

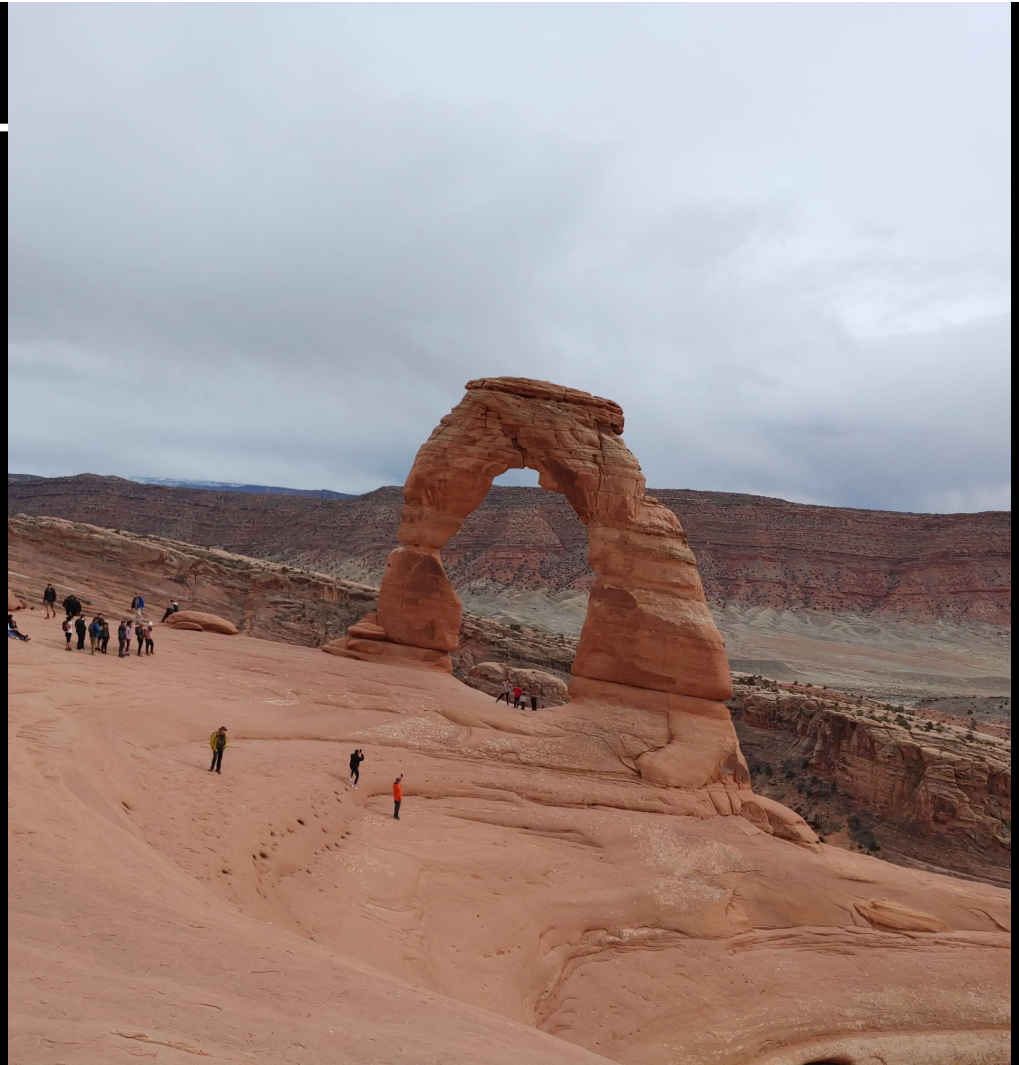
---

# MODEL EVALUATION

Elena Razenkova

Karaganda, Kazakhstan August 2025

---



---

## PLAN FOR TODAY

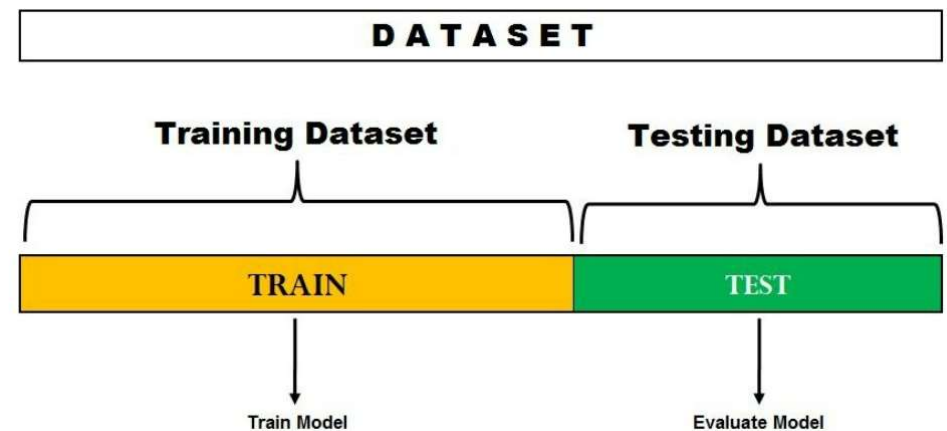
- Strategies for obtaining test data
  - Confusing matrix
  - Test statistics
  - Example for Black-Capped Chickadee
  - R code
- 



# STRATEGIES FOR OBTAINING TEST DATA

---

- Training (calibration) and test (evaluation) data
- Preferable that test data are different from training data
- Collected independently
- Use data from different regions
- Data at different spatial resolution
- Data from different time periods



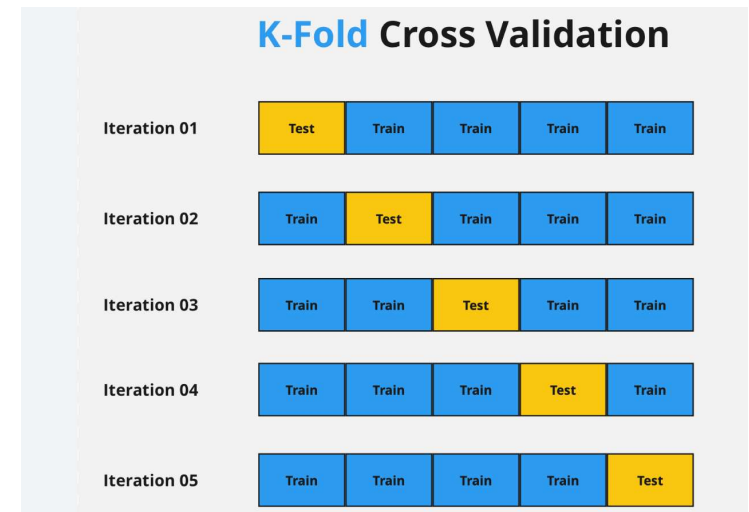
# STRATEGIES FOR OBTAINING TEST DATA

---

## Data partitioning

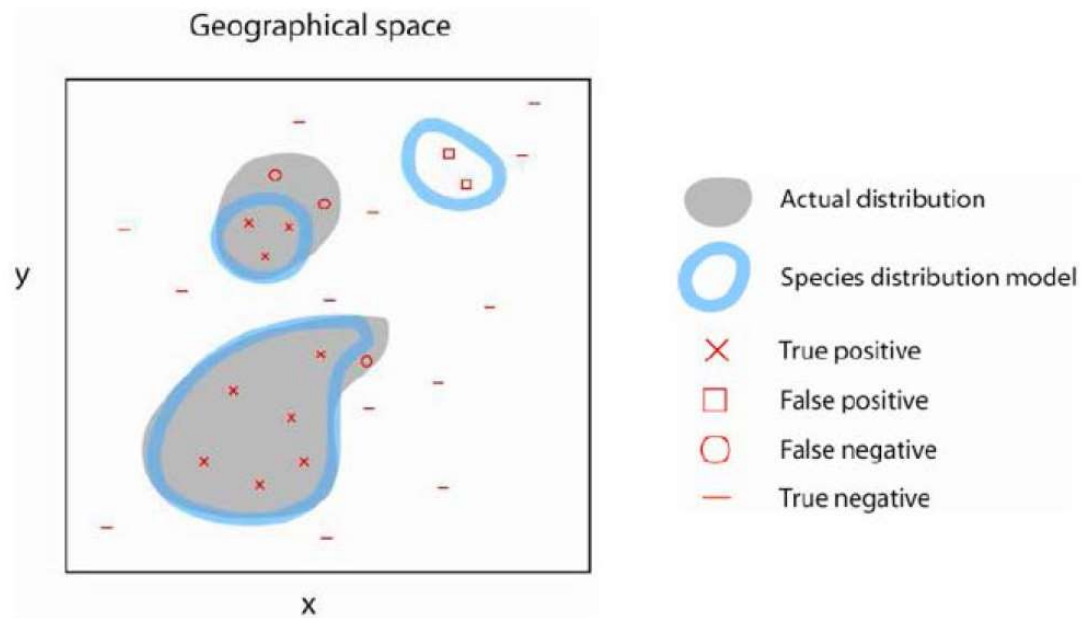
- One time split randomly
- Spatially divided data
- 70% training data and 30% testing data
- Bootstrapping split (with replacement) - >  
Validation statistics reported as the mean and a range from the set of bootstrap samples
- “Randomization” sampling without replacement
- K-fold partitioning

If sample size is very small -> k-fold partitioning,  
k=the number of occurrence



# CONFUSION MATRIX (ERROR MATRIX)

---

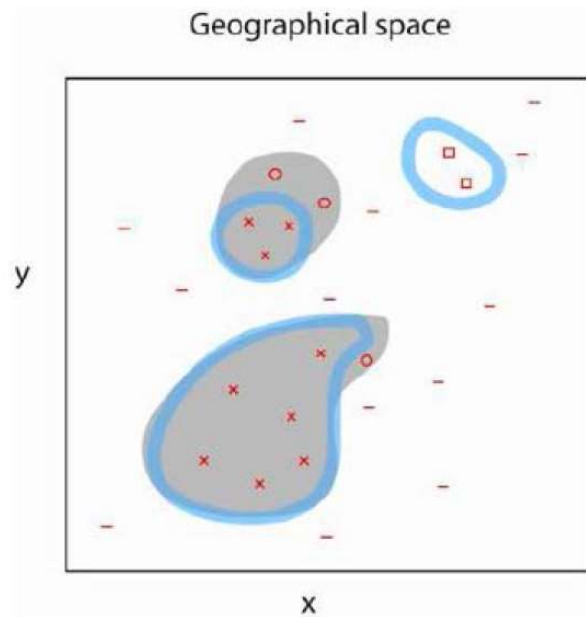








	recorded present	recorded absent
predicted present	<i>a (true positive)</i>	<i>b (false positive)</i>
predicted absent	<i>c (false negative)</i>	<i>d (true negative)</i>

	recorded present	recorded absent
predicted present	9	2
predicted absent	3	13

# TEST STATISTICS DERIVED FROM THE CONFUSION MATRIX

Only for presence/absence data



-  Actual distribution
-  Species distribution model
-  True positive
-  False positive
-  False negative
-  True negative

	recorded present	recorded absent
predicted present	<i>a (true positive)</i>	<i>b (false positive)</i>
predicted absent	<i>c (false negative)</i>	<i>d (true negative)</i>

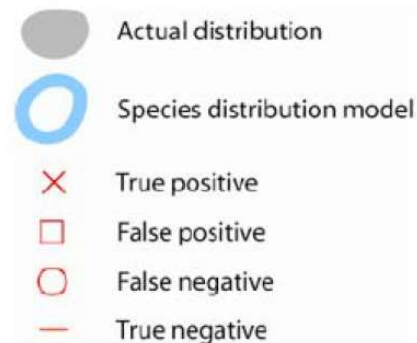
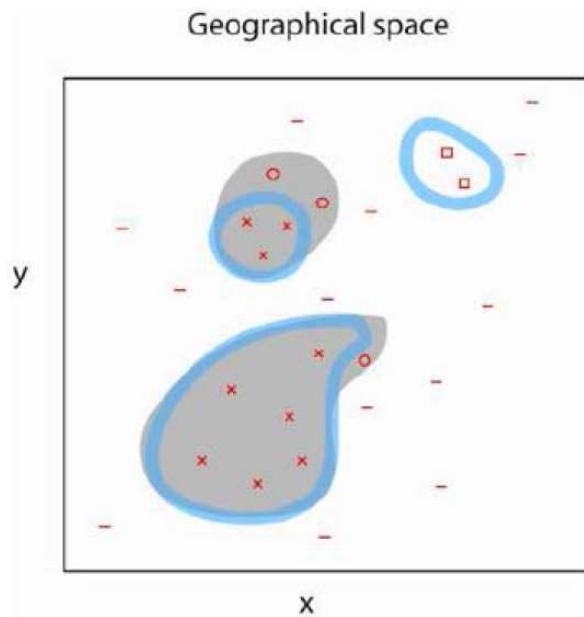
$$\text{Accuracy} = (a+d)/(a+b+c+d)$$

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



# TEST STATISTICS DERIVED FROM THE CONFUSION MATRIX

Only for presence data



	recorded present
predicted present	<i>a (true positive)</i>
predicted absent	<i>c (false negative)</i>

$$\text{Sensitivity} = a / (a + c)$$

$$\text{Omission} = c / (a + c)$$

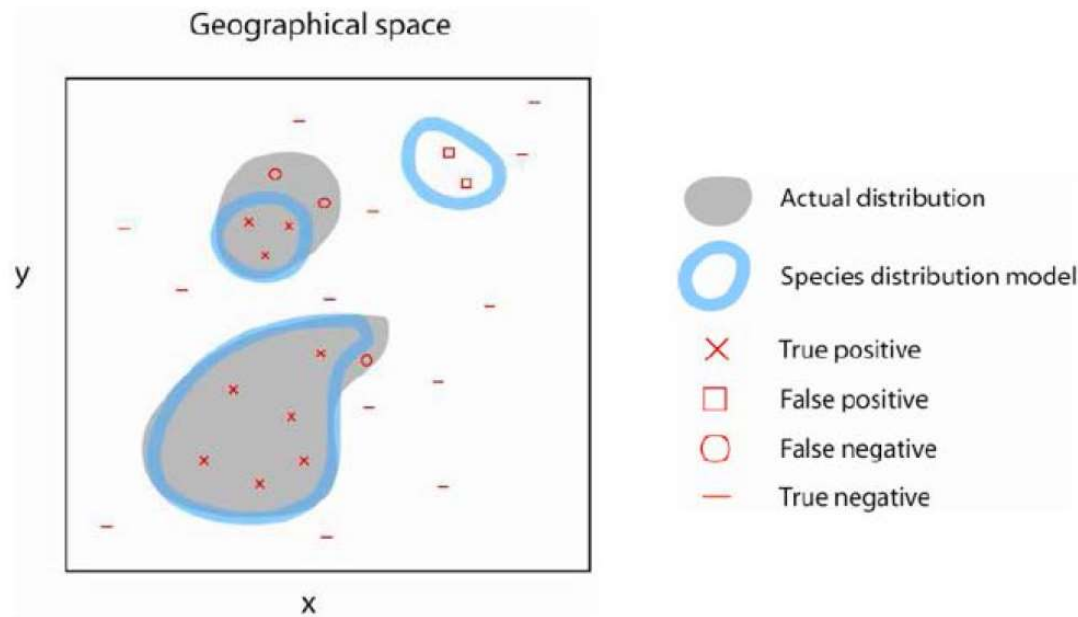
A serious disadvantage of these tests is that it is possible to achieve very high sensitivity (and low omission) simply by predicting that the species is present at an excessively large proportion of the study area.

Test for statistical significance

# TEST STATISTICS DERIVED FROM THE CONFUSION MATRIX

---

Only for presence data



	recorded present	recorded absent
predicted present	<i>a (true positive)</i>	<i>b (false positive)</i>
predicted absent	<i>c (false negative)</i>	<i>d (true negative)</i>

$$\text{Specificity} = d / (b + d)$$

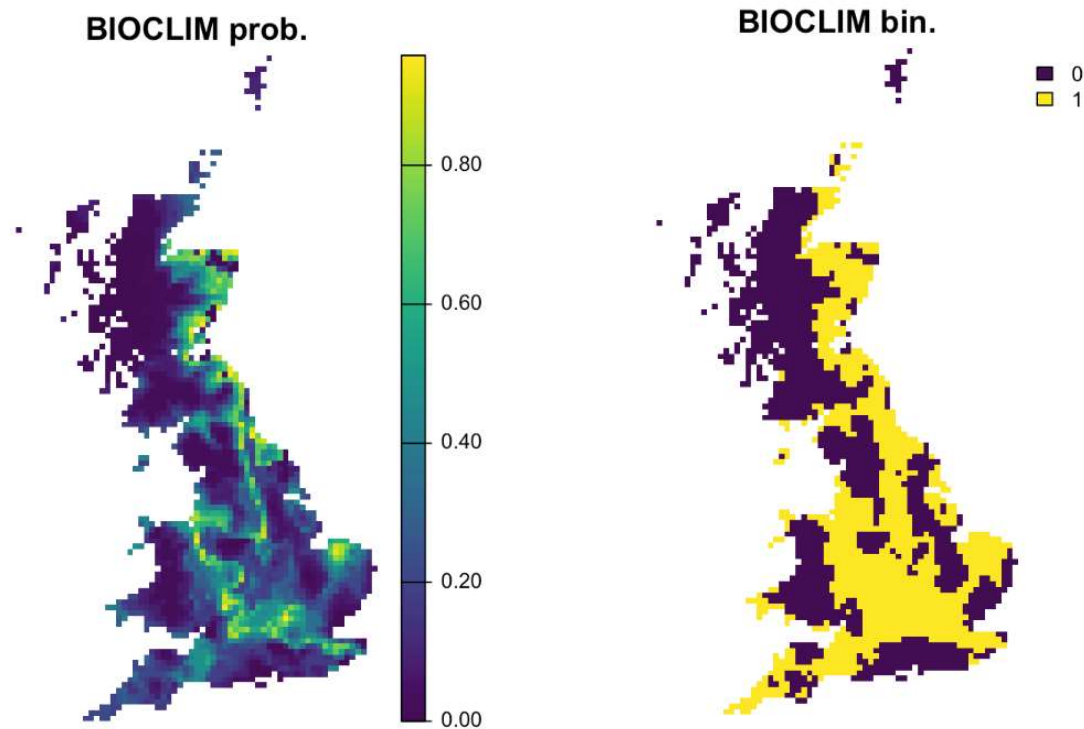
An important measure used in setting decision thresholds and in the ROC analysis.



# CONVERTING CONTINUOUS INTO BINARY OUTPUT

---

- **Binary model prediction (i.e., suitable or unsuitable habitat)**
- **Our output was in a form of probabilities of occurrence**
- **Into binarity prediction!**
- **Select threshold of occurrence**



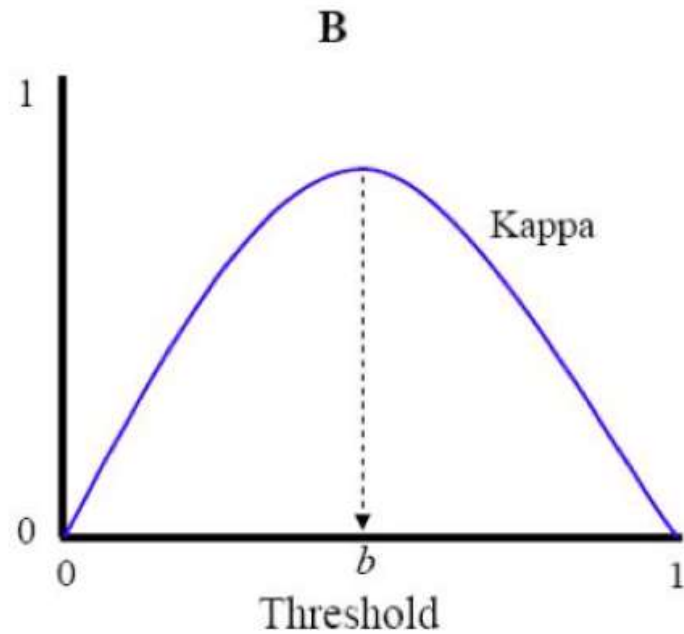
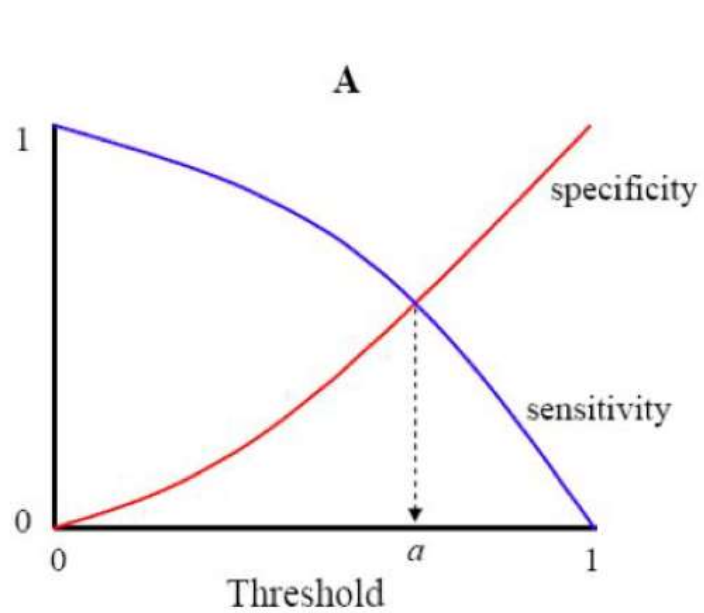
# SELECTING THRESHOLDS OF OCCURRENCE

Table 4. Some published methods for setting thresholds of occurrence			
Method	Definition	Species data type <sup>1</sup>	Reference(s)
Fixed value	An arbitrary fixed value (e.g. probability = 0.5)	presence-only	Manel et al., 1999 Robertson et al., 2001
Lowest predicted value	The lowest predicted value corresponding with an observed occurrence record	presence-only	Pearson et al., 2006 Phillips et al., 2006
Fixed sensitivity	The threshold at which an arbitrary fixed sensitivity is reached (e.g. 0.95, meaning that 95% of observed localities will be included in the prediction)	presence-only	Pearson et al., 2004
Sensitivity-specificity equality	The threshold at which sensitivity and specificity are equal	presence and absence	Pearson et al., 2004
Sensitivity-specificity sum maximization	The sum of sensitivity and specificity is maximized	presence and absence	Manel et al., 2001
Maximize Kappa	The threshold at which Cohen's Kappa statistic is maximized	presence and absence	Huntley et al., 1995 Elith et al., 2006
Average probability/suitability	The mean value across model output	presence-only	Cramer, 2003
Equal prevalence	Species' prevalence (the proportion of presences relative to the number of sites) is maintained the same in the prediction as in the calibration data.	presence and absence	Cramer, 2003

Pearson 2007

# SELECTING THRESHOLDS OF OCCURRENCE

---



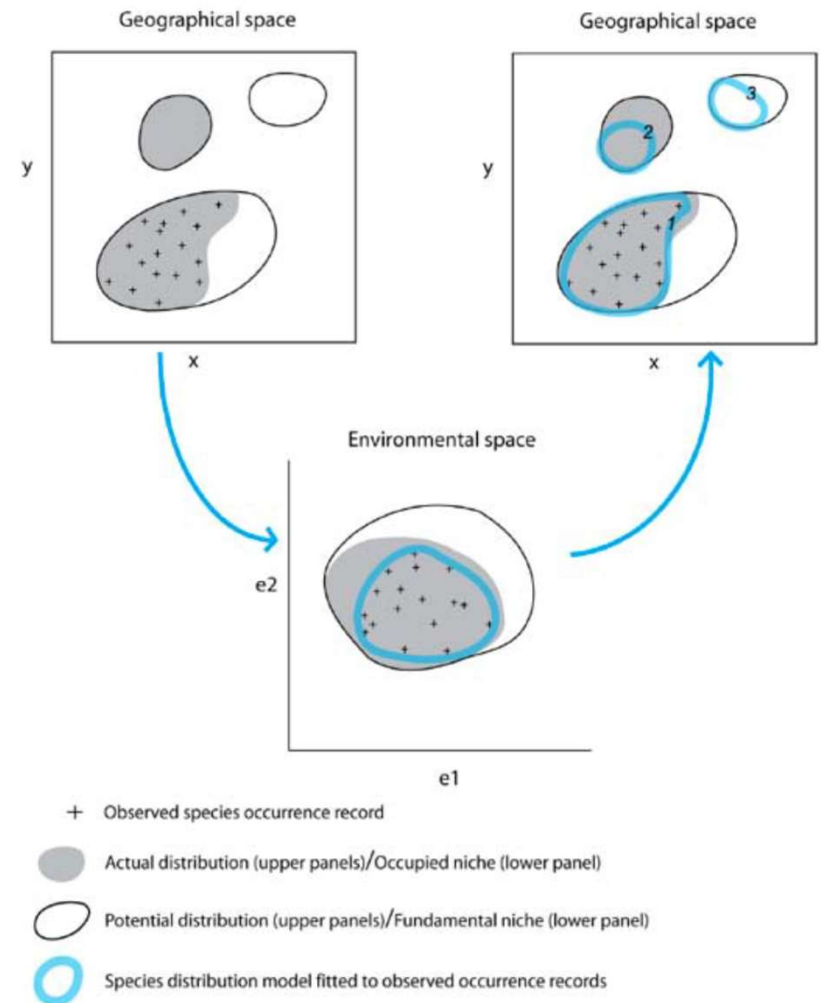
---

Pearson 2007

# CHOOSING A SUITABLE TEST STATISTIC

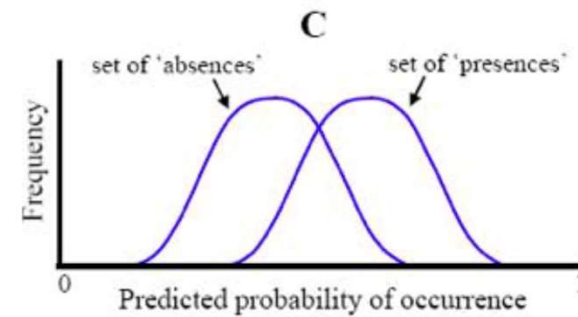
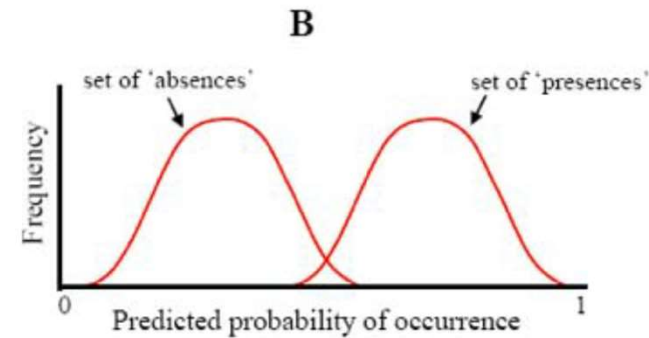
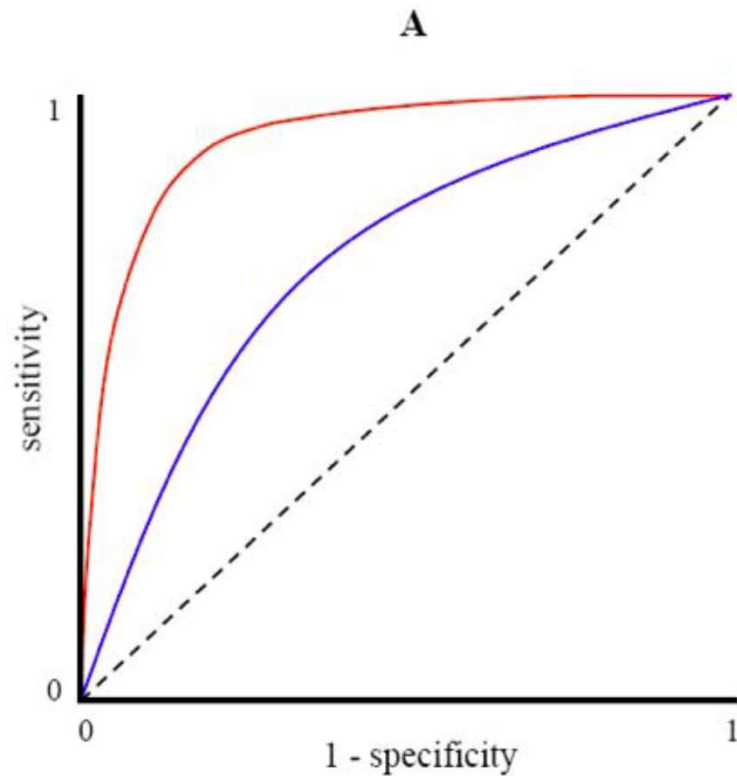
**Depends on how the model will be applied:**

- If the aim is to predict actual distribution (presence and absence) -> Accuracy or Kappa (Area 3 are considered model errors)
- If the aim is to estimate the potential distribution (presence-only) -> Sensitivity and statistical significance



# RECEIVER OPERATING CHARACTERISTICS

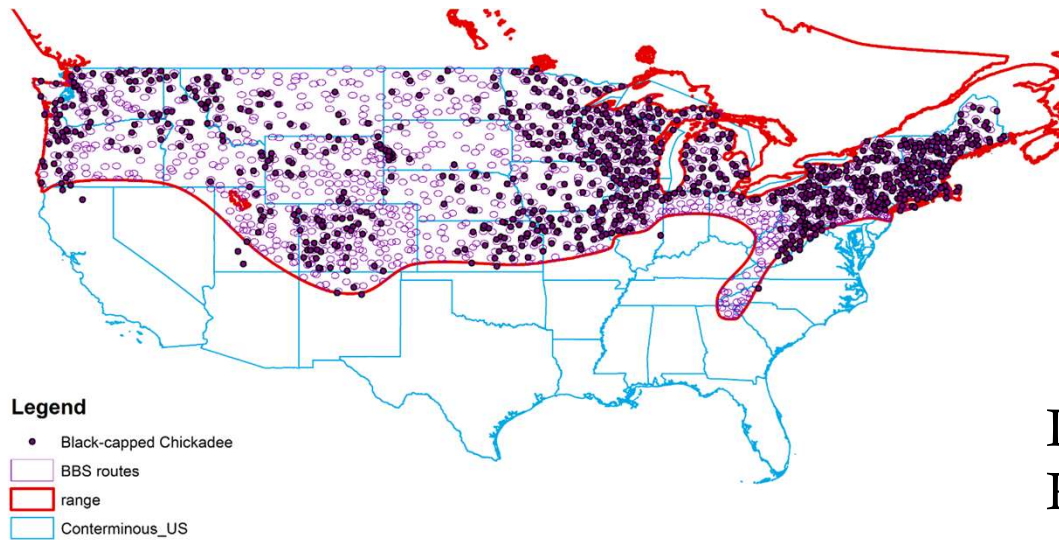
---



Pearson 2007

# EXAMPLE FOR BLACK-CAPPED CHICKADEE

---



Data:

Presence points: 867

Absence points : 624

75% for training

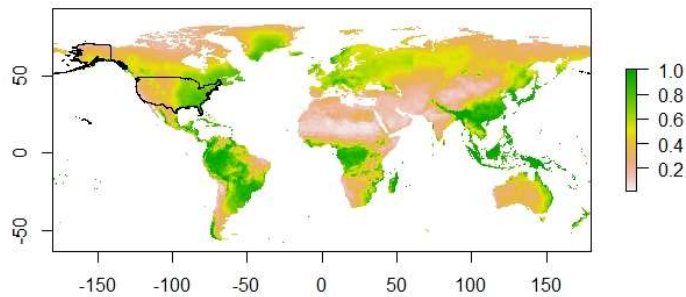
25% for validation



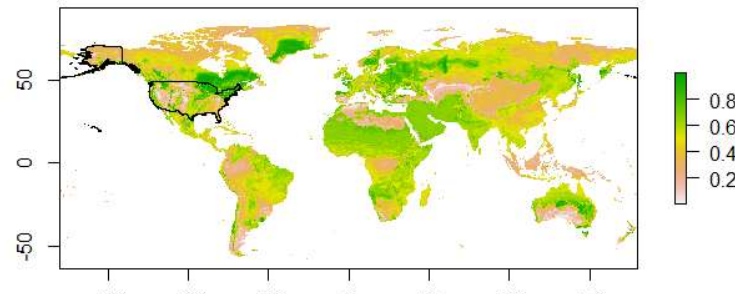
# EXAMPLE FOR BLACK-CAPPED CHICKADEE

---

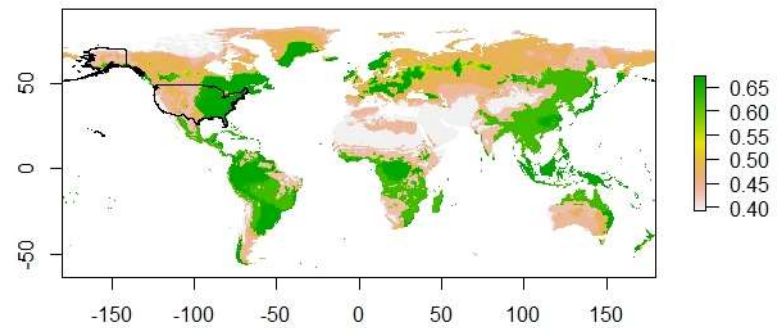
GLM



RF



BRT



---

# THANK YOU

Elena Razenkova  
[erazenkova@gmail.com](mailto:erazenkova@gmail.com)

---

