AMERICAN MUSEUM OF NATURAL HISTORY
CENTER FOR BIODIVERSITY AND CONSERVATION

Week 5:
Evaluating niche-based distribution models

Peter Galante

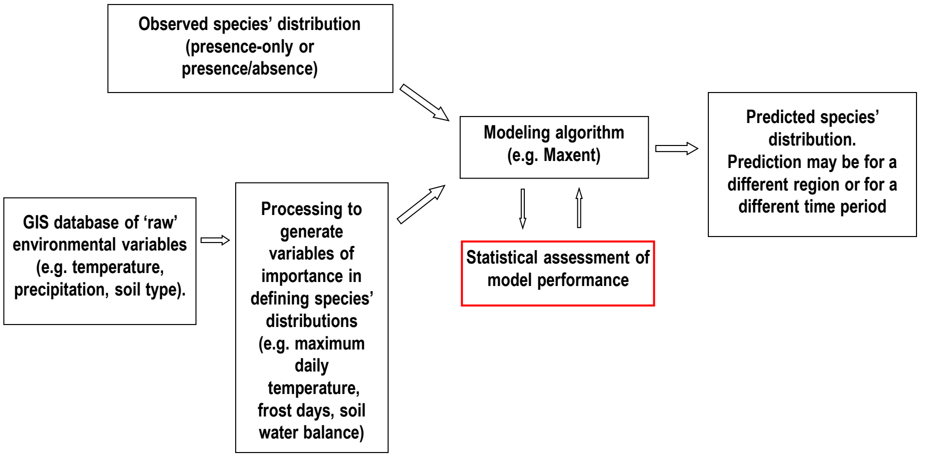
Outline:

1. Maxent model tuning
2. Model validation
3. Model evaluation
4. Evaluation statistics
5. Model complexity and advanced considerations

(thanks to Richard Pearson, Rob Anderson, and Mary Blair for content)

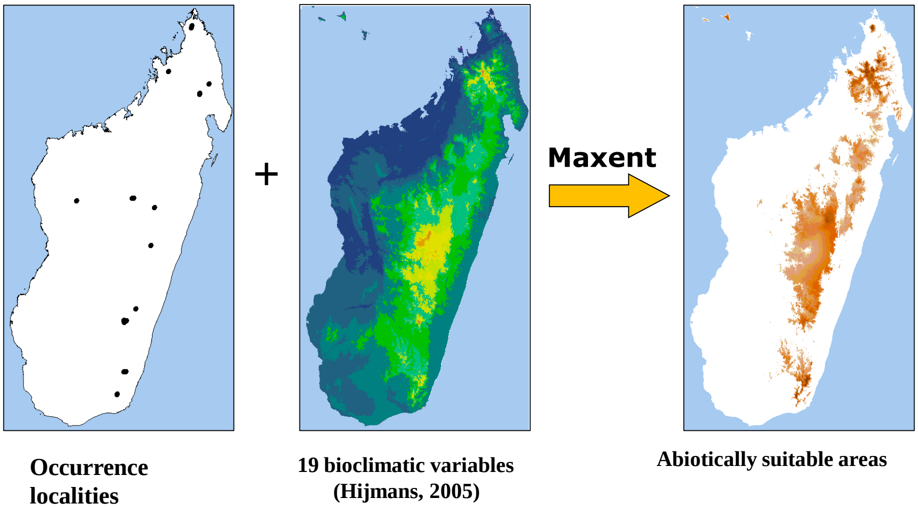
1

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



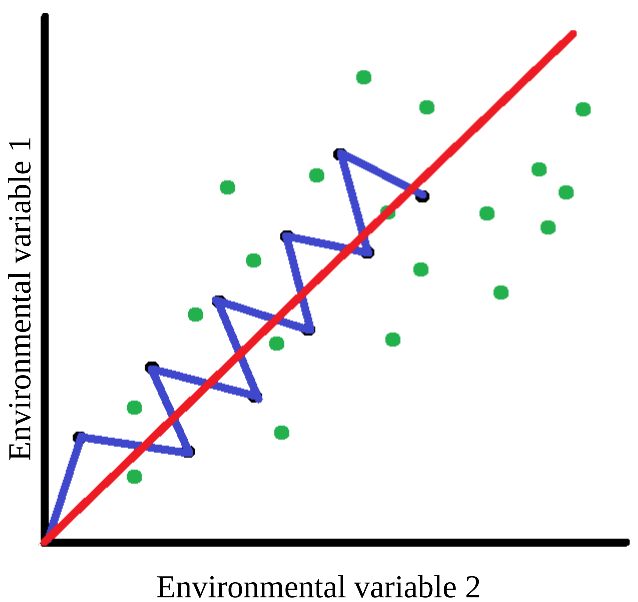
2

Niche Models



3

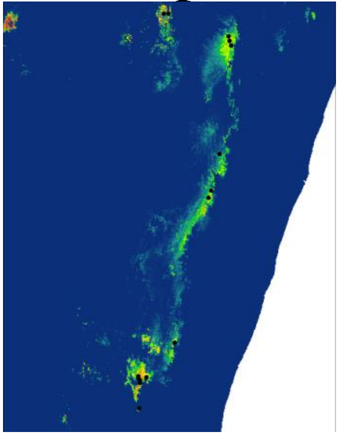
Model Complexity vs. Generality



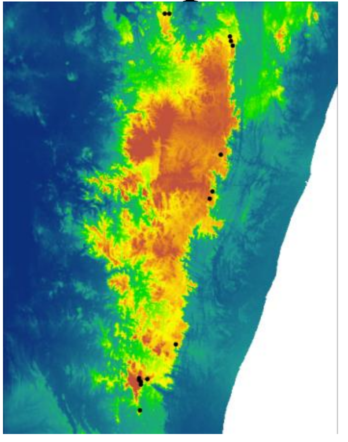
4

Model Selection

Complex



Simple



5

Model Tuning

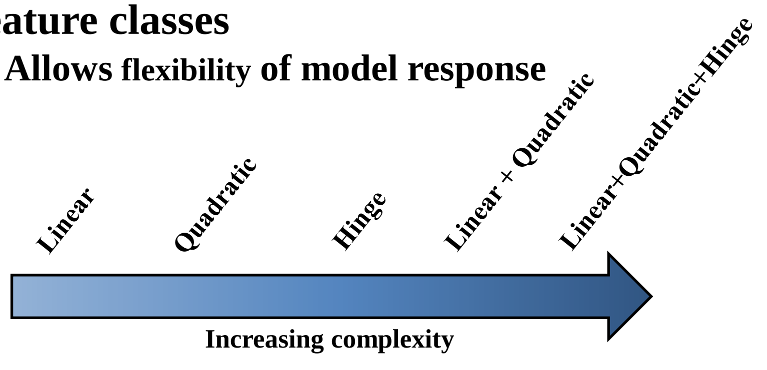
Two settings in Maxent (in ENMeval):

Regularization multiplier:

Increasing regularization increases penalties on model complexity 0 → ?

Feature classes

Allows flexibility of model response



6

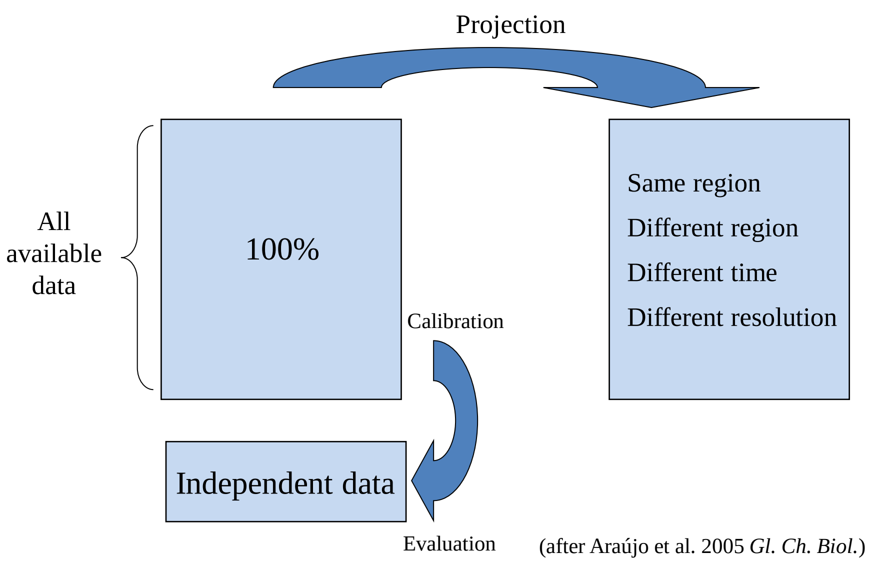
AMERICAN MUSEUM OF NATURAL HISTORY
CENTER FOR BIODIVERSITY AND CONSERVATION

Outline:

1. Maxent model tuning
2. Model validation
3. Model evaluation
4. Evaluation statistics
5. Model complexity and advanced considerations

7

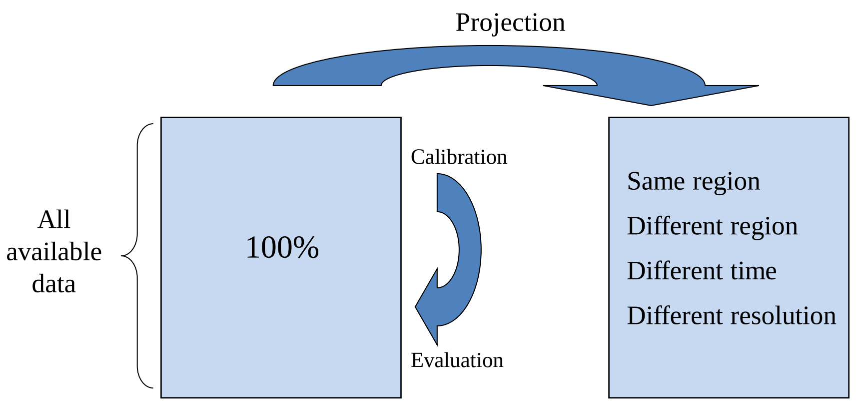
Model calibration and evaluation strategies: independent validation



(after Araújo et al. 2005 *Gl. Ch. Biol.*)

8

Model calibration and evaluation strategies: resubstitution

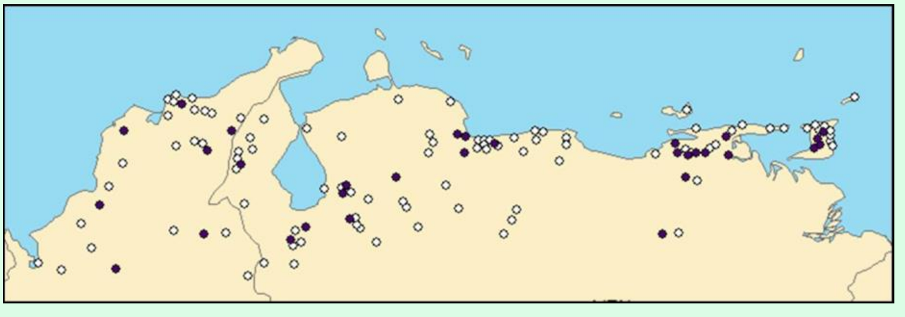


(after Araújo et al. 2005 *Gl. Ch. Biol.*)

9

Random subsets

random split-sample approach: easy test, cannot detect overfitting to bias

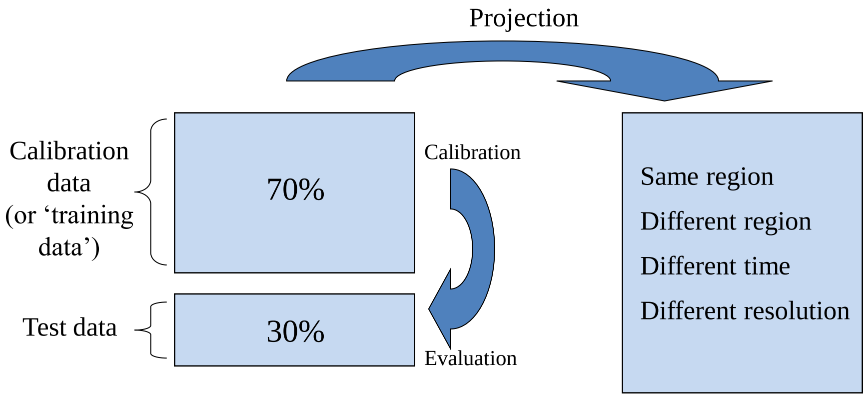


White: calibrate the model

Black: evaluate the model

10

Model calibration and evaluation strategies: data splitting

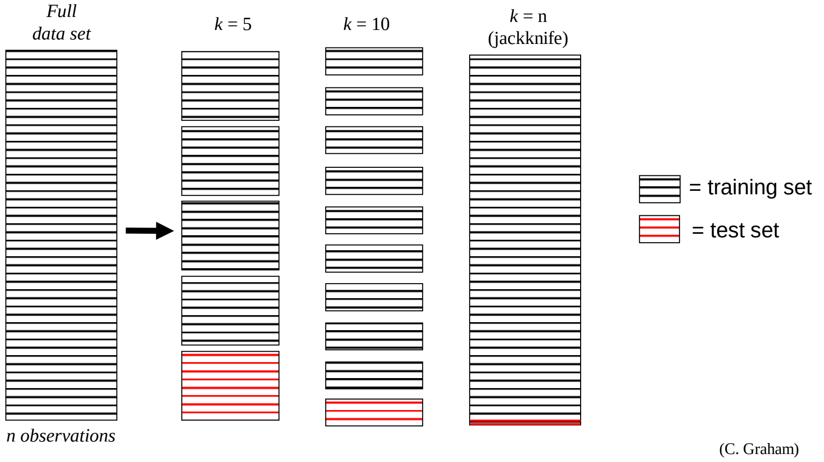


(after Araújo et al. 2005 *Gl. Ch. Biol.*)

11

Model calibration and evaluation strategies: cross validation/ *k*-fold partitioning

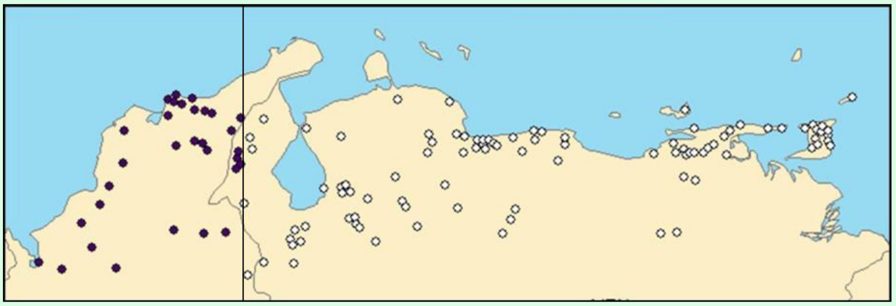
1. Split data randomly into *k* roughly equal-sized parts. Take turns using each part as a test set and the other $k - 1$ parts for model training.
2. Compute test statistic each time. Cross-validation estimate of predictive performance is the average of the *k* tests.



(C. Graham)

12

Spatial subsets

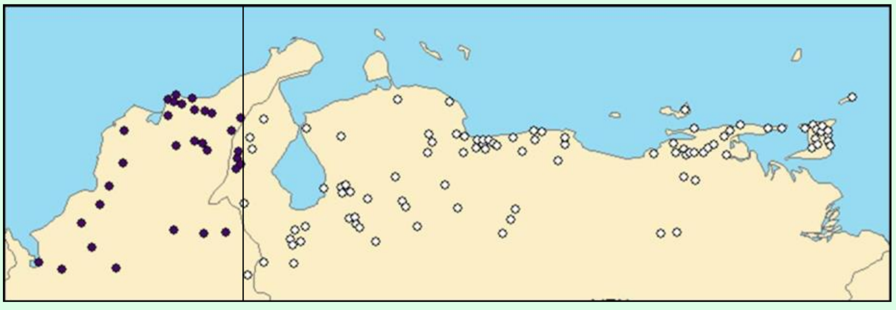


White: calibrate the model

Black: evaluate the model

13

Spatial subsets



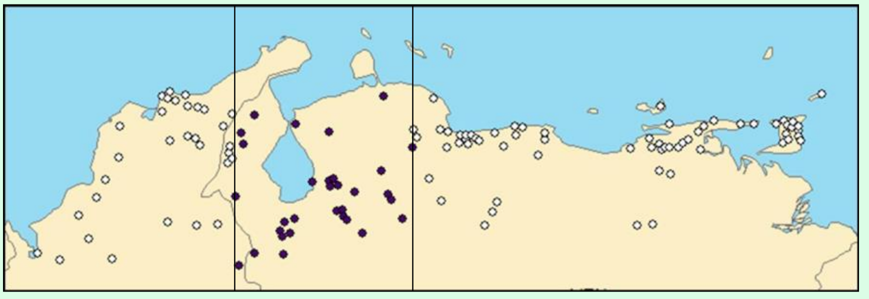
White: calibrate the model

Black: evaluate the model

Radosavljevic & Anderson 2014

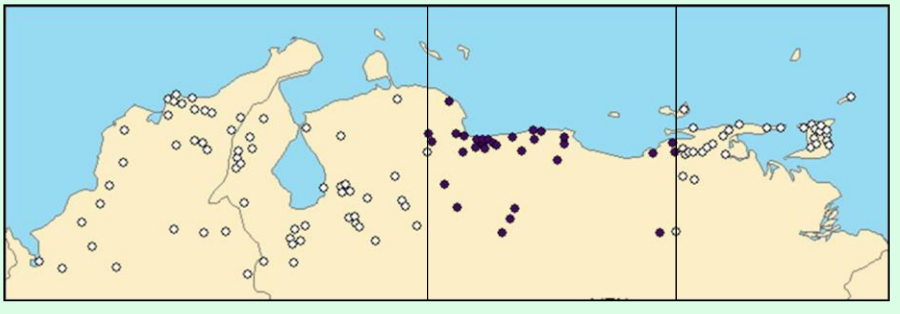
14

Bin B for evaluation



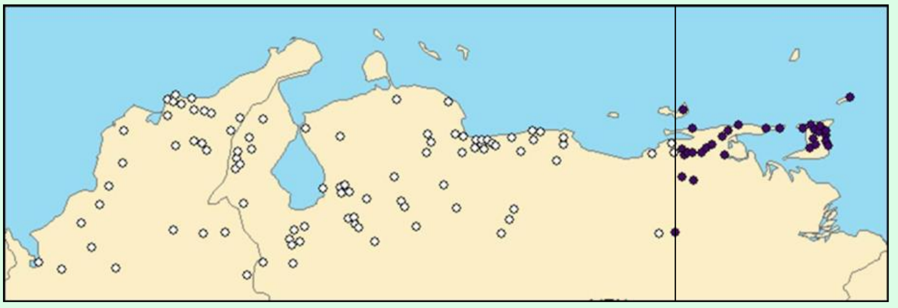
15

Bin C for evaluation

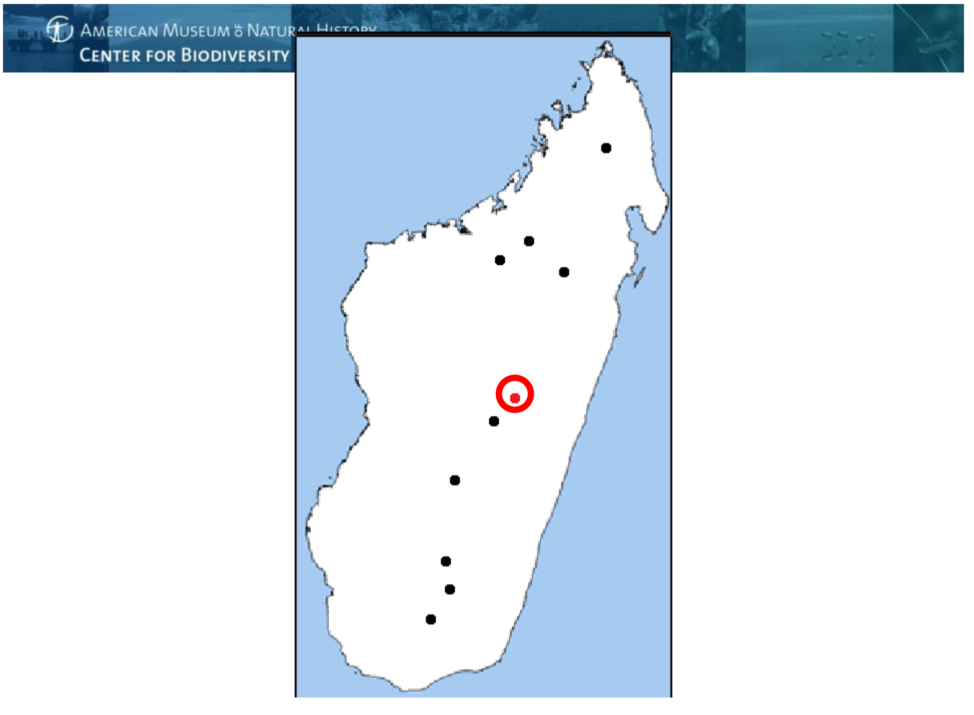


16

Bin D for evaluation



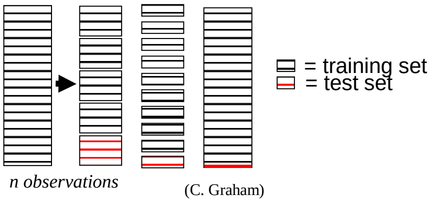
17



18

Model calibration and evaluation strategies: cross validation/ *k*-fold partitioning

1. Split data randomly into *k* roughly equal-sized parts. Take turns using each part as a test set and the other *k* – 1 parts for model training.
2. Compute test statistic each time. Cross-validation estimate of predictive performance is the average of the *k* tests.



X

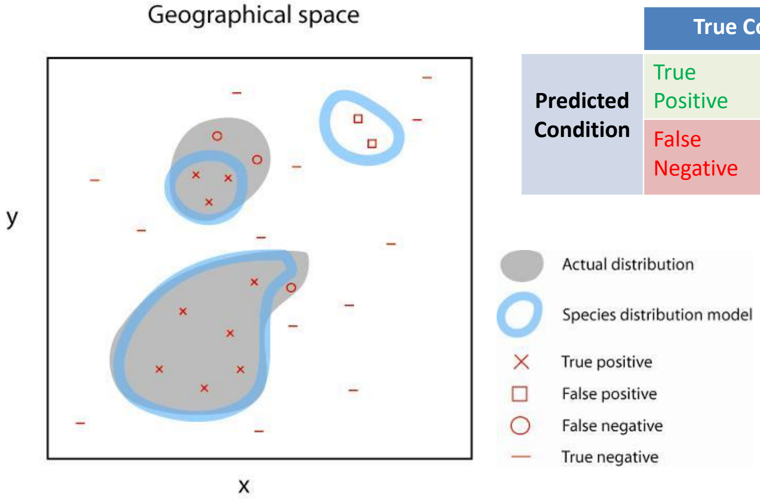
Each Feature Class and Regularization Multiplier combination

AMERICAN MUSEUM OF NATURAL HISTORY
CENTER FOR BIODIVERSITY AND CONSERVATION

Outline:

1. Maxent model tuning
2. Model validation
3. Model evaluation
4. Evaluation statistics
5. Model complexity and advanced considerations

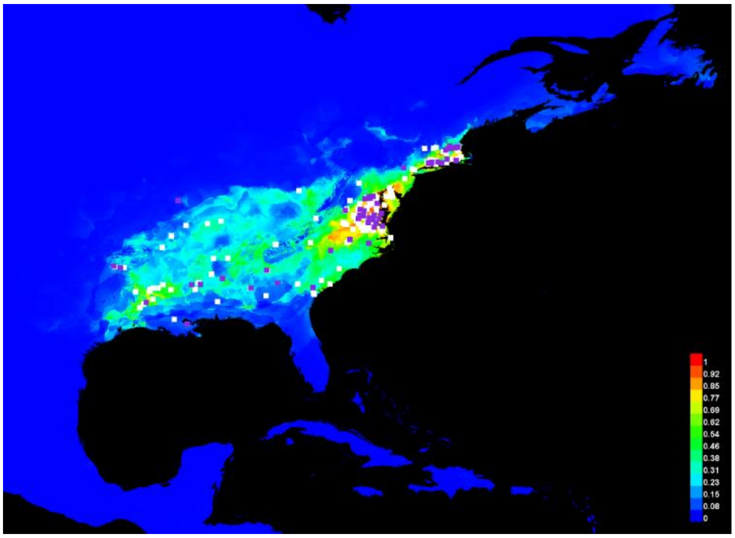
The four types of results that are possible
when testing a distribution model



	True Condition	
Predicted Condition	True Positive	False Positive
	False Negative	True Negative

21

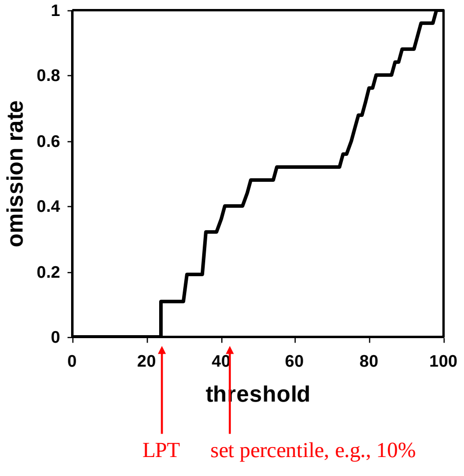
'Continuous' model output



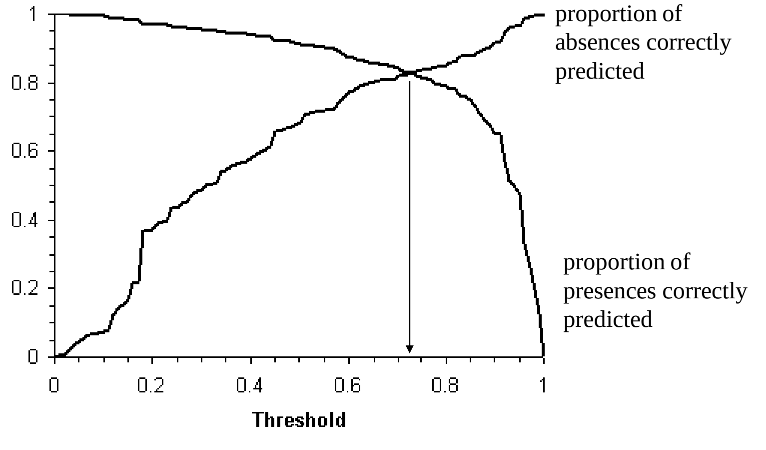
Maxent model for the marbled salamander

22

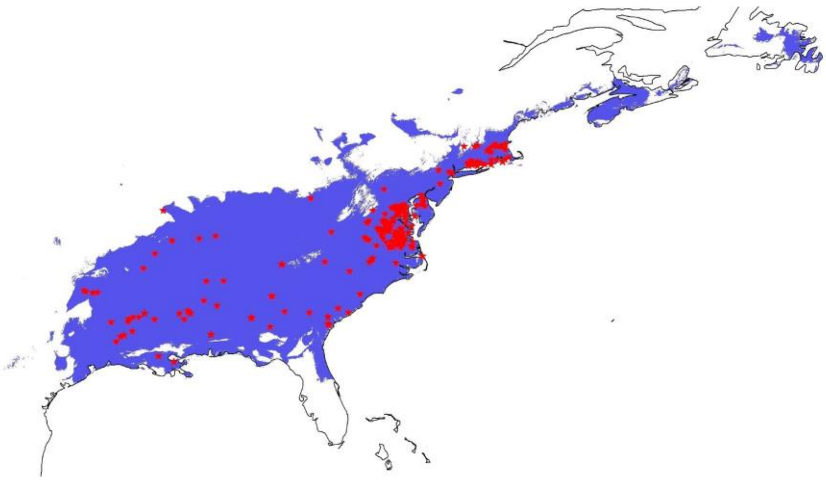
Selecting a threshold with presence-only data



Selecting a threshold with presence/absence data



Thresholded model output



Maxent model for the marbled salamander with threshold applied

25

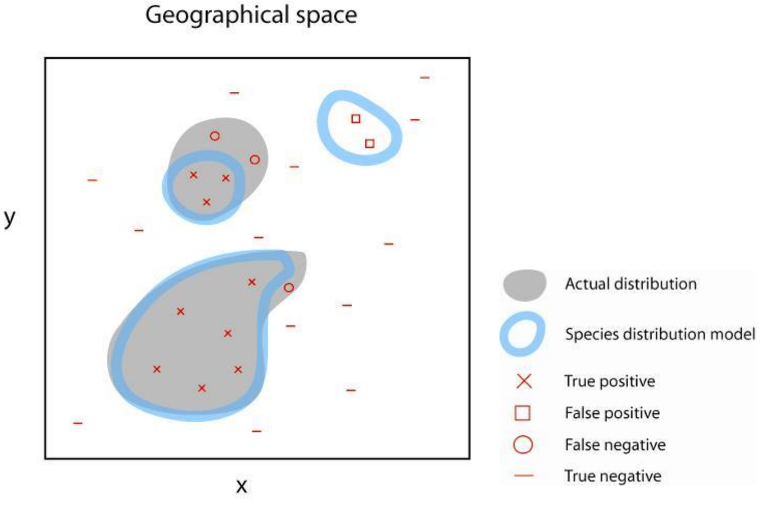
AMERICAN MUSEUM & NATURAL HISTORY
CENTER FOR BIODIVERSITY AND CONSERVATION

Outline:

1. Maxent model tuning
2. Model validation
3. Model evaluation
4. Evaluation statistics
5. Model complexity and advanced considerations

26

The four types of results that are possible
when testing a distribution model



27

Presence-absence confusion matrix

	Recorded present	*Recorded (or assumed) absent*
Predicted present	a (true positive)	b (false positive)
Predicted absent	c (false negative)	d (true negative)

28

Presence-only test statistics

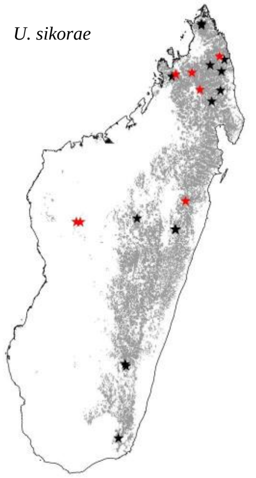
	Recorded present	*Recorded (or assumed) absent*
Predicted present	a (true positive)	b (false positive)
Predicted absent	c (false negative)	d (true negative)

Proportion of observed presences correctly predicted (or ‘sensitivity’, or ‘true positive fraction’):

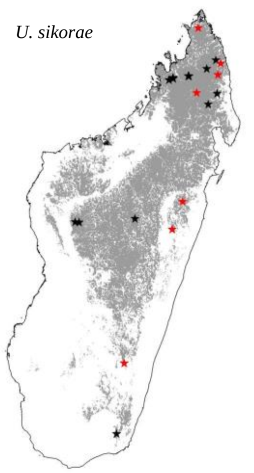
$$a/(a + c)$$

29

Presence-only test statistics: testing for statistical significance

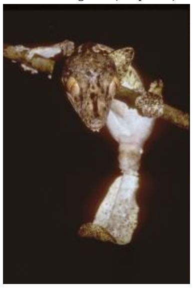


Success rate: 4 from 7
Proportion predicted present: 0.231
Binomial $p = 0.0546$



Success rate: 6 from 7
Proportion predicted present: 0.339
Binomial $p = 0.008$

Leaf-tailed gecko (*Uroplatus*)



30

Presence-absence test statistics

	Recorded present	*Recorded (or assumed) absent*
Predicted present	a (true positive)	b (false positive)
Predicted absent	c (false negative)	d (true negative)

Proportion of observed presences correctly predicted (or ‘sensitivity’):

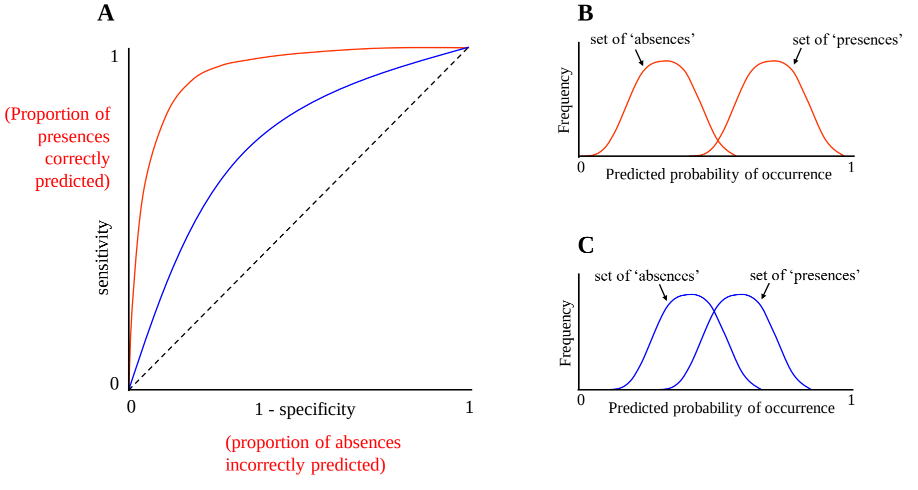
$$a/(a + c)$$

Proportion of observed absences correctly predicted (or ‘specificity’):

$$d/(b + d)$$

31

The Receiver Operating Characteristic (ROC) Curve



(check out: http://www.anaesthetist.com/mnm/stats/roc/Findex.htm)

32

So, what is a ‘good’ result?

Some *__subjective__* guidelines (after Swets 1988 *Science)* :

- 0.5 – 0.7: poor discrimination
- 0.7 – 0.9: reasonable discrimination
- 0.9 – 1.0: very good discrimination

However, note that the Maxent software generates AUC statistics using ‘background’ rather than absence data, so the maximum achievable AUC score is <1 (see Phillips et al. 2006).

AMERICAN MUSEUM of NATURAL HISTORY
CENTER FOR BIODIVERSITY AND CONSERVATION

Outline:

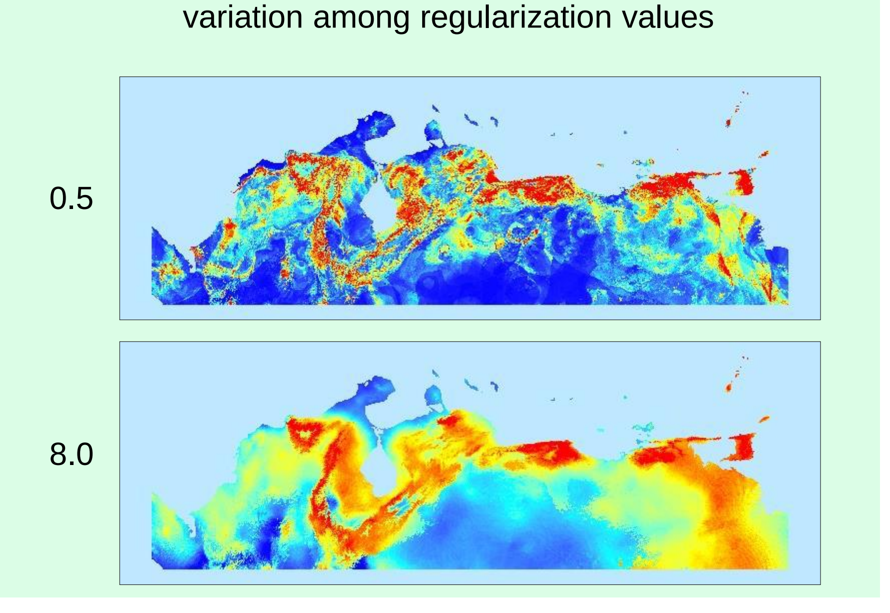
1. Maxent model tuning
2. Model validation
3. Model evaluation
4. Evaluation statistics
5. Model complexity and advanced considerations

Advanced considerations

- Model complexity:
  - Occurrence data suffer from (among other things):
    - Biased sampling across geography
    - Which may also be biased in environmental space
- Overly complex models may overfit to this bias (or to noise)
  - What is overly complex?
    - Number of variables, feature classes considered, level of regularization

35

Overfitting



36

Advanced considerations

- Background selection

37

Advanced considerations

- ENMEval R package
  - Try different regularization values, design spatially independent training and test datasets, feature classes, number of variables (Muscarella et al. 2014)
  - SDMToolbox for ArcGIS also does some of this (Brown et al. 2014)
- spThin R package
  - thin occurrence records re: spatial autocorrelation bias (Aiello-Lammens et al. 2015)
- Bias layer for Maxent
  - for known sampling bias masking (Phillips et al. 2009)

38