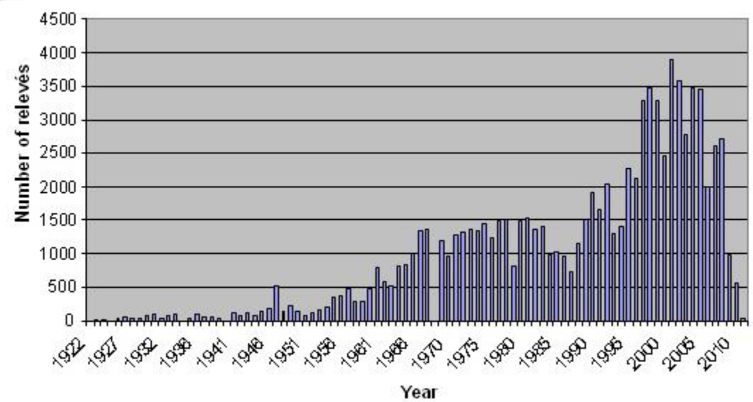
Biodiversity Surveys

Pro: Spatial accuracy is heterogeneous, although can be good

Con: Generally biased or purposive samping



Czech national Phytosociological Database



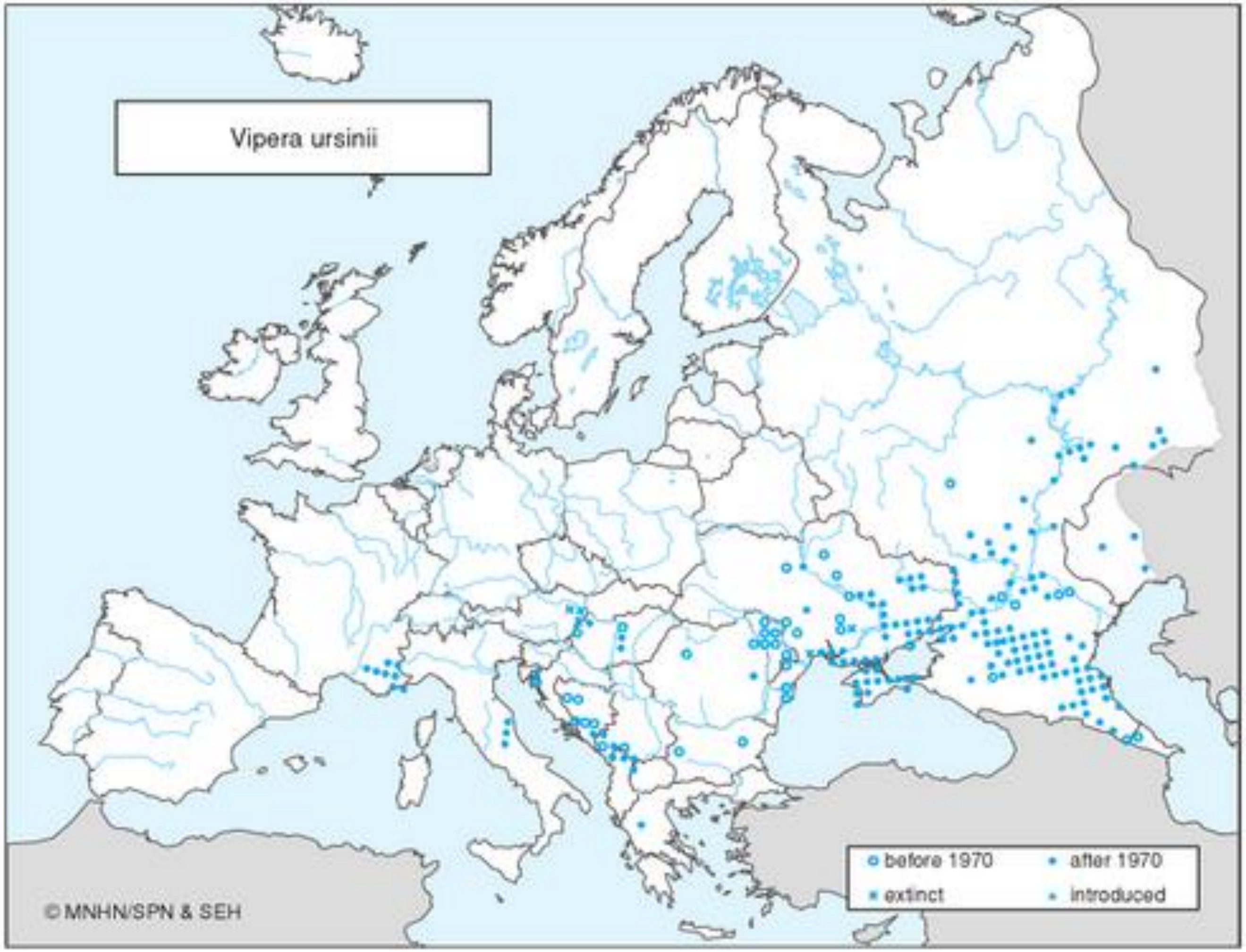


Fig. 2.3. Example of "dot map" for *Vipera ursinii* from *The Atlas of Amphibians and Reptiles in Europe* (Gasc *et al.* 1997) (http://www.seh-herpetology.org/). Each dot represents a 50 × 50 km UTM grid square. Copyright © Muséum National d'Histoire Naturelle & Service du Petrimone Naturel and Societas Europaea Herpetologica. Used with permission.

There is some sampling design behind this? PROBABLY NOT

Is the data valid? PROBABLY YES

From Franklin 2009

Problems associated with biodiversity databases

1. **Low spatial accuracy**: location and coordinates (if existing) are generally imprecise

2. **Unknown sampling design**: generally biased or purposive, but in general not reported

How this affects our data:
- Incomplete distributions (bias)
- Undersampling
- Pseudo-replication
- Spatial autocorrelation of samples
- Low spatial accuracy of the analyses

Next week we will go back to the spatial issues

How to solve these limitations?

Georeferenciating: it takes time but it allow us to measure spatial uncertainty

Resampling: to have some control of the data (e.g. analyzing subsets separately)

Adaptative sampling: resampling after a first assessment

Evaluating bias: using spatial information

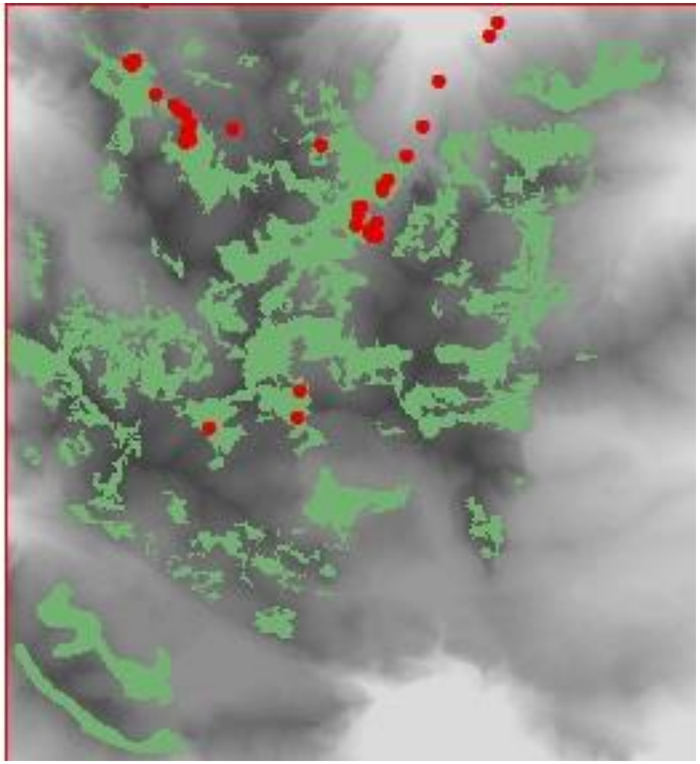
Measuring spatial autocorrelation

Georeferencing

The main challenge of biological collections is the assignment of geographic coordinates to millions of historical records (Baker & al., 1998)

Spatial autocorrelation

Oversampling of aras produces pseudo-replication and further overfiting of the models.



What to do?

Sampling (or resampling) according to spatial criteria

Assessing spatial autocorrelation (e.g. Moran's *I*) after modeling

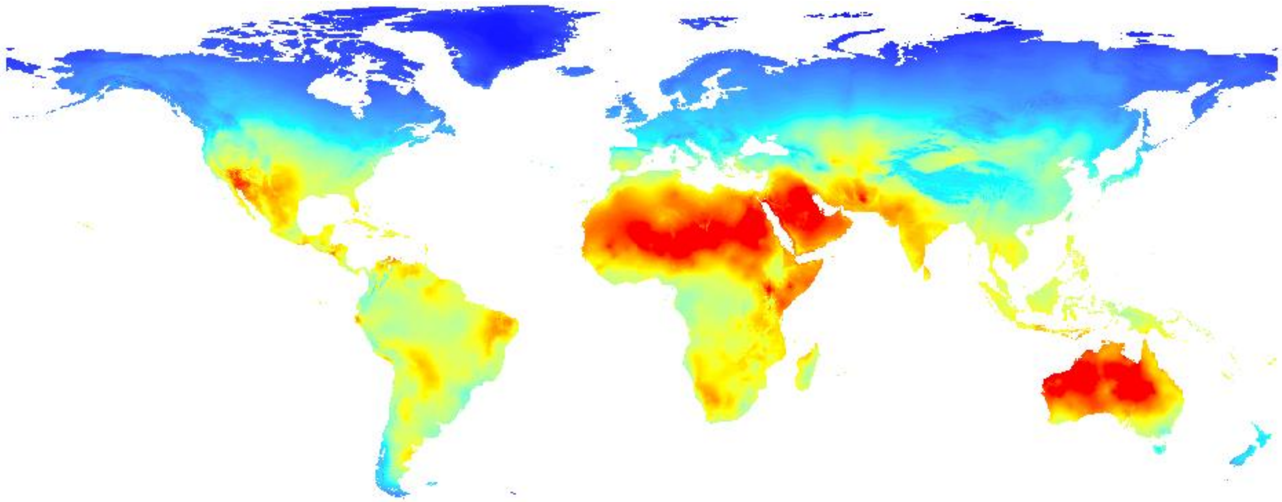
Part 1:

MAPPING

ENVIRONMENTAL VARIABLES

In the ecological space, what factors are important for the distribution of species?

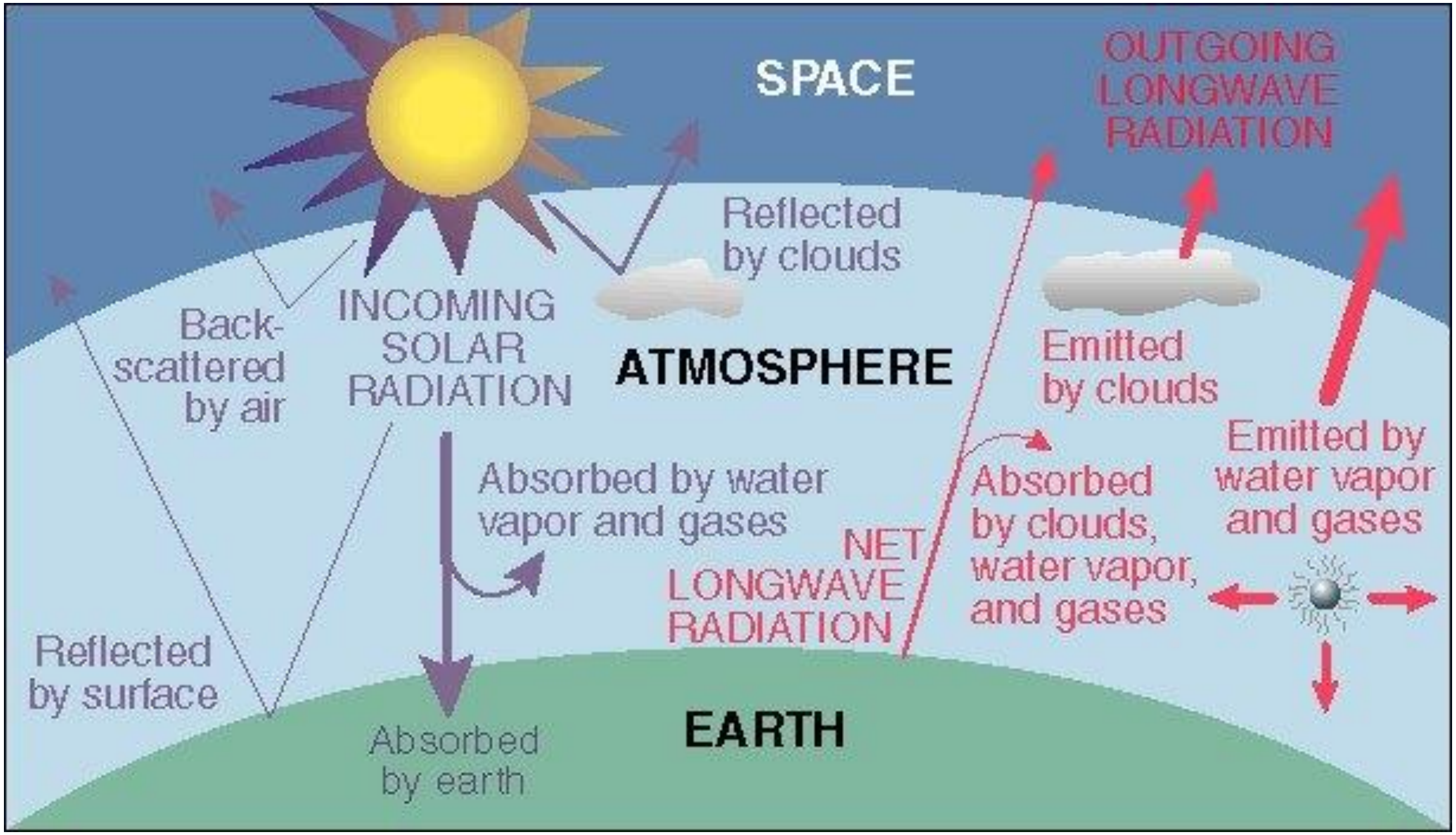
At the **macro-distributional** scale, ultimate controlling factors have to do with energy requirements of species.
Energy requirements are, in turn, determined by physiology and morphology



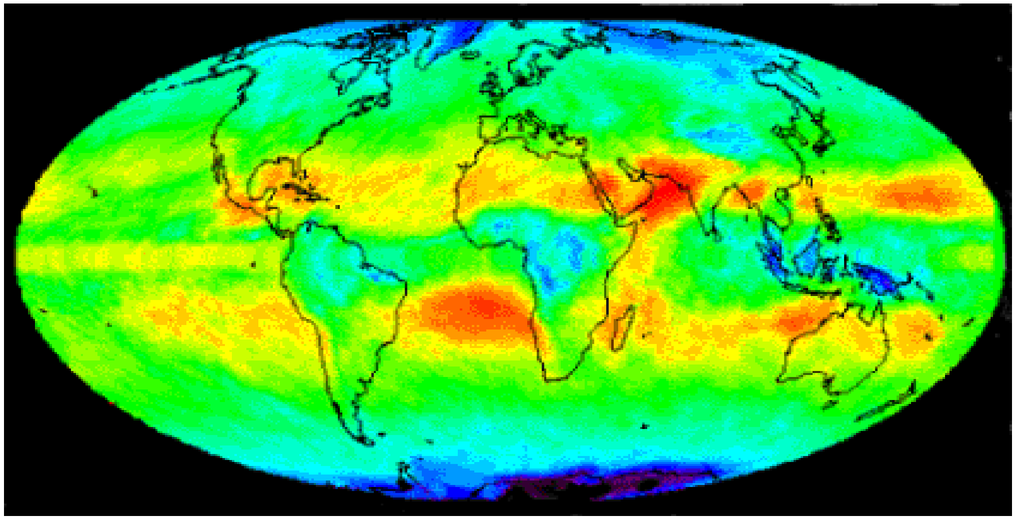
World Potential Evapotranspiration

http://ruig.grid.unep.ch/?p=95

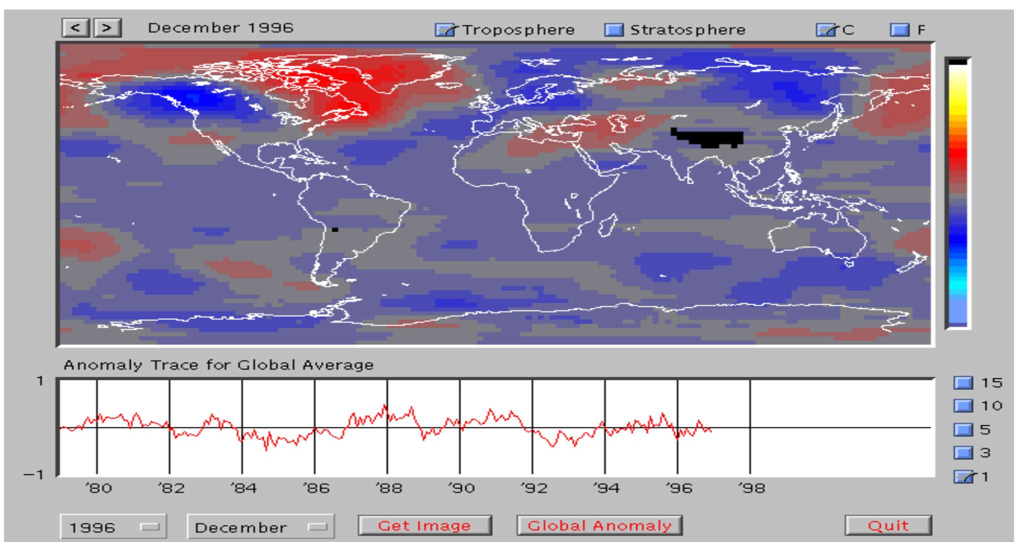
And… what is the primary source of energy for the Earth?



Solar energy brings

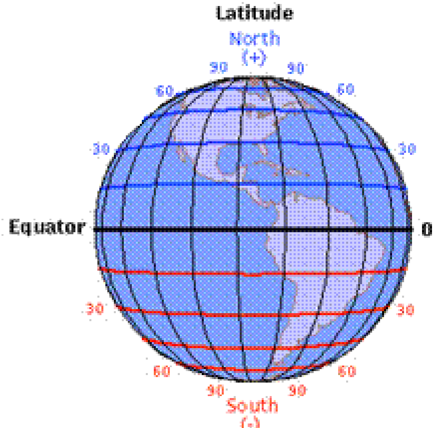


Light
(quantity and quality)



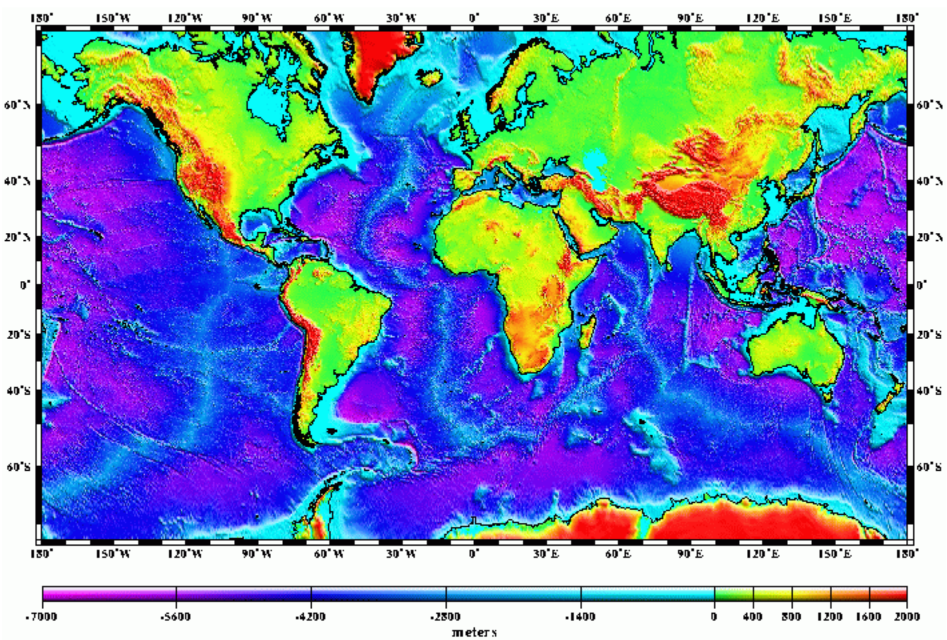
Heat

What factors affect solar radiation and temperature?



Microsoft Illustration

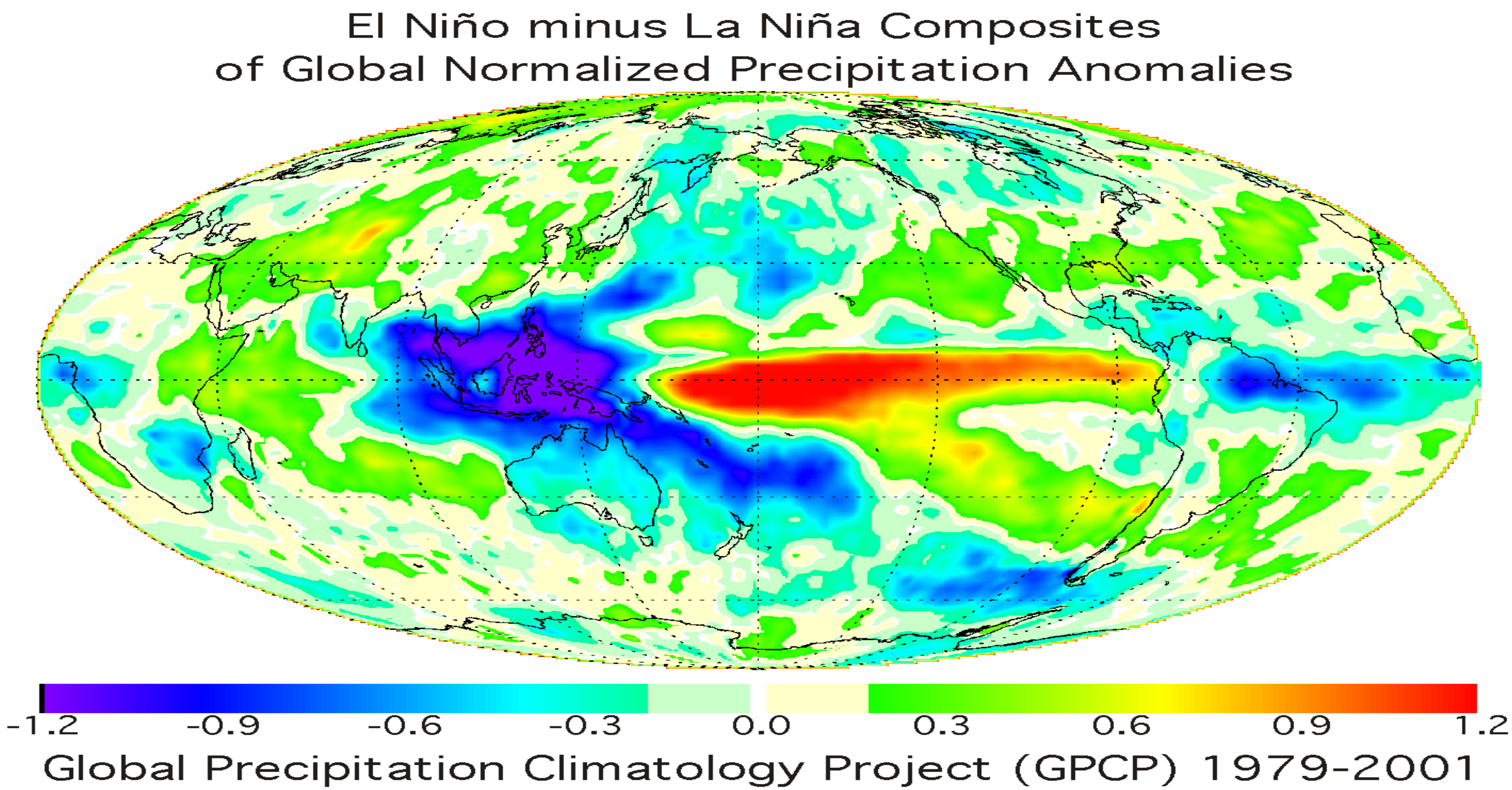
Latitude



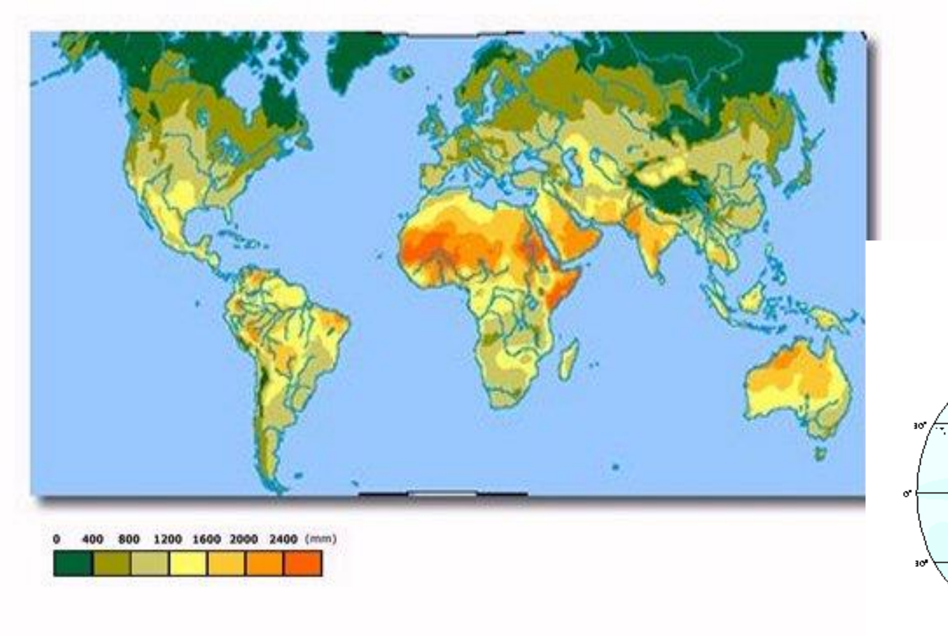
Topography:
- Elevation
- Slopes
- Exposure

But the factory of primary production (vegetation)

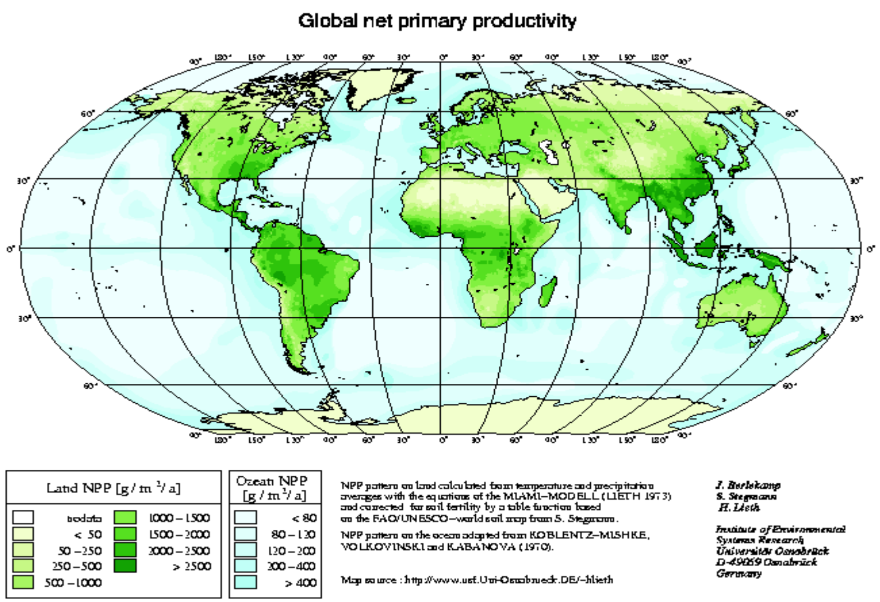
... also needs WATER



There are also functional manifestations of the interplay of all these factors



evapotranspiration



productivity

In sum

There are distal factors that determine directly or indirectly the distribution of all species (at broad spatial scales):

-> Amount of light
-> Amount of heat
-> Amount of water
-> Topography

NOTE: Energy and water income is dynamic in time. For some questions regarding the eco-geographic distribution of species, the time dimension is crucial

Implications

When modelling individual species, more proximal variables become relevant:

-> **Soil types**

-> **Evapotranspiration**

-> **Primary productivity**

-> **Light quality**

-> **Number of frost days**

NOTE: A necessary field of research is needed to achieve a better understanding of the inclusion of different types of variables in the modelling process, as well as the effect of redundancy on model quality.

But EACH organism has its own requirements

E.g.: interactions, parasitisms…….

Global Change Biology

Global Change Biology (2013), doi: 10.1111/gcb.12226

Finding the appropriate variables to model the distribution of vector-borne parasites with different environmental preferences: climate is not enough

ANTÓN PÉREZ-RODRÍGUEZ*, SOFÍA FERNÁNDEZ-GONZÁLEZ*, IVÁN DE LA HERA*,† and JAVIER PÉREZ-TRIS*

*Departamento de Zoología y Antropología Física, Universidad Complutense de Madrid, Madrid, E-28040, Spain, †Departamento de Zoología y Biología Celular Animal, Universidad del País Vasco (UPV/EHU), Vitoria-Gasteiz, E-01006, Spain

Abstract

Understanding how environmental variation influences the distribution of parasite diversity is critical if we are to anticipate disease emergence risks associated with global change. However, choosing the relevant variables for modelling current and future parasite distributions may be difficult: candidate predictors are many, and they seldom are statistically independent. This problem often leads to simplistic models of current and projected future parasite distributions, with climatic variables prioritized over potentially important landscape features or host population attributes. We studied avian blood parasites of the genera *Plasmodium*, *Haemoproteus* and *Leucocytozoon* (which are viewed

Types of environmental variables

In statistical terms, there are two main variables:

QUANTITATIVE
elevation, temperature, precipitation, etc.

QUALITATIVE (CATEGORICAL)
Soil type, land cover, vegetation

->They will be used differently in the modeling process

Types of environmental variables

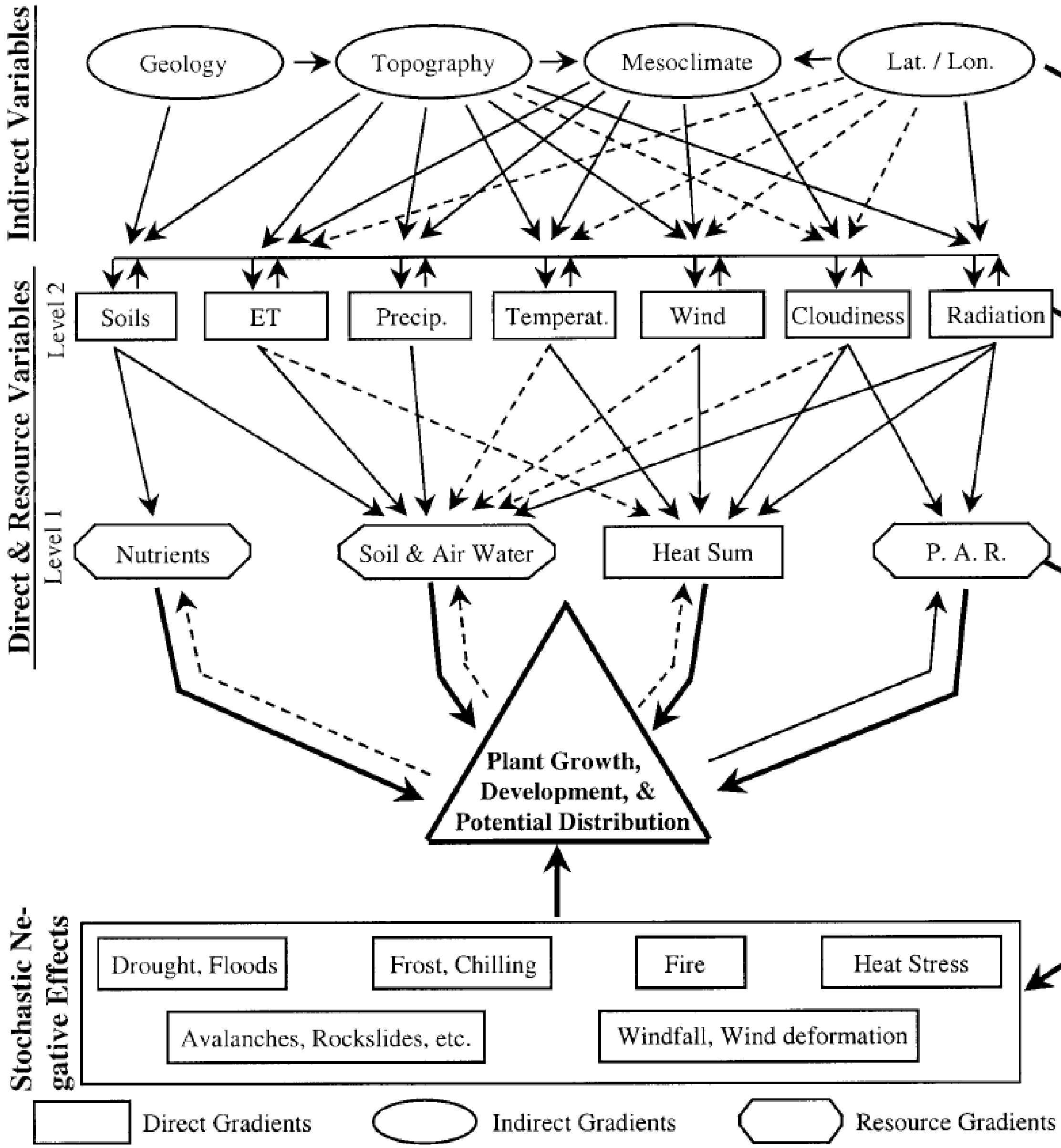
In practice, we should distinguish the most appropiate variables for our study case, and especially the **SCALE**

Broad scale studies:
More focused on DIRECT variables, mostly climatic:
Temperature, precipitacion, solar radiation, evapotranspiration

Local scale studies:
More focused on INDIRECT variables, mostly topogaphic:
Elavation, slope aspect, exposition, topograhical índices, etc.

ENVIRONMENTAL VARIABLES

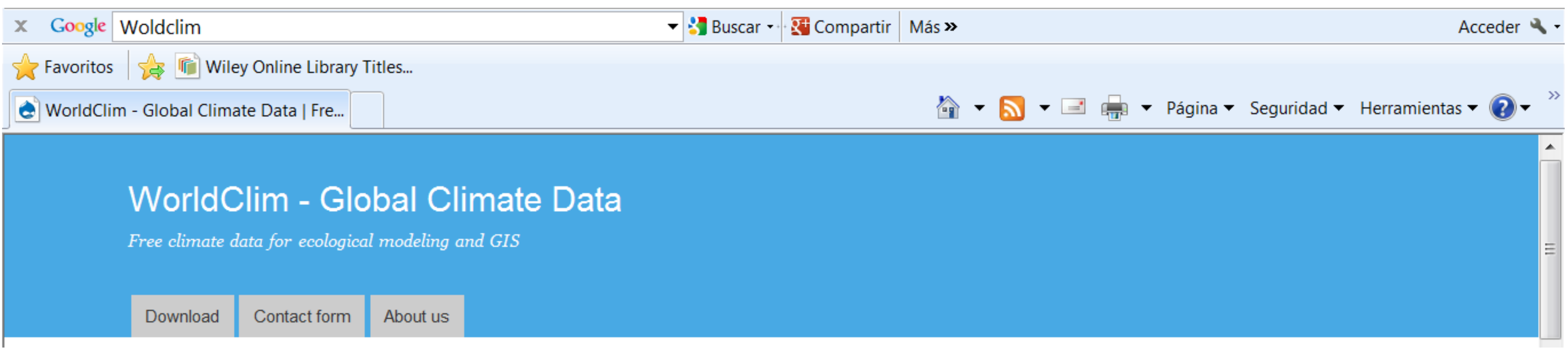
Conceptual model of relationships between resources, direct and indirect variables, and their influence on plant performance (from Guisan & Zimmerman 2000)



WORLDCLIM

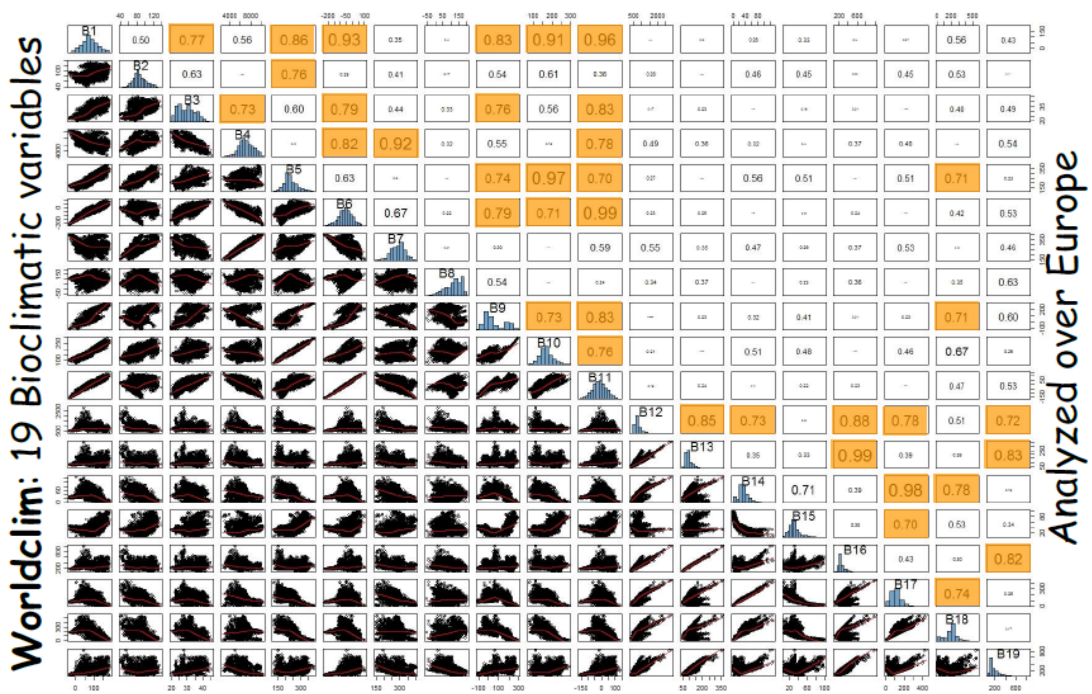
Averaged from long-term (30yr) series of Temp and Prec

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)"

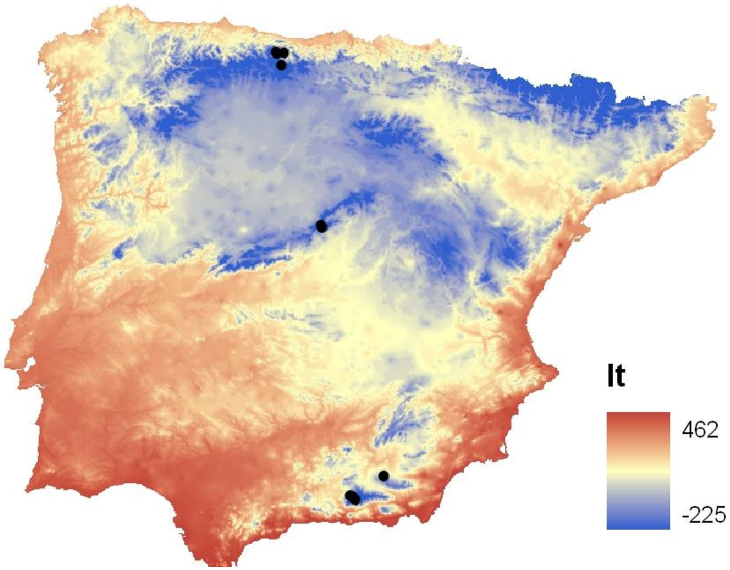


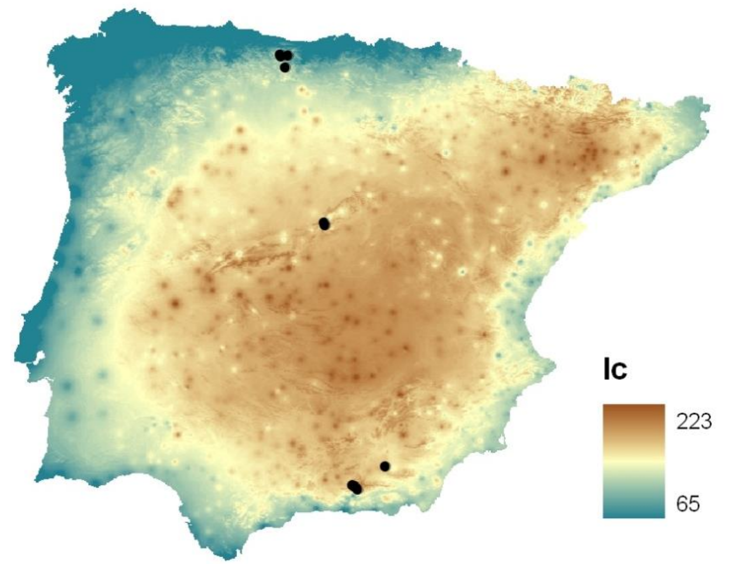
WORLDCLIM (www.worldclim.org)

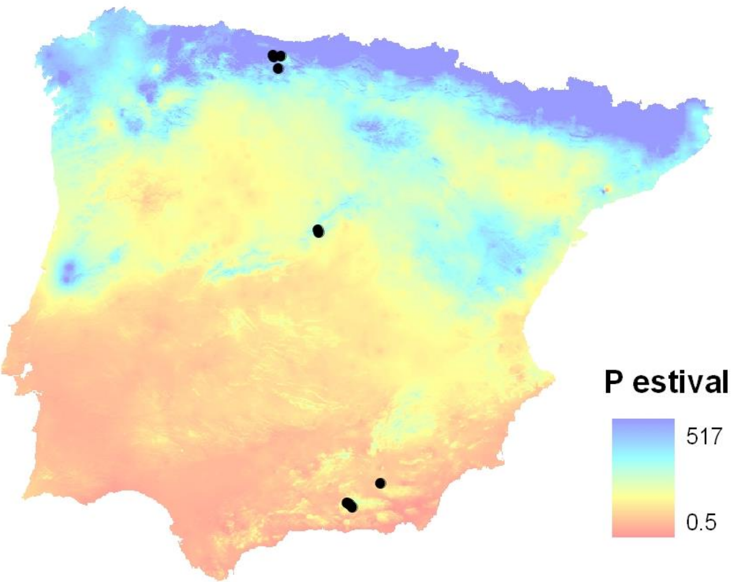
BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (P2/P7) (* 100)
BIO4 = T Seasonality (standard deviation *100)
BIO5 = Max T of Warmest Month
BIO6 = Min T of Coldest Month
BIO7 = T Annual Range (P5-P6)
BIO8 = Mean T of Wettest Quarter
BIO9 = Mean T of Driest Quarter
BIO10 = Mean T of Warmest Quarter
BIO11 = Mean T of Coldest Quarter
BIO12 = Annual Prec
BIO13 = Prec of Wettest Month
BIO14 = Prec of Driest Month
BIO15 = Prec Seasonality
BIO16 = Prec of Wettest Quarter
BIO17 = Prec of Driest Quarter
BIO18 = Prec of Warmest Quarter
BIO19 = Prec of Coldest Quarter

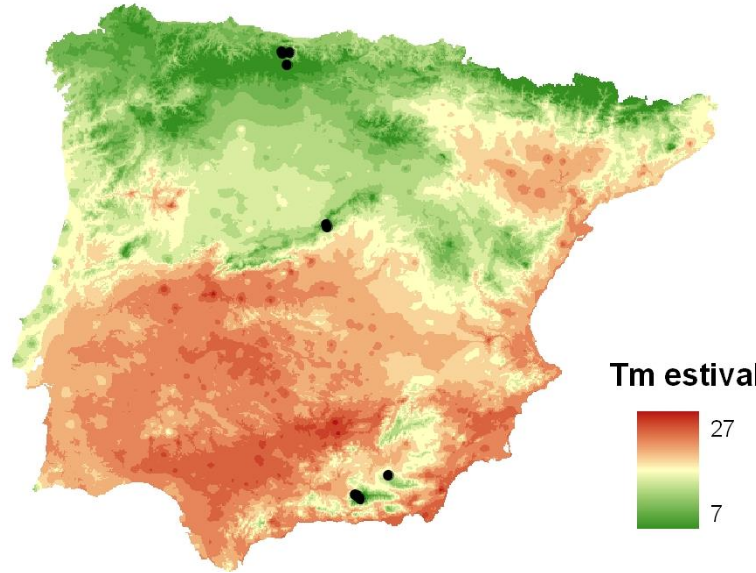


National data at higher resolution (e.g. Spain, 200m x 200m)



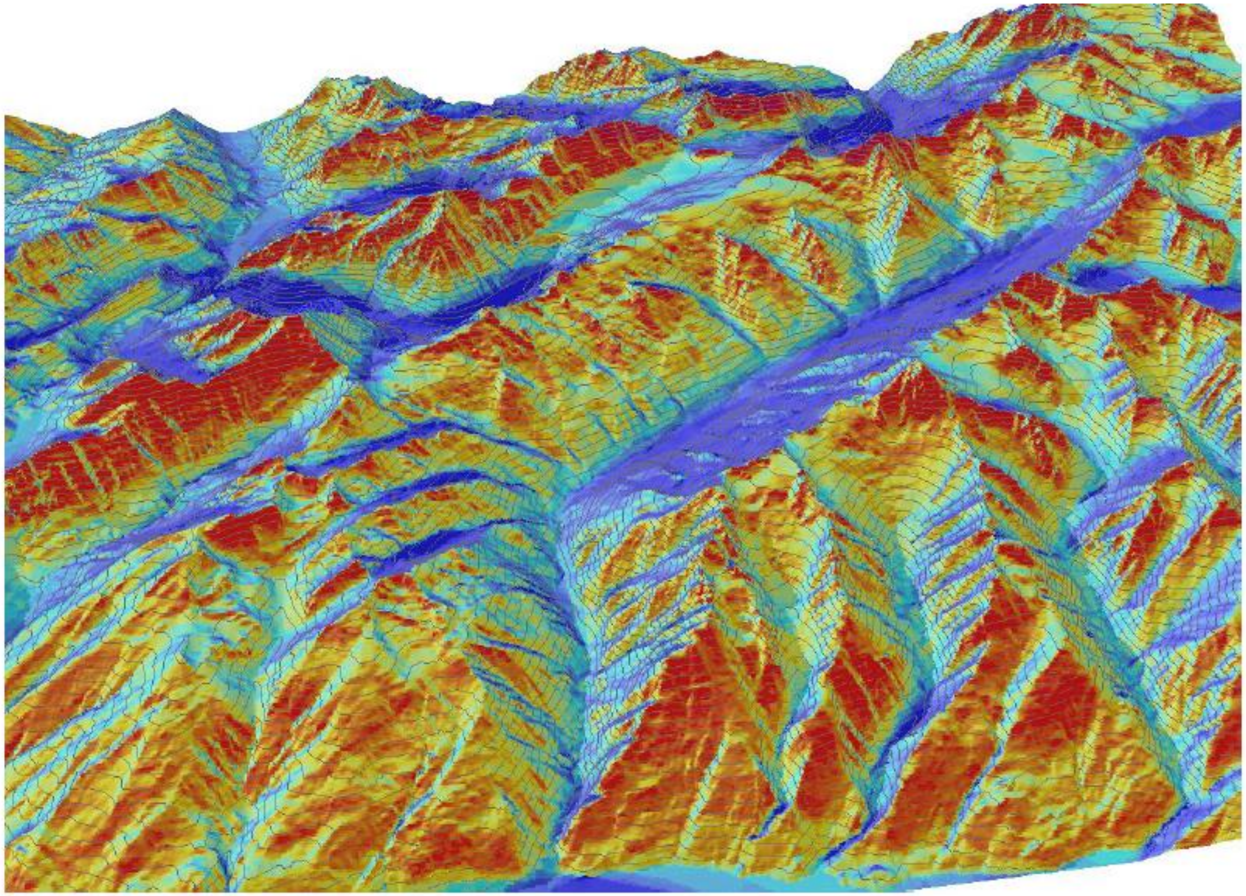






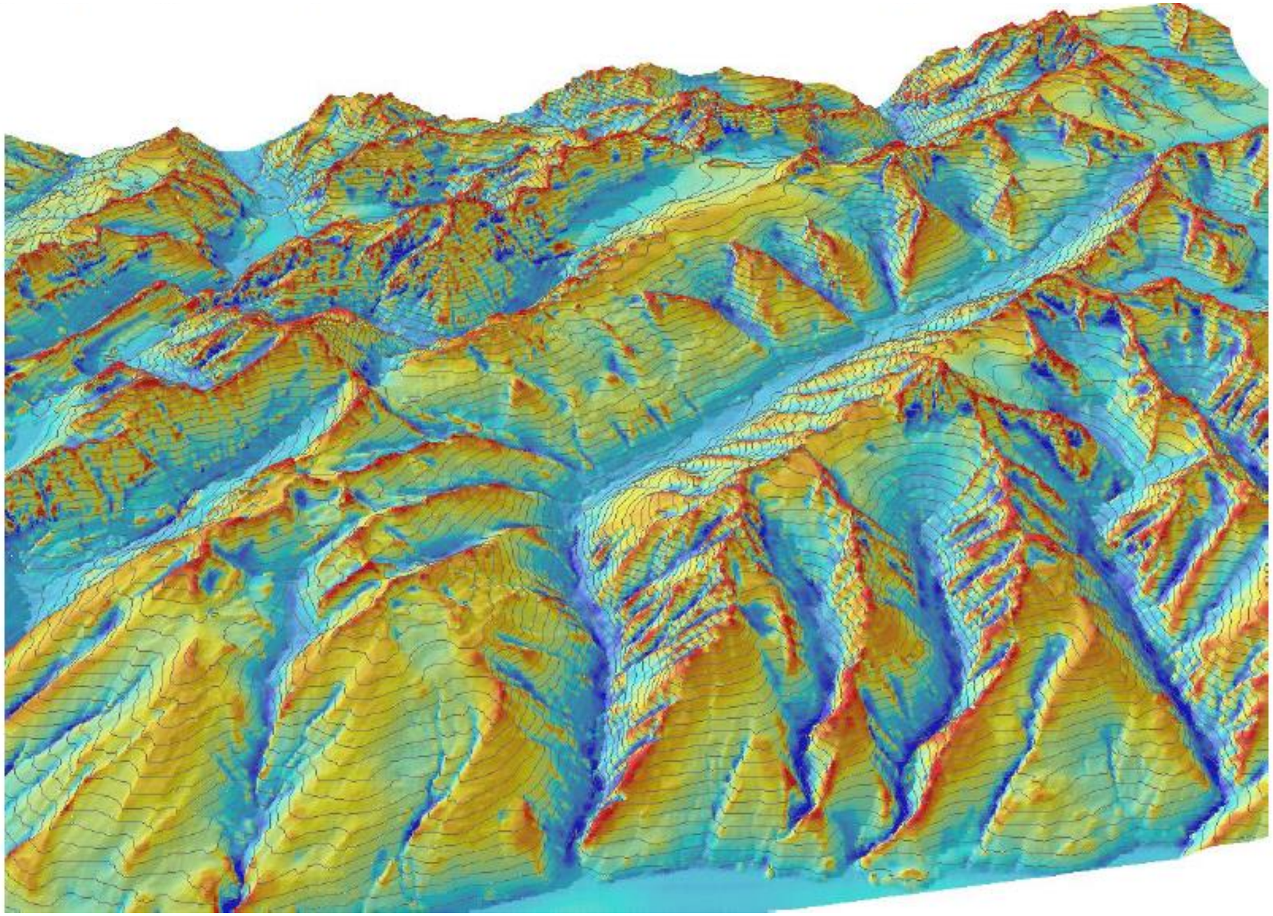
TOPOGRAHY

Potential solar radiation (from the Digital Elevation Model)
Indirect variable reflecting heat acumulation

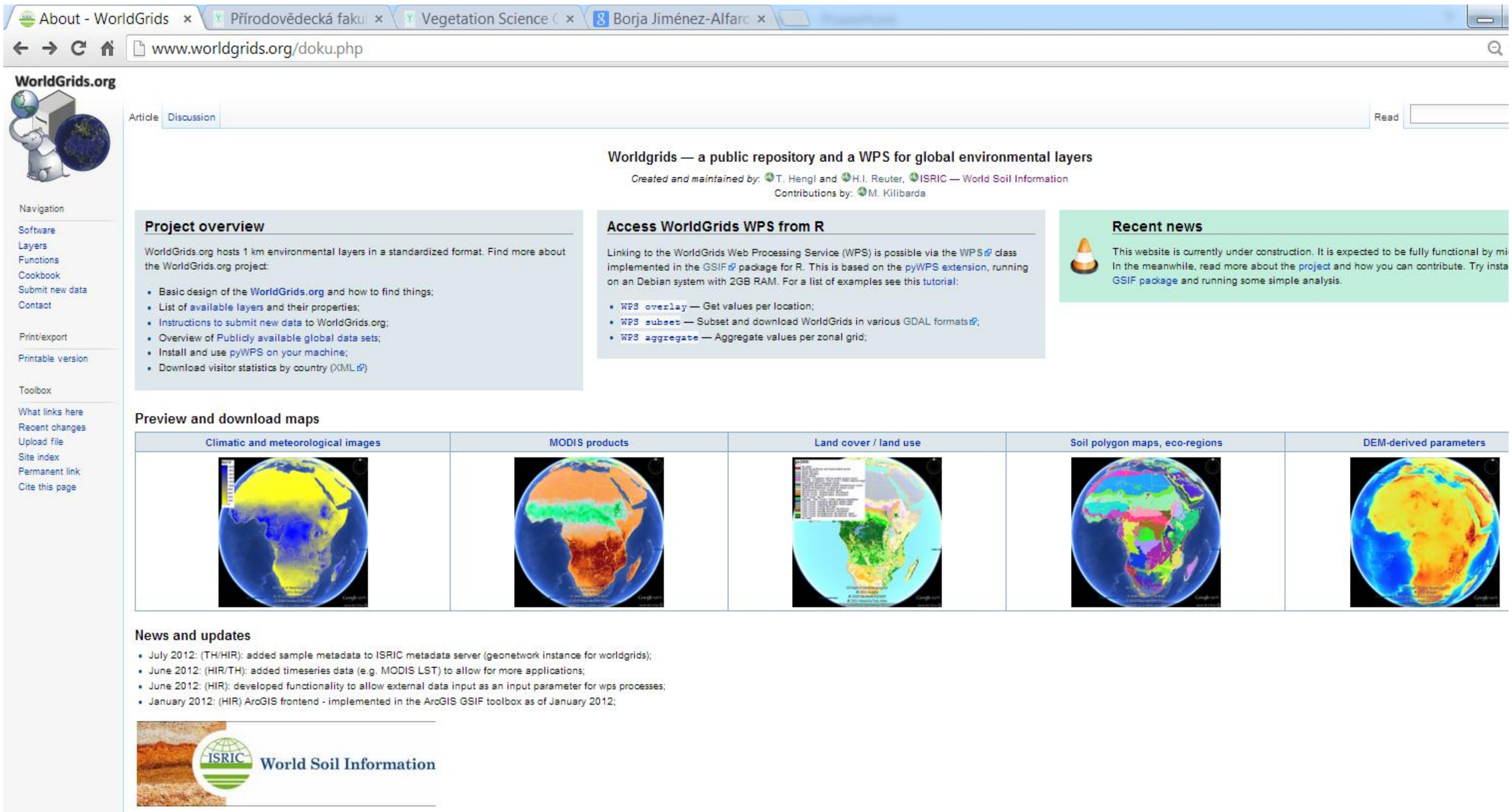


TOPOGRAHY

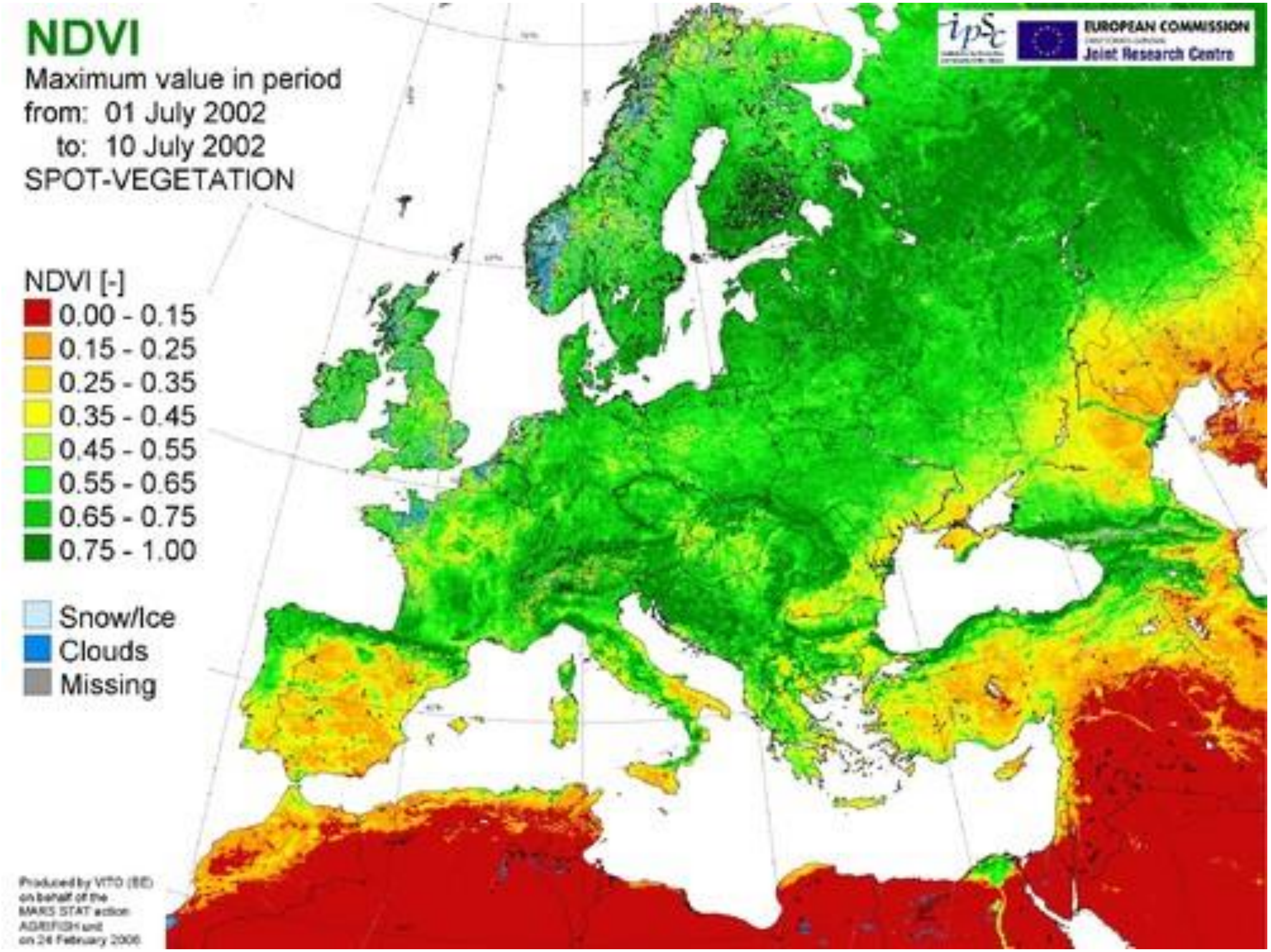
Topographic Position Index (from the Digital Elevation Model)
Indirect variable reflecting moisture and wind exposure



MORE VARIABLES (AT DIFFERENT SCALES)



MORE VARIABLES (AT DIFFERENT SCALES)

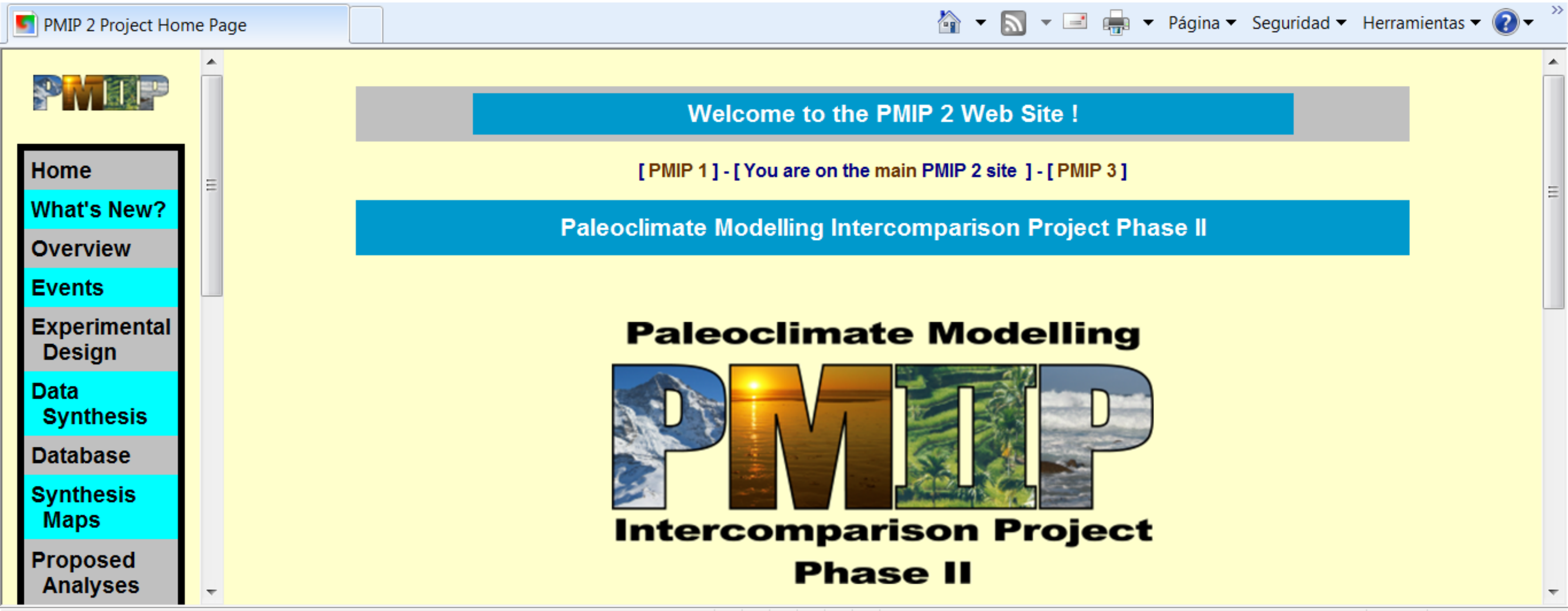


MORE VARIABLES (AT DIFFERENT SCALES)

A map of **land use** in Europe. Yellow: cropland and arable, light green: grassland and pasture, dark green: forest, light brown: tundra or bogs, unshaded areas: other (including towns and cities).

Paleoclimatic models

Last inter-glacial (LIG; ~120,000 - 140,000 years BP)
Mid-Holocene (~6000 BP)



Paleoclimatic models

Last glacial maximum (LGM; ~21,000 years BP)

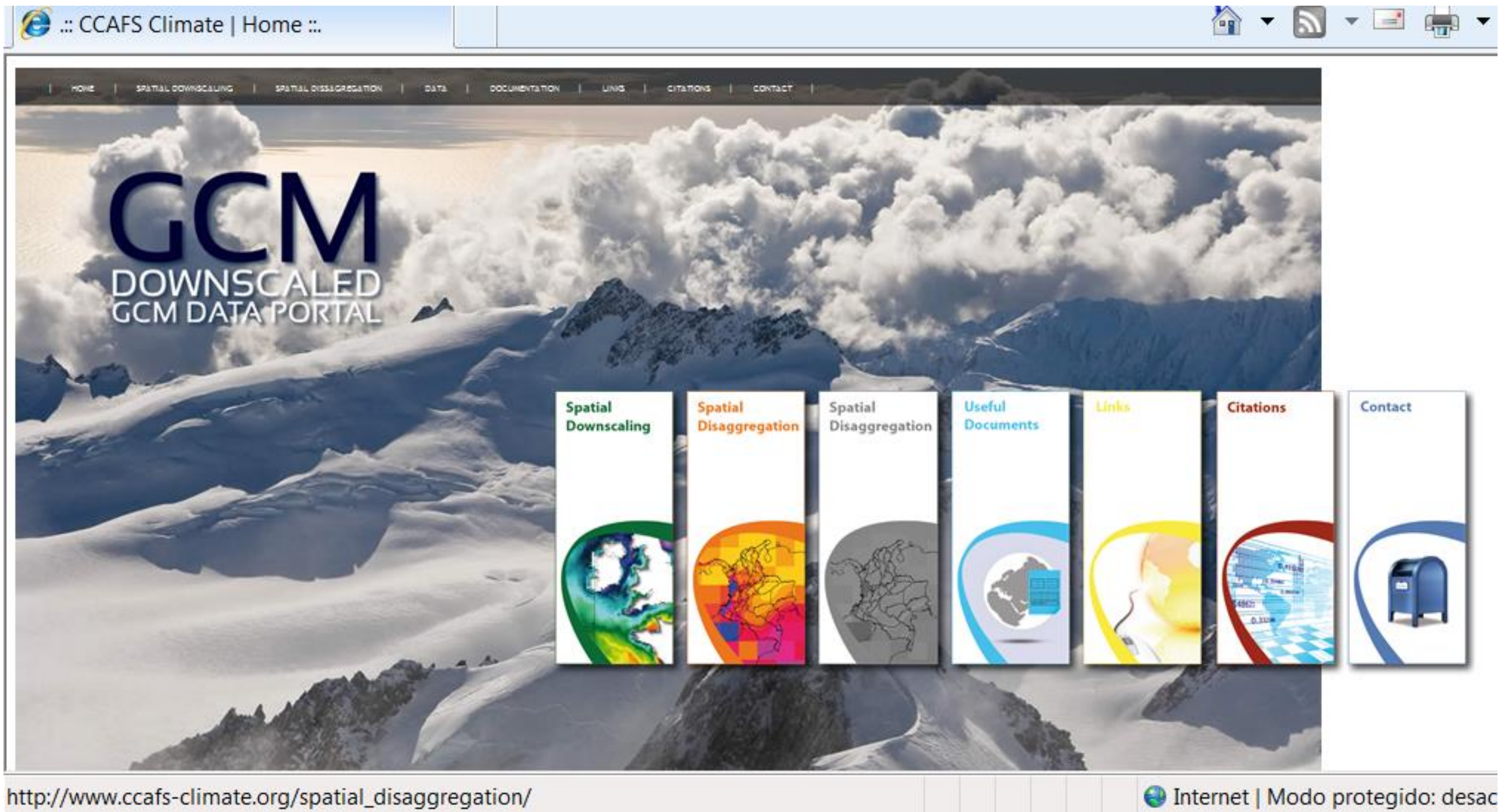
Science 24 March 2006:
Vol. 311 no. 5768 pp. 1751-1753
DOI: 10.1126/science.1120808

REPORT

Simulating Arctic Climate Warmth and Icefield Retreat in the Last Interglaciation

Bette L. Otto-Bliesner[1,*], Shawn J. Marshall[2], Jonathan T. Overpeck[3], Gifford H. Miller[4], Aixue Hu[1], CAPE Last Interglacial Project members

Climate future projections



NEXT WEEK

Part 1 – Mapping

Dealing with occurrence data
Environmental variables
Spatial issues and PRACTICE with GIS (October 18)

Part 2 – Modeling

Background theory (niche concept)
Modeling methods
Maximum Entropy and PRACTICE with MaxEnt

Part 3 – Mapping and Modeling

Model implementation and evaluation
Applications and future challenges
Using your own data and GROUP PRACTICE