Species Distribution Modeling (SDM) Workshop, Antananarivo

January 31st to Feb 2nd 2018

Instructors:

PRIFYSGOL BANGOR UNIVERSITY

Dr Alison Cameron

REBIOMA
REseau pour la BIOdiversité de MAdagascar

Dimby Razafimpahanana
Rija Rajaonson
Tsiky Rabetrano

Aknowledgements:

Richard Pearson, Ned Horning, Catherine Graham, Robert Anderson, Miguel Nakamura, Jesse O'Hanley, Eleanor Sterling, Chris Raxworthy, Wilfried Thuiller, Jane Elith, Bette Loiselle, Robert Hijmans, Town Peterson, Craig Moritz, Ken Kozak, Steven Phillips

and many others!

REPC MADAGASCAR

Réseau des Educateurs et Professionnels de la Conservat à Madagascar

GBIF

MACARTHUR

The John D. and Catherine T. MacArthur Foundation

Lecture 1
An introduction to niche-based species distribution modeling: theory and practice

Dr Alison Cameron

a.cameron@bangor.ac.uk

1884

PRIFYSGOL BANGOR UNIVERSITY

Some published uses of species' distribution models

Use	Example reference(s)
Species' invasions	*Vences et al, 2017. *Duttaphrynus melanostictus* in Madagascar *Kelly et al, 2014. Aquatic plants.
Reintroductions	Pearce and Lindenmayer 1998 Lentini et al, 2018 Kakapoo
Guiding field surveys to accelerate detection of unknown distributional areas and undiscovered species	Raxworthy 2003 Madagascar
Climate change	*Hannah et al. 2008, *Kelly et al, 2014. Aquatic plants Brown & Yoder 2015 Lemurs Madagascar
Species Delimitation	Raxworthy 2007 Madagascar
Conservation prioritization and reserve selection	*Kremen et. al. 2008. Madagascar
Impacts of land cover change	*Kelly et al, 2014. Aquatic plants Pearson et al. 2004
IUCN assessments	*Leach et al, 2015.

(This table has evolved from Guisan and Thuiller 2005)

The main steps to build & validate a species distribution model (SDM)

L2

Observed species' distribution
(a list of localities where the species has been observed, and sometimes also localities where the species is known to be absent)

L2

Database of 'raw' environmental variables (e.g. temperature, precipitation, soil type).

Data usually stored in a GIS

Processing to generate variables of importance in defining species' distributions (e.g. maximum daily temperature, frost days, soil water balance)

L3 & P1, P2

Modeling algorithm
e.g. Maxent, GARP, BioClim, Domain, artificial neural network, generalized linear model, regression tree

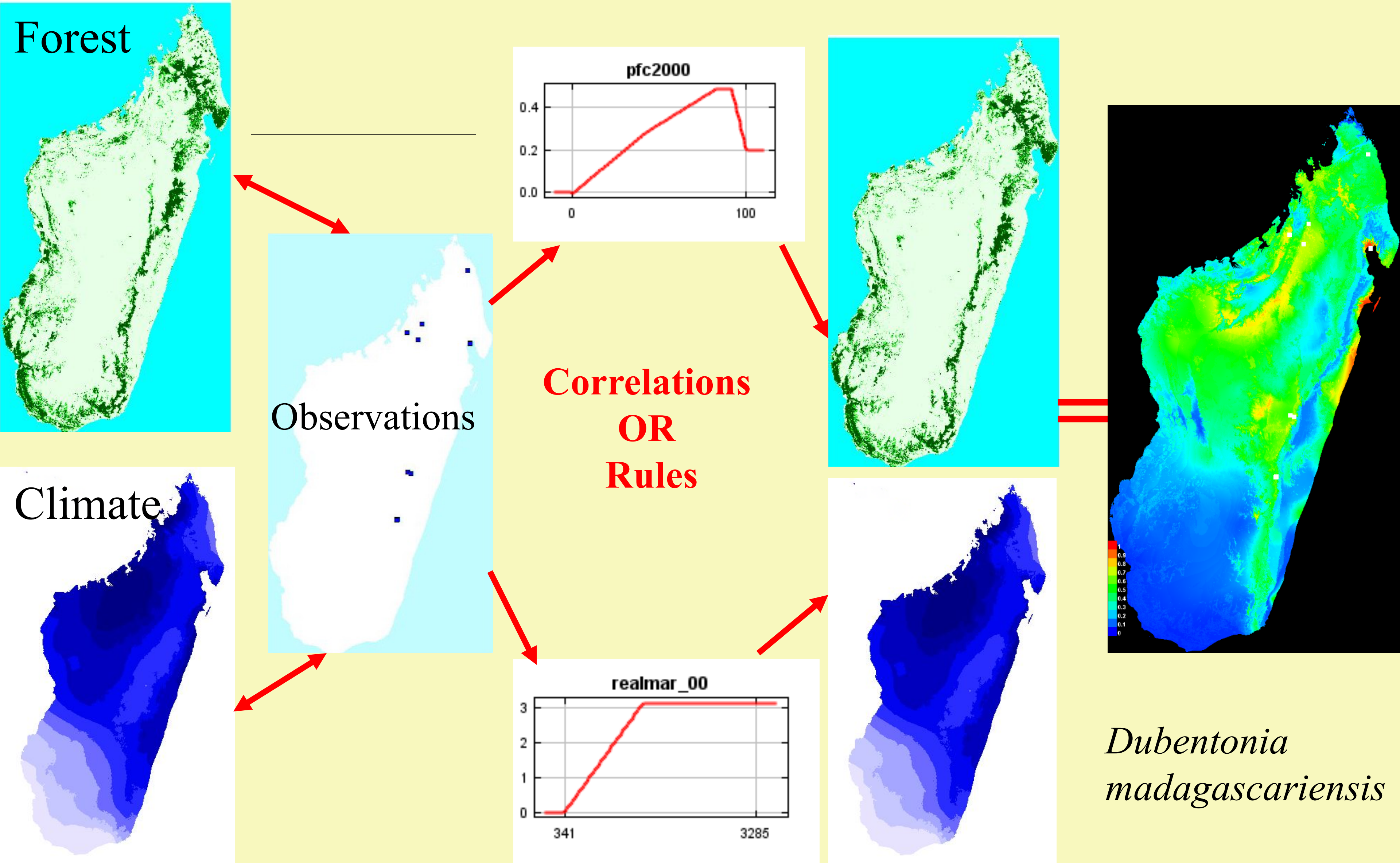
Model testing
(statistical assessment of predictive ability, using test such as AUC or Kappa)

L4, L5, P1, P2

Predicted species' distribution.
Prediction may be for a different region (e.g. for an invasive species) or for a different time period (e.g. under future climate change)

Case Studies

The main steps to build ~~& validate~~ a species distribution model (SDM)



Workshop Agenda

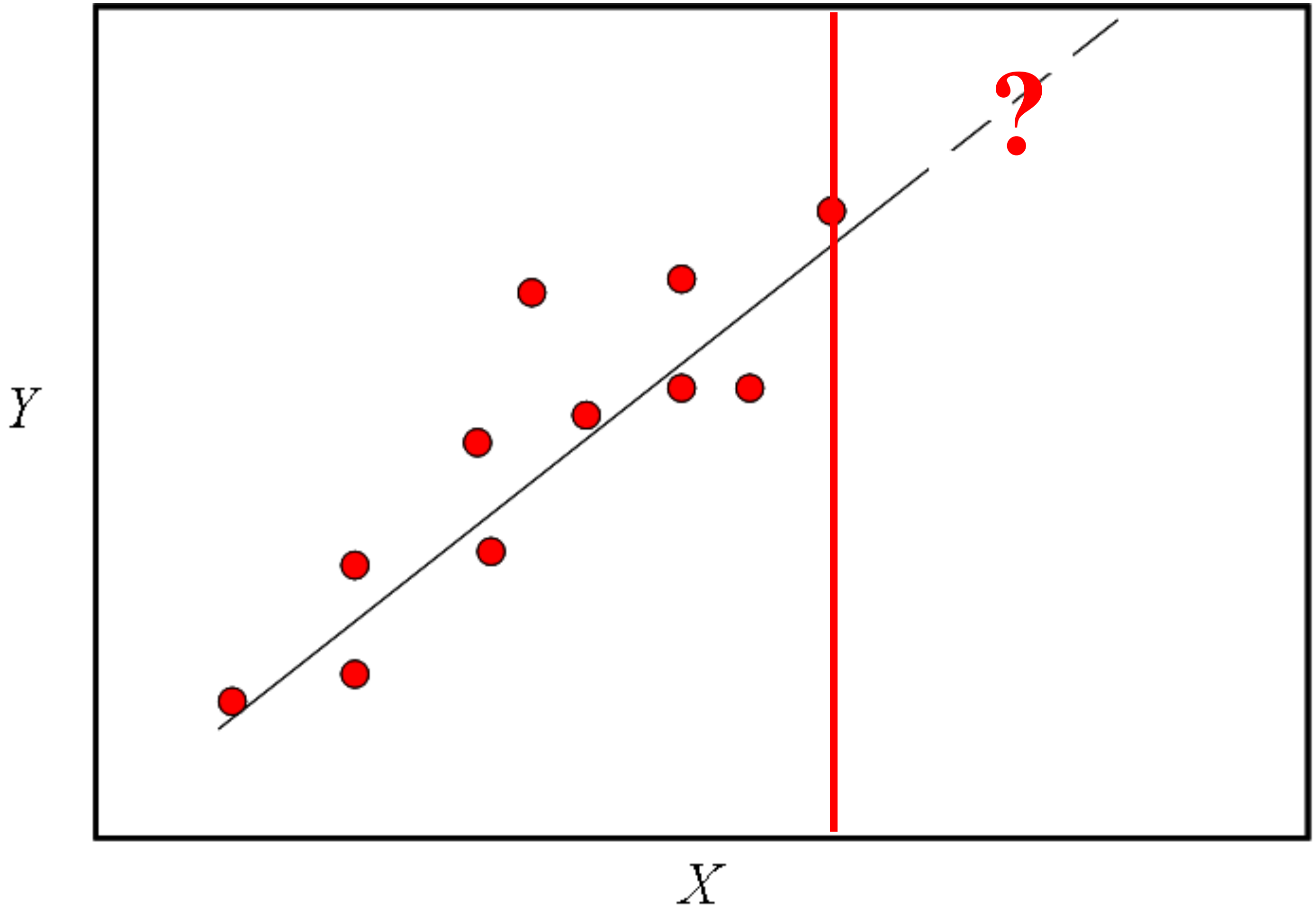
	9 – 9.45	9.45 - 10.30		11 – 11.45	11.45 –12.30		1.30 – 2.15	2.15 - 3		3.30 - 4.15	4.15 – 5.00
Wednesday 31st Jan	Introduction to AIA and Workshop objectives Brian	**L1: Intro** Overview of principles & applications of ENM/ SDM. Alison	BREAK	**Discussion** >Individual goals >Choose Case Studies for Friday Alison	**L2: Data** >Sources >Manag't >Quality >Geo-referencing Tsiky & Dimby	LUNCH	**Intro to Practical 1** Data & Exploration Modelling >DIVA >Other tools Alison	**Practical 1** Data evaluation using DIVA REBIOMA	BREAK	**Practical 1** BIOCLIM & Model Evaluation Using DIVA REBIOMA	**Flexi time** Work on your own data, read the case study papers, or discuss in groups
Thursday 1st Feb	**L3: Modelling Algorithms** Alison	**L4: Model Evaluation** Alison		**Intro to Practical 2** Maxent Alison	**Practical 2:** Modelling Ants Using Maxent GUI REBIOMA		**Practical 2:** Discussion about outputs REBIOMA	**Practical 2:** Using Maxent's command line REBIOMA		**Case Study 1** Conservation Planning >GAP analysis >Terrestrial >Marine Alison & Dimby	**Flexi time**
Friday 2nd Feb	**L5: Thresholds & Overp-rediction** Alison	**Case Study 2** TBC Alison		**Case Study 3** TBC Alison	Case Study Discussion Alison		**CLINIC** Advance your skills, plan your modelling project, and/or work on your own project Alison, Dimby, Tsiky			Debate on current BID progress and also the African Malaise Transect. Brian	Close and Photos Brian

Approaches:
Correlative vs. Mechanistic

- Assume current distribution gives a good indicator of ecological requirements

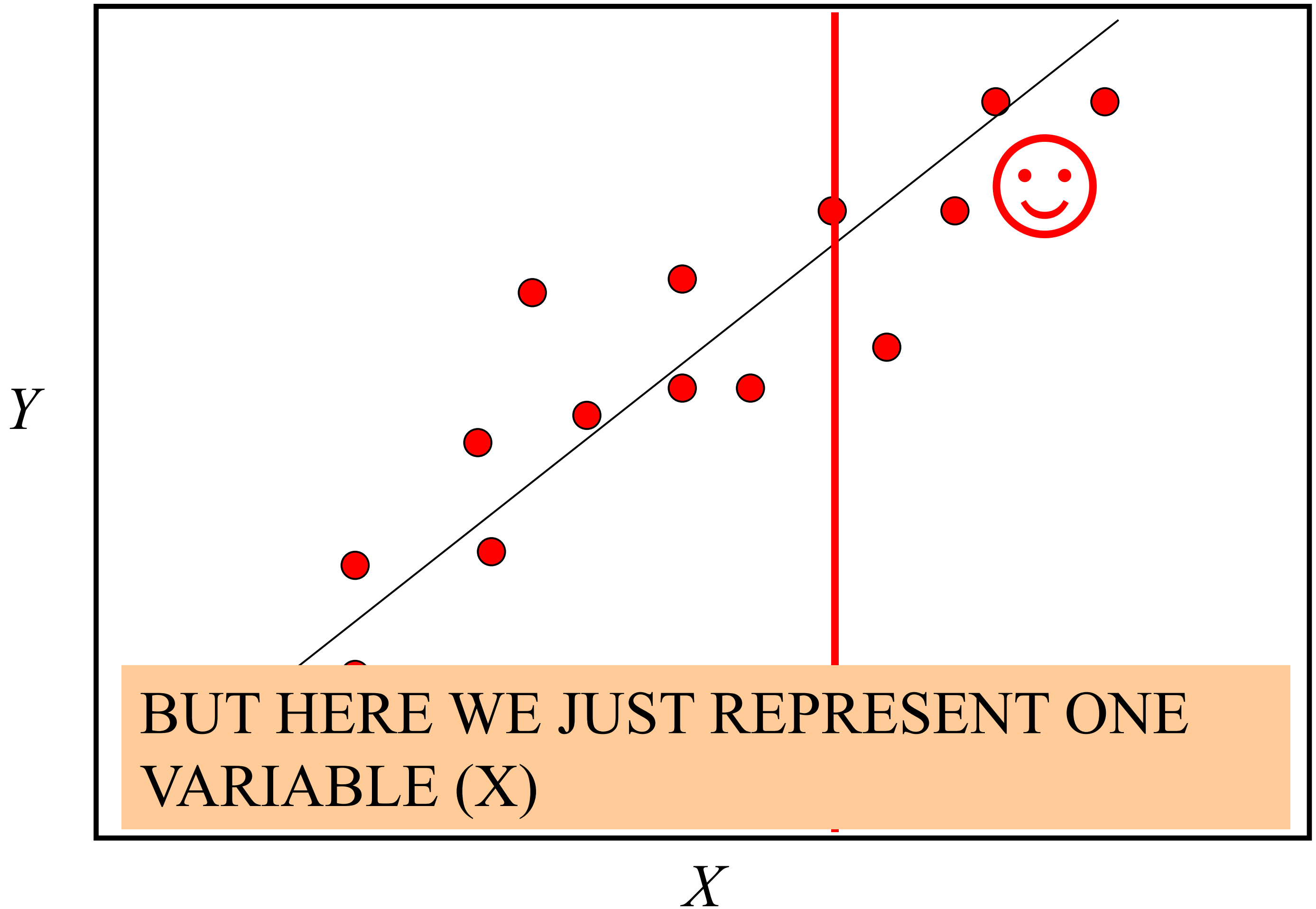
- Do not rely on observed occurrence records
- Require detailed physiological data

e.g. simple linear regression



X = Abiotic variable, Y = Frequency of species presence, or relative abundance

e.g. simple linear regression



X = Abiotic variable, Y = Frequency of species presence, or relative abundance

Generally: the "niche" is the term used to describe *multiple* conditions that limit a species distribution, and within which viable populations can be maintained.

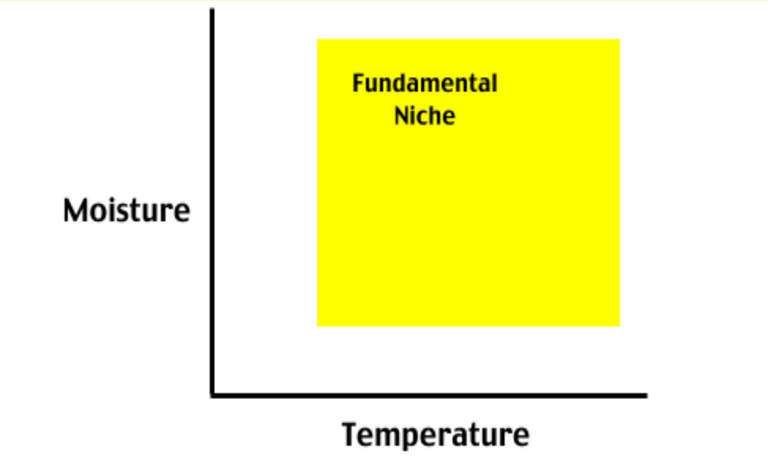
To complicate things there are two domains in which we must consider species' distributions: ecological & geographical

Ecological: The function or position of an organism or population within an ecological community

Geographical: The particular area within a habitat occupied by an organism

G. Evelyn Hutchinson (1944-58)

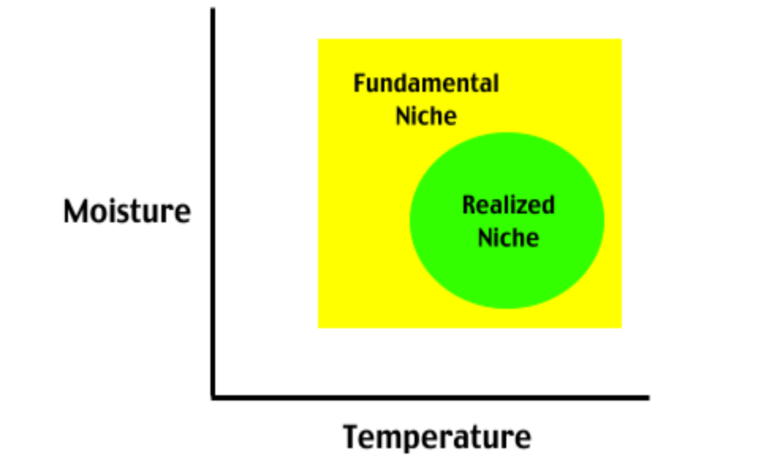
"The term niche... defined as the sum of all the environmental factors acting on the organism; the niche is thus a region of an *n*-dimensional hyper-space..." (1944).



"…if this procedure could be carried out with all X_n variables, both physical and biological, the fundamental niche of species would completely define its ecological properties.

G. Evelyn Hutchinson (1944-58)

However: **The range of conditions in which a species could feasibly live are often greater than those where it actually lives, and this is typically caused by biotic interactions.**



Fundamental niche**: all aspects of the** ***n*****-dimensional hypervolume in the absence of other species.**

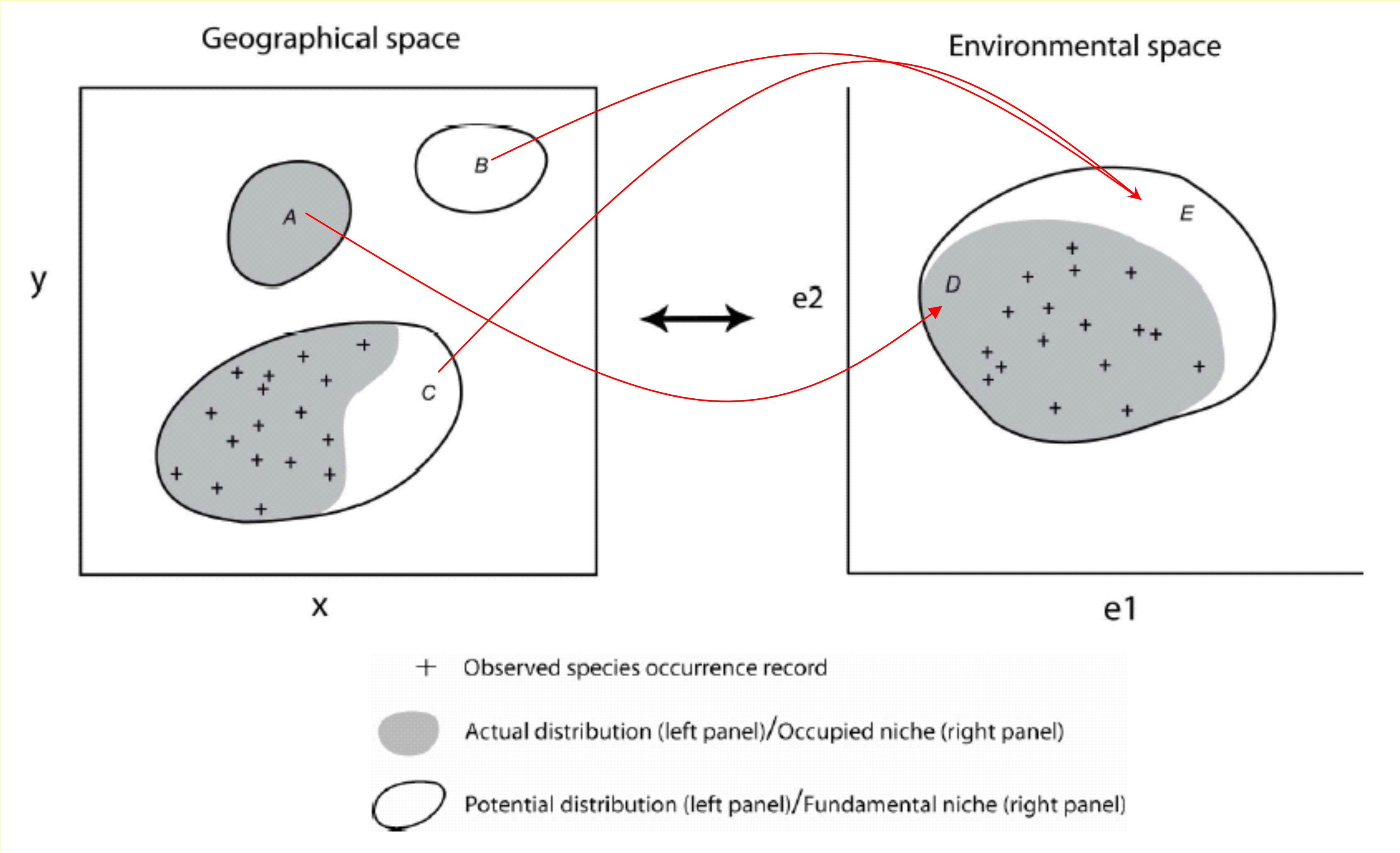
Realised niche**: the part of the fundamental niche to which the species was restricted due to interspecific interactions.**

What niche concept should we use?

- Hutchinson's **fundamental niche** : a set of abiotic parameters describing the area where the species can persist
  - e.g. temperature, precipitation, soil type – often limited by physiology
- A **colonizable or realised niche**: the set of environmental conditions where both biotic and abiotic conditions are favorable
  - e.g. competition, food availability, nesting sites – often limited by life history and behaviour
- A **occupied niche** or, the proportion of the colonisable niche where the species actually occurs
  - e.g. biogeograpical barriers – often limited by dispersal abilities

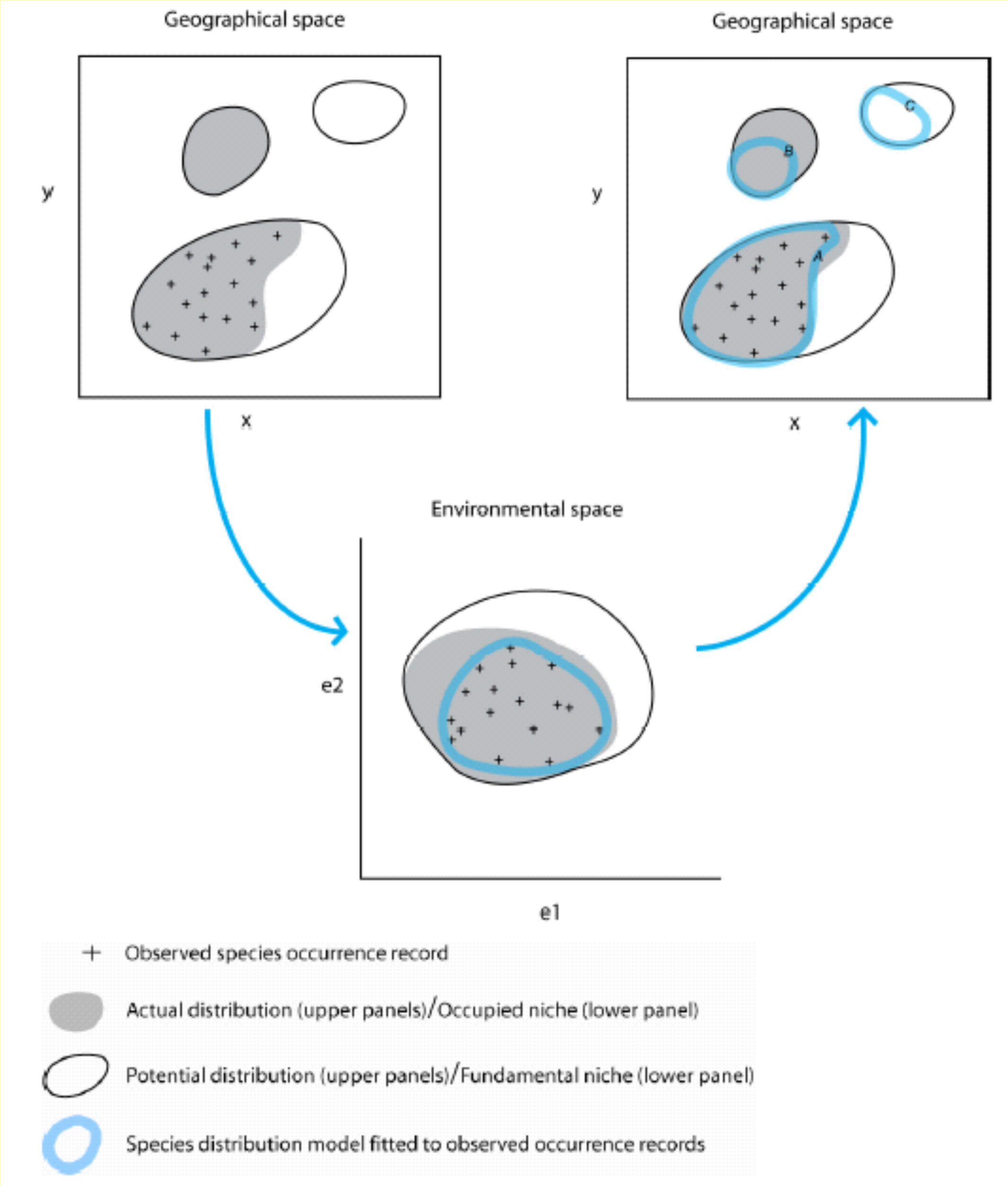
It depends on your question, and you need to think about them all.

The relationship between species' position in environmental space and geographical space



(see Pearson NCEP/REPC module 2007)

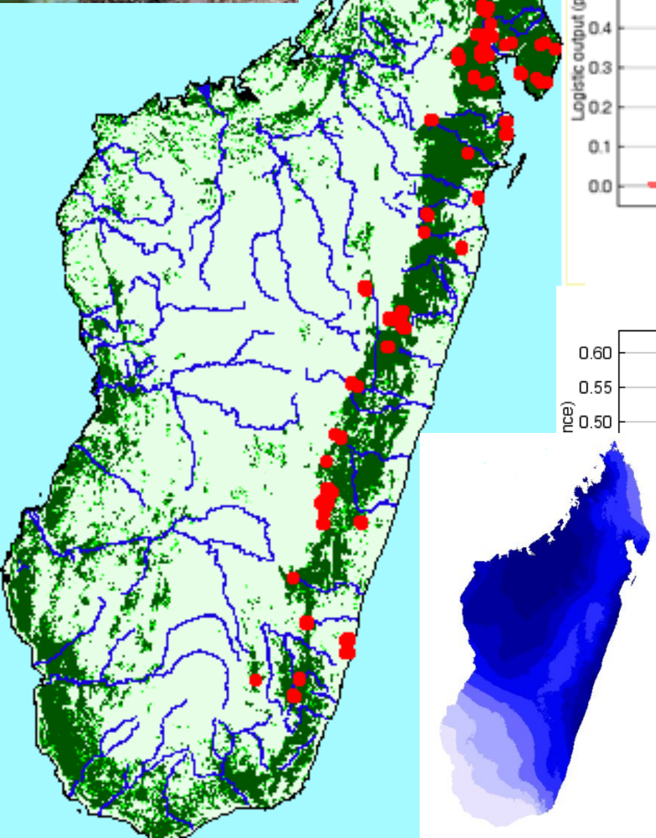
The general species' distribution modeling approach



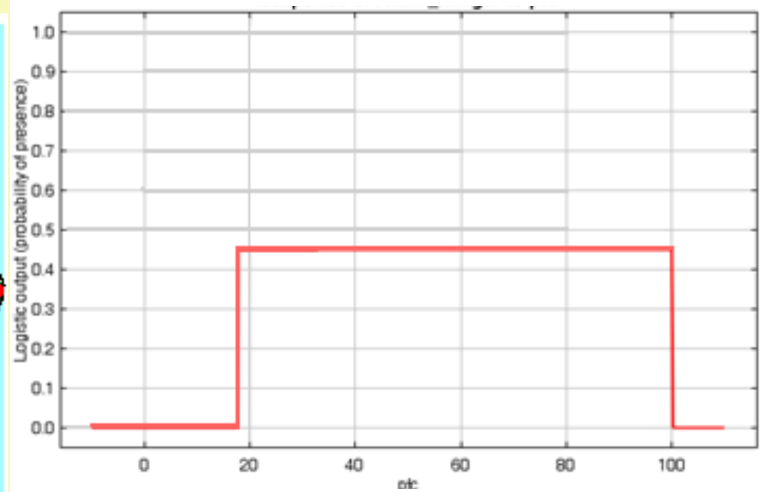
(see Pearson NCEP/REPC module 2007)

Maximum Entropy (MAXENT) Phillips *et al.*

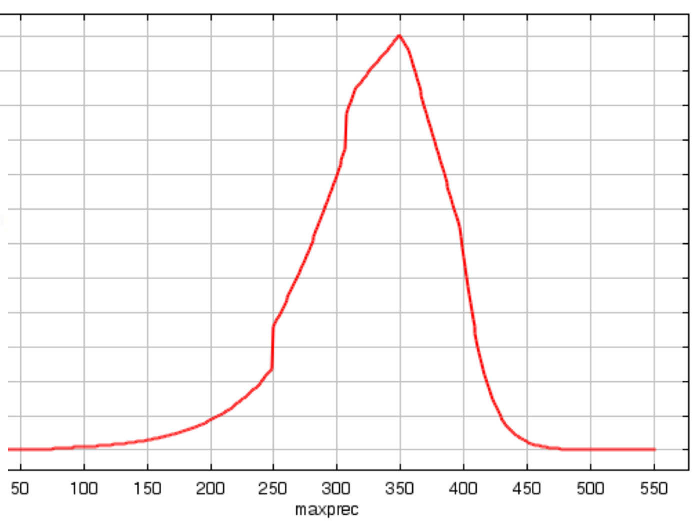
Geographical Space → **Environmental Space** → **Geographical Space**

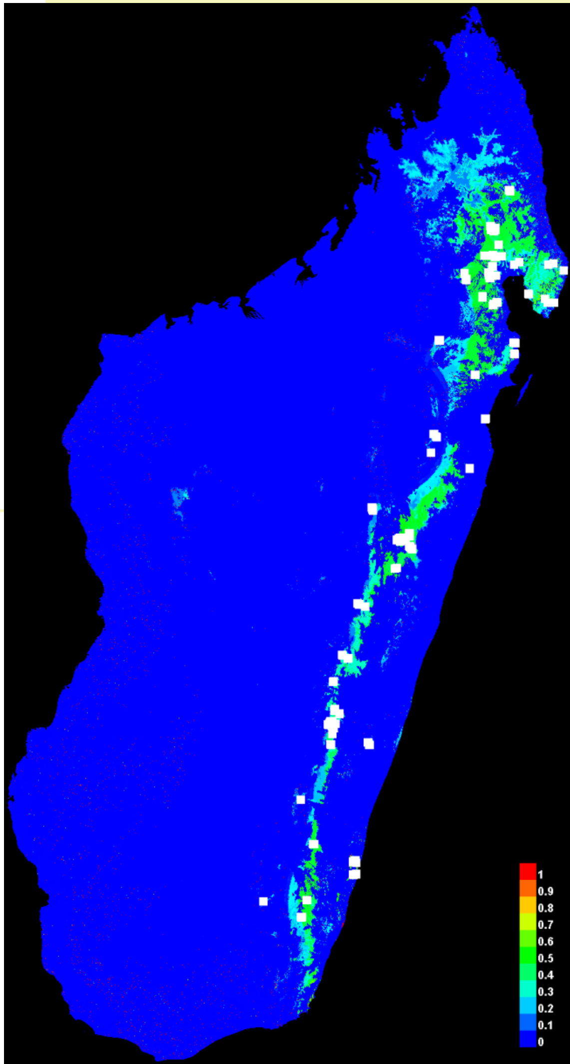


% FOREST COVER



MEAN ANNUAL RAINFALL





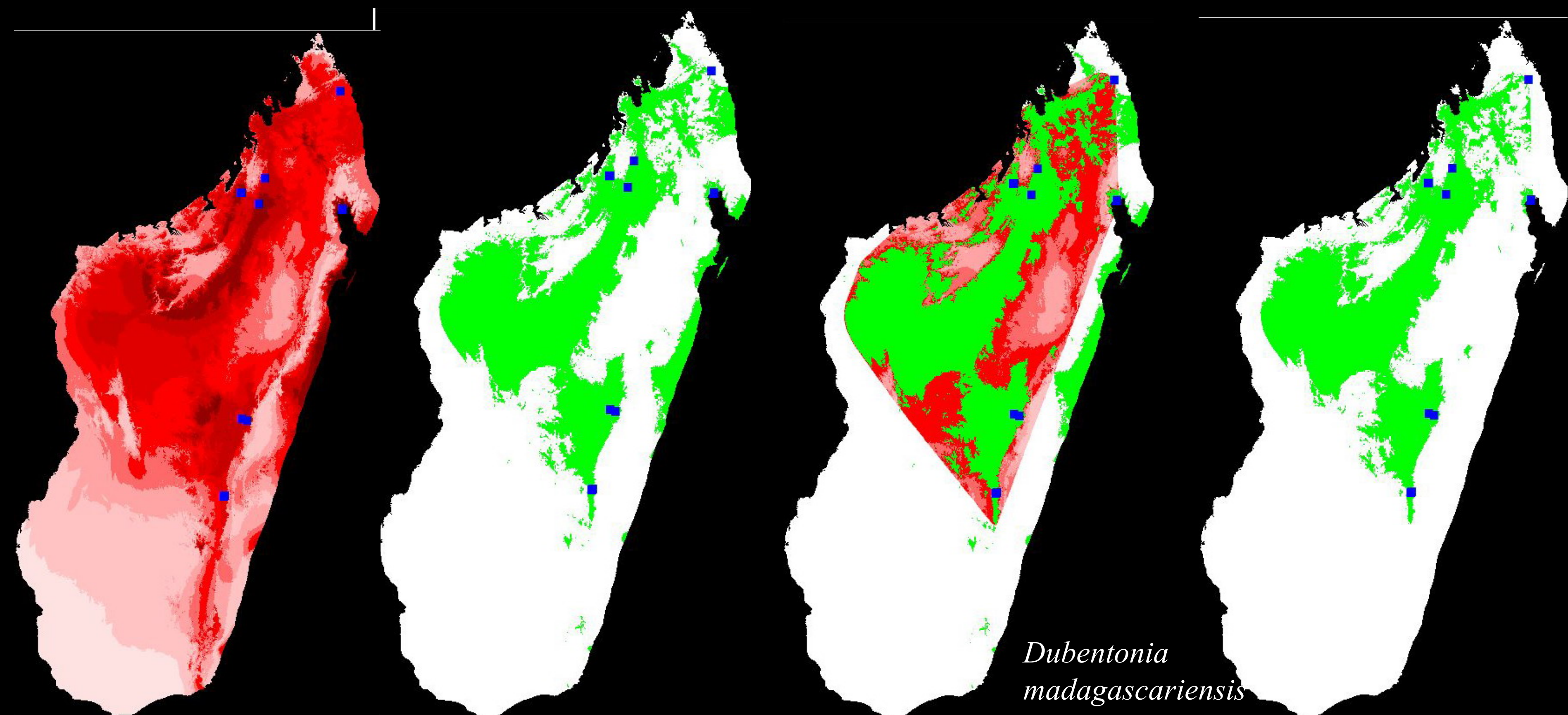
If you understand the niche: you can improve your model to suit your needs

a. Continuous prediction (Maxent) of a fundamental niche?

b. Threshold fundamental niche (presence/absence)

c. Incorporate dispersal limitations.

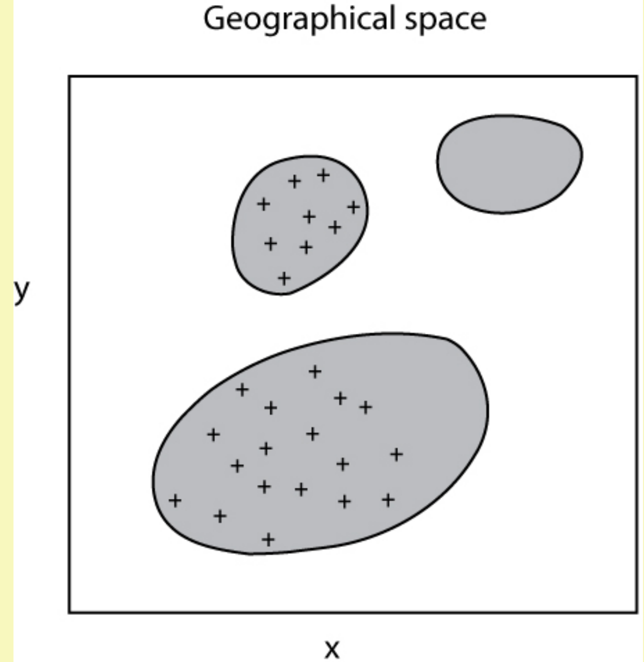
d. Realised niche? Occupied niche?

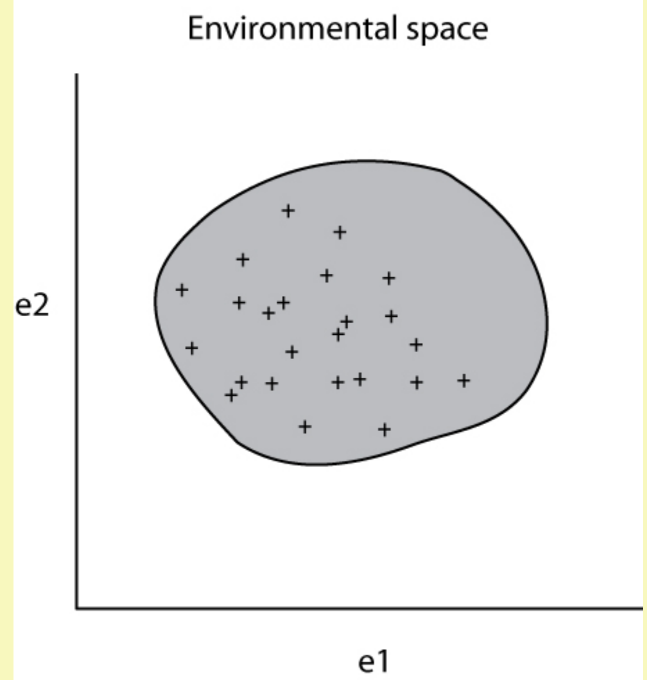


Two key factors determine the degree to which observed localities can be used to estimate the niche or distribution:

- ***Equilibrium*: A species is said to be at equilibrium with current environmental conditions if it occurs in all suitable areas, whilst being absent from all unsuitable areas. The degree to which a species is at equilibrium depends both on biotic interactions (e.g. competitive exclusion from an area) and dispersal ability.**

- ***Sampling adequacy*: The extent to which the observed occurrence records provide a sample of the environmental space**





e.g. Equilibrium between climate and species' distributions

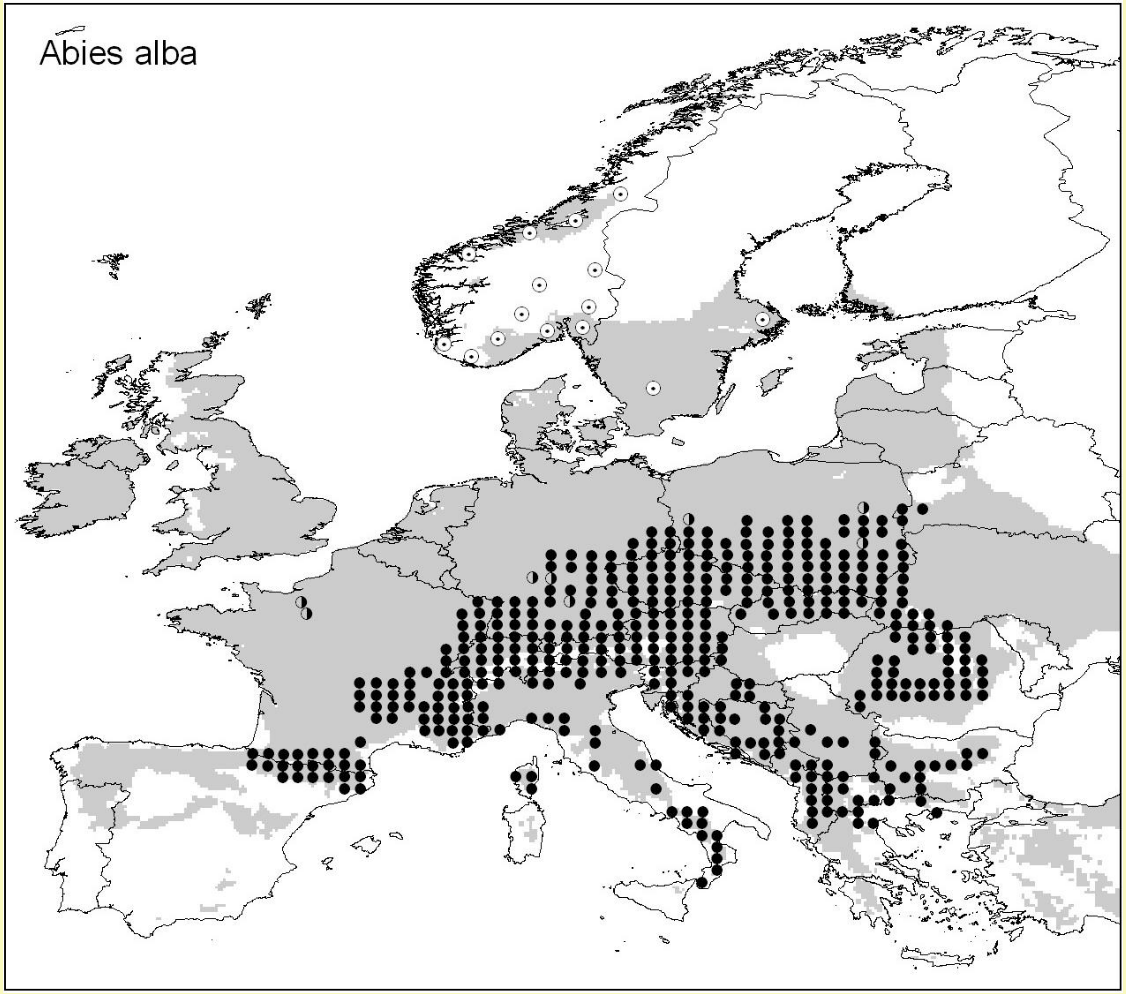
Degree of range filling (Svenning and Skov 2004, *Ecol. Letters*):

Mean realized/potential

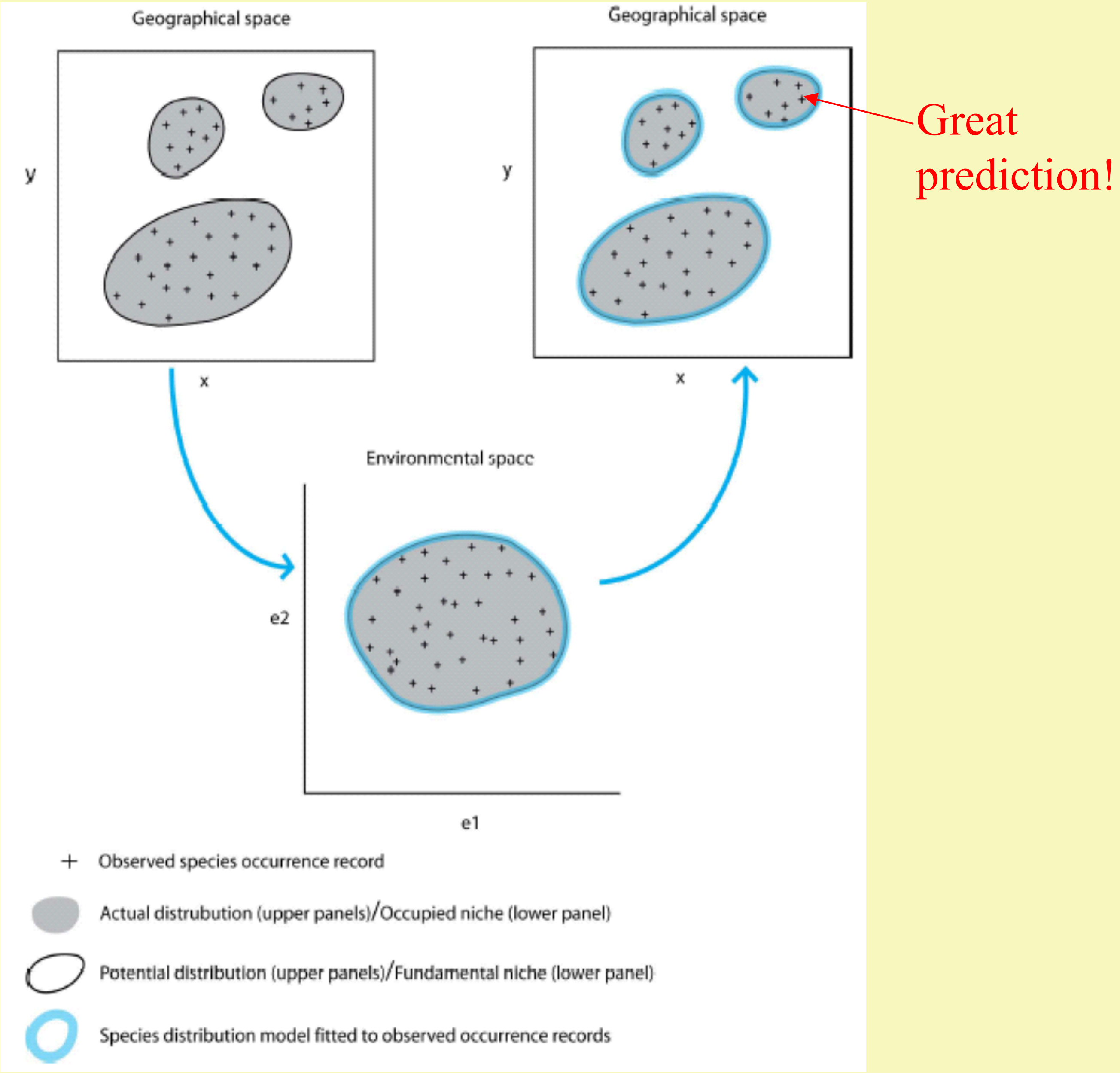
range size ratio

for 55 tree species

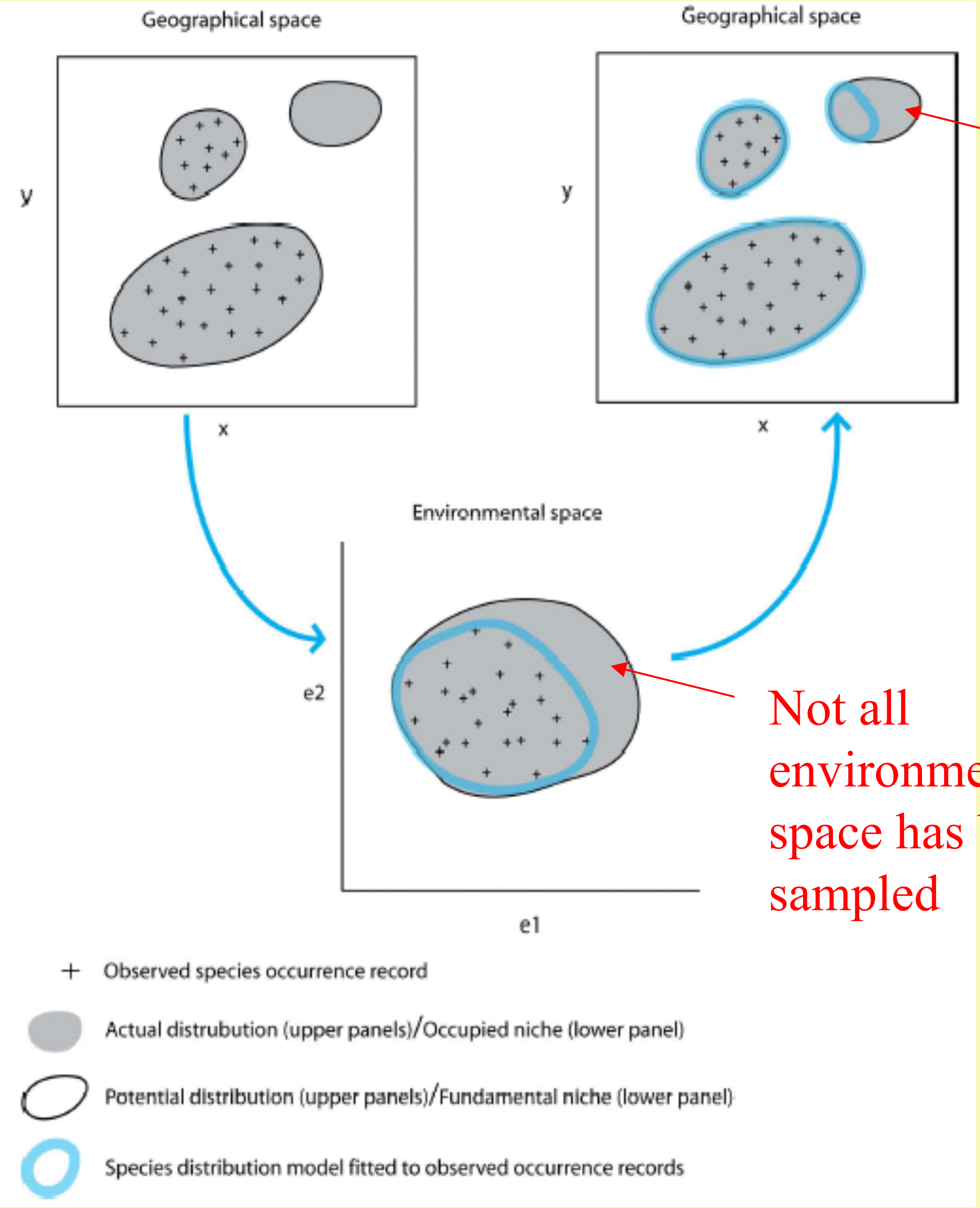
in Europe = 38.3%



Suppose high equilibrium & excellent sampling



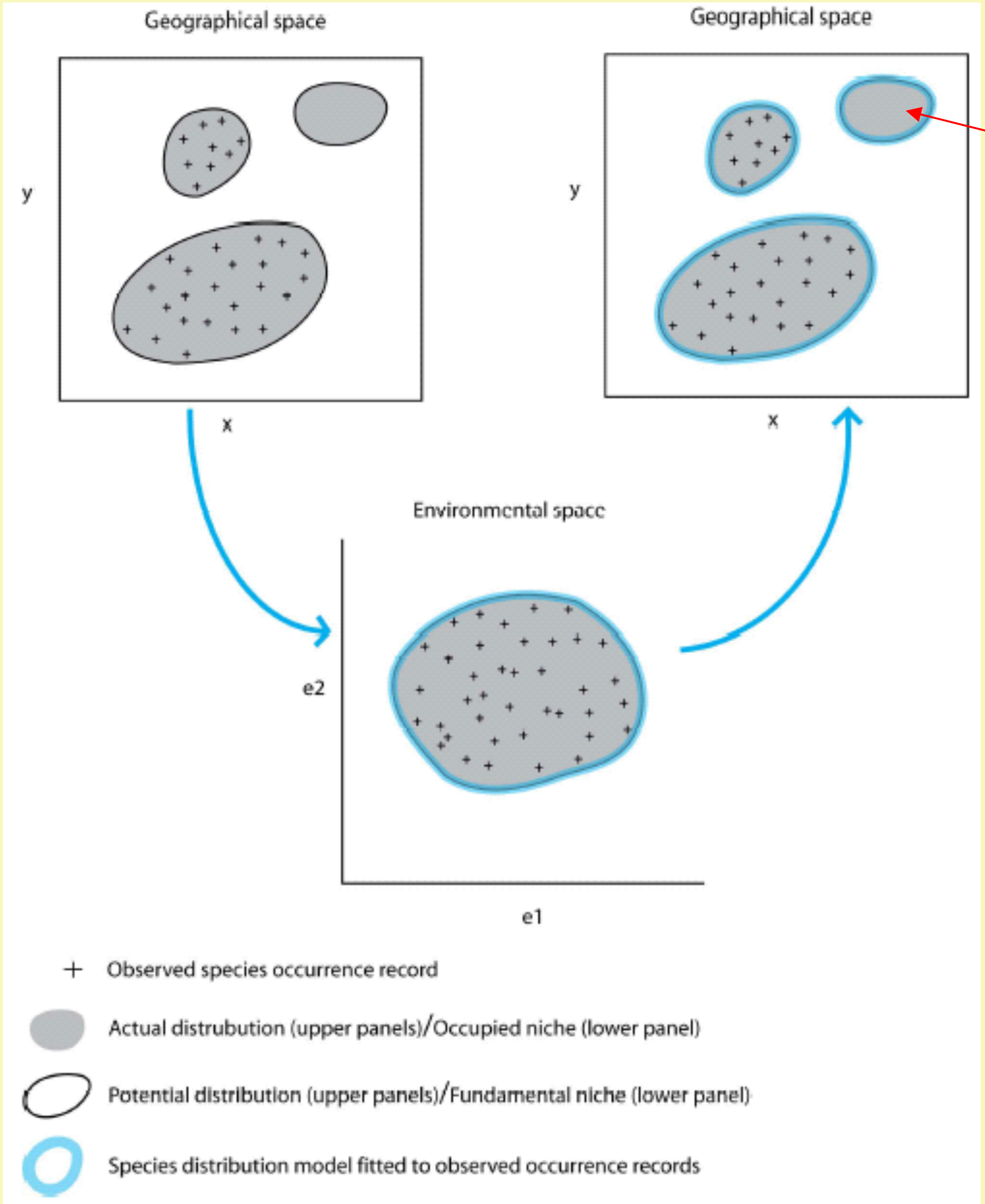
Suppose high equilibrium: but poor sampling in geographical space



So, you might not predict all of the distribution

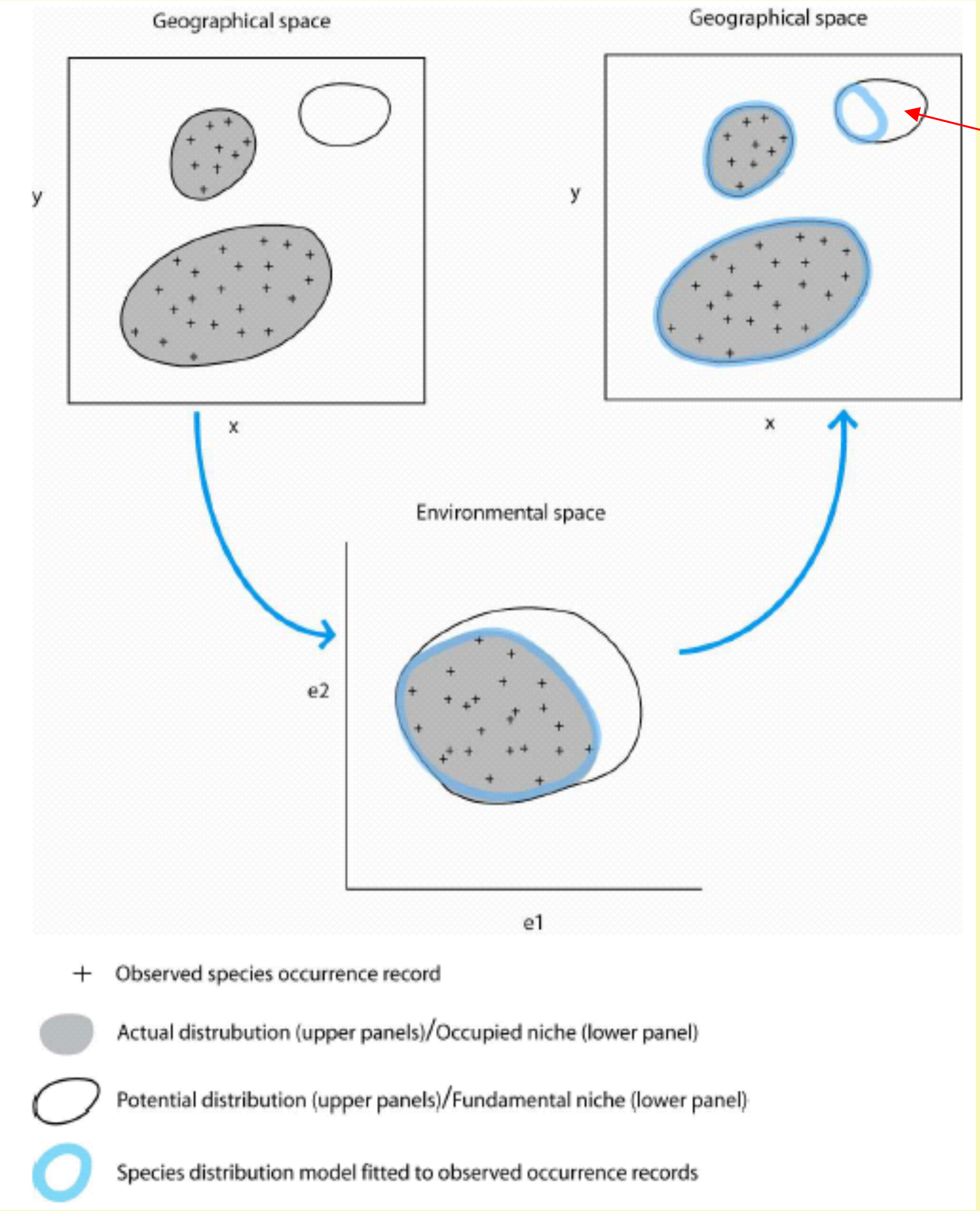
Not all environmental space has been sampled

Suppose high equilibrium, and poor sampling in geographical space, but good sampling in environmental space



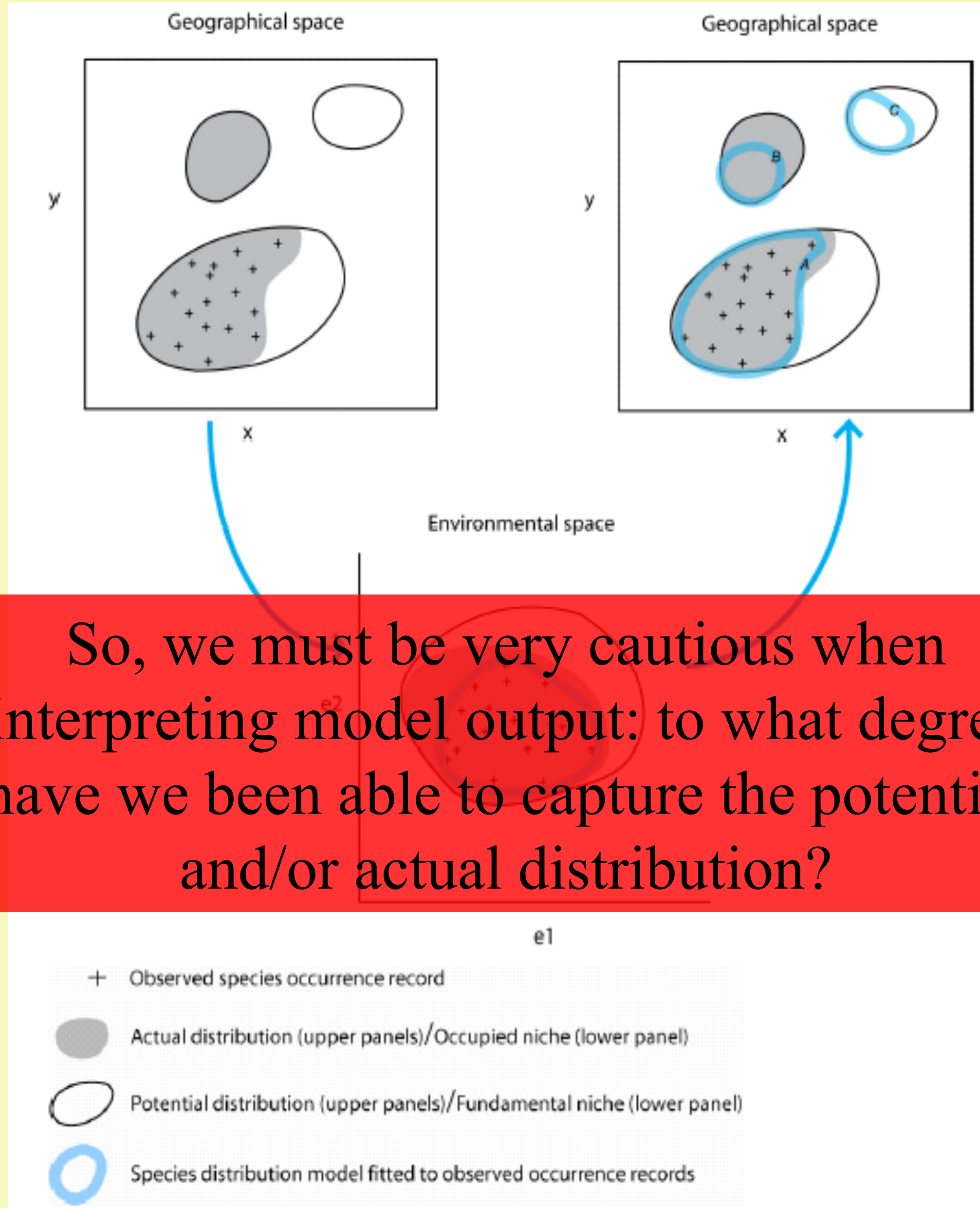
Great prediction!

Suppose low equilibrium but good sampling



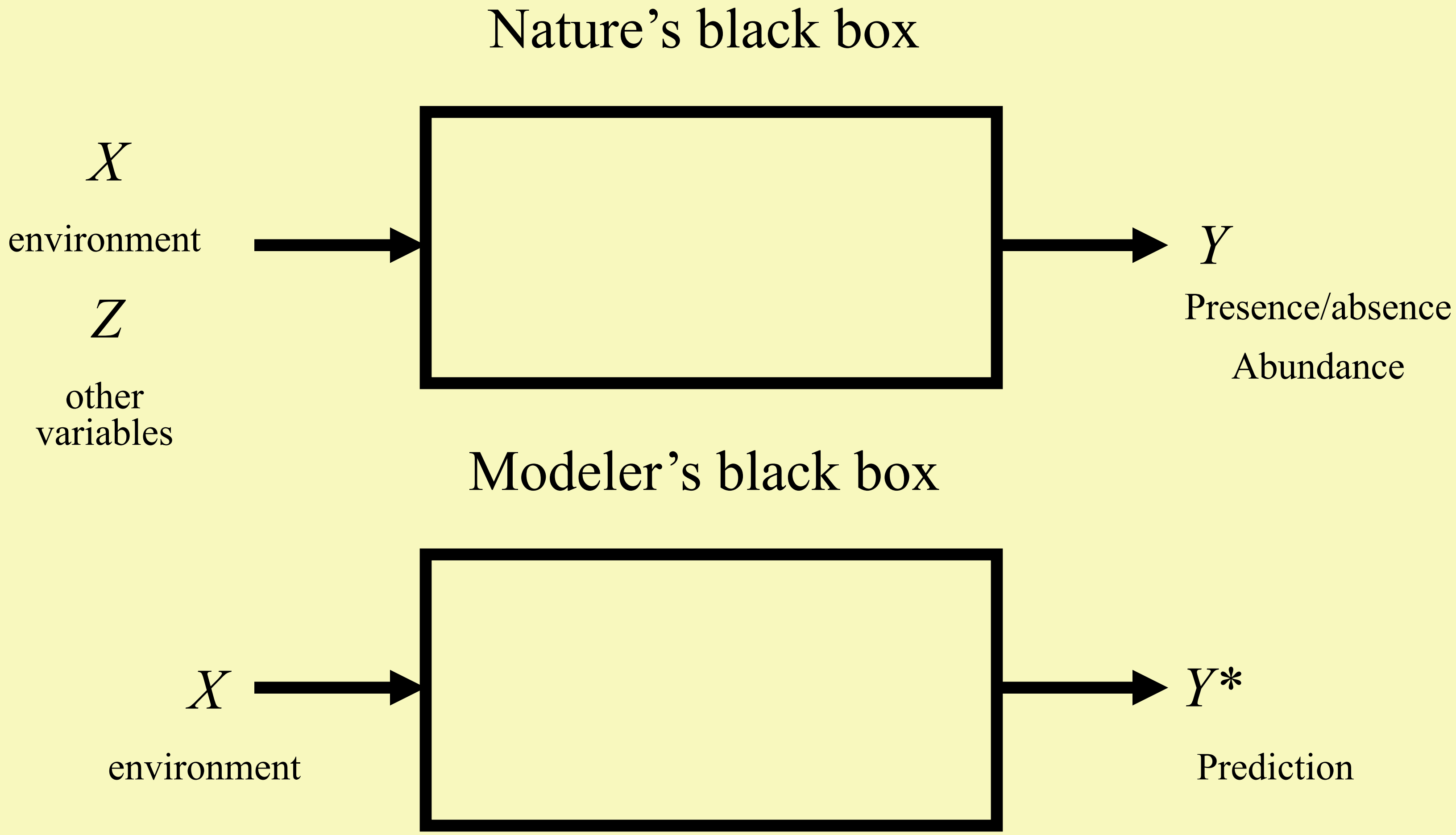
Area of “over prediction”

Back to our first example: in reality we have a combination of dis-equilibrium and incomplete sampling



So, we must be very cautious when interpreting model output: to what degree have we been able to capture the potential and/or actual distribution?

Another way to think about it:



Why Y and Y^* do not agree: (a) we do not know natures black box, (b) we do not use Z, (c) We use X although Nature may not.

The role of GIS

- The large datasets of biological and environmental data that are used in distribution modeling are ideally suited to being stored, viewed and manipulated in a GIS.
- Common tasks: visualisation for cleaning, changing projections, reformatting spatial resolution, and interpolating point locality data to a grid.
- GIS is also crucial for visualizing model results and carrying out additional processing of model output.

The role of GIS

- However, the distribution modeling itself is not usually done within the GIS framework. With few exceptions, the distribution model does not ‘see’ geographical coordinates; instead, the model operates in niche space.
- Some GIS platforms incorporate distribution modeling tools

    DIVA GIS: www.diva-gis.org

    ArcGIS with SDMToolbox: http://sdmtoolbox.org/download

Caution! The use and misuse of models

- ***Garbage in, garbage out*:** if the occurrence records used to build a correlative species' distribution model do not provide useful information as to the environmental requirements of the species, then the model cannot provide useful output.
- ***Model extrapolation*:** 'Extrapolation' refers to the use of a model to make predictions for environmental values that are beyond the range of the data used to calibrate the model. Model extrapolation should be treated with a great deal of caution.
- ***The lure of complicated technology*:** Many approaches to modeling species' distributions utilize complex computational technology (e.g. artificial neural networks and genetic algorithms) along with huge GIS databases of digital environmental layers. "Wow – it must be correct!" Remember that a model can only be useful if the theoretical underpinnings on which it is based are sound.