Species Distribution Models – Putting Theory into Practice

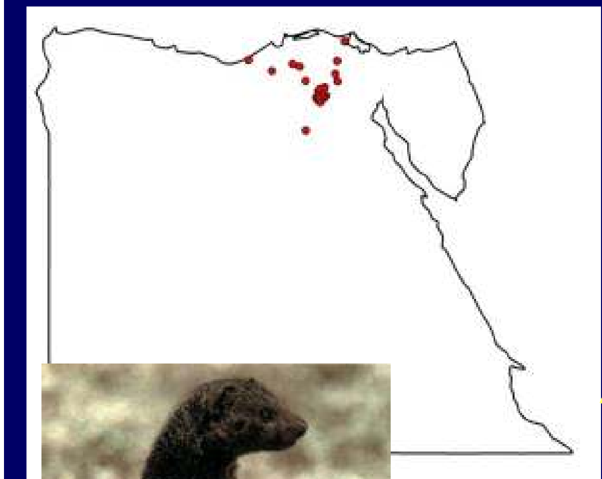
Tim Newbold

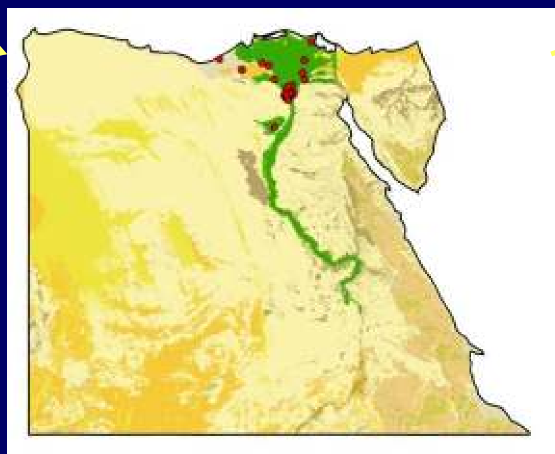
Outline

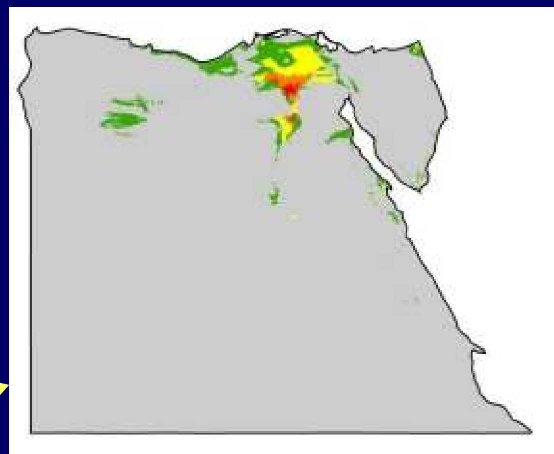
- How to build a distribution model
- Testing the accuracy of predictions
- Conservation applications
- Challenges

How to Build a Distribution Model

Data on where the species is known to occur → + Data on the environment at these locations → A model to relate them – what does the species require? → Produce a map of predicted distribution

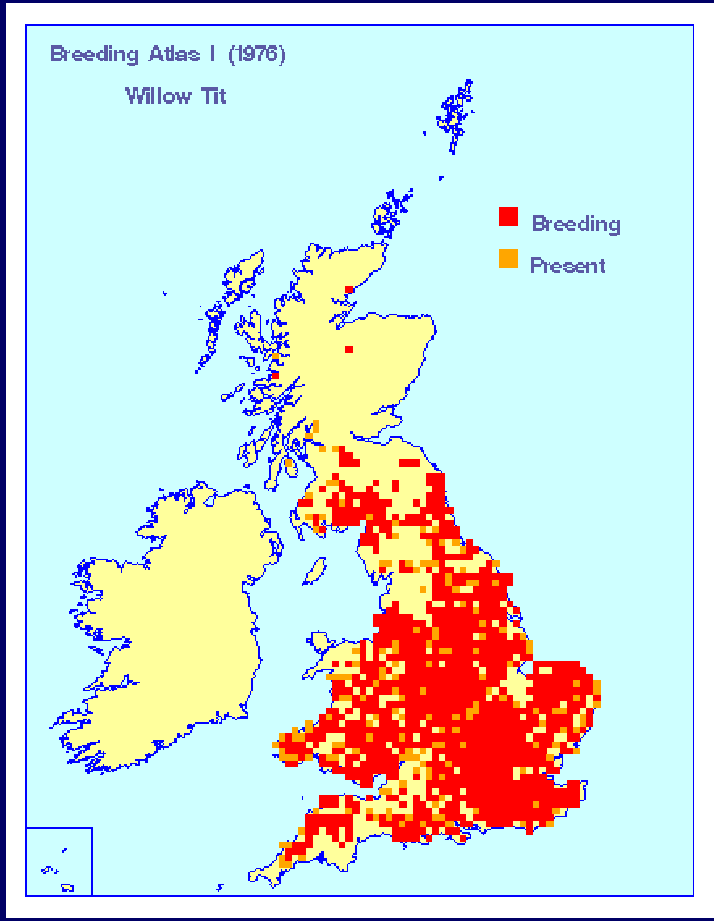






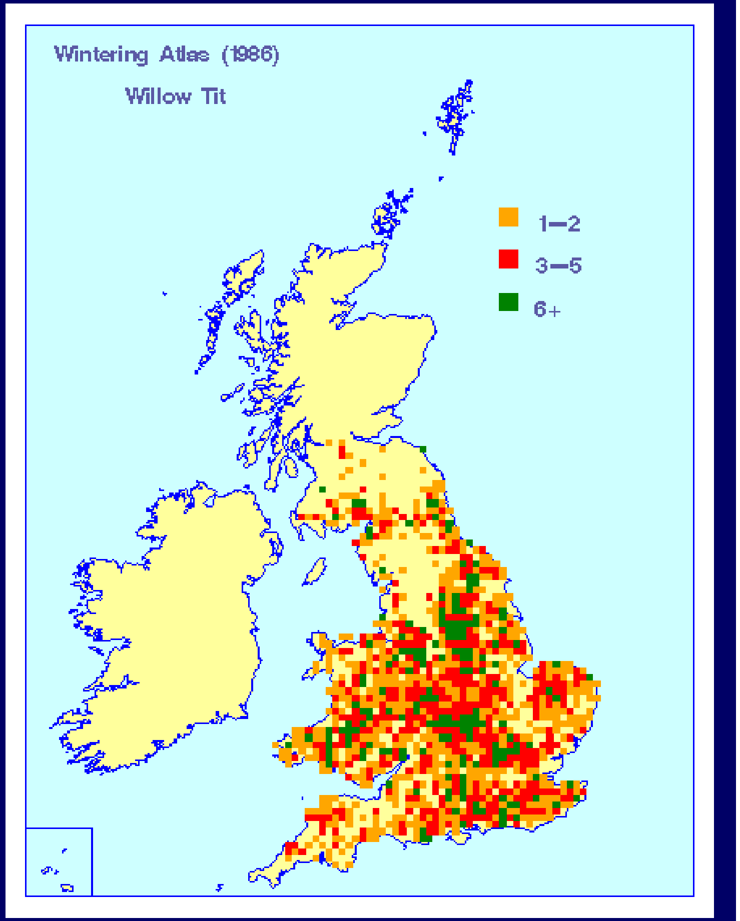
How to Build a Distribution Model

- Species occurrence data…



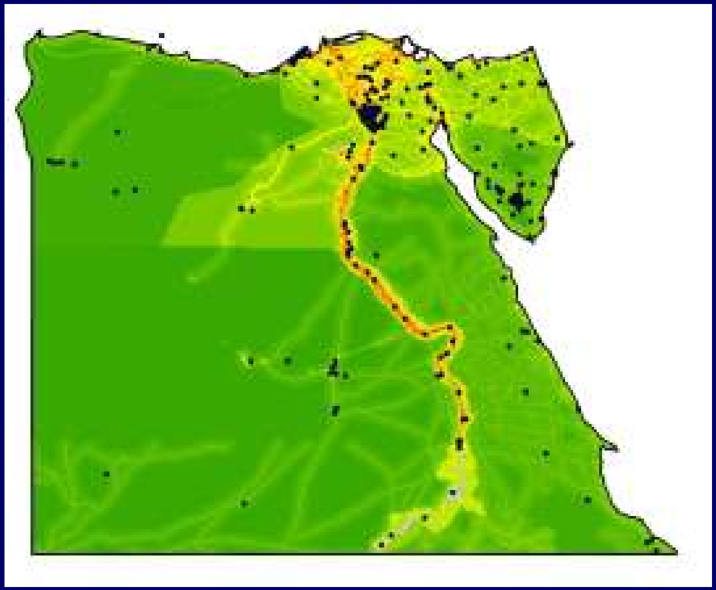
How to Build a Distribution Model

- Species occurrence data…
- …or abundance data



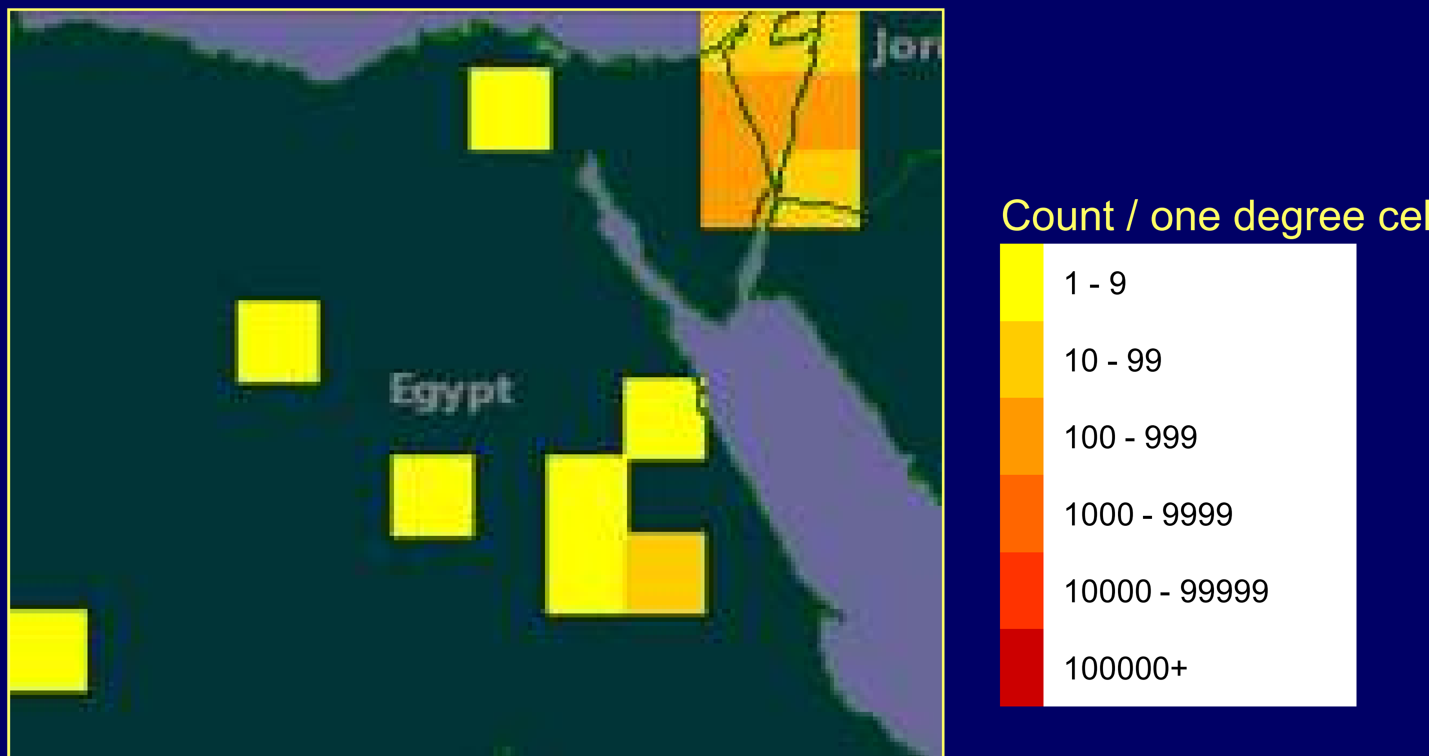
How to Build a Distribution Model

- **Records from museums or collections** (e.g. Raxworthy et al., 2003)
  - Widely available
  - Errors in locations
  - Errors in species identification
  - Only records of presence
  - Bias
  - Graham et al. (2004)



How to Build a Distribution Model

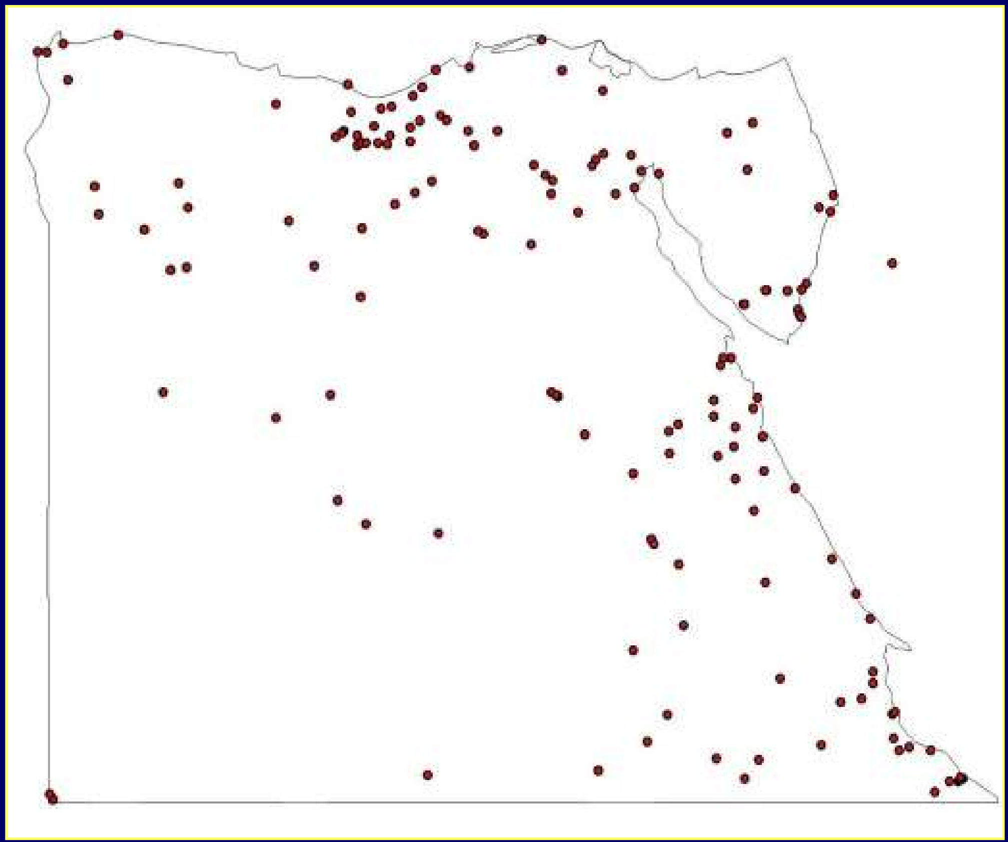
- **Records from museums or collections**
  - **Global Biodiversity Information Facility (www.gbif.org)**



Gazella dorcas

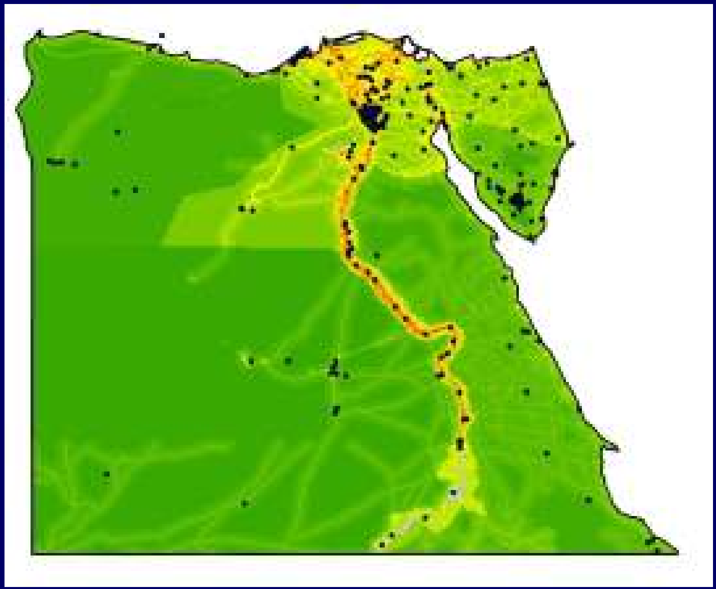
How to Build a Distribution Model

- **Records from museums or collections**
  - **Detailed country-level data**



How to Build a Distribution Model

- Records from museums or collections (e.g. Raxworthy et al., 2003)
  - Widely available
  - Errors in locations
  - Errors in species identification
  - Only records of presence
  - Bias
  - Graham et al. (2004)



How to Build a Distribution Model

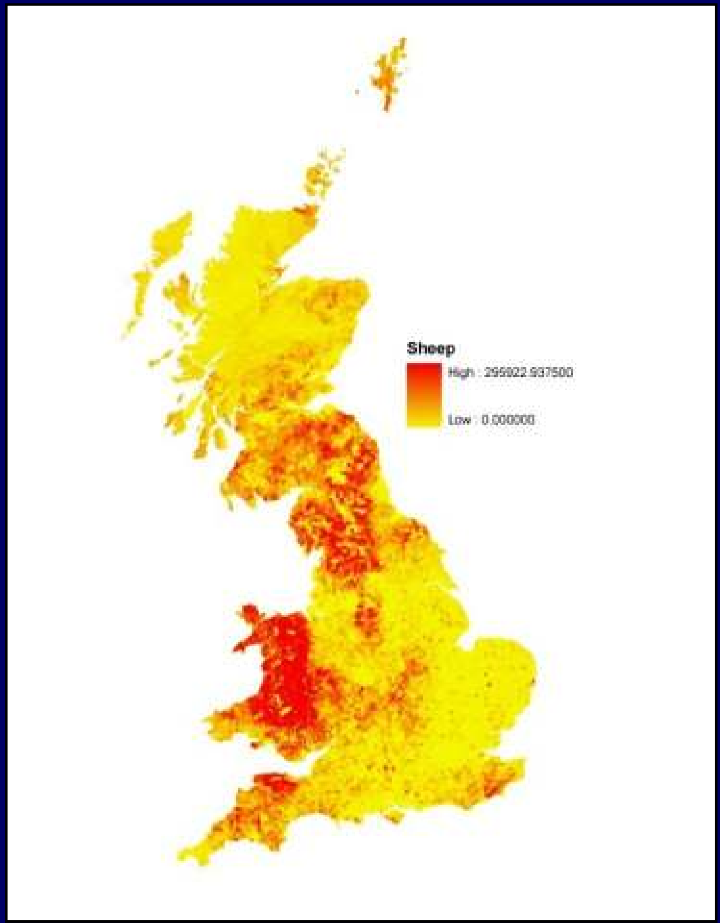
- Bias in species data
  - Sites closer to roads & cities
  - More sites in protected areas than expected by chance
  - Soberon et al. (2000)
  - Reddy & Davalos (2003)

How to Build a Distribution Model

- **Records from specially-designed surveys** (e.g. Fleishman et al., 2001)
  - Choose sites to cover the whole range of environments
  - GPS technology
  - Wintle et al. (2005)
  - Presences and absences
  - Reliability of absence data (species may be undetected in suitable habitat) Hirzel et al. (2002)

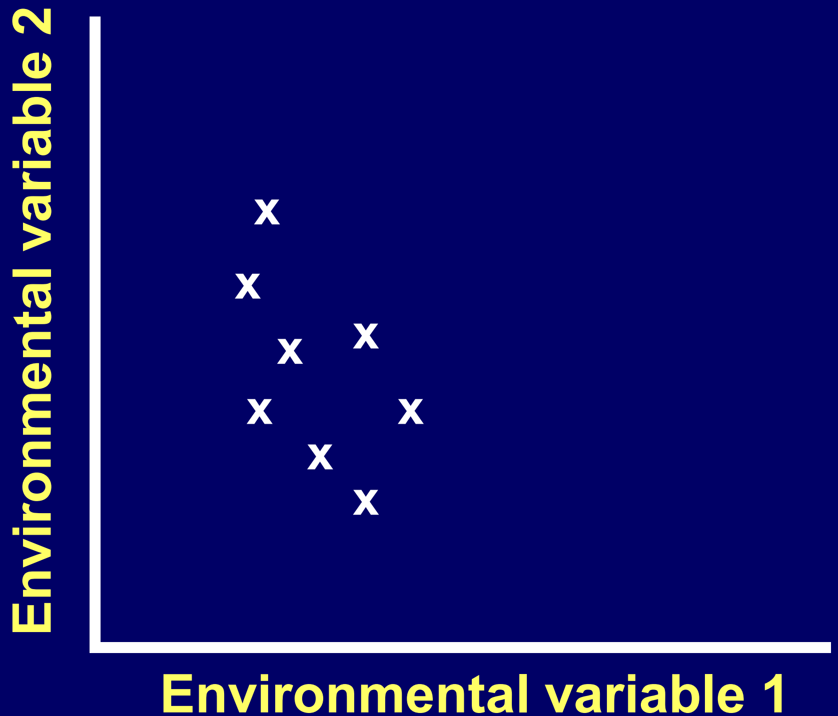
How to Build a Distribution Model

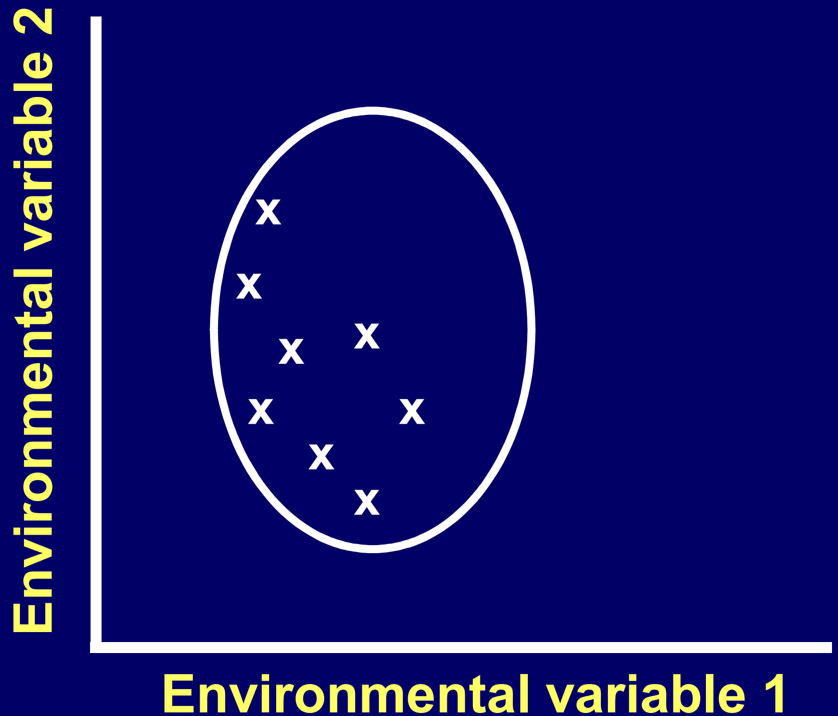
- **Environmental data**
  - – Climate
  - – Habitat
  - – Topography
  - – Agriculture
- **Remote Sensing**



How to Build a Distribution Model

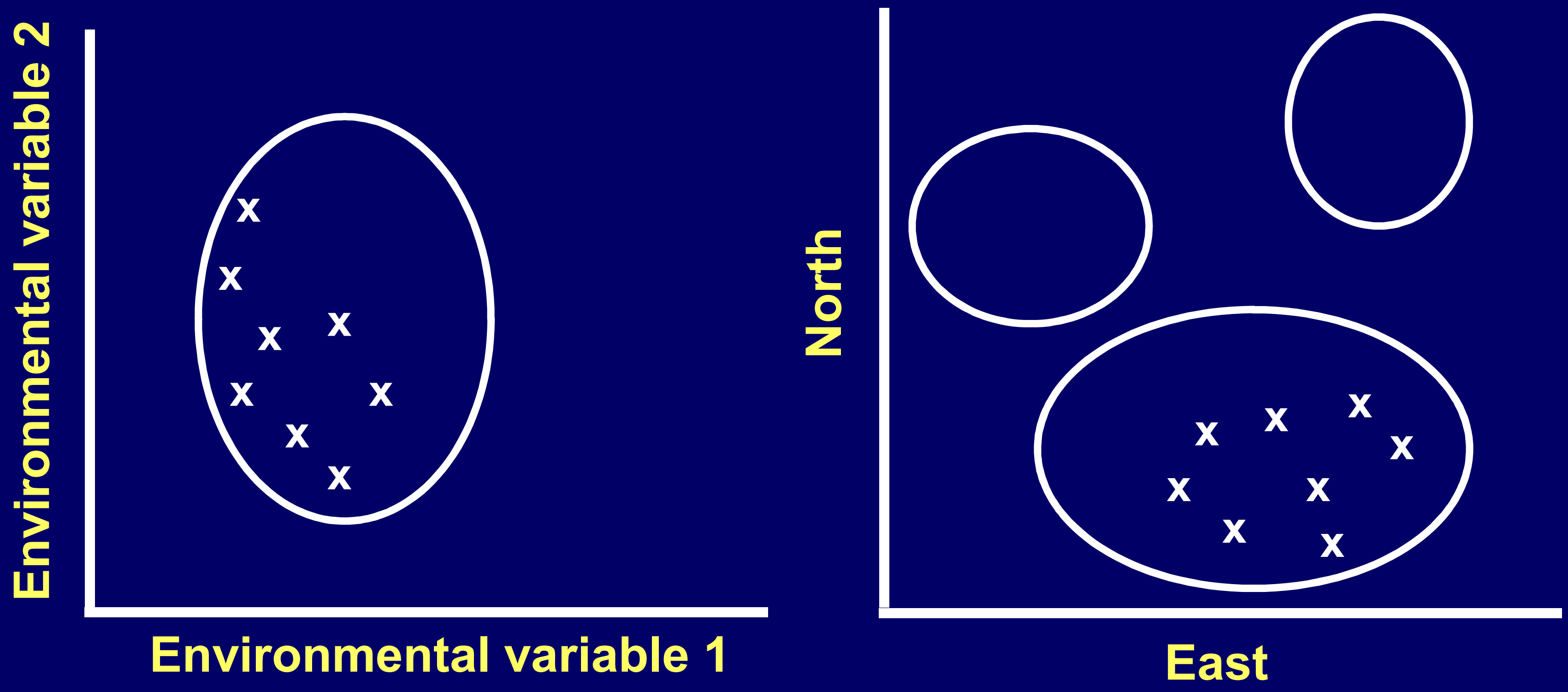
- **Models**
  - **Bioclimatic envelope models** (Nix, 1986)





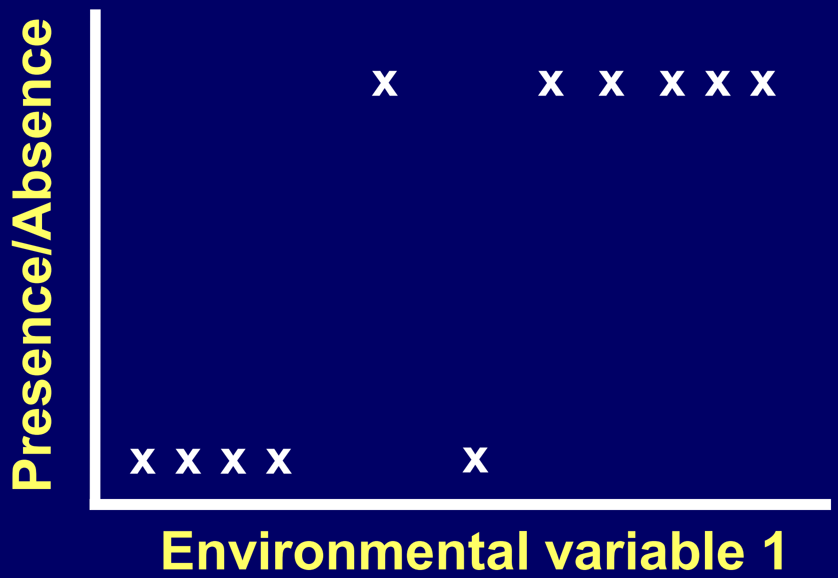
How to Build a Distribution Model

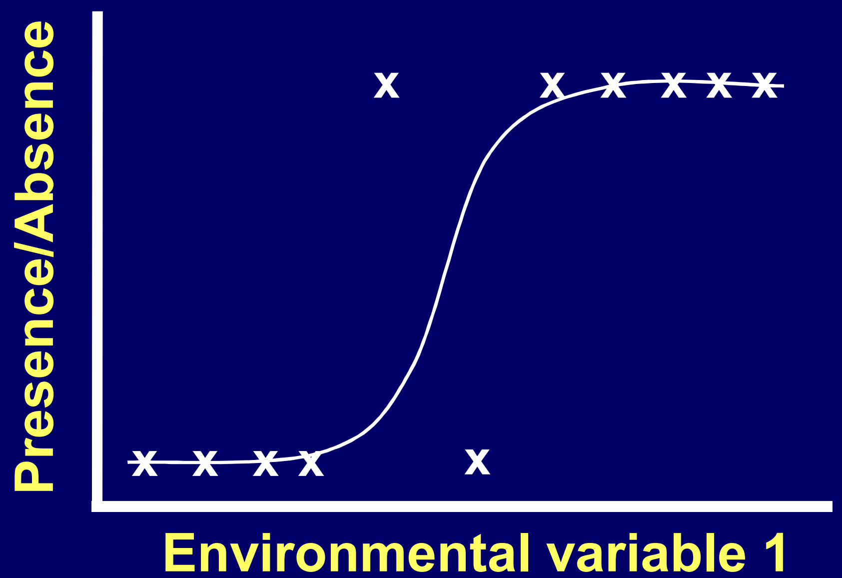
- **Models**
  - Projected into geographical space



How to Build a Distribution Model

- **Models**
  - Statistical Models
  - General linear models – normal error distribution
  - Generalized linear models – other error structures e.g. Poisson (abundance), binomial (presence/absence) – McCullagh & Nelder (1989)





How to Build a Distribution Model

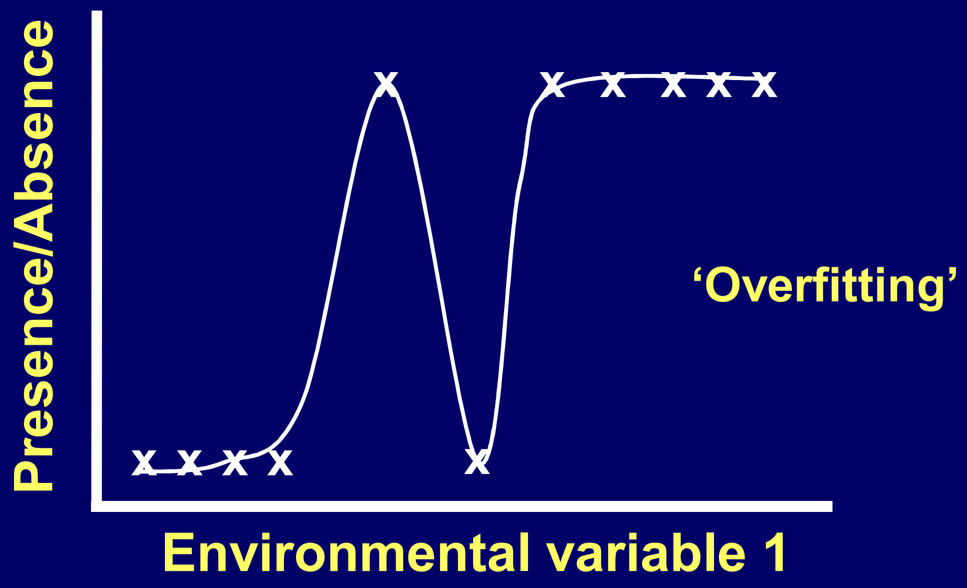
- Models
  - Newer software e.g. Maxent, Genetic Algorithm for Ruleset Process (GARP)
  - ‘Black box’ methods
  - Generally give good results (Elith et al., 2006)

Testing the Predictions

- Predictions are tested against known species occurrences
- Presences should have high predicted probability
- Absences should have low probability

Testing the Predictions

- **Using model data may give over-optimistic estimate of accuracy** (Araujo et al., 2005)



Testing the Predictions

- Split the data before modelling (e.g. Pearson et al., 2002)
- Reserve some for model testing
- Cross-validation

- **Cross-validation**

Species	Longitude	Latitude
Agrodiaetus loewii	34.11644	24.98213
Agrodiaetus loewii	34.00336	28.49633
Agrodiaetus loewii	33.70000	30.60000
Agrodiaetus loewii	33.90000	29.80000
Agrodiaetus loewii	33.23000	29.00000
Agrodiaetus loewii	34.04000	28.46000
Agrodiaetus loewii	32.48360	25.28298
Agrodiaetus loewii	31.41422	30.18839
Agrodiaetus loewii	31.29796	30.02098
Agrodiaetus loewii	31.36520	29.88784
Agrodiaetus loewii	31.29429	29.95519
Agrodiaetus loewii	34.23926	28.80688
Agrodiaetus loewii	31.53365	30.08264
Agrodiaetus loewii	33.06748	30.06376
Agrodiaetus loewii	32.20481	29.77734
Agrodiaetus loewii	34.12095	24.98984
Agrodiaetus loewii	32.63883	29.86738
Agrodiaetus loewii	33.60231	28.71970
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.67405	28.68044
Agrodiaetus loewii	33.93776	28.54843
Agrodiaetus loewii	33.93269	28.53318
Agrodiaetus loewii	33.95000	28.51667
Agrodiaetus loewii	33.57400	28.75170
Agrodiaetus loewii	30.51567	25.33337
Agrodiaetus loewii	31.30000	30.05000
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.91701	28.60719

Reserved Test Data

- **Cross-validation**

Reserved Test Data

Species	Longitude	Latitude
Agrodiaetus loewii	34.11644	24.98213
Agrodiaetus loewii	34.00336	28.49633
Agrodiaetus loewii	33.70000	30.60000
Agrodiaetus loewii	33.90000	29.80000
Agrodiaetus loewii	33.23000	29.00000
Agrodiaetus loewii	34.04000	28.46000
Agrodiaetus loewii	32.48360	25.28298
Agrodiaetus loewii	31.41422	30.18839
Agrodiaetus loewii	31.29796	30.02098
Agrodiaetus loewii	31.36520	29.88784
Agrodiaetus loewii	31.29429	29.95519
Agrodiaetus loewii	34.23926	28.80688
Agrodiaetus loewii	31.53365	30.08264
Agrodiaetus loewii	33.06748	30.06376
Agrodiaetus loewii	32.20481	29.77734
Agrodiaetus loewii	34.12095	24.98984
Agrodiaetus loewii	32.63883	29.86738
Agrodiaetus loewii	33.60231	28.71970
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.67405	28.68044
Agrodiaetus loewii	33.93776	28.54843
Agrodiaetus loewii	33.93269	28.53318
Agrodiaetus loewii	33.95000	28.51667
Agrodiaetus loewii	33.57400	28.75170
Agrodiaetus loewii	30.51567	25.33337
Agrodiaetus loewii	31.30000	30.05000
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.91701	28.60719

- **Cross-validation**

Reserved Test Data

Species	Longitude	Latitude
Agrodiaetus loewii	34.11644	24.98213
Agrodiaetus loewii	34.00336	28.49633
Agrodiaetus loewii	33.70000	30.60000
Agrodiaetus loewii	33.90000	29.80000
Agrodiaetus loewii	33.23000	29.00000
Agrodiaetus loewii	34.04000	28.46000
Agrodiaetus loewii	32.48360	25.28298
Agrodiaetus loewii	31.41422	30.18839
Agrodiaetus loewii	31.29796	30.02098
Agrodiaetus loewii	31.36520	29.88784
Agrodiaetus loewii	31.29429	29.95519
Agrodiaetus loewii	34.23926	28.80688
Agrodiaetus loewii	31.53365	30.08264
Agrodiaetus loewii	33.06748	30.06376
Agrodiaetus loewii	32.20481	29.77734
Agrodiaetus loewii	34.12095	24.98984
Agrodiaetus loewii	32.63883	29.86738
Agrodiaetus loewii	33.60231	28.71970
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.67405	28.68044
Agrodiaetus loewii	33.93776	28.54843
Agrodiaetus loewii	33.93269	28.53318
Agrodiaetus loewii	33.95000	28.51667
Agrodiaetus loewii	33.57400	28.75170
Agrodiaetus loewii	30.51567	25.33337
Agrodiaetus loewii	31.30000	30.05000
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.91701	28.60719

- **Cross-validation**

Reserved Test Data

Species	Longitude	Latitude
Agrodiaetus loewii	34.11644	24.98213
Agrodiaetus loewii	34.00336	28.49633
Agrodiaetus loewii	33.70000	30.60000
Agrodiaetus loewii	33.90000	29.80000
Agrodiaetus loewii	33.23000	29.00000
Agrodiaetus loewii	34.04000	28.46000
Agrodiaetus loewii	32.48360	25.28298
Agrodiaetus loewii	31.41422	30.18839
Agrodiaetus loewii	31.29796	30.02098
Agrodiaetus loewii	31.36520	29.88784
Agrodiaetus loewii	31.29429	29.95519
Agrodiaetus loewii	34.23926	28.80688
Agrodiaetus loewii	31.53365	30.08264
Agrodiaetus loewii	33.06748	30.06376
Agrodiaetus loewii	32.20481	29.77734
Agrodiaetus loewii	34.12095	24.98984
Agrodiaetus loewii	32.63883	29.86738
Agrodiaetus loewii	33.60231	28.71970
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.67405	28.68044
Agrodiaetus loewii	33.93776	28.54843
Agrodiaetus loewii	33.93269	28.53318
Agrodiaetus loewii	33.95000	28.51667
Agrodiaetus loewii	33.57400	28.75170
Agrodiaetus loewii	30.51567	25.33337
Agrodiaetus loewii	31.30000	30.05000
Agrodiaetus loewii	33.86655	28.52327
Agrodiaetus loewii	33.91701	28.60719

Testing the Predictions

- **Split the data before modelling (e.g. Pearson et al., 2002)**
- **Reserve some for model testing**
- **Cross-validation**
- **Biases in sampling → over-optimistic accuracy measures**

Testing the Predictions

- Collect new data
- Ideal but often impractical

Testing the Predictions

- **Statistics**
  - **Confusion matrix** (Fielding & Bell, 1997)
  - **Measures numbers of species presences and absences correctly predicted**

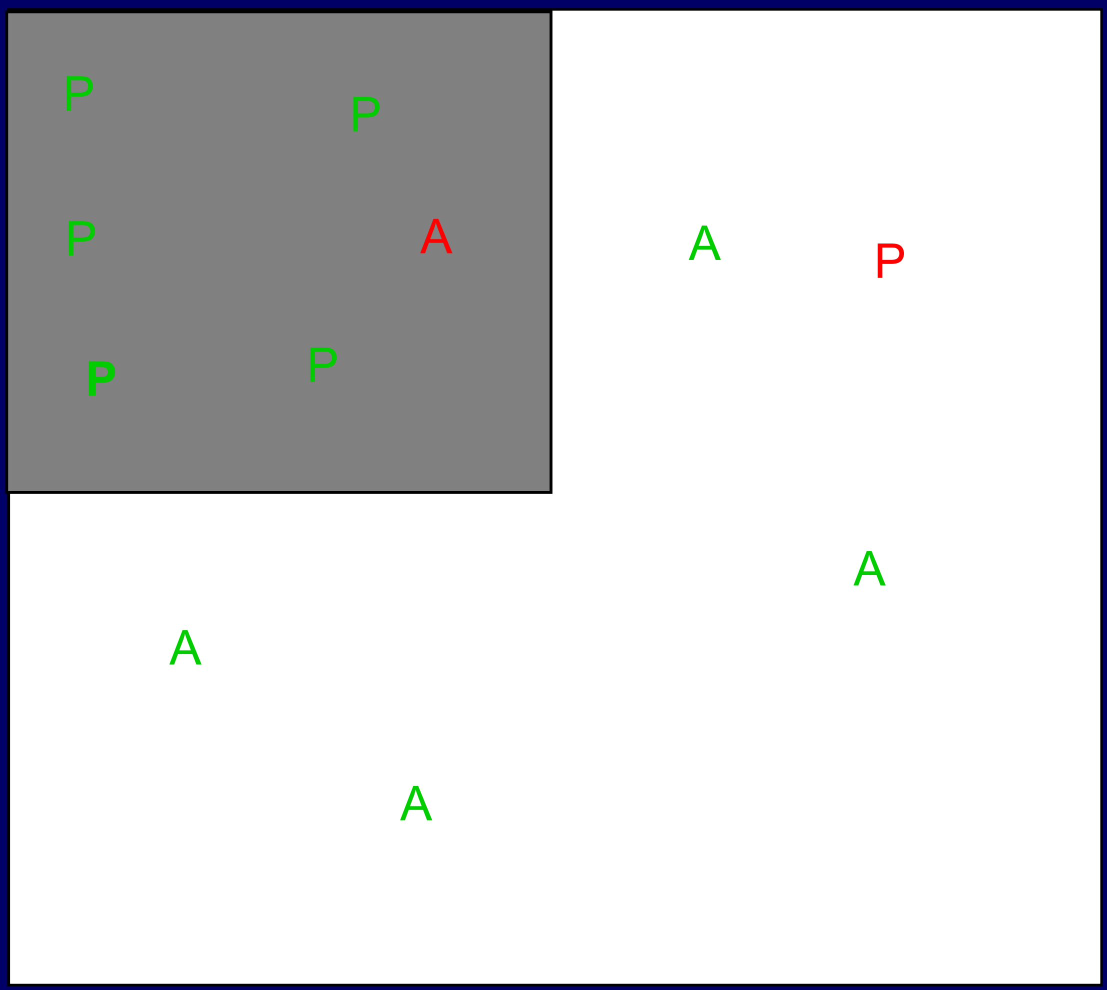
	Actually Present	Actually Absent
Predicted Present	a	b
Predicted Absent	c	d

Testing the Predictions

	Actually Present	Actually Absent
Predicted Present	a	b
Predicted Absent	c	d

- **Statistics**
  - **True positive rate** $\dfrac{a}{a+c}$
  - **True negative rate** $\dfrac{d}{b+d}$
  - **Correct classification rate (CCR)** $\dfrac{a+d}{a+b+c+d}$

Testing the Predictions



Present

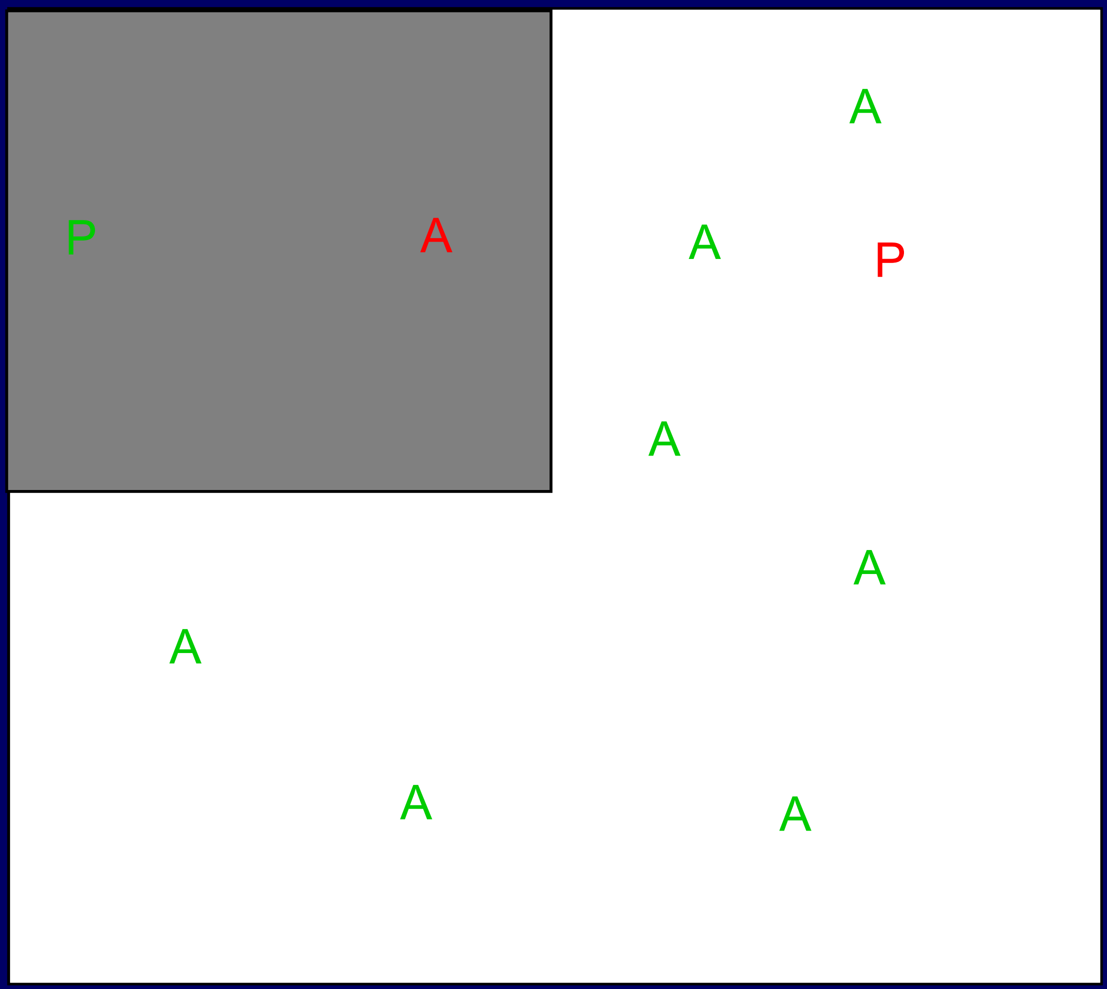
Absent

P – Presences

A – Absences

CCR = 80%

Testing the Predictions



Present

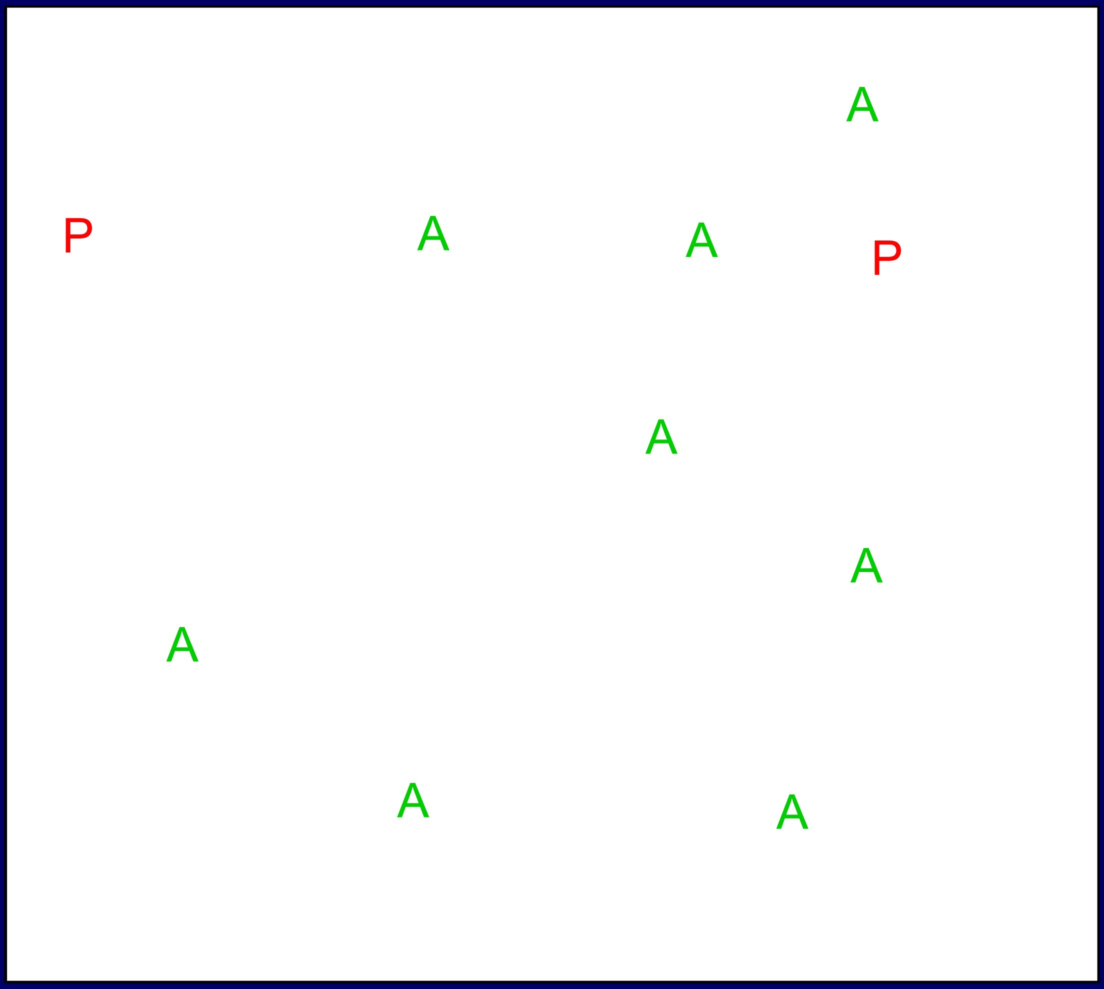
Absent

P – Presences

A – Absences

CCR = 80%

Testing the Predictions



Present

Absent

P – Presences

A – Absences

CCR = 80%

Testing the Predictions

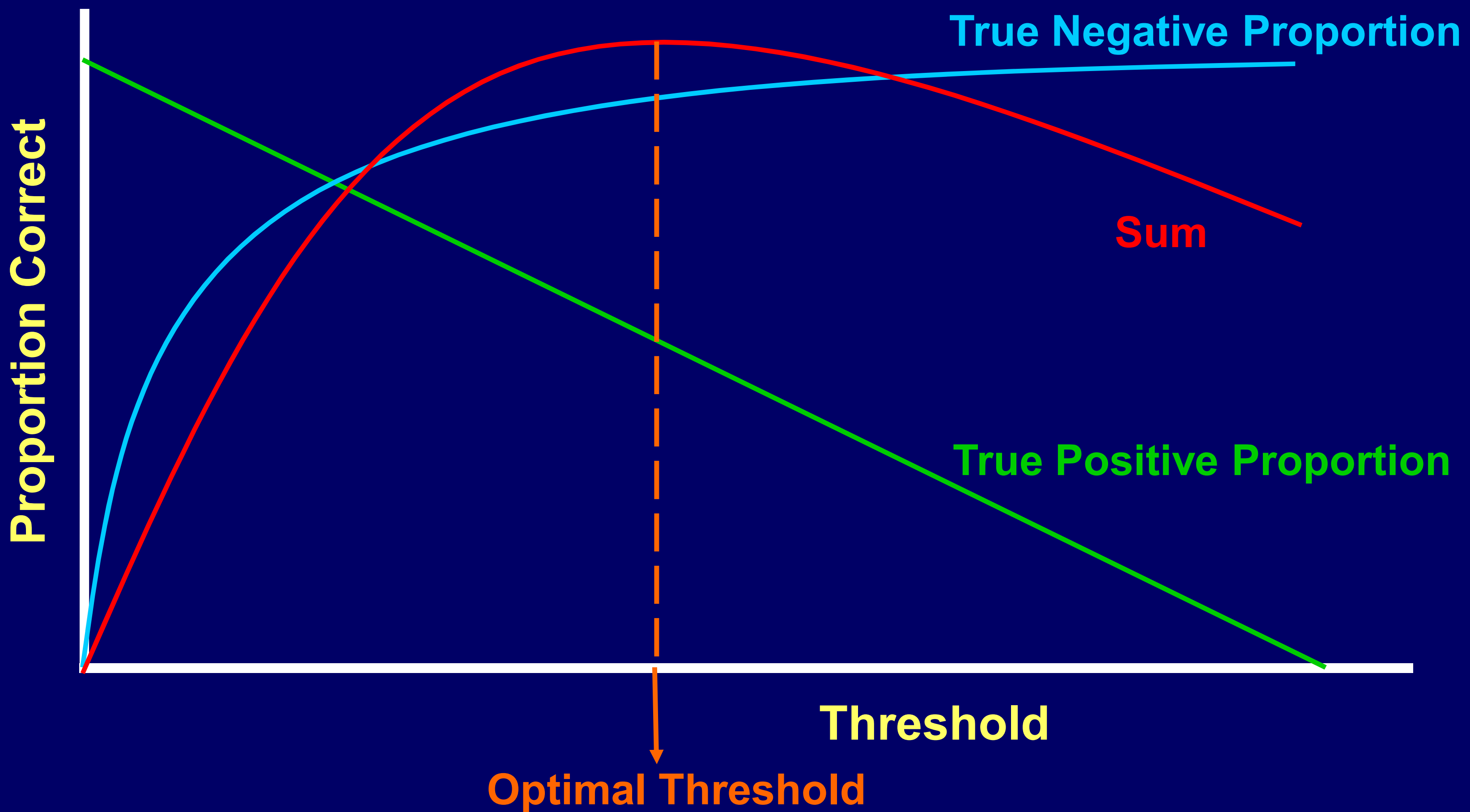
- **Statistics**
  - **But if only 5% of records are presences, can achieve 95% accuracy if whole area is predicted absent**
  - **Kappa statistic (Cohen, 1960) accounts for chance agreement:**

$$k = \frac{[(a+d) - (((a+c)(a+b) + (b+d)(c+d))/n)]}{[n - (((a+c)(a+b) + (b+d)(c+d))/n)]}$$

Testing the Predictions

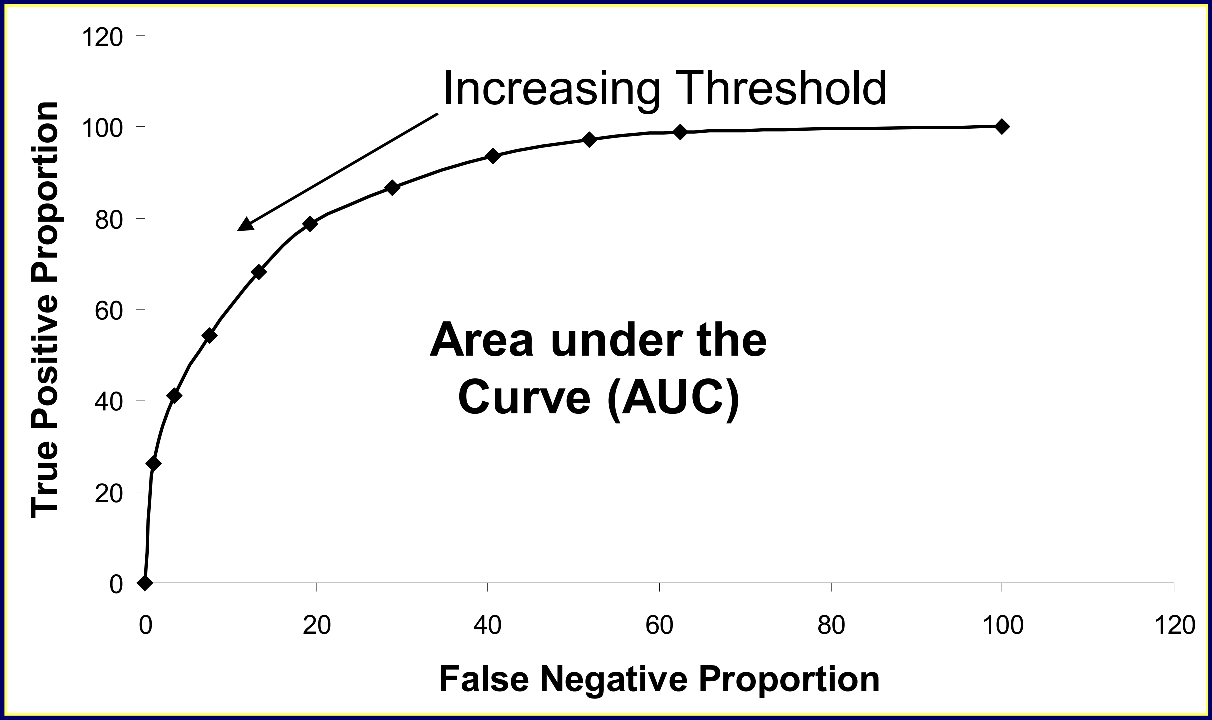
- Statistics
  - Confusion matrix measures require a threshold to turn predictions into presence/absence
  - Arbitrary thresholds are inappropriate
  - Some objective methods – e.g. maximize proportions of presences and absences correctly predicted (Liu et al., 2005)

Testing the Predictions



Testing the Predictions

- Statistics
  - Some don't require thresholds
  - E.g. Receiver Operating Characteristic (ROC) curves

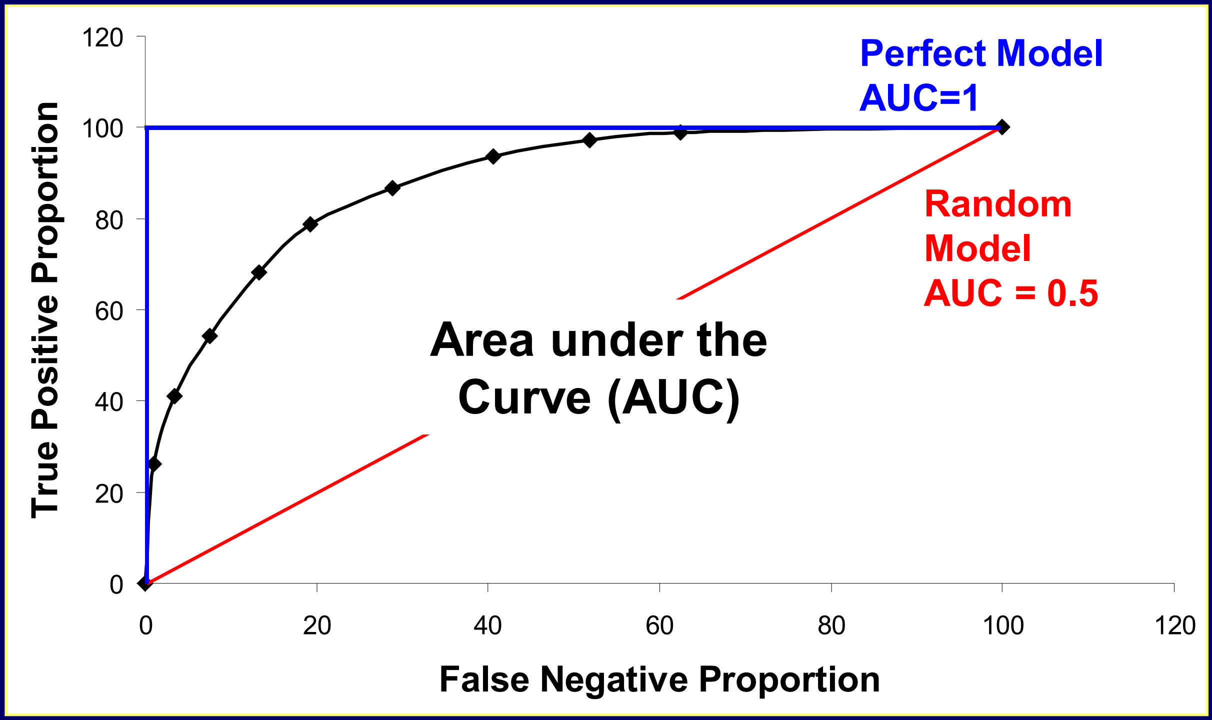


Testing the Predictions

- Statistics
  - Area Under the Curve (AUC) measures the probability that a random presence will have a higher model value than a random absence
  - See Fielding & Bell (1997) for complete discussion

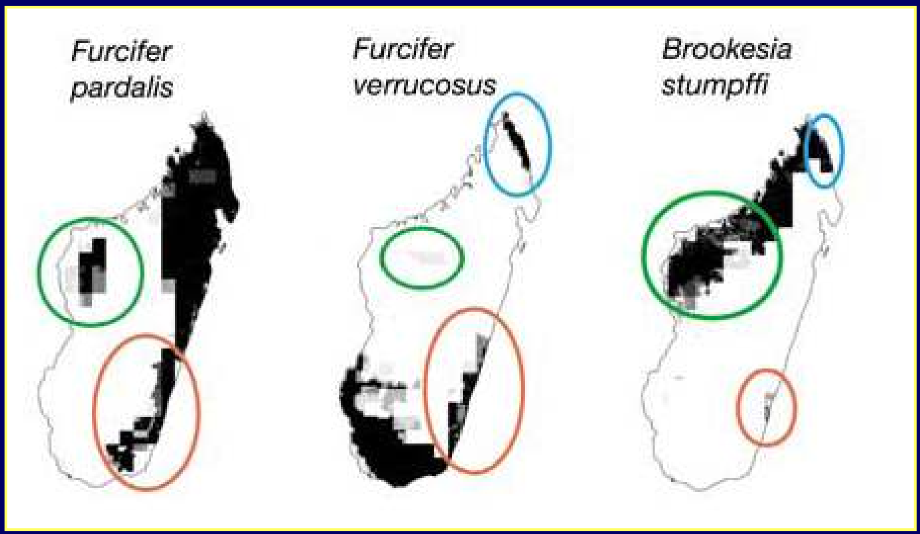
Testing the Predictions

- **Statistics**
  - Area under the ROC curve (AUC)



Applications

- **Directing future surveys**
  - **Raxworthy et al. (2003)**
  - **Modelled Madagascan chameleons**
  - **Found 7 new species**



Applications

- **Directing future surveys**
  - **Guisan et al. (2006)**
  - **Modelled Alpine sea holly (*Eryngium alpinum*)**
  - **Found 7 new populations**

Applications

- Predicting the impact of invasive species
  - Model requirements in native range
  - Apply to environment in new areas
  - Predict invaded range
  - Can evaluate against known invasions

Applications

- Predicting the impact of invasive species
  - Thuiller et al. (2005)
  - 96 S. African plant species
  - Predicted known invasions well

Applications

- **Predicting the impact of invasive species**
  - **Barred owl (*Strix varia*) predicted to invade range of endangered northern spotted owl (*Strix occidentalis*)**
  - **Known to hybridize**
  - **Peterson & Robins (2003)**

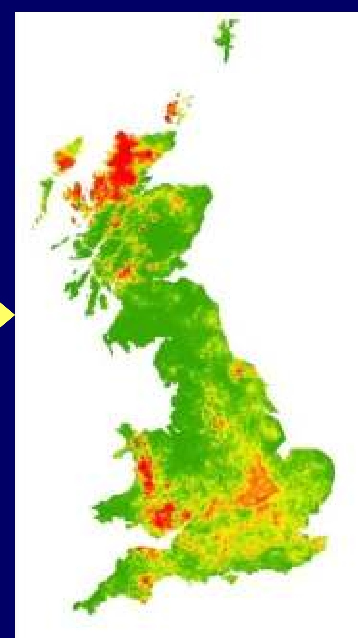
Applications

- **Guiding Species Reintroductions**
  - Klar et al. (2008)
  - European wildcats (*Felis silvestris*)
  - Large suitable patches for introduction

Applications

- Predicting the effects of climate change
  - Build model with species and environment data from one time period
  - Project it onto environment data for a second time period

1972-1977



1998-2002

Applications

Catherine Bennett in G2 In G2 In G2 Plus Online & jobs

55p
Thursday
January 8 2004
Published in London
and Manchester
guardian.co.uk

The Guardian

An unnatural disaster

- Global warming to kill off 1m species
- Scientists shocked by results of research
- Third of life forms doomed by 2050

Paul Brown
Environment correspondent

Climate change over the next 50 years is expected to drive a quarter of land animals and plants into extinction, according to the first comprehensive study into the effect of higher temperatures on the natural world.

The sheer scale of the disaster facing the planet shocked those involved in the research. They estimate that more than 1 million species will be lost by 2050.

The results are described as "terrifying" by Chris Thomas, professor of conservation biology at Leeds University, who is lead author of the research from four continents published today in the magazine Nature.

Much of that loss — more than one in 10 of all plants and animals — is already irretralia, Central and South America, and South Africa, showed that species living in mountainous areas had a greater chance of survival because they could simply move uphill to get cooler.

Those in flatter areas such as Brazil, Mexico and Australia, were more vulnerable, faced with the impossible task of moving thousands of miles to find suitable conditions.

Birds, which had the greatest chance of escape, could in theory move to a more suitable climate but the trees and other habitat they needed for survival could not keep pace and all would die.

Professor Thomas said: "When scientists set about research they hope to come up with definite results, but what we found we wish we had not. It was far, far worse than we thought, and what we have discovered may even be an

Continued

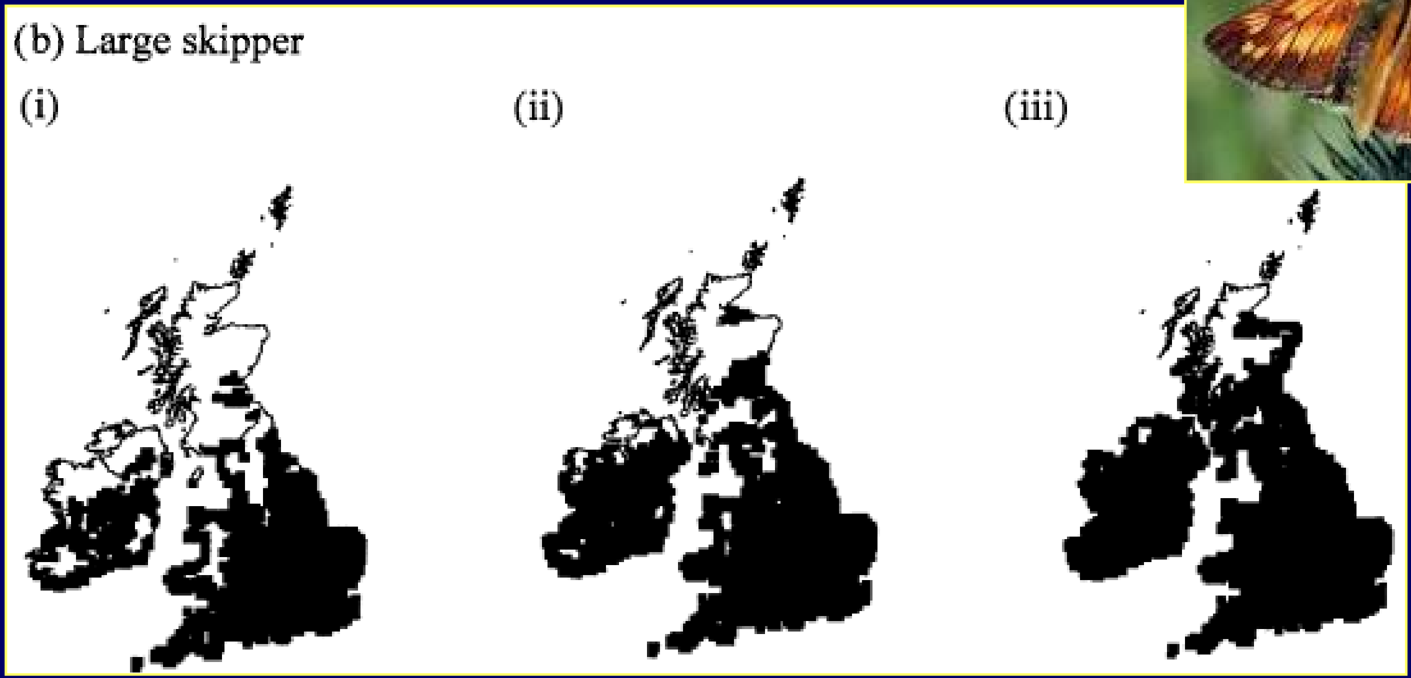
Under threat

Applications

- Predicting the effects of climate change
  - Berry et al. (2002)
  - 54 species
  - Current distributions for Europe
  - 2050s predictions in Britain

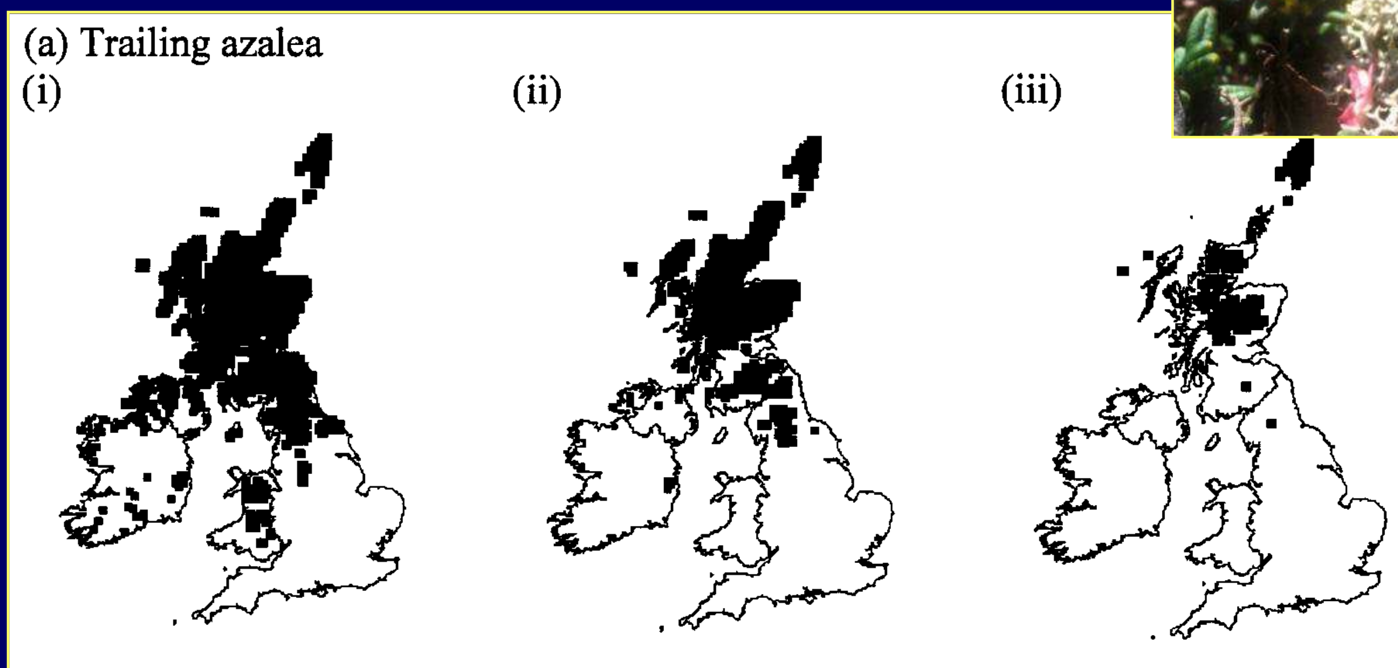
Applications

- **Predicting the effects of climate change**
  - **Some expansions**



Applications

- **Predicting the effects of climate change**
  - **Some contractions**

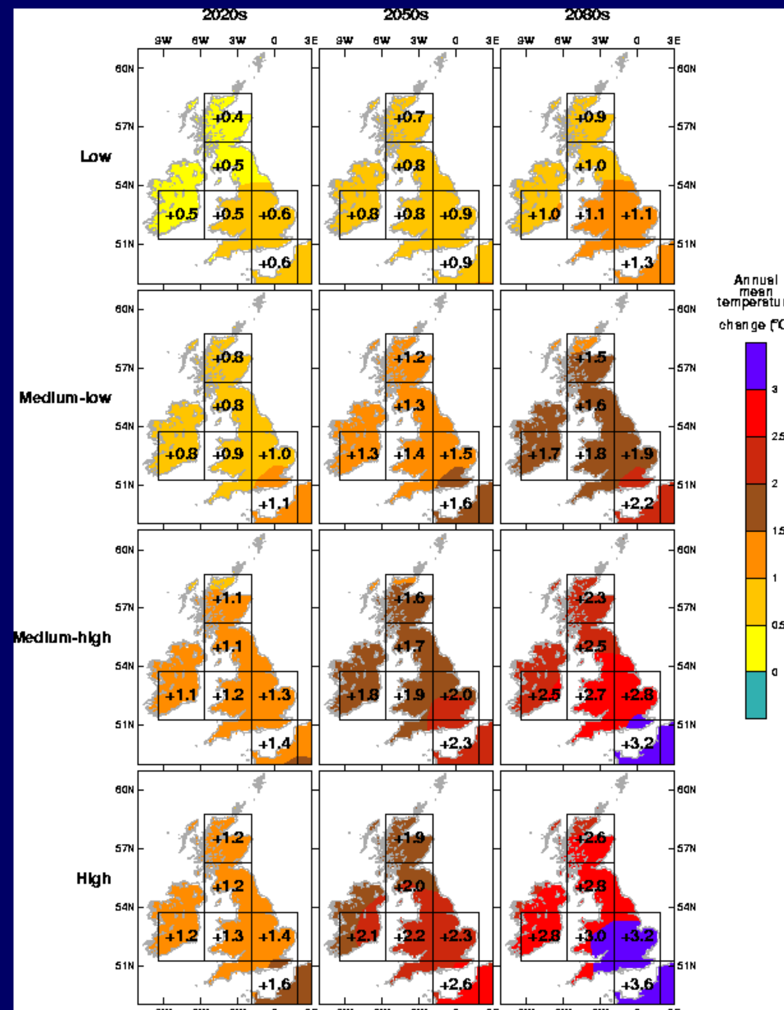


Applications

- Predicting the effects of climate change
  - Usually tested for current time period
  - Responses to the environment may change (adaptation, species interactions etc.)
  - Models are only correlative

Applications

- **Predicting the effects of climate change**
  - **Uncertainty**
  - **Climate change scenarios** (e.g. Hulme & Jenkins, 1998)

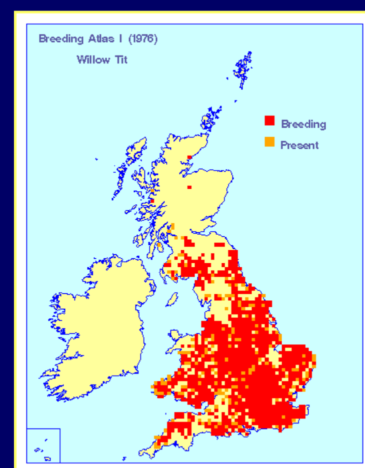


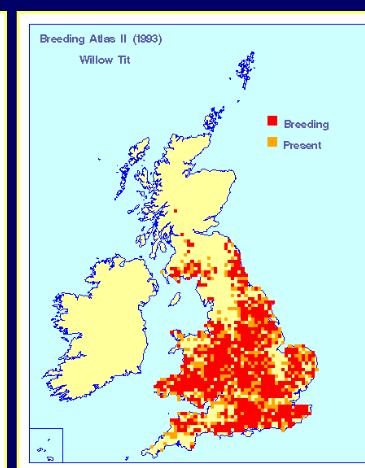
Applications

- Predicting the effects of climate change
  - Usually tested for current time period
  - Responses to the environment may change (adaptation, species interactions etc.)
  - Models are only correlative
  - Uncertainty about future changes
  - Solution – predict changes that have already happened.

Applications

- **Predicting the effects of climate change**
  - Araujo et al. (2005)
  - BTO breeding atlas data for 116 British birds
  - Modelled using 1968-72 data, projected to 1988-91
  - Predictions matched 1988-91 data fairly well (mean AUC = 0.78)





Caveats

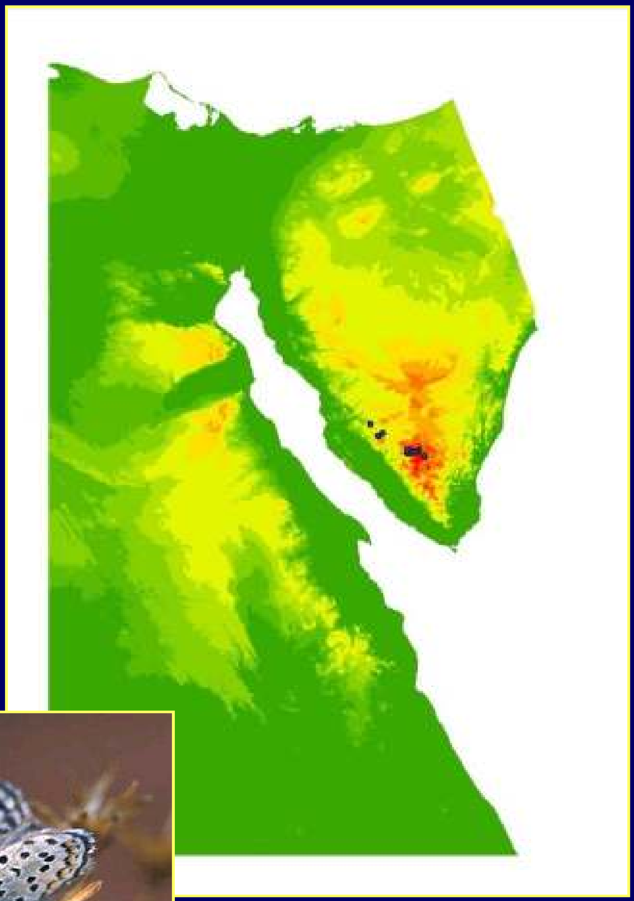
- Interactions among species
  - Competition
  - Predation
  - Other e.g. woodpeckers and owl nest sites (Heikkinen et al., 2007)

Caveats

- **Interactions among species**
  - **Incorporate other species' distributions**
  - **e.g. Leathwick & Austin (2001) – tree distns. including *Nothofagus* (southern beech) abundance**
  - **Araujo & Luoto (2007) – clouded Apollo butterfly and its host plant**

Caveats

- Dispersal history
  - Individuals may be prevented from reaching some suitable areas
  - Individuals may occupy some unsuitable areas close to suitable habitat
  - Leads to spatial autocorrelation



Caveats

- Spatial autocorrelation
  - Sites close together are more similar than those further apart
  - Autocorrelation in species distns. can occur because environment is autocorrelated (extrinsic)…
  - …or because of dispersal, speciation & extinction (intrinsic)

Caveats

- **Spatial autocorrelation**
  - **Solutions:**
  - **Include longitude, latitude and interaction terms in models**
  - **Autocovariate term – a weighted average of species presence or abundance in adjacent squares**
  - **More complex statistical methods**
  - **Reviewed in Dormann et al. (2007)**

Conclusions

- Species distribution models are a powerful tool for conservation
- They can predict a great deal about sites that have never been visited
- However, they have important shortcomings
- Predictions must be treated carefully
- Field surveys are still important