

Session 4: Introduction to niche-based distribution models

Mary Blair

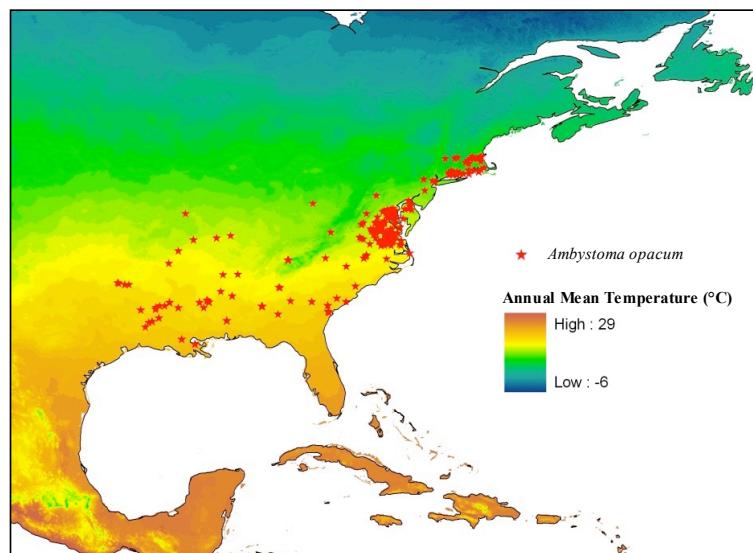
Outline:

1. Theory and concepts
2. Making a model: the nitty-gritty

With content courtesy of: Richard Pearson and Steven Phillips

1

So far...



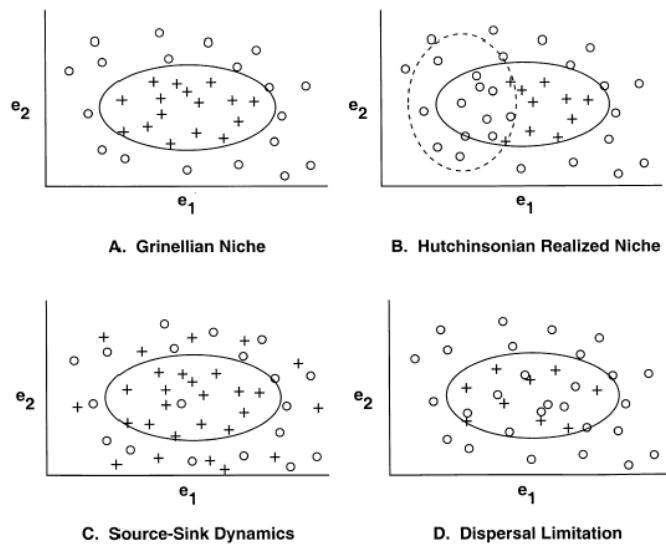
2

Correlative vs. Mechanistic approaches

- Assume current distribution gives a good indicator of ecological requirements
- Do not rely on observed occurrence records
- Require detailed physiological data

3

Theoretical underpinnings: niche theory



(Pulliam, 2000, *Ecology Letters*)

4

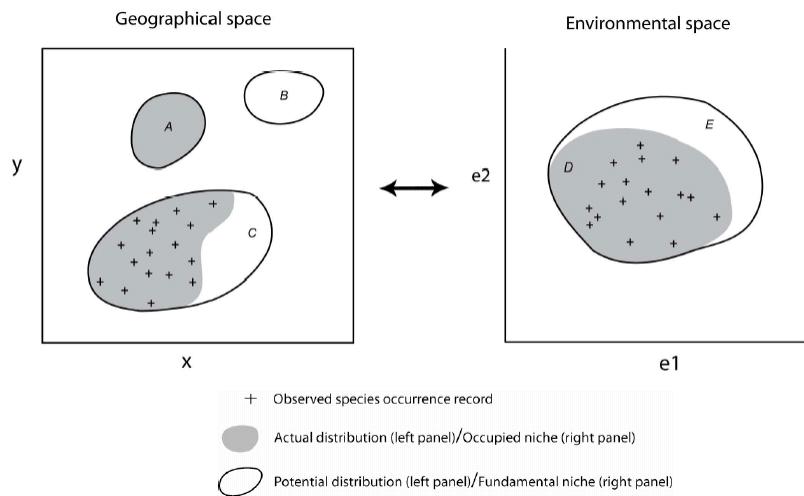
2

Factors impacting the geographic range of species

- The abiotic environment (Grinnell/Hutchinson's fundamental niche)
 - temperature
 - precipitation
 - soil type
- The biotic community (Elton's niche)
 - food webs and ecological networks
- History and Geography
 - dispersal

5

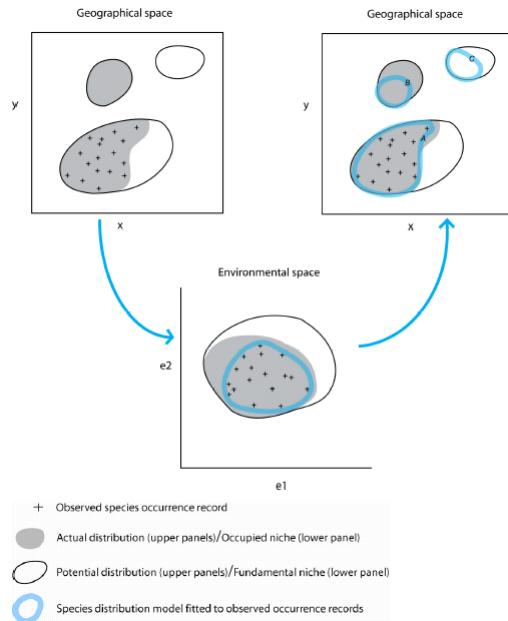
Illustration of the relationship between species' position in environmental space and geographical space



6

3

Illustration of the general species' distribution modeling approach

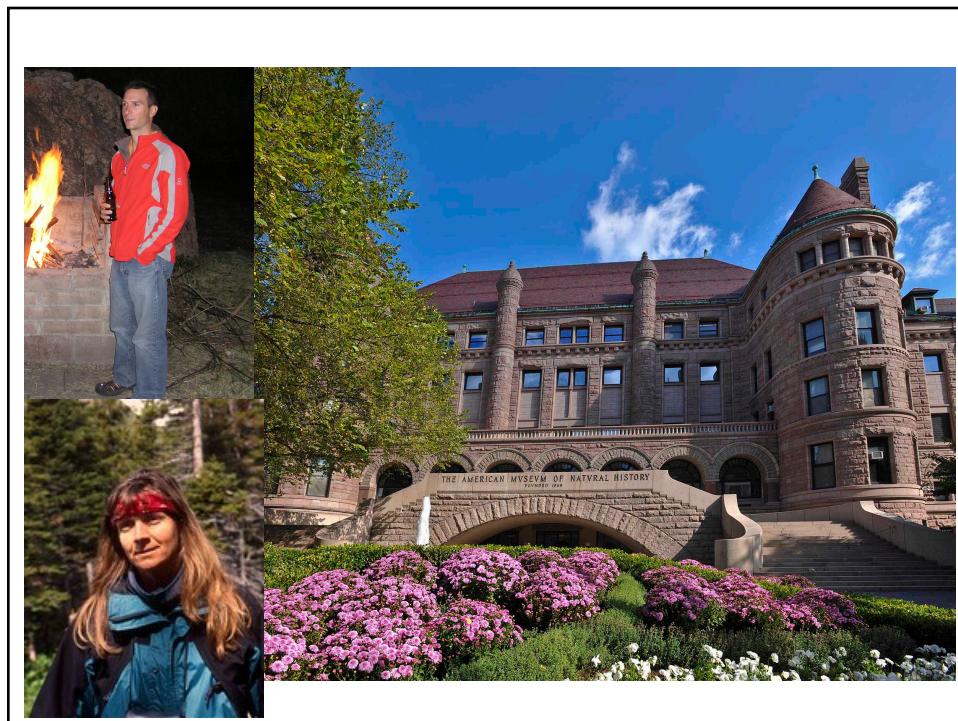


7

MACHINE LEARNING



8



9

Available online at www.sciencedirect.com

 ELSEVIER

SCIENCE @ DIRECT®

Ecological Modelling 190 (2006) 231–259

www.elsevier.com/locate/ecolmodel

ECOLOGICAL
MODELLING

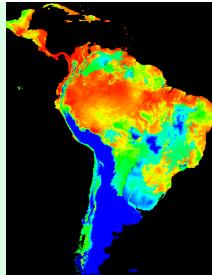
Maximum entropy modeling of species geographic distributions

Steven J. Phillips^{a,*}, Robert P. Anderson^{b,c}, Robert E. Schapire^d

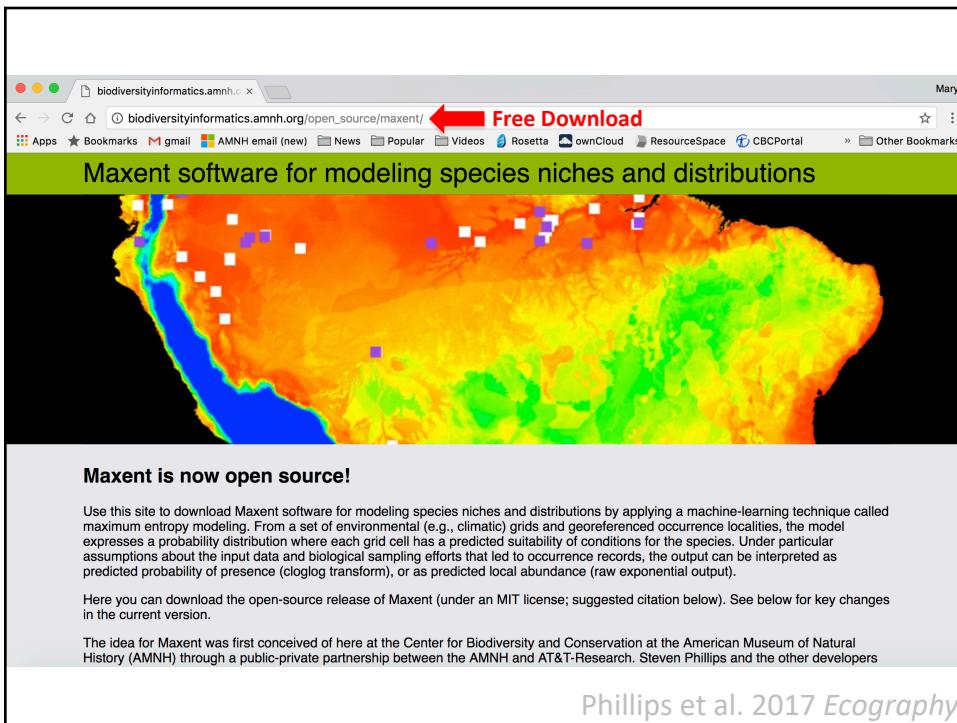
Maxent

Bradypus variegatus

Climatic, topographic, and vegetational variables



10



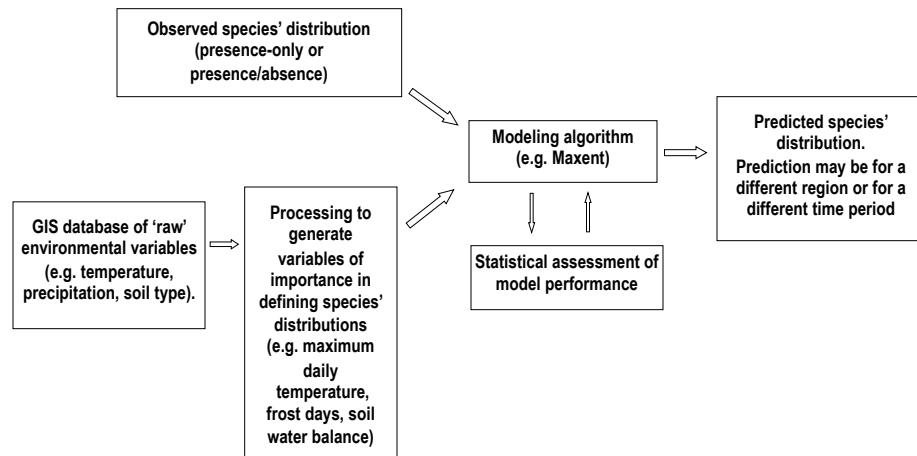
11

Making a model: the nitty-gritty

- i. The modeling process
- ii. Species data types
- iii. Model algorithms

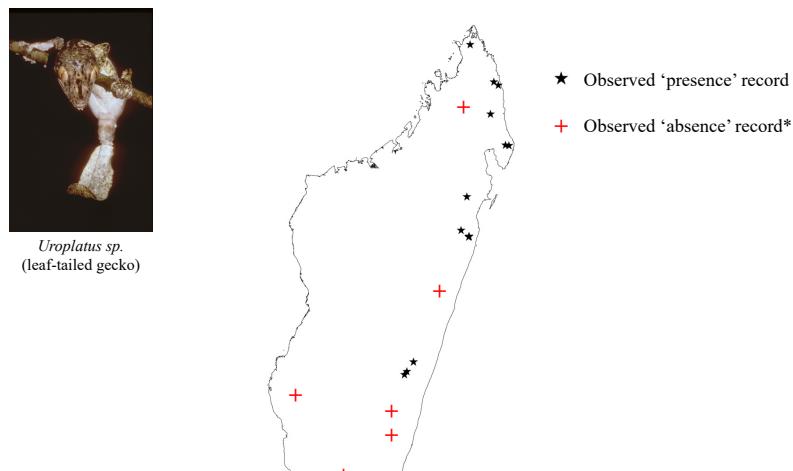
12

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



13

Species' distribution data: presence-only or presence/absence?



14

When is an absence really an absence?

A locality may be classified as ‘absent’ for a number of reasons:

1. The species could not be detected, even though it was present
'False absence'
2. The species was absent, even though the environment is suitable (e.g. due to dispersal limitation)
'False absence'
3. The environment is truly unsuitable for the species
'True absence'

Take care when using ‘absence’ data



15

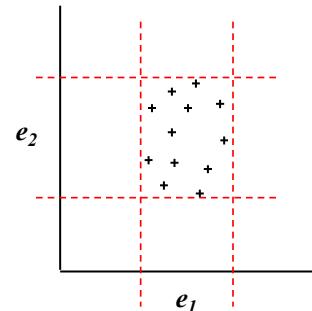
Some approaches that have been applied:

Method(s)	Model/software name	Species data type
Climatic envelope	BIOCLIM	Presence-only
Gower Metric	DOMAIN	Presence-only
Ecological Niche Factor Analysis (ENFA)	BIOMAPPER	Presence/background
Maximum Entropy	MAXENT	Presence/background
Genetic algorithm	GARP	Presence/pseudo-absence
Regression: Generalized linear model (GLM) and Generalized additive model (GAM)	GRASP	Presence/absence
Artificial Neural Network (ANN)	SPECIES	Presence/absence
Classification and regression trees (CART), GLM, GAM and ANN	BIOMOD	Presence/absence
Boosted regression trees	(implemented in R)	Presence/absence
Multivariate adaptive regression splines (MARS)	(implemented in R)	Presence/absence

16

BIOCLIM: a presence-only approach

- Only requires presence records
- Simple and intuitive ‘box’ model
- Gives equal weight to all variables
- Does not account for potential interactions between variables
- Cannot use categorical variables
- No extrapolations

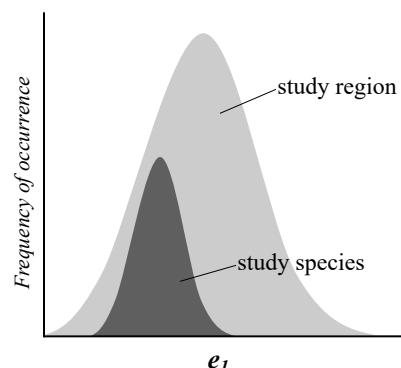


See: Lindenmayer et al. 1991 *J. Biogeog.* 18: 371-383.
Diva GIS: www.diva-gis.org

17

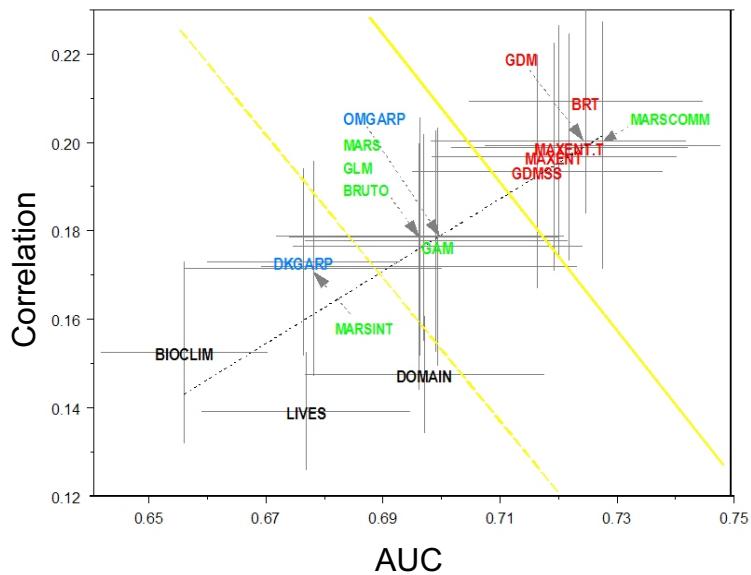
Maxent: a presence-background approach

- Only requires presence records
- Powerful machine learning method, precise mathematical definition
- Does not give equal weight to all variables
- Can account for potential interactions between variables (product features)
- Can use both continuous and categorical environmental variables
- Intuitive extrapolation approach ('clamping')



18

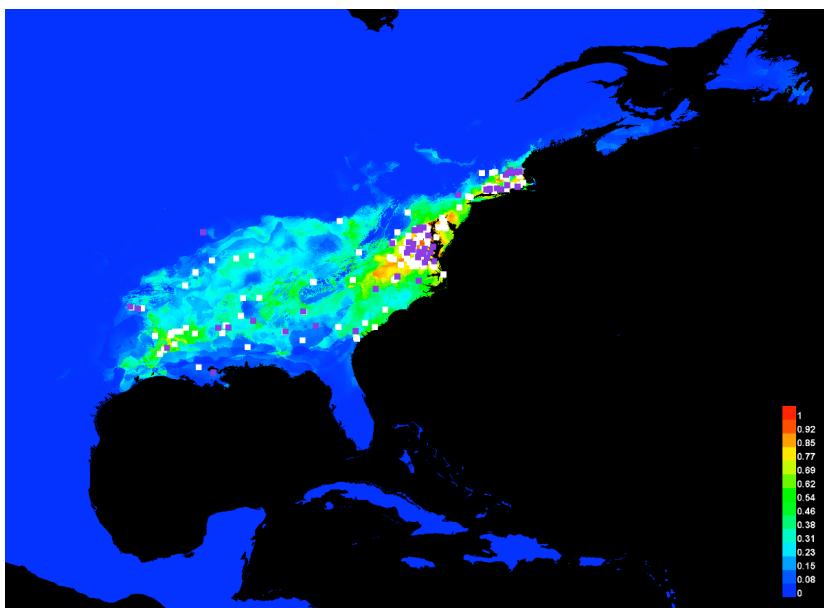
Which modeling method is best?



(from Elith *et al.*, *Ecography* 2006)

19

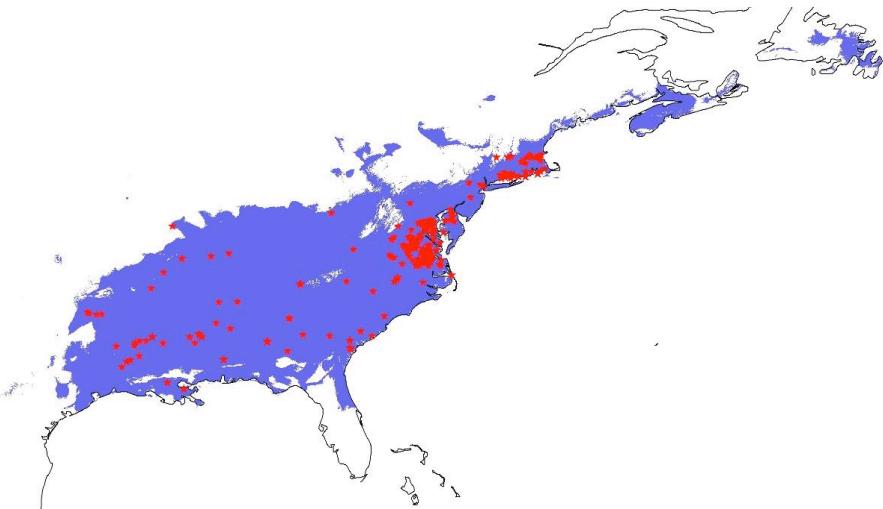
A Maxent model for the marbled salamander



20

10

A Maxent model for the marbled salamander



21

The basics of Maxent

We start with:

1. A map divided into cells (raster)
2. Environmental variables, with values in each cell
3. Occurrence records, which are samples from an unknown probability distribution that defines the species distribution

The task is to estimate the unknown probability distribution

(S. Phillips)

22

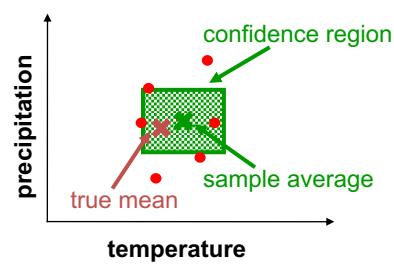
Maximum Entropy

- Estimates *an unknown probability distribution* by fitting a statistical model of *maximum entropy* (i.e., most spread out, closest to uniform) subject to *constraints*

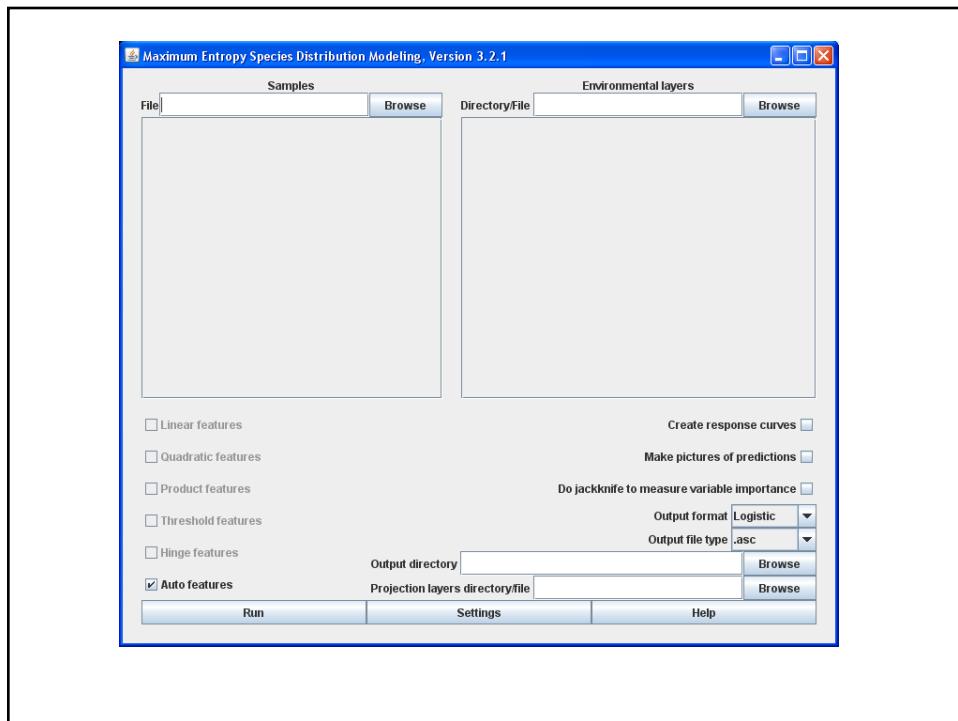
23

Maxent features & constraints

Feature class	Description	Constraints
Linear features	Variable itself	Mean
Quadratic features	Square of variable	Variance
Product features	Product of two variables	Covariance
Threshold/hinge features	...	Fit an arbitrary response
Discrete features (categorical)	Variable itself	Proportion



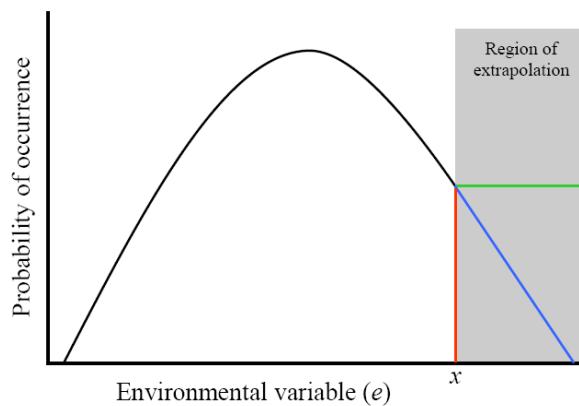
24



25

Caution! The use and misuse of models

- *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.



26

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.
- *The lure of complicated technology:* Many modeling approaches 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.