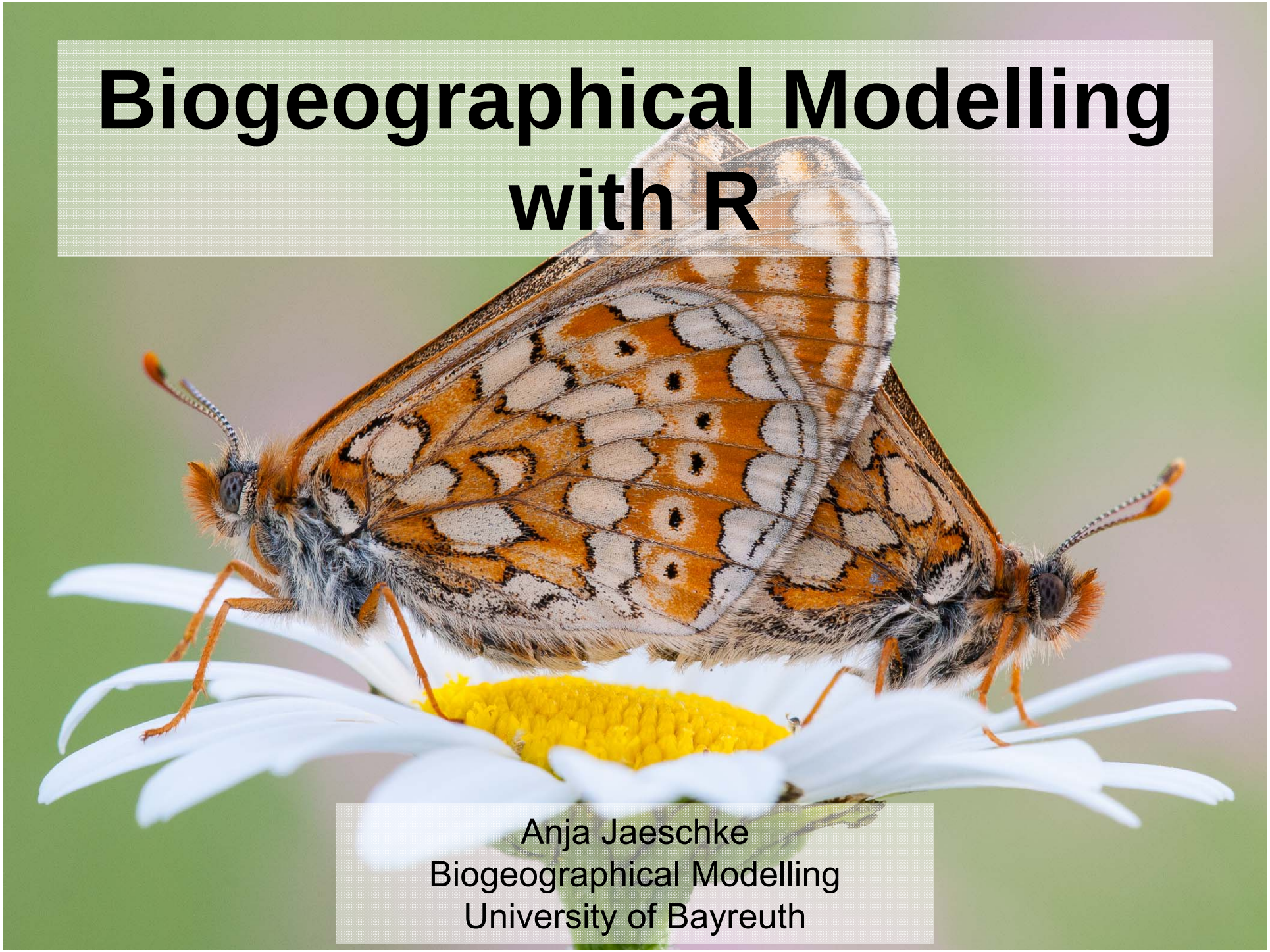
# Biogeographical Modelling with R



Anja Jaeschke
Biogeographical Modelling
University of Bayreuth

# Part I

## What is a

## Species Distribution Model?

# Why predicting species‘ distributions?

- Guiding field surveys to find populations of rare species
- Supporting conservation prioritization and reserve selection
- Guiding reintroduction of endangered species
- Projecting potential impacts of climate or land cover change
- Predicting species‘ invasion
- Assessing disease risk
- Species delimination
- Testing ecological theory
- Comparing paleodistributions and phylogeography

Pearson, R.G. 2007. Species’ Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# Why predicting species‘ distributions?

- Guiding field surveys to find populations of rare species
- Supporting conservation prioritization and reserve selection
- Guiding reintroduction of endangered species
- Projecting potential impacts of climate or land cover change
- Predicting species‘ invasion
- **Assessing disease risk**
- Species delimination
- Testing ecological theory
- Comparing paleodistributions and phylogeography

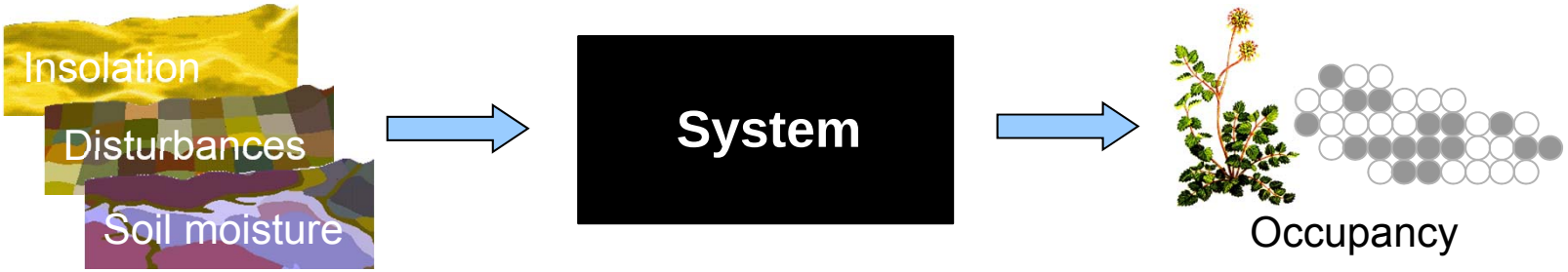
Pearson, R.G. 2007. Species’ Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Aim: Estimating the actual or potential geographic distribution of a species

- Characterize suitable environmental conditions
- Identify where they are distributed in space

➢ Fundamental strategy common to most distribution models




Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Suitable environmental conditions may be characterized using either a mechanistic or a correlative approach

Correlative models: estimate the environmental conditions that are suitable for a species by associating known species' occurrence records with suites of environmental variables that can reasonably be expected to affect the species' physiology and probability of persistence

Assumes that the observed distribution of a species provides useful information as to the environmental requirements of that species

"Since spatially explicit occurrence records are available for a large number of species, the vast majority of species' distribution models are correlative."

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Build and validate a correlative species' distribution model:

1) known species' occurrence records

2) a suite of environmental variables

- At this stage care should be taken to ensure that data are checked for errors

Example: Simply plotting the species' occurrence records in a GIS → identify records that are distant from other occupied sites

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

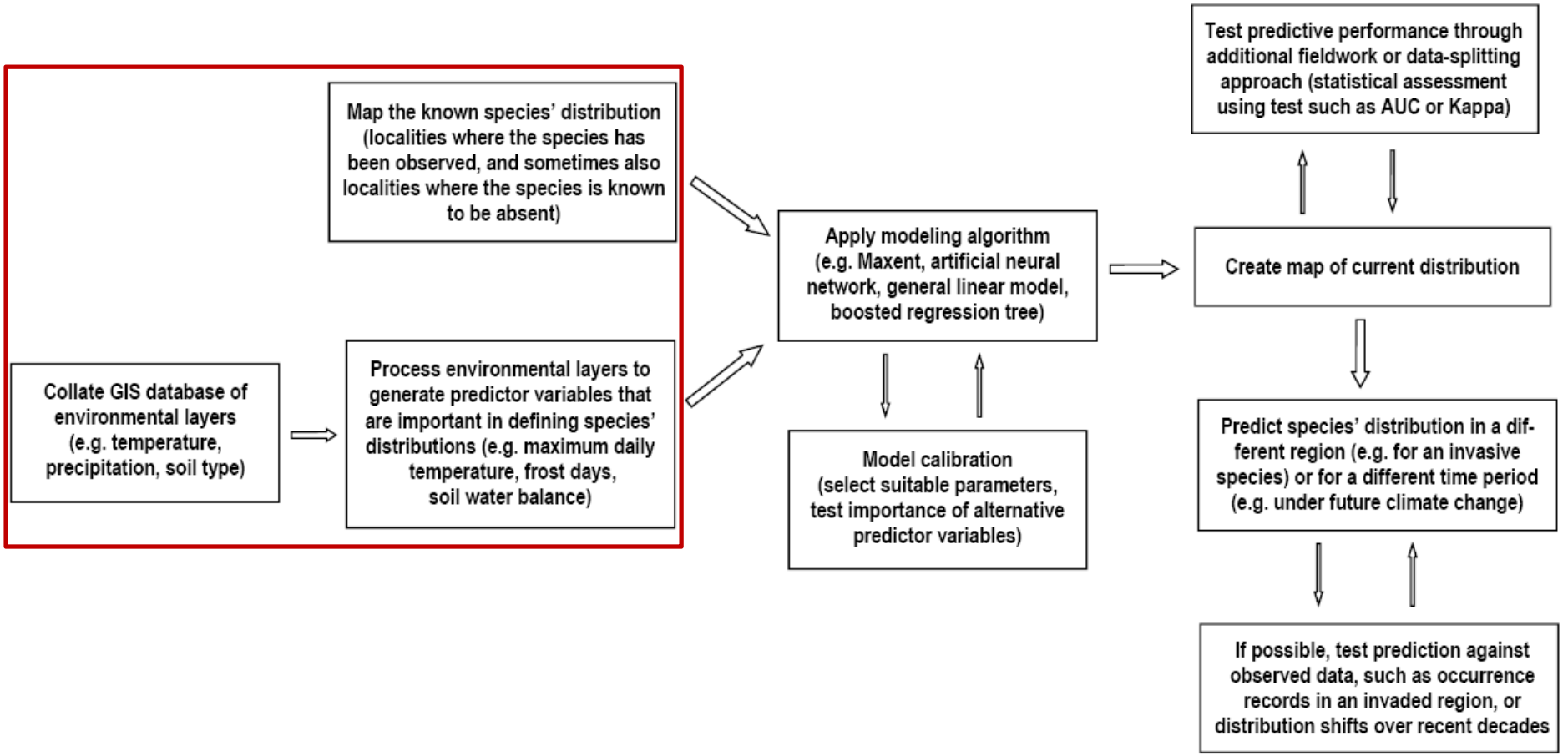



**Figure 1.** Flow diagram detailing the main steps required for building and validating a correlative species distribution model.

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Occurrence records and environmental variables are entered into a modelling algorithm that aims to identify environmental conditions that are associated with species occurrence

Example: A plant species has only been recorded at localities with mean monthly precipitation above 60mm and soil clay content above 40%

- In practice, we usually seek algorithms that are able to integrate more than two environmental variables, since species are in reality likely to respond to multiple factors

- Algorithms that can incorporate interactions among variables are also preferable

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Example: A more accurate description of the plant's requirements may be that it can occur at localities with mean monthly precipitation between 60mm and 70mm if soil clay content is above 60%, and in wetter areas (>70mm) if clay content is as low as 40%

Depending on the method used, various decisions and tests will need to be made at this stage to ensure the algorithm gives optimal results

The relative importance of alternative environmental predictor variables may also be assessed to select which variables are used in the final model

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

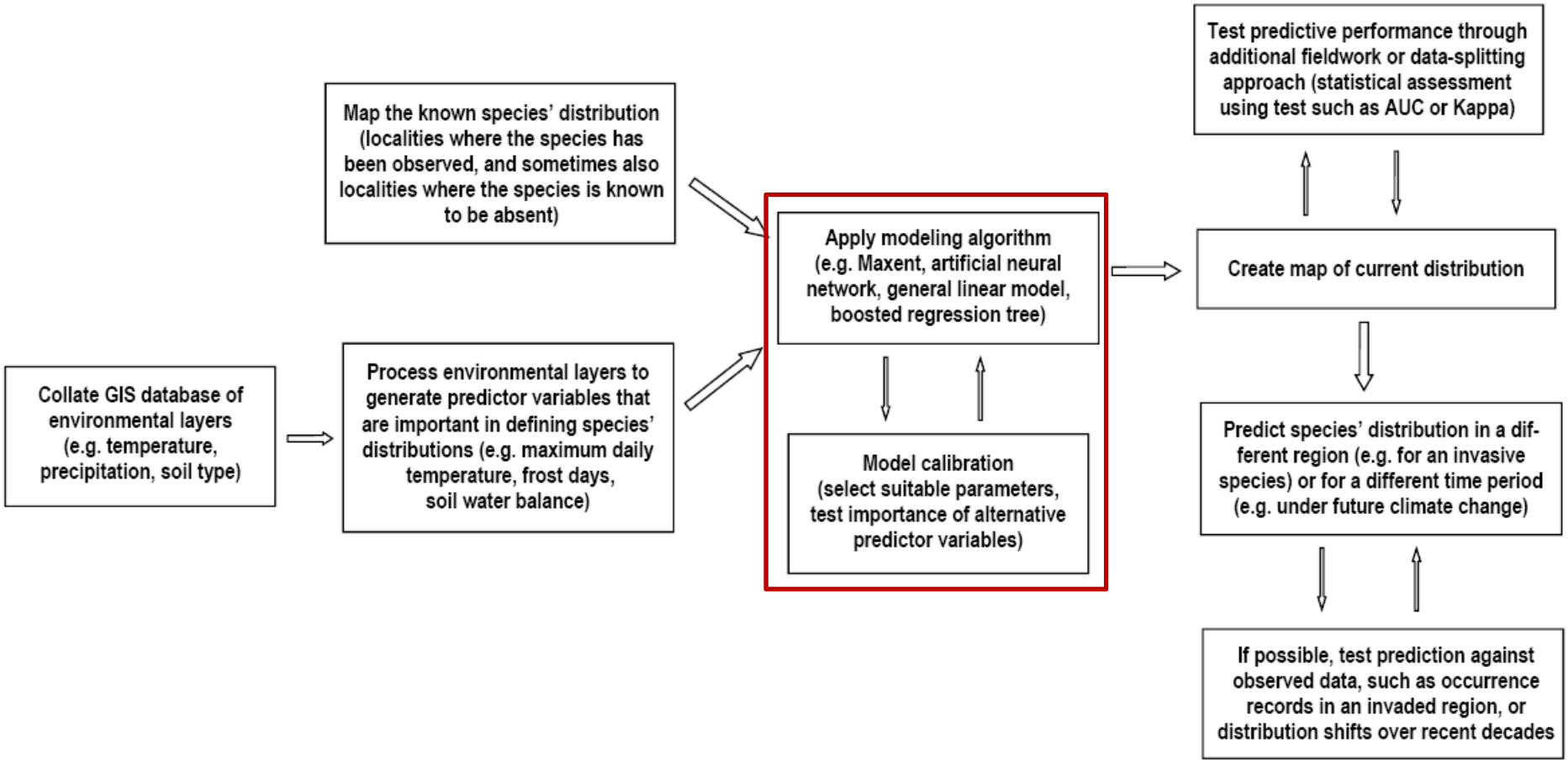



**Figure 1.** Flow diagram detailing the main steps required for building and validating a correlative species distribution model.

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Having run the modeling algorithm, a map can be drawn showing the predicted species' distribution

The ability of the model to predict the known species' distribution should be tested at this stage

A set of species occurrence records that have not previously been used in the modeling should be used as independent test data

The ability of the model to predict the independent data is assessed using a suitable test statistic

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

A number of modelling algorithms predict a continuous distribution of environmental suitability (i.e. a prediction between 0 and 1), it is sometimes useful to convert model output into a prediction of suitable (1) or unsuitable (0)

- Necessary step before applying many test statistics → methods for setting a threshold probability, above which the species is predicted as present, are needed

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

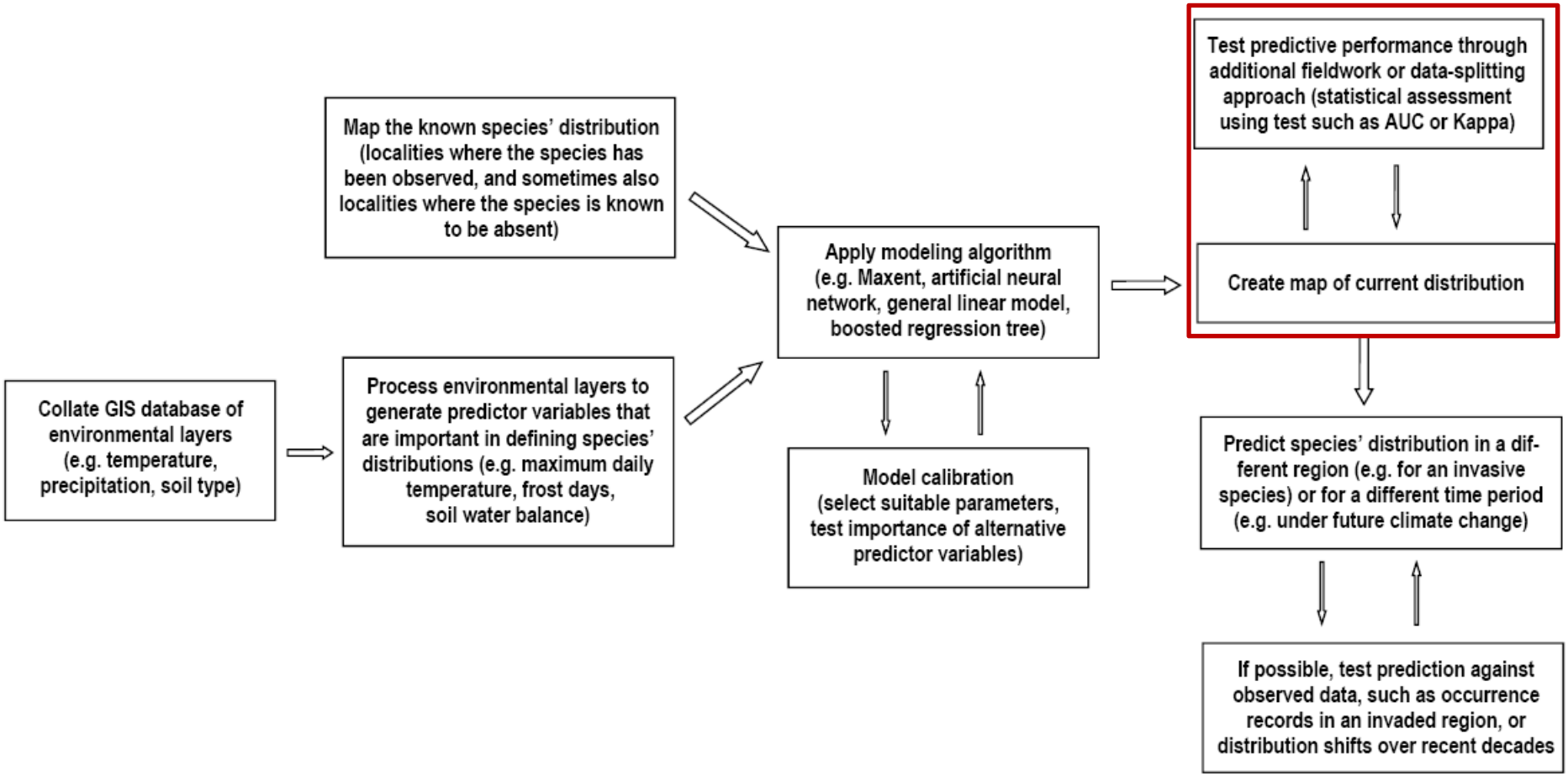



**Figure 1.** Flow diagram detailing the main steps required for building and validating a correlative species distribution model.

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

Now the model can be used to predict species' occurrence in areas where the distribution is unknown

Thus, a set of environmental variables for the area of interest is input into the model and the suitability of conditions at a given locality is predicted

In many cases the model is used to 'fill the gaps' around known occurrences

In other cases, the model may be used to predict species' distributions in new regions (e.g. to study invasion potential) or for a different time period (e.g. to estimate the potential impacts of future climate change)

Ideally, model predictions into different regions or for different time periods should be tested against observed data

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

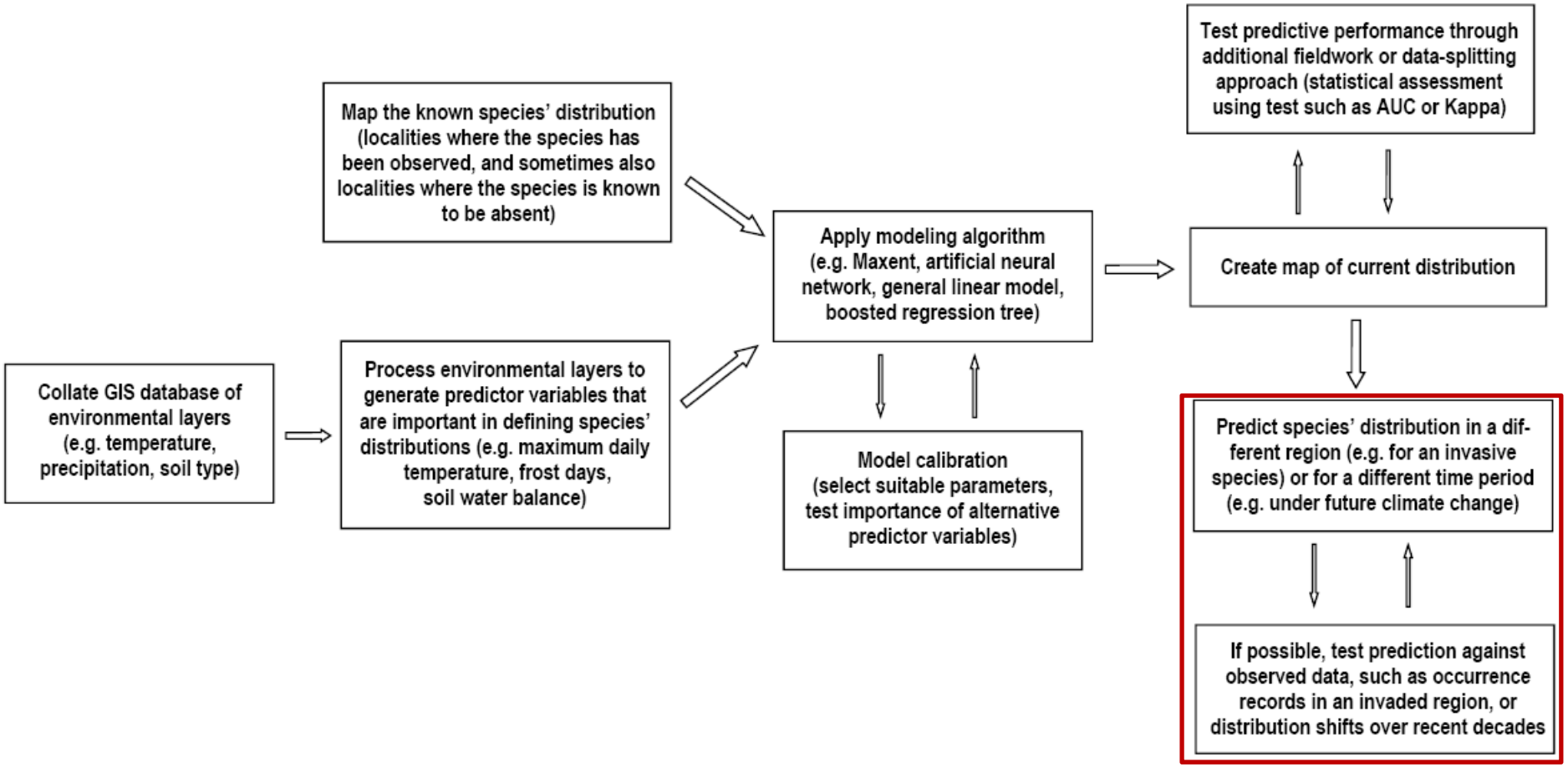



**Figure 1.** Flow diagram detailing the main steps required for building and validating a correlative species distribution model.

Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# What is a species distribution model?

This modelling approach has been variously termed ‘species distribution’, ‘ecological niche’, ‘environmental niche’, ‘habitat suitability’ and ‘bioclimate envelope’ modelling

Use of the term ‘species distribution modeling’ is widespread but somewhat misleading since it is actually the distribution of suitable environments that is being modelled, rather than the species’ distribution per se

Regardless of the name used, the basic modelling process is essentially the same and the theoretical underpinnings of the models are similar

It is essential that these theoretical underpinnings are properly understood in order to interpret model outputs accurately

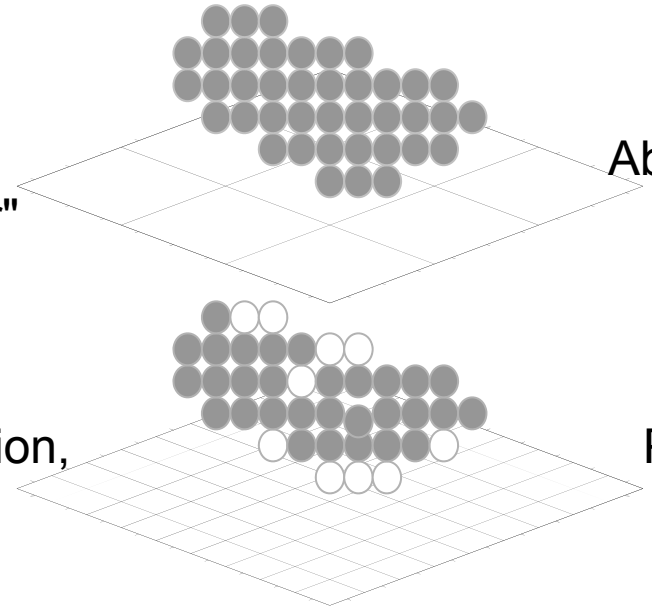
Pearson, R.G. 2007. Species’ Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

# Part II

# Methodology of SDMs

# Concept: landscape filters on a hierarchy of scales

Regional species pool



Filter I - Resources,
"habitat quality-filter"

Abiotic environmental factors,
disturbance regime,
resources

Filter II - Configuration,
scale of interaction
"metapop-filter"

Patch area, patch isolation,
habitat connectivity,
dispersal capabilities

Filter III –
Biotic interactions

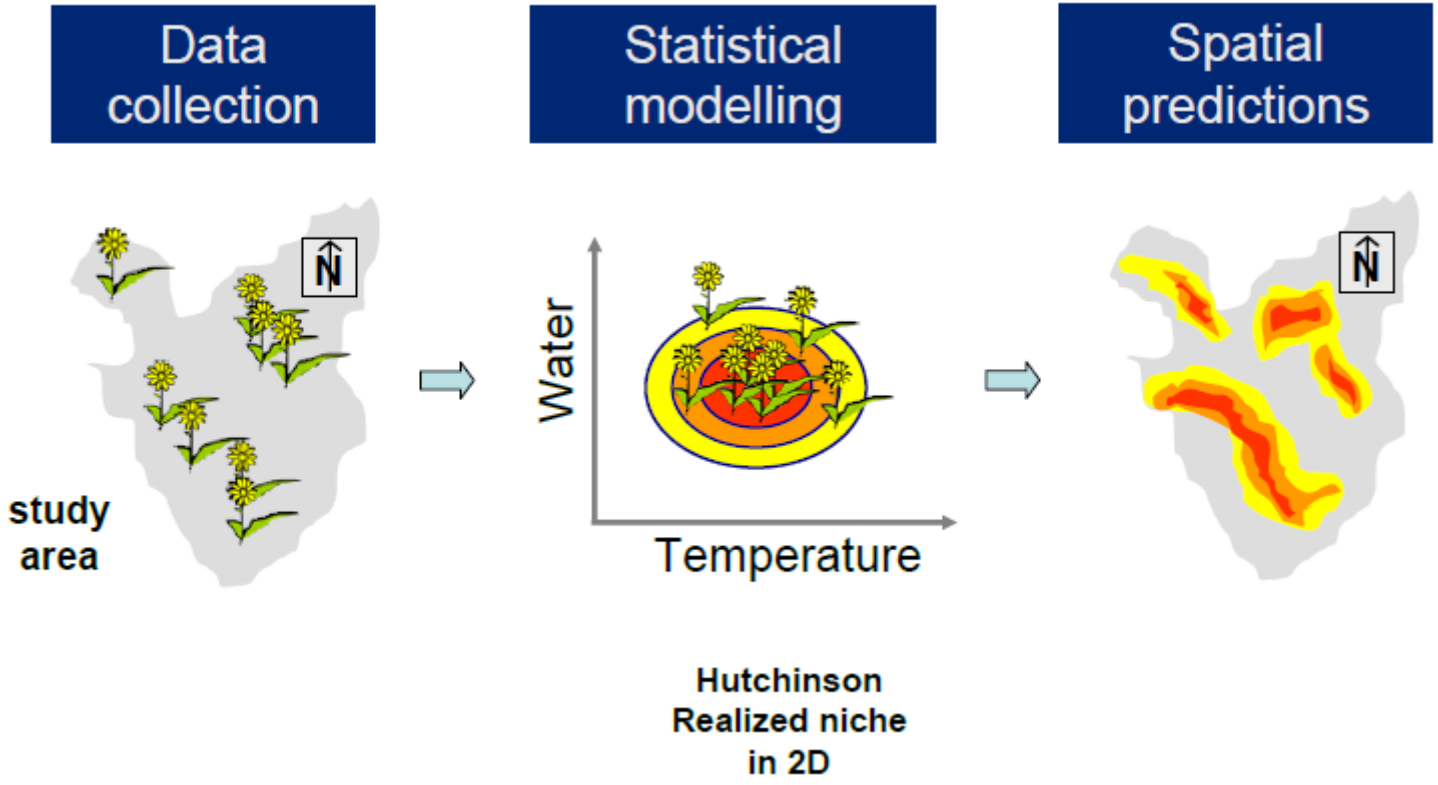
Predation, competition,
facilitation etc.

Observed species

Each filter corresponds
to a set of species traits!

Schröder & Reineking, 2004. UFZ-Bericht 9/2004: 5-26.

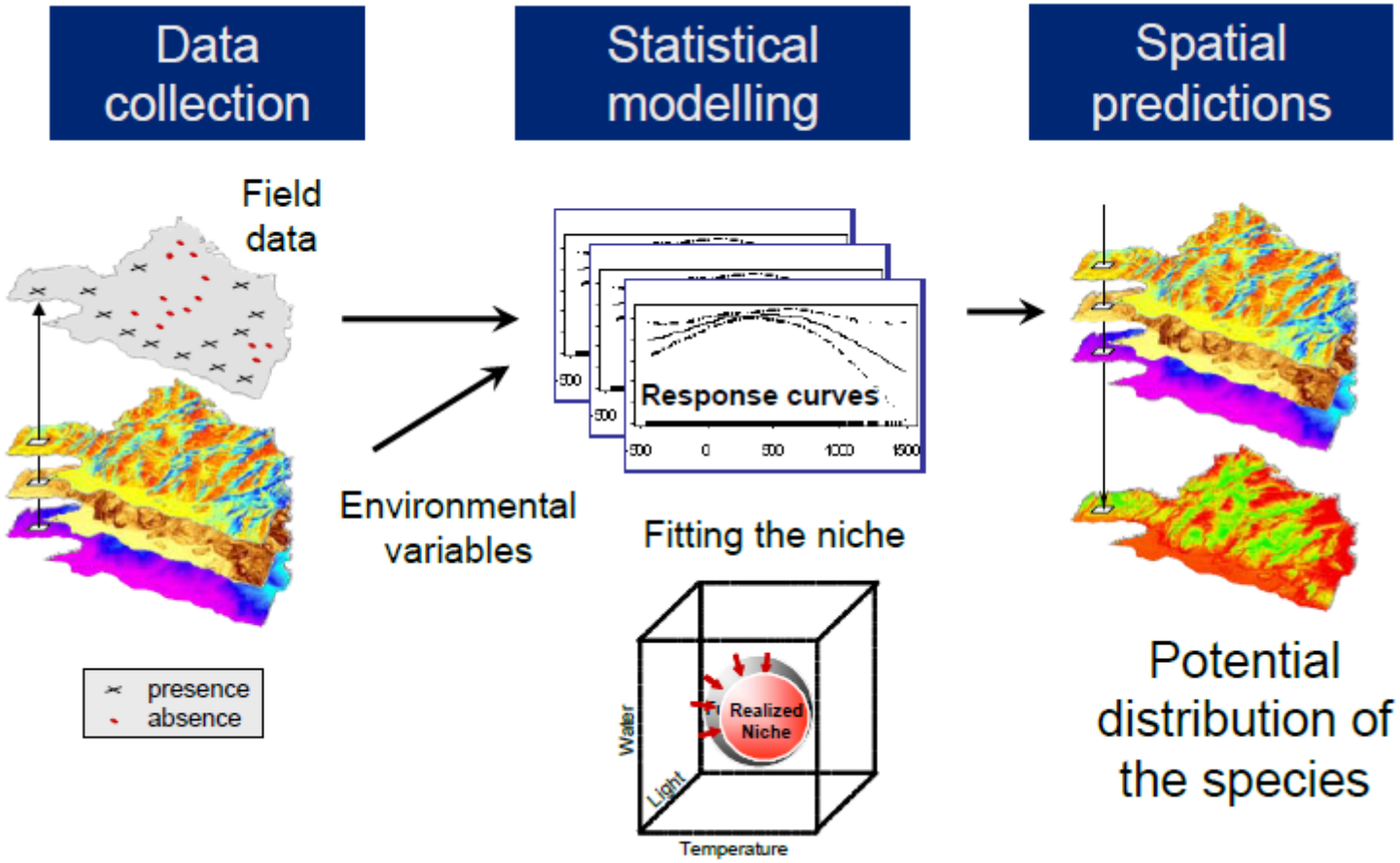
# SDMs: Principle




**Hutchinson Realized niche in 2D**

Guisan & Zimmermann (2000) Ecological Modelling

# SDMs: Principle




Guisan et al. (2006) J. Appl. Ecol. | Guisan & Zimmermann (2000) Ecological Modelling

## SDMs: Steps

- Research questions
- Data collection (species, environment)
- Preprocessing: Selection of candidate variable sets
- Selection of model algorithm → Model fitting
- Quantification of variable importance
- Projection
- Model evaluation (Validation on test data; performance criteria) → Model calibration
- Model selection

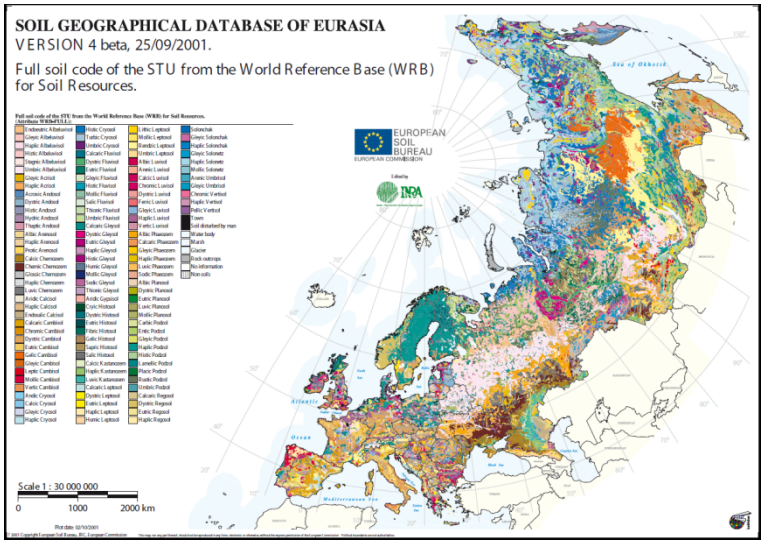
# SDMs: Preprocessing

Parameter set: climate, land use, soil parameter, …

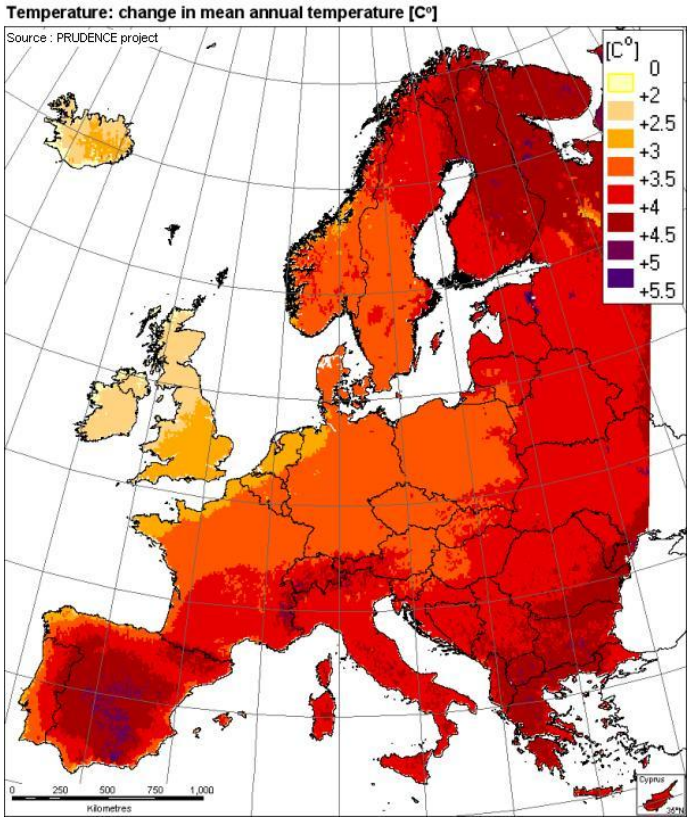
Which climate model / emission scenario (B1, A2, A1B, A1FI,…)?

How many variables? – 2-10 low correlating variables

Integrating expert knowledge?!




Map source: http://eusoils.jrc.ec.europa.eu/ESDB_Archive/sgdbe/wrb-fulla3.pdf




Map source: http://peseta.jrc.ec.europa.eu/docs/ClimateModel.html

# SDMs: Selection of model algorithm

There are many different statistical species distribution models but they can be classified into two groups

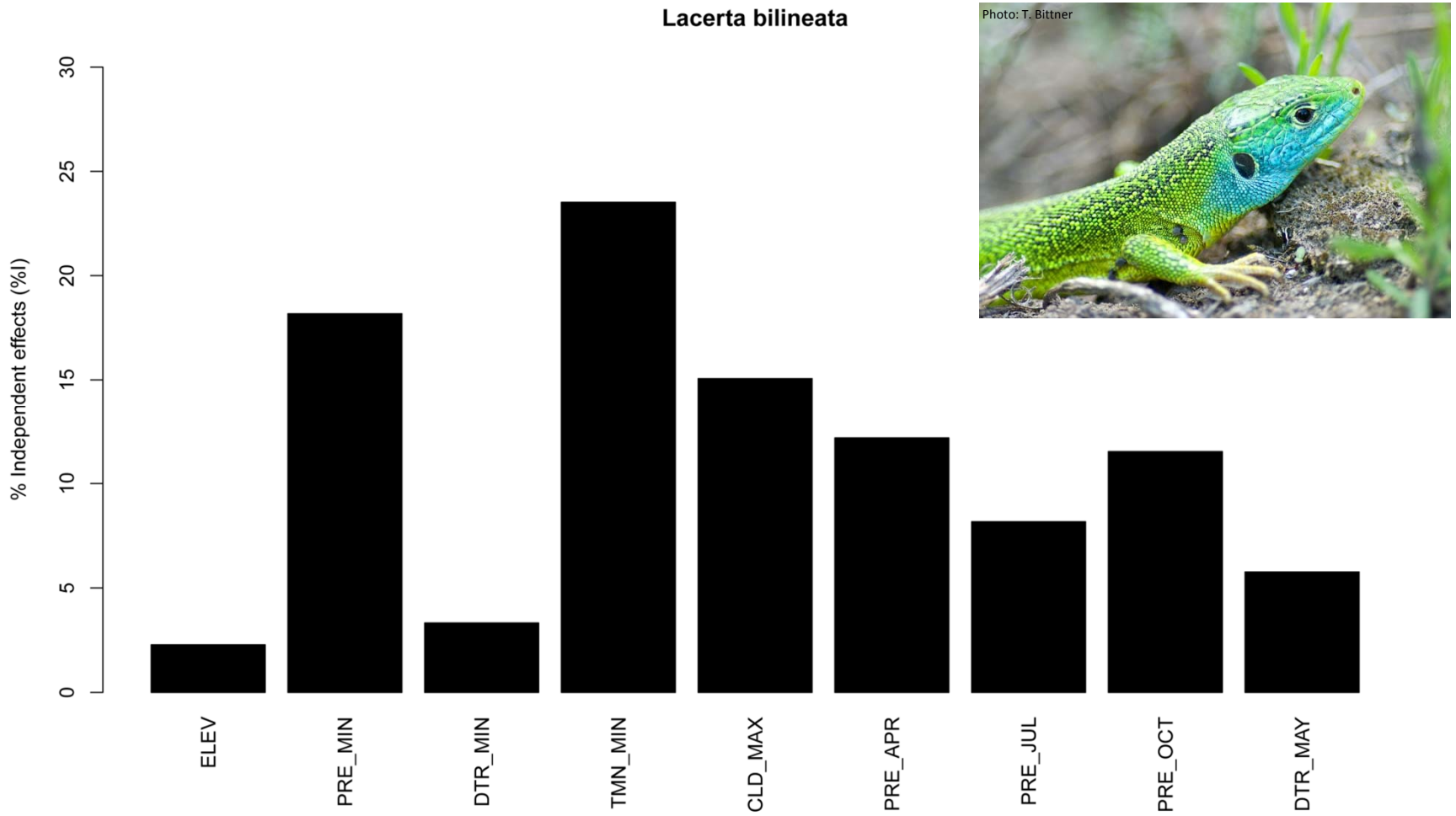
| Presence-only methods | Presence-absence (or Pseudo-A or background) |
|---|---|
| Bioclim/SRE | Entropy models (Maxent) |
| DOMAIN | Regression-like models (GLM, GAM, MARS) |
| Habitat | Classification-like models (CART) |
| ENFA | Machine learning models (ANN, GARP) |
| PCA species | Averaging models (RF, BRT) |

SRE: Species Range Envelope; ENFA: Environmental Niche Factor Analysis; Maxent: Maximum entropy; GLM: Generalized Linear Models; GAM: Generalized Additive Models; MARS: Multivariate Adaptive Regression Splines; CART: Classification and Regression Trees; ANN: Artificial Neural Networks; GARP: Genetic Algorithm for Rule-set Predictions; RF: Random Forests; BRT: Boosted Regression Trees

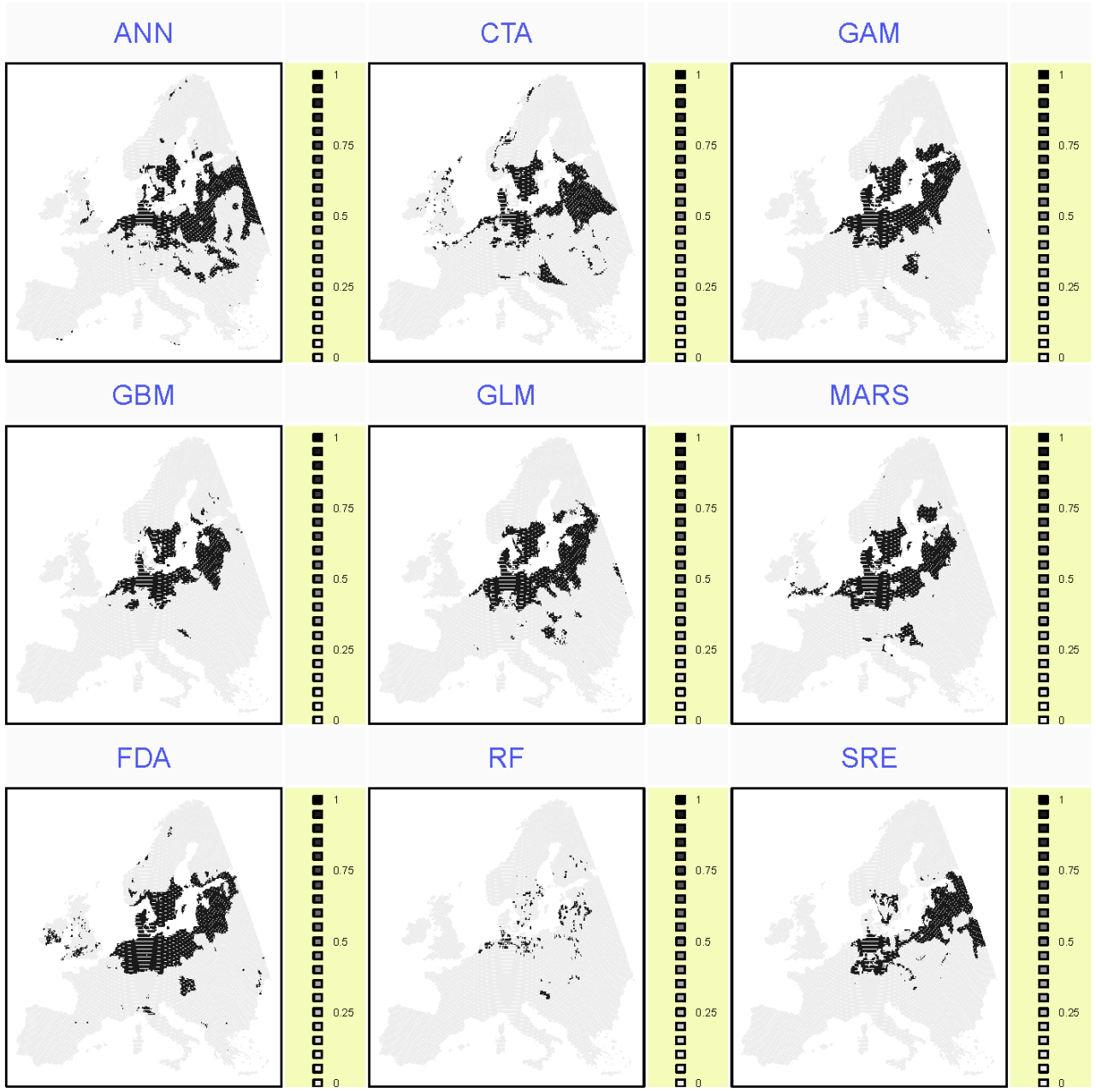
Elith, J. et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

# SDMs: Variable importance

- Importance of environmental variables?
- Which is the most important one? → ecological meaning?

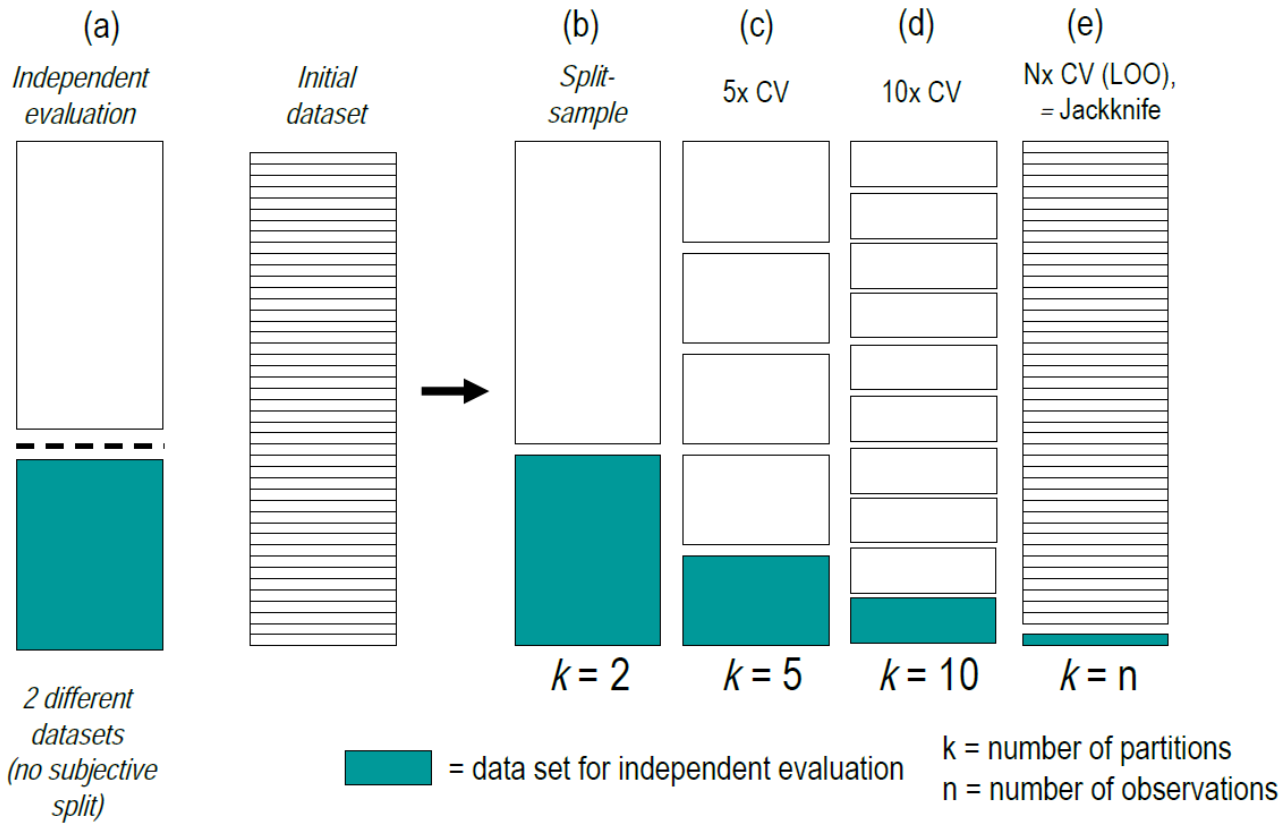


# SDMs: Prediction of current distribution



# SDMs: Model evaluation – validation options

- One-time data-splitting: e.g. 70% training data, 30% test data (80/20)
- Cross validation: e.g. 10 fold → data set divided in 10 parts
- Jackknife: data set divided in n parts (n = data)



# SDMs: Model evaluation – performance criteria

**Classification matrix**

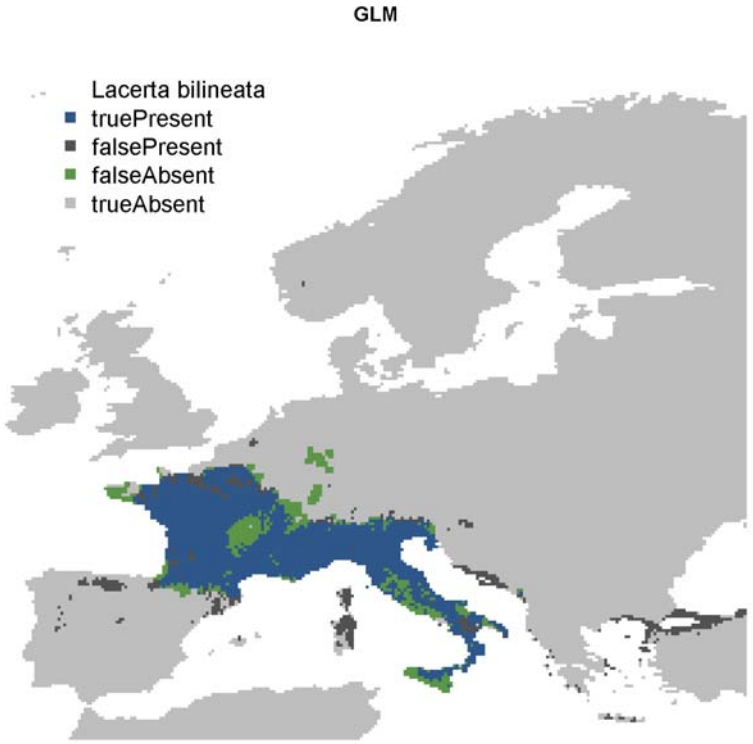
| | | Observed | | |
|---|---|---|---|---|
| | | **1** | **0** | |
| **Predicted** | **1** | *a* | *b* | *a+b* |
| | **0** | *c* | *d* | *c+d* |
| | | *a+c* | *b+d* | *n* |

% correct: (*a*+*d*)/*n*
Sensitivity: *a*/(*a*+*c*)
Specificity: *d*/(*b*+*d*)

Sensitivity:
Proportion of correctly classified presences

Specificity:
Proportion of correctly classified absences

% correct:
Proportion of correct predictions




Schröder, B. & Richter, O. (1999/2000): ZÖN 8: 195-205

# SDMs: Model evaluation – performance criteria

AUC (Area under the receiver operating characteristic curve):

- Threshold-independent method of evaluating the performance of presence/absence models
- The true positive rate (sensitivity) is plotted against the false positive rate (1- specificity) as the threshold varies from 0 to 1
- The ROC-plot for a poor model (whose predictive ability is the equivalent of random assignment) will lie near the diagonal, where the true positive rate equals the false positive rate for all thresholds

Swets' (1988) interpretation scale:

AUC = 1 : perfect prediction

AUC > 0.9 : good predictions

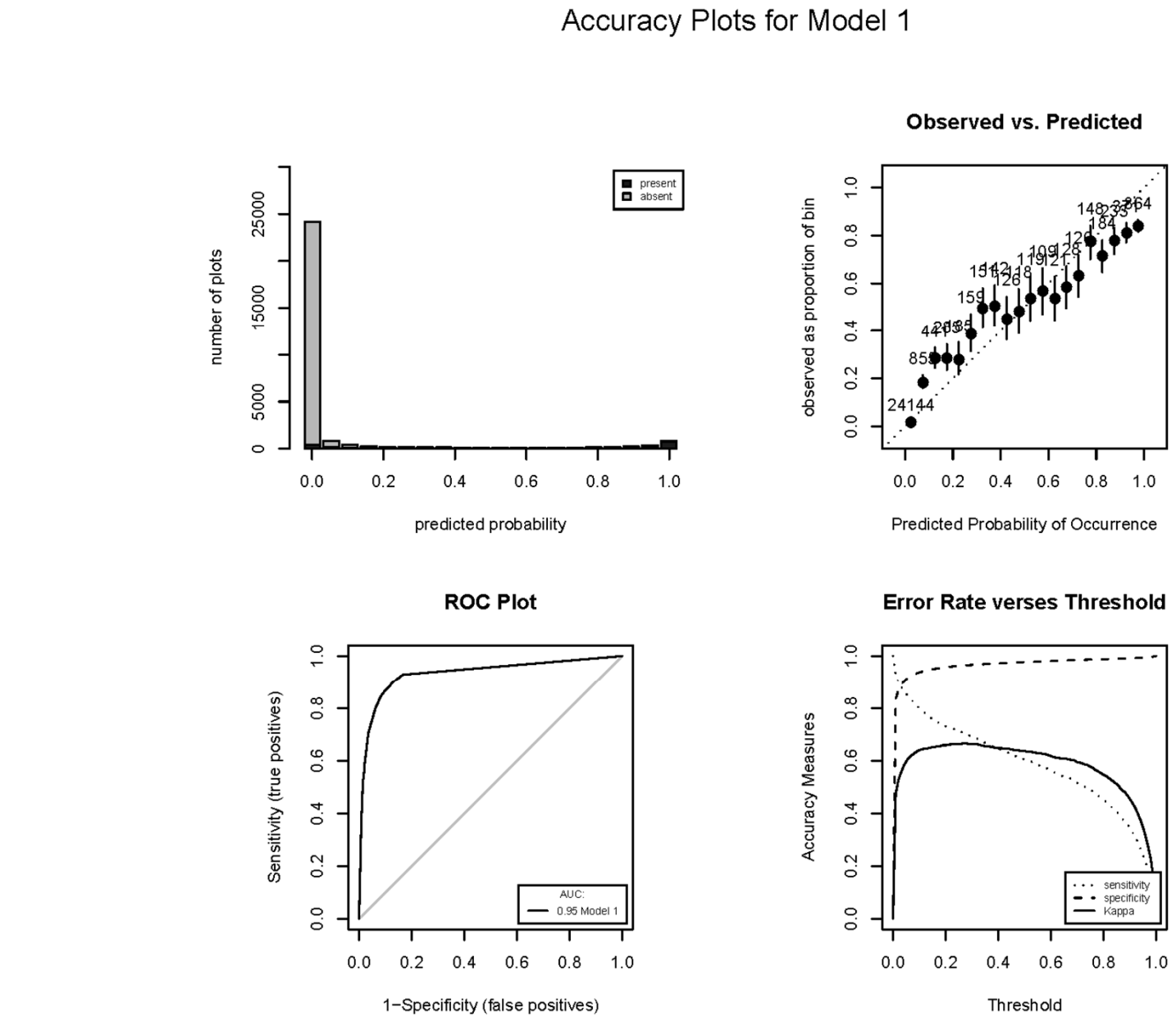
0.7 < AUC < 0.9 : useful predictions

0.5 < AUC < 0.7 : poor predictions

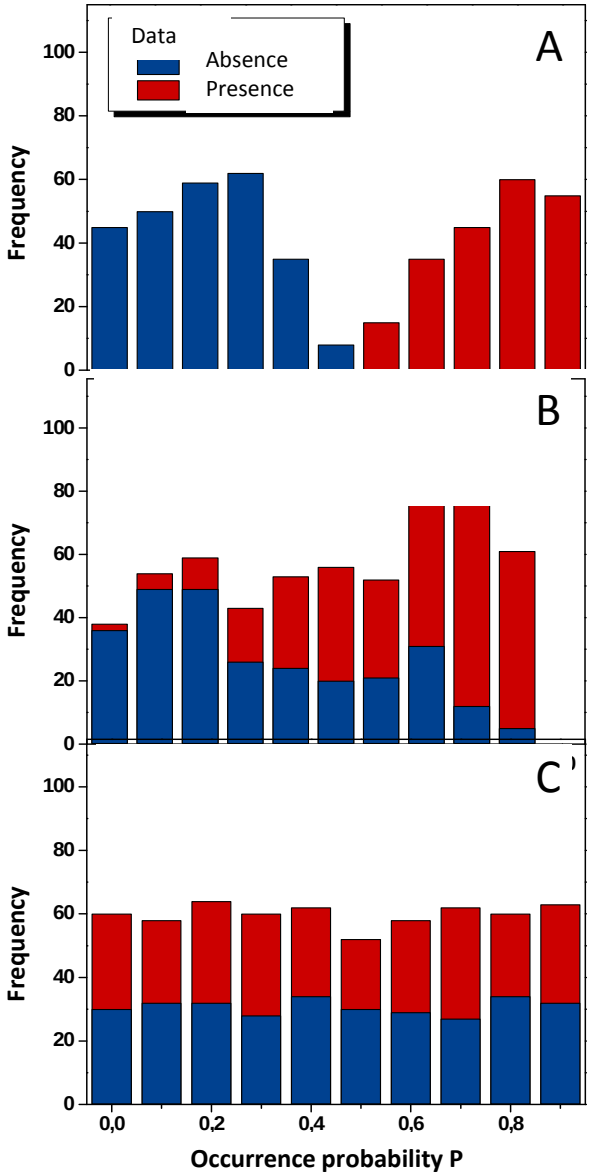
0.5 = not different than random

0.5 > AUC > 0 : counter-predictions

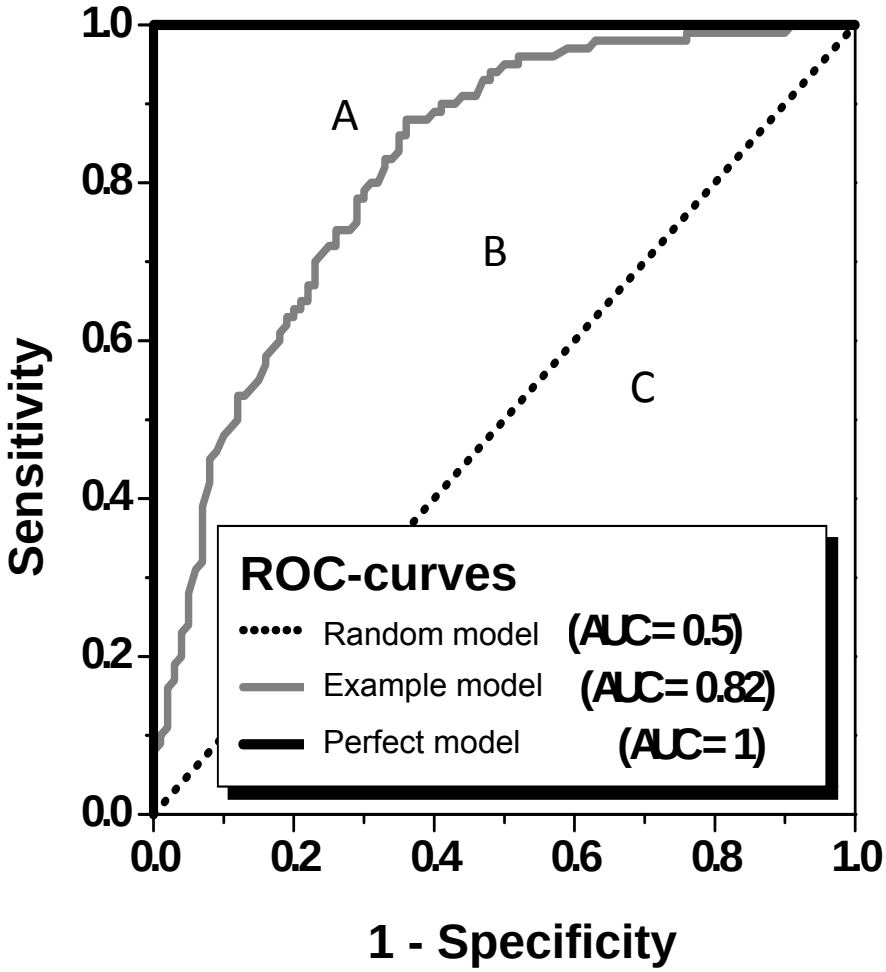
# SDMs: Model evaluation – performance criteria



# SDMs: Model evaluation – performance criteria




Relationship between classification diagram & ROC-curve



# SDMs: Model selection

## Novel methods ir occurrence data

Jane Elith*, Catherine H. C
Robert J. Hijmans, Falk Hu
Bette A. Loiselle, Glenn Ma
A. Townsend Peterson, Steve
Jorge Soberón, Stephen Wil

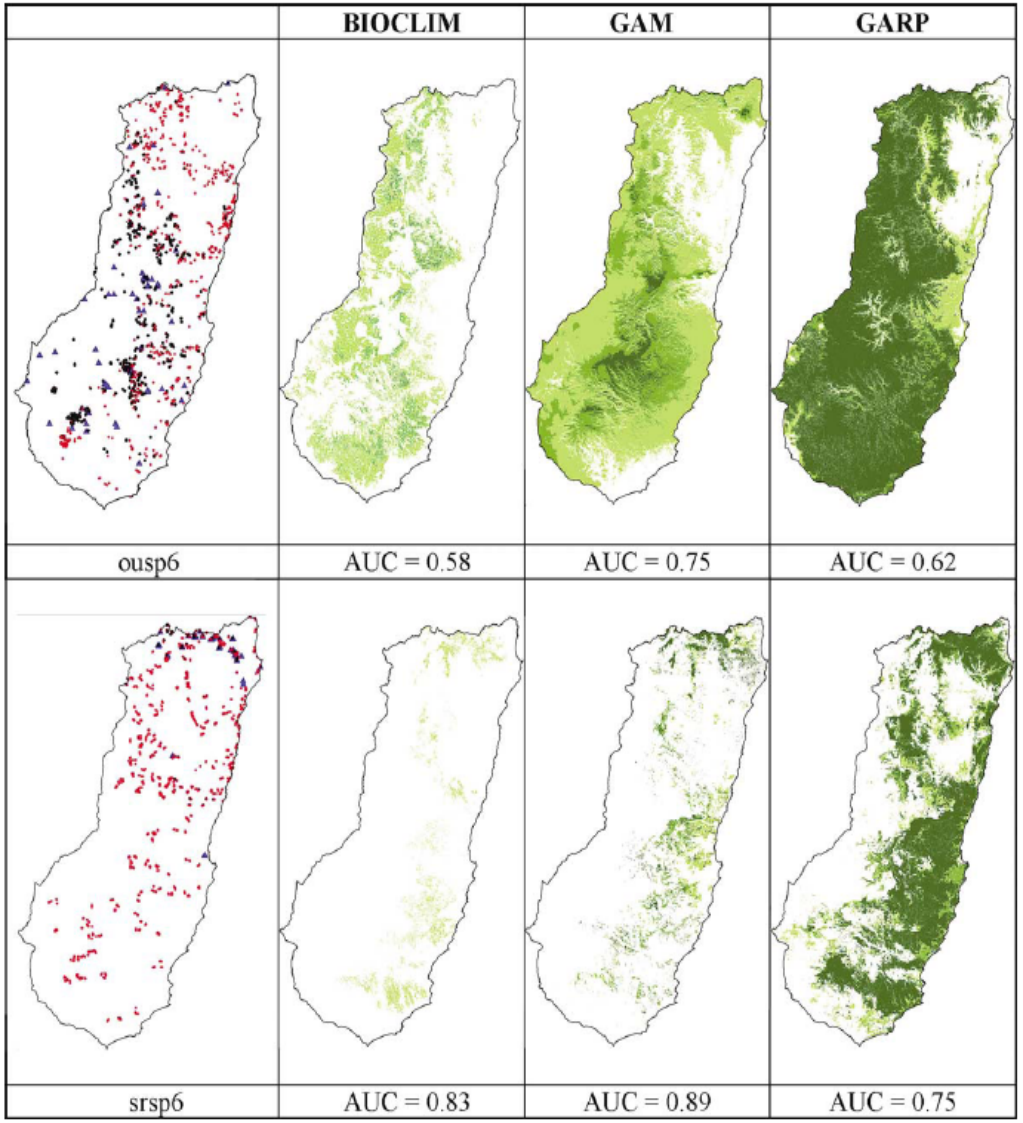
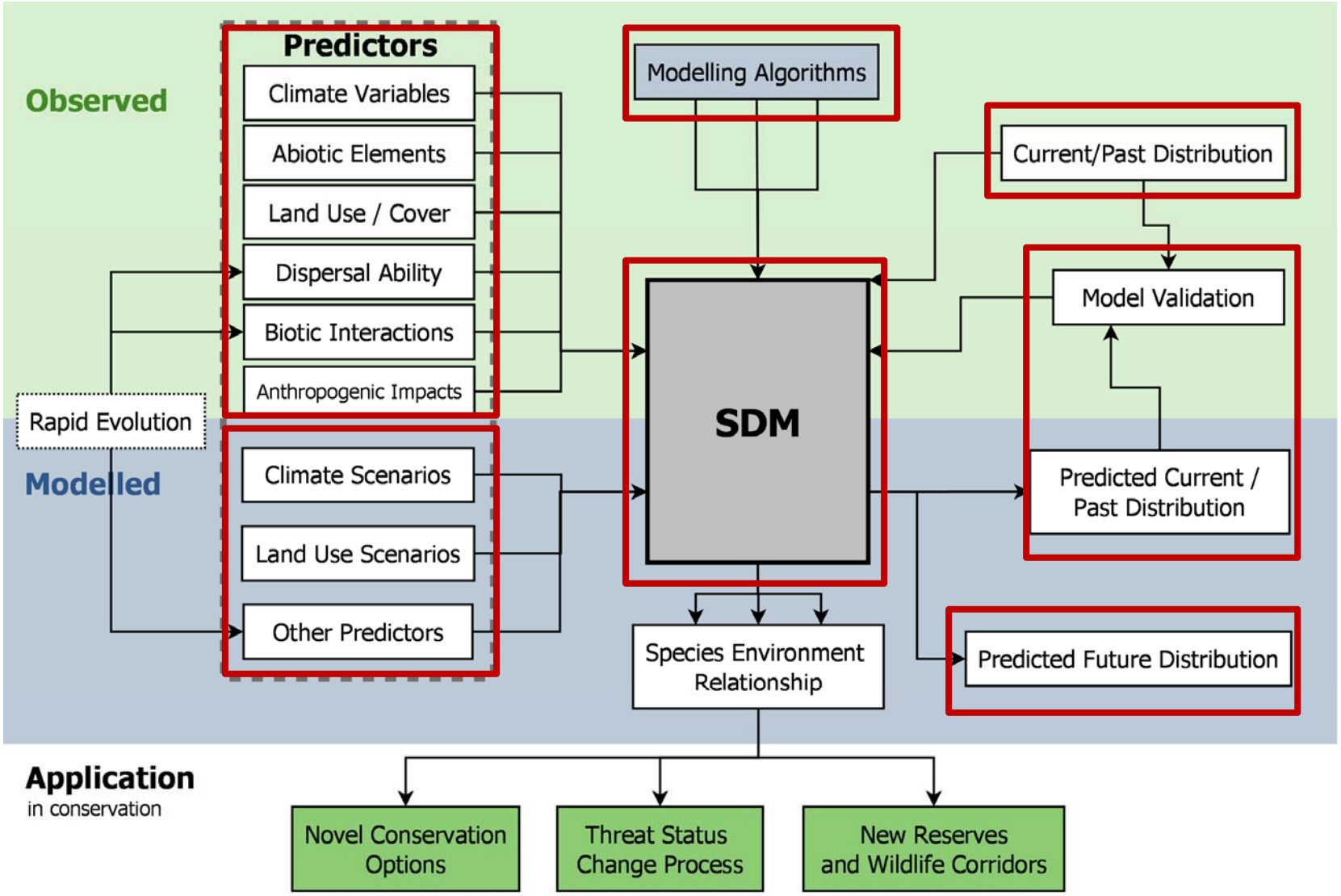



Fig. 1. Maps for two species from NSW for each of three selected techniques. Details: ousp6, *Poa sieberiana* (53 records for modelling and 512 presence/797 absence for evaluation); srsp6 *Ophioscincus truncatus* (79 model, 74/932 eval). The first column shows modelling sites (grey triangles) and evaluation sites: presence = black circle, absence = black cross.

Elith, J., Graham, C. H. et al. 2006. Ecography 29: 129-151.

# SDMs: Model


Observed
Predictors
Climate Variables
Abiotic Elements
Land Use / Cover
Dispersal Ability
Biotic Interactions
Anthropogenic Impacts
Rapid Evolution
Modelled
Climate Scenarios
Land Use Scenarios
Other Predictors
Modelling Algorithms
SDM
Species Environment Relationship
Current/Past Distribution
Model Validation
Predicted Current / Past Distribution
Predicted Future Distribution
Application
in conservation
Novel Conservation Options
Threat Status Change Process
New Reserves and Wildlife Corridors

# Part III

# What is a Process-Based Model?

# Process-based models

"When constructing a model, one is constantly trading off the degree of precision, generality and realism […]. It is not possible to include all details of a system and still have a useful predictive tool.

For example, a one-to-one scale map of a city may include all details, but ceases to be useful as a guide for finding the nearest hotel.

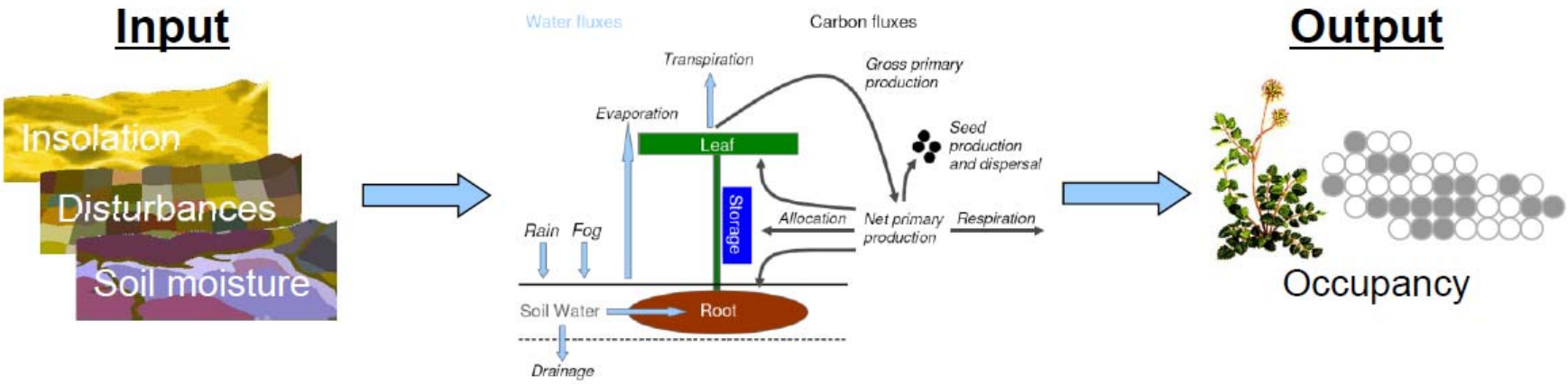
As a result, models are always false in some aspects of their representation of a system, and there is no one correct model that links a theory to a particular system […]."

- Statistical models do not explicitly include important ecological processes such as demographic relationships or interspecific interactions that may also limit geographic range

Cuddington et al. 2013 Ecosphere 4: 1-12

# Process-based models

- Synonym: mechanistic models, process models, biogeochemical models
- Aim to incorporate physiologically limiting mechanisms in a species' tolerance to environmental conditions
- Require detailed understanding of the physiological response of species to environmental factors



# Process-based models

- Developed to model key growth process(es) and fundamental causes of productivity such as:

– Photosynthesis and respiration

– Carbon allocation

– Nutrient cycles

– Climate effects

- Take into account at the physiological level plant responses to site factors

http://www.forestmodels.com/process-based-forest-models.html

# Process-based models

- Are the most appropriate approach for the majority of management questions
- Are built on explicit assumptions about how a system works
- Assumptions are grounded in ecological theory
- Because these models are based on causal mechanisms rather than correlation, the confidence in extrapolating beyond known data is enhanced
- However, there is always uncertainty about how an ecological process will interact with novel global change conditions

Cuddington et al. 2013 Ecosphere 4: 1-12

# Process-based models

- Essential scientific tools:

– Providing formalized statements of hypotheses

– And a framework that encapsulates disparate pieces of information and knowledge

- Behaviour of a system is derived from a set of functional components and their interactions with each other and the system environment, through physical and mechanistic processes occurring over time

- Functional components are chosen at a specified level of hierarchy, customarily one level below the level of the entire system

- Model system can be regarded as an analog of the real system at a specified level of hierarchy

Mäkelä et al. 2000 Tree Physiology 20: 289-298

# Part IV

# Modelling algorithms and their implementation in R

# SDMs: Selection of model algorithm

There are many different statistical species distribution models but they can be classified into two groups

| Presence-only methods | Presence-absence (or Pseudo-A or background) |
|---|---|
| Bioclim/SRE | Entropy models (Maxent) |
| DOMAIN | Regression-like models (GLM, GAM, MARS) |
| Habitat | Classification-like models (CART) |
| ENFA | Machine learning models (ANN, GARP) |
| PCA species | Averaging models (RF, BRT) |

SRE: Species Range Envelope; ENFA: Environmental Niche Factor Analysis; Maxent: Maximum entropy; GLM: Generalized Linear Models; GAM: Generalized Additive Models; MARS: Multivariate Adaptive Regression Splines; CART: Classification and Regression Trees; ANN: Artificial Neural Networks; GARP: Genetic Algorithm for Rule-set Predictions; RF: Random Forests; BRT: Boosted Regression Trees

Elith, J. et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

# Modelling algorithms

There are many different statistical species distribution models but they can be classified into two groups

| Presence-only methods | Presence-absence (or Pseudo-A or background) |
|---|---|
| Bioclim/SRE | Entropy models (Maxent) |
| DOMAIN | Regression-like models (GLM, GAM, MARS) |
| Habitat | Classification-like models (CART) |
| ENFA | Machine learning models (ANN, GARP) |
| PCA species | Averaging models (RF, BRT) |

SRE: Species Range Envelope; ENFA: Environmental Niche Factor Analysis; Maxent: Maximum entropy; GLM: Generalized Linear Models; GAM: Generalized Additive Models; MARS: Multivariate Adaptive Regression Splines; CART: Classification and Regression Trees; ANN: Artificial Neural Networks; GARP: Genetic Algorithm for Rule-set Predictions; RF: Random Forests; BRT: Boosted Regression Trees

Elith, J. et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

# Modelling algorithms: Maxent

- Maximum Entropy
- General-purpose method for making predictions or inferences from incomplete information
- General approach for presence-only modeling of species distributions
- Estimate a target probability distribution by finding the probability distribution of maximum entropy (i.e., that is most spread out, or closest to uniform), subject to a set of constraints that represent our incomplete information about the target distribution

S.J. Phillips et al. (2006) Ecological Modelling 190: 231–259

# Modelling algorithms: Maxent

Advantages:

1. Requires only presence data and environmental information

2. Can utilize continuous and categorical data, and can incorporate interactions between different variables

3. Efficient deterministic algorithms to converge to the optimal (maximum entropy) probability distribution

4. The Maxent probability distribution has a concise mathematical definition

5. Over-fitting can be avoided

6. The output is continuous (allowing fine distinctions)

S.J. Phillips et al. (2006) Ecological Modelling 190: 231–259

# Modelling algorithms: Maxent

Disadvantages:

1. Not as mature a statistical method as GLM or GAM

2. Amount of regularization requires further study as does its effectiveness in avoiding over-fitting compared with other variable-selection methods

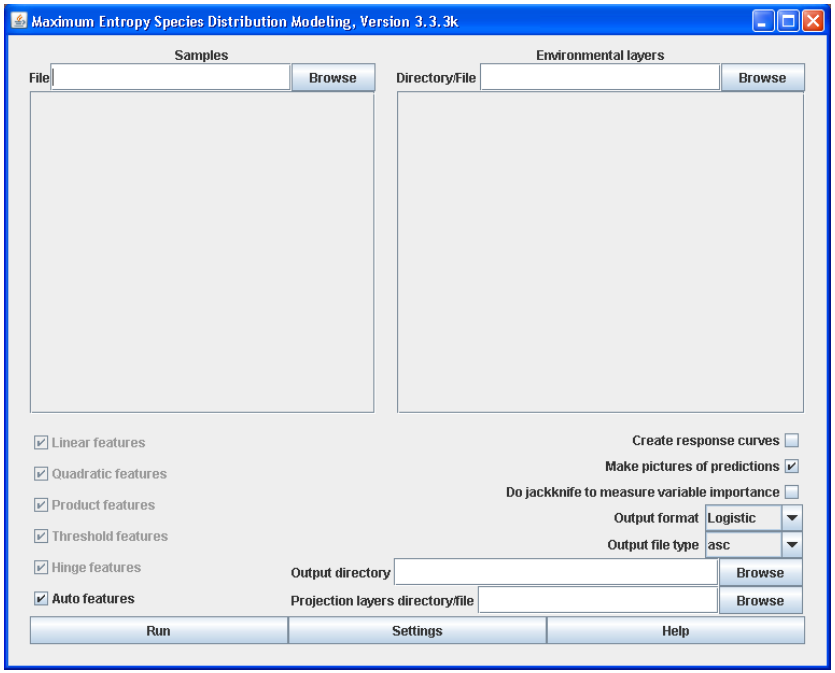
3. Uses an exponential model for probabilities, which is not inherently bounded above (→ take care when extrapolating to another study area or to future or past climatic conditions)

4. Special-purpose software required

S.J. Phillips et al. (2006) Ecological Modelling 190: 231–259

# Modelling algorithms: Maxent

## User interface




## Implementation in R

Package *dismo*

Function *maxent*

Based on java → put the file 'maxent.jar' in the 'java' folder of this package

maxent(x, p, ...)

# x: Predictors as Raster object, SpatialGridDataFrame or data.frame

# p: Occurrence data as data.frame, matrix, SpatialPoints object or vector

# Modelling algorithms: GLM

- Generalized Linear Model

When using GLMs?

- When the variance is not constant and/or when the errors are not normally distributed

- When the response variable is:

– Count data expressed as proportions (e.g. logistic regression)

– Count data that are not proportions (e.g. log linear models of counts)

– Binary response variables (e.g. dead or alive)

– Data on time-to-death where the variance increases faster than linearly with the mean (e.g. time data with gamma errors)

## Modelling algorithms: GLM

- Analyze models that have a particular kind of nonlinearity and particular kinds of nonnormally distributed (but still independent) errors

- Combine a range of nonnormal error distributions with the ability to work with some reasonable nonlinear functions

- By far the two most common GLMs are Poisson regression, for count data, and logistic regression, for survival/failure data

- The class of nonnormal errors that GLMs can handle is called the exponential family (Poisson, bionomial, Gamma and normal distributions)

# Modelling algorithms: GLM

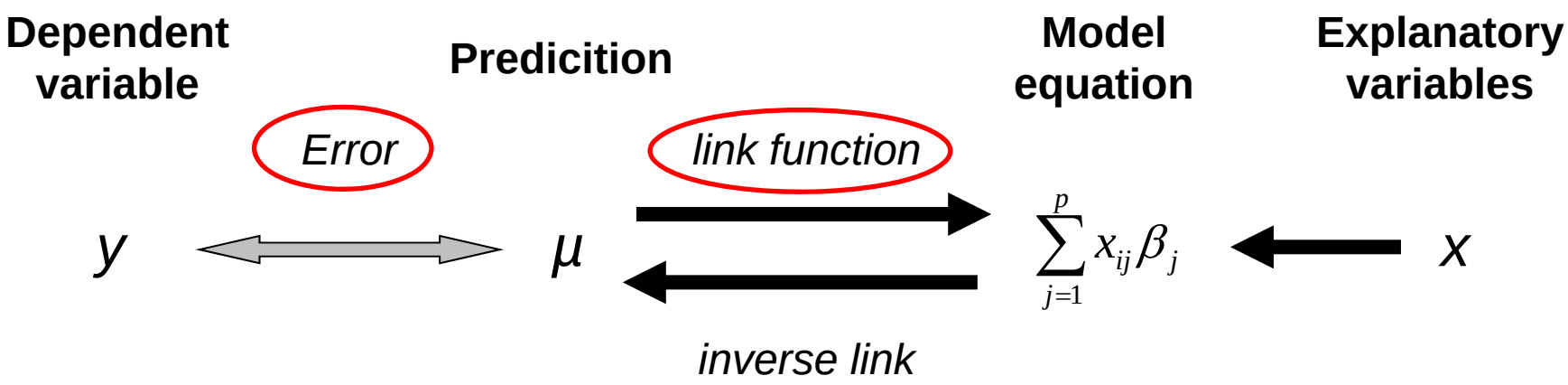
- Every GLM has three components:

1. A response variable distribution (the error structure)

2. A linear predictor η (that involves the explanatory variables)

3. A link function g (that connects the linear predictor to the natural mean of the response variable)

- The usual GLM assumes that the observations are independent



# Modelling algorithms: GLM

Implementation in R

Function *glm*

The error structure is defined by the family directive, used as part of the model formula like this:

```
glm(y ~ z, family = poisson)
```

Canonical link functions as the default options

| Family | Link | Name |
|---|---|---|
| Normal | $\eta = \mu$ | Identity link |
| Poisson | $\eta = \log\mu$ | Log link |
| Binomial | $\eta = \log(P/(1-P))$ | Logit link |
| Gamma | $\eta = \mu^{-1}$ | Reciprocal link |

## Modelling algorithms: GAM

- Generalized Additive Model

- Like GLMs in that they can have different error structures and different link functions to deal with count data or proportion data

Difference to GLM:

- Relationship between y and a continuous variable x is not specified by some explicit functional form

- Instead, non-parametric smoothers are used to describe the relationship

- Useful for relationships that exhibit complicated shapes, like hump-shaped curves

# Modelling algorithms: GAM

Implementation in R

Package *gam*

Function *gam*

- Looks like a glm, except that the relationships we want to be smoothed are prefixed by s (smoothing splines):

```
gam(y ~ s(w) + s(x) + s(z), family = poisson)
```

## Modelling algorithms: RF

- Random Forest
- Recursive partitioning method particularly well-suited to small *n* large *p* problems
- Involve a set of classification (or regression) trees that are calculated on random subsets of the data, using a subset of randomly restricted and selected predictors for each split in each classification tree
- Better examine the contribution and behavior that each predictor has
- Conditional variable importance → Difference of the model accuracy before and after the random permutations, averaged over all trees in the forest, tells us how important that predictor is for determining the outcome

## Modelling algorithms: RF

Implementation in R

Package *randomForest*

Function *randomForest*

```
randomForest(y ~ x, data, ntree, importance = TRUE)
```

# x: Data frame or matrix of predictors

# y: Response vector (if a factor, classification is assumed)

# data: Optional data frame containing the variables in the model

# ntree: Number of trees to grow

# importance: Assess importance of predictors

# Modelling algorithms: BRT

- Boosted Regression Trees
- Aim to improve the performance of a single model by fitting many models and combining them for prediction
- Uses two algorithms: regression trees are from the classification and regression tree (decision tree) group of models, and boosting builds and combines a collection of models
- Boosting is a method for improving model accuracy, based on the idea that it is easier to find and average many rough rules of thumb, than to find a single, highly accurate prediction rule

Elith et al. (2008) Journal of Animal Ecology 77: 802–813

# Modelling algorithms: BRT

- BRT approach differs fundamentally from traditional regression methods that produce a single 'best' model → instead using the technique of boosting to combine large numbers of relatively simple tree models adaptively, to optimize predictive performance

- boosting is unique because it is sequential: it is a forward, stagewise procedure

- In boosting, models (e.g. decision trees) are fitted iteratively to the training data, using appropriate methods gradually to increase emphasis on observations modelled poorly by the existing collection of trees

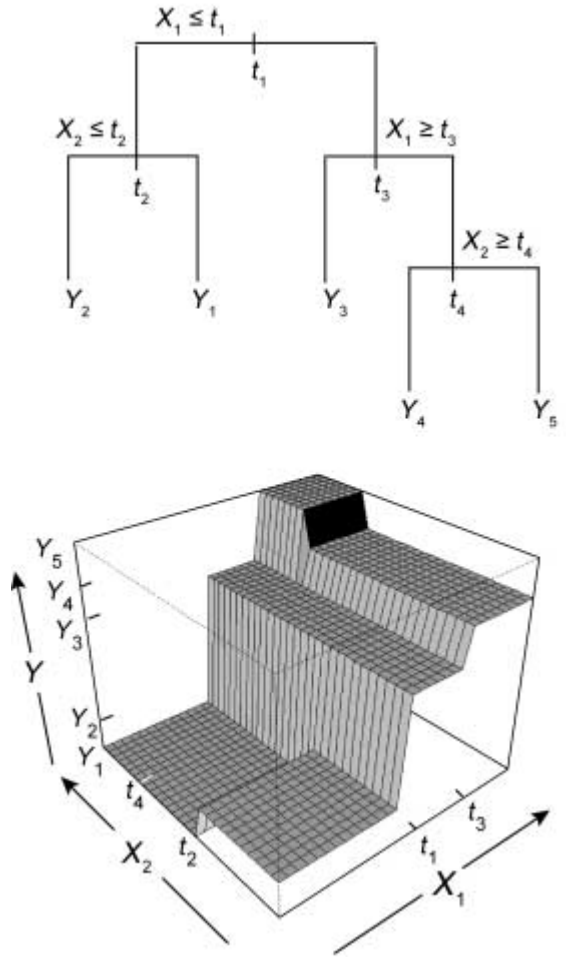



**Fig. 1.** A single decision tree (upper panel), with a response $Y$, two predictor variables, $X_1$ and $X_2$ and split points $t_1$, $t_2$, etc. The bottom panel shows its prediction surface (after Hastie *et al*. 2001)

Elith et al. (2008) Journal of Animal Ecology 77: 802–813

## Modelling algorithms: BRT

Implementation in R

Package *gbm*

Function *gbm*

```
gbm.step(data, gbm.x, gbm.y, family = "bernoulli",
tree.complexity = 5, learning.rate = 0.05, bag.fraction = 0.5)
```

# data: Optional data frame containing the variables in the model

# gbm.x: Predictors

# gbm.y: Response

# family: Name of the distribution

# bag.fraction: Proportion of observations used in selecting variables

# tree.complexity: Complexity of individual trees

# learning.rate: Weight applied to inidivudal trees

# Modelling algorithms: Ensemble approach

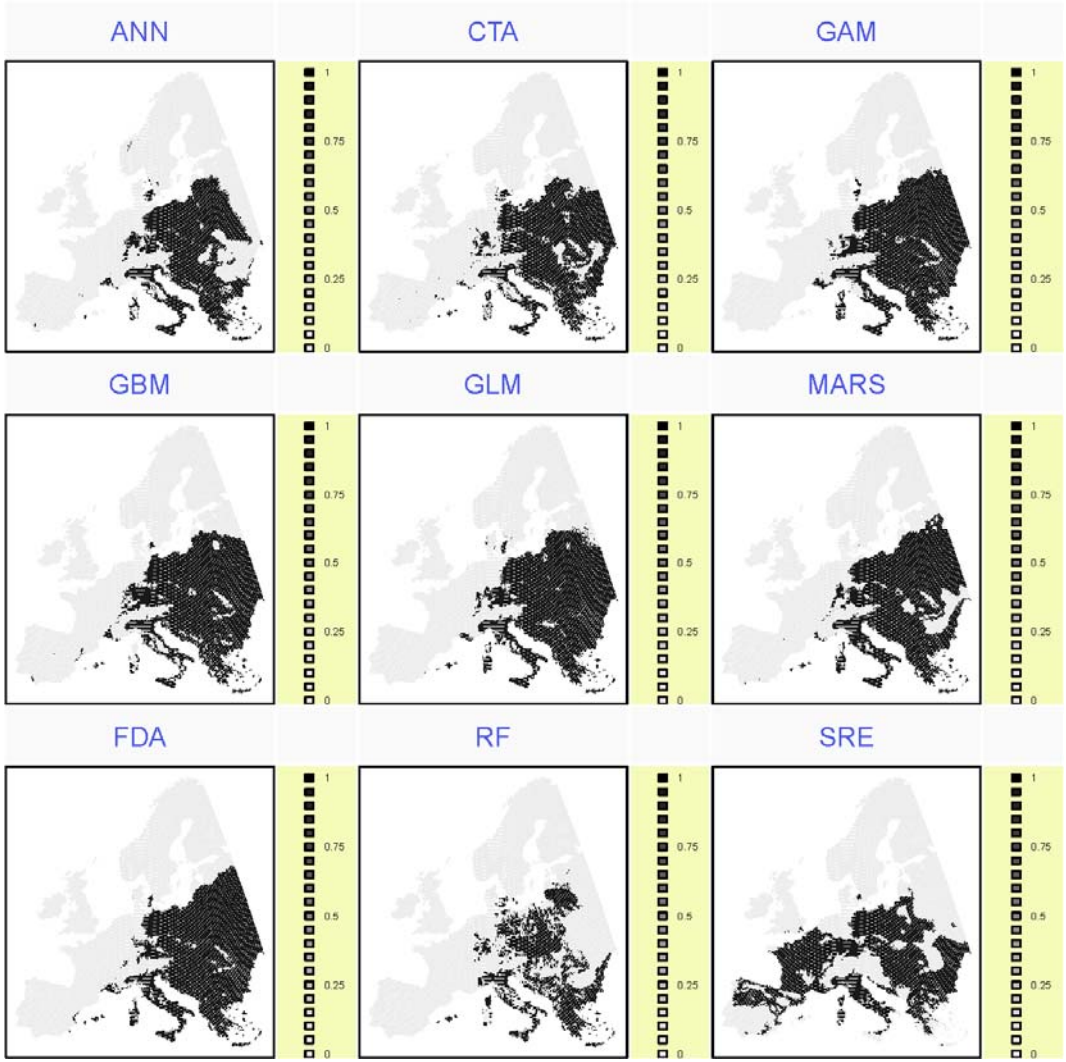
## A lot of algorithms out there

- Which is the best? (if any)
- Is there any model better than the others in a consistent way?
- How to compare them in a comprehensive framework?
- How to use them all together?
- What is the uncertainty associated with the use of one particular technique?

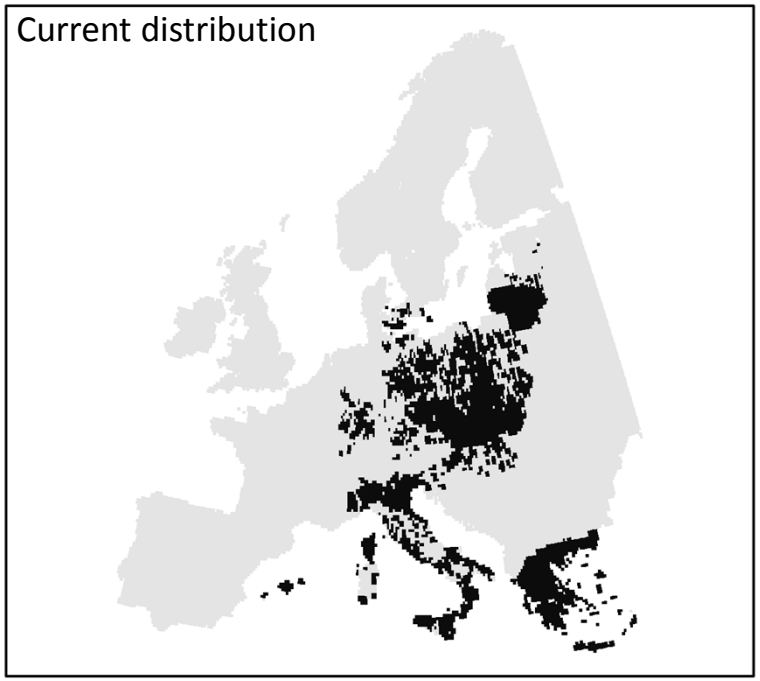
➢ BIOMOD (R-Package), since July 2012: biomod2 (replaces BIOMOD)

# Modelling algorithms: Ensemble approach

Green toad (*Bufo viridis*)

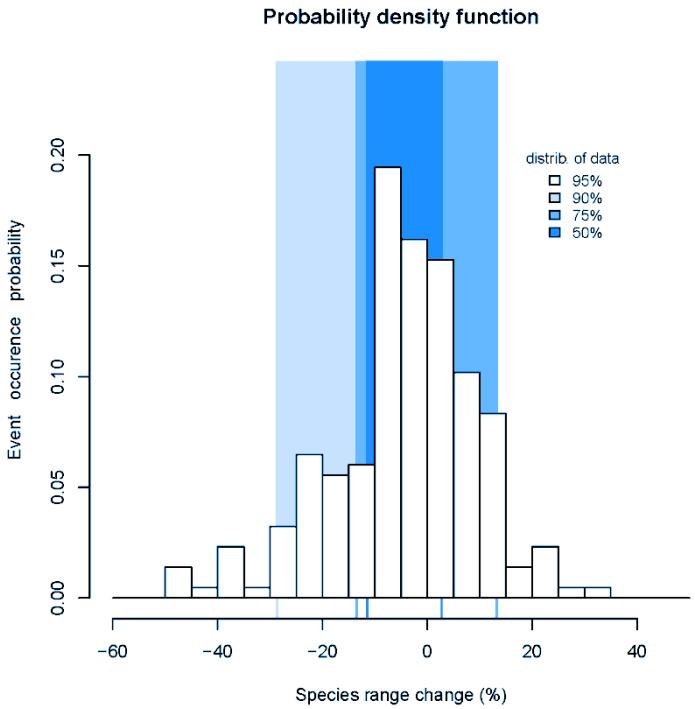


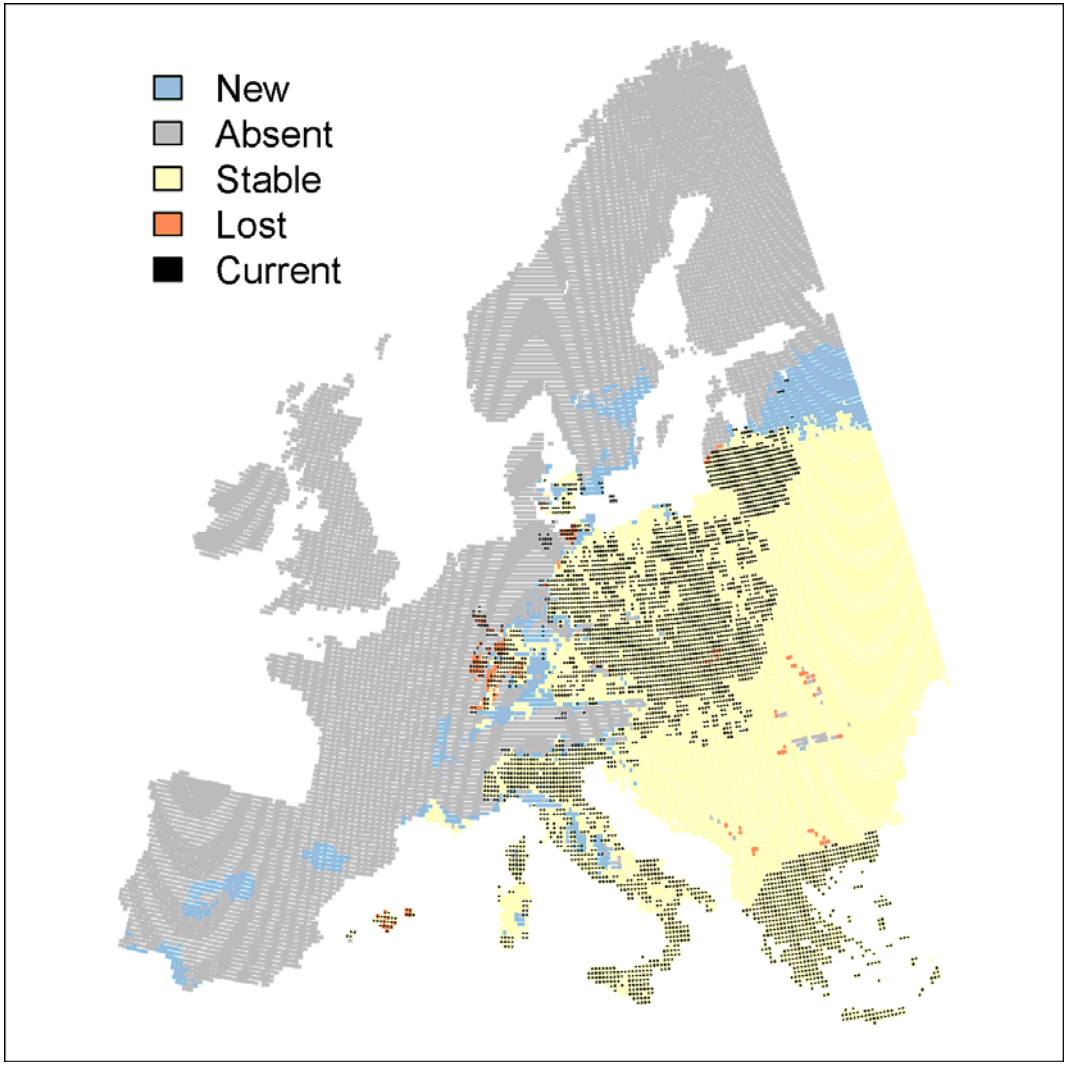




# Modelling algorithms: Ensemble approach

- HadCM3, A2, 2021-50
- No and full dispersal
- Range of predictions



# Part V

# Summary and Outlook

## Summary: Species distribution models

- Suitable for a wide range of applications: conservation, invasive species, ecological theory, future projections, …
- Estimate the actual or potential geographic distribution of a species by correlating the observed occurrence with environmental variables
- Several necessary steps (data, preprocessing, choice of algorithm, climate model, …) with different uncertainties
- A lot of modelling algorithms with different characteristics → possibility to build ensembles

# Summary: Process-based models

- Incorporate physiologically limiting mechanisms
- Developed to model key growth processes
- Appropriate approach for management questions
- Analog of the real system at a specified level of hierarchy

<u>However:</u>

- Require high quantity of detailed data
- Processes are not necessarily always relevant to explain problems at hand
- Models are inherently incomplete

# Outlook

Hybrid models / Dynamic Range Models

- Developed to overcome the problem that SDMs do not take into account processes

- Statistically estimate both range dynamics and the underlying environmental response of demographic rates from species distribution data

➢ Process-based statistical approach

## Outlook

→ Formulate a model that describes the link between environmental variation and biogeographical data in three submodels

1. A demographic response model

→ Describes how spatio-temporal variation in the environment translates into spatio-temporal variation in birth, death and dispersal

2. A range dynamics model

→ Describes how spatio-temporal variation in population growth and dispersal determine the spatio-temporal distribution of a local population size

## Outlook

3. An observation model

→ Describes how the variation in population size and demographic rates is sampled to obtain the available data (e.g. presence/absence maps or time series of local abundances)

Submodels form a hierarchy in the sense that the predictions of the demographic response model are input for the range dynamics model, whose output is in turn input for the observation model

# Outlook

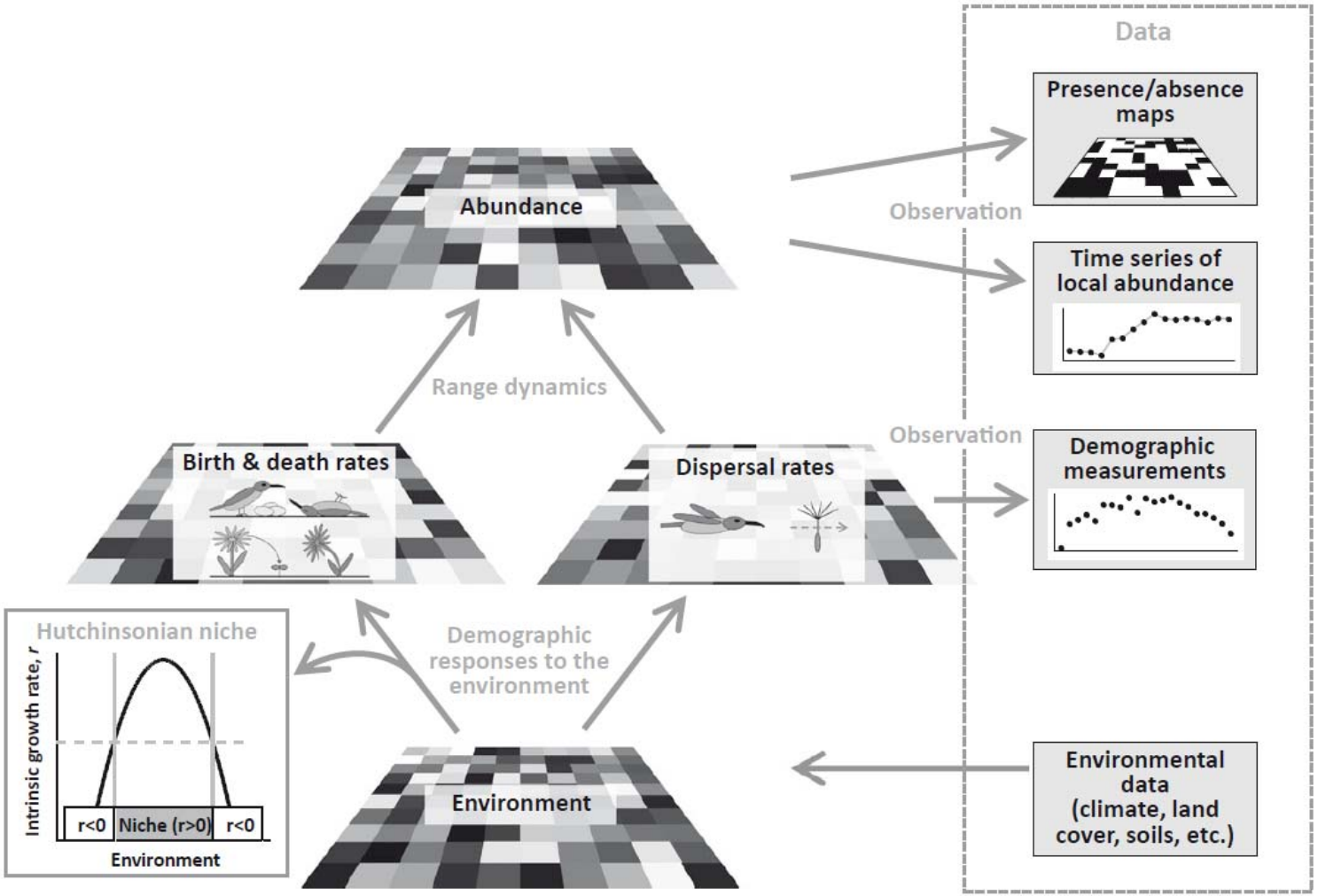
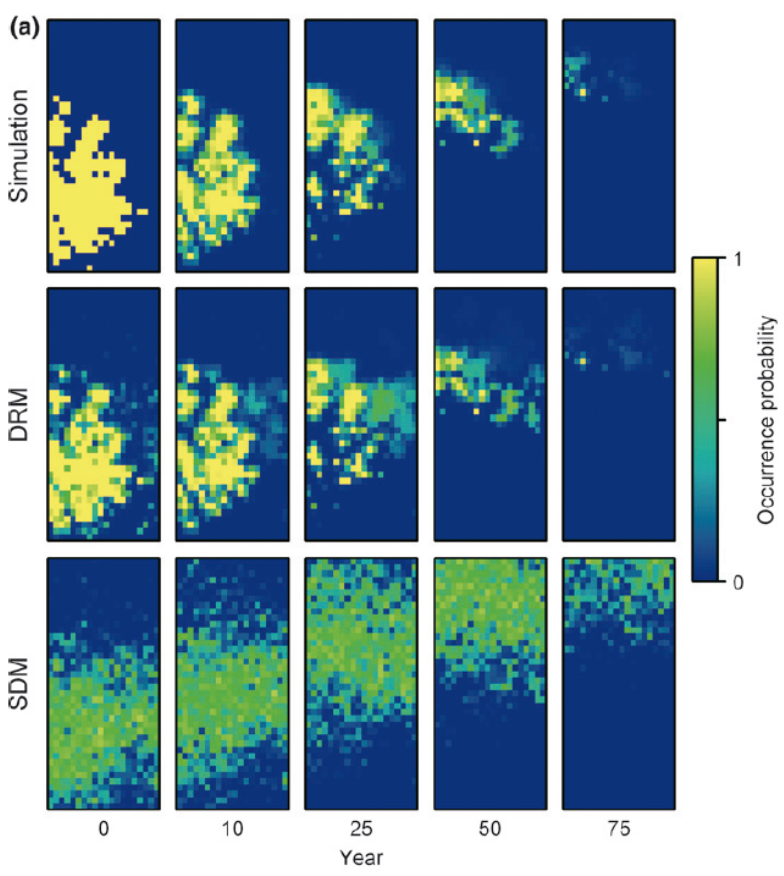



Fig.: The demographic basis of Hutchinsonian niches, range dynamics and biogeographical data. Demographic response functions translate spatio-temporal variation of the environment into variation of the fundamental demographic rates of birth, death and dispersal. […] Range dynamics then result from the dynamics of local populations that are coupled by dispersal. […]

Schurr et al. (2012) J. Biogeogr. 39: 2146-2162

# Outlook

The sample from the posterior can be used to calculate niche estimates (b), to forecast range dynamics (a), and to quantify the uncertainty in these forecasts (c)



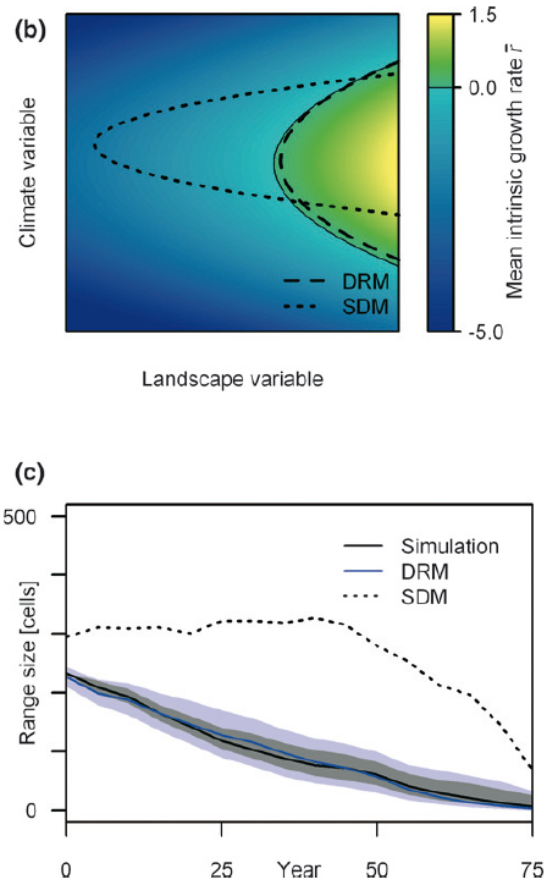




Fig.: An example analysis demonstrating DRMs can be used to estimate Hutchinsonian niches and to forecast range dynamics. (a) 'True' (simulated) versus predicted range dynamics of a hypothetical study species. (b) Estimates of the species' Hutchinsonian niche. (c) Dynamics of future range size for replicate simulations of the true model versus the corresponding forecasts of the DRM and SDM.

Schurr et al. (2012) J. Biogeogr. 39: 2146-2162

# Outlook

Advantages of Hybrid models / DRMs:

- Can statistically synthesize different types of biogeographical and demographic data

- Can infer spatio-temporal range dynamics in equilibrium and non-equilibrium conditions

- Provide estimates of a species' realized Hutchinsonian niche that are founded in ecological theory

- Yield fully probabilistic forecasts of future range dynamics under environmental change that transparently quantify the involved uncertainty

## Outlook

However:

- Currently, such process-based statistical analyses are only applicable to species with simple life cycles for which sufficient data exist

- Application to a broader range of study species and study systems requires substantial efforts in demographical research

# Thank you for your attention!