

## Biodiversity dynamics course



Leibniz Institute for Zoo  
and Wildlife Research

IN THE FORSCHUNGSVERBUND BERLIN E.V.

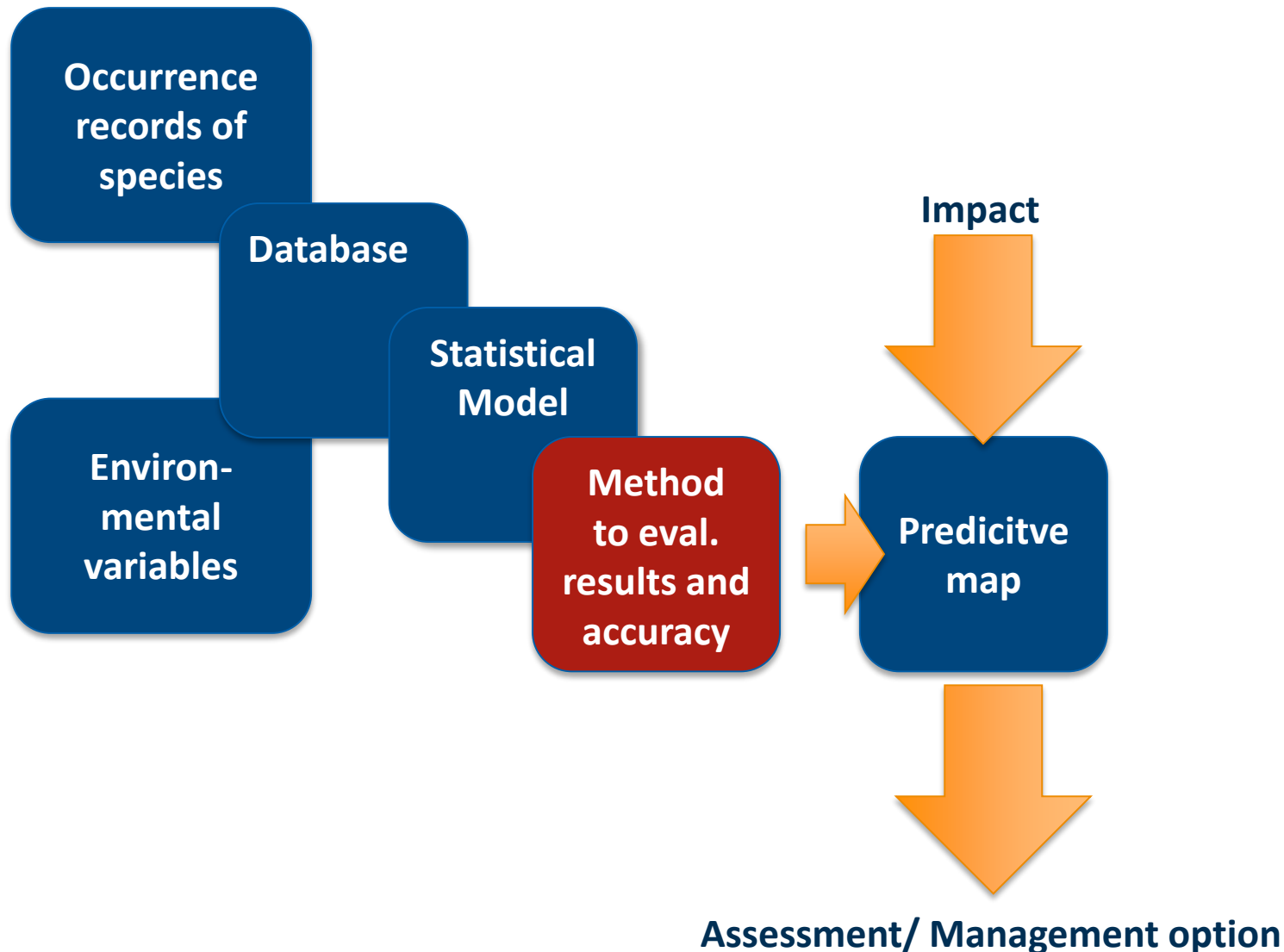
Member of the

Leibniz  
Leibniz Association

# Model evaluation & threshold selection

Stephanie Kramer-Schadt

## Flow Chart - What we need:

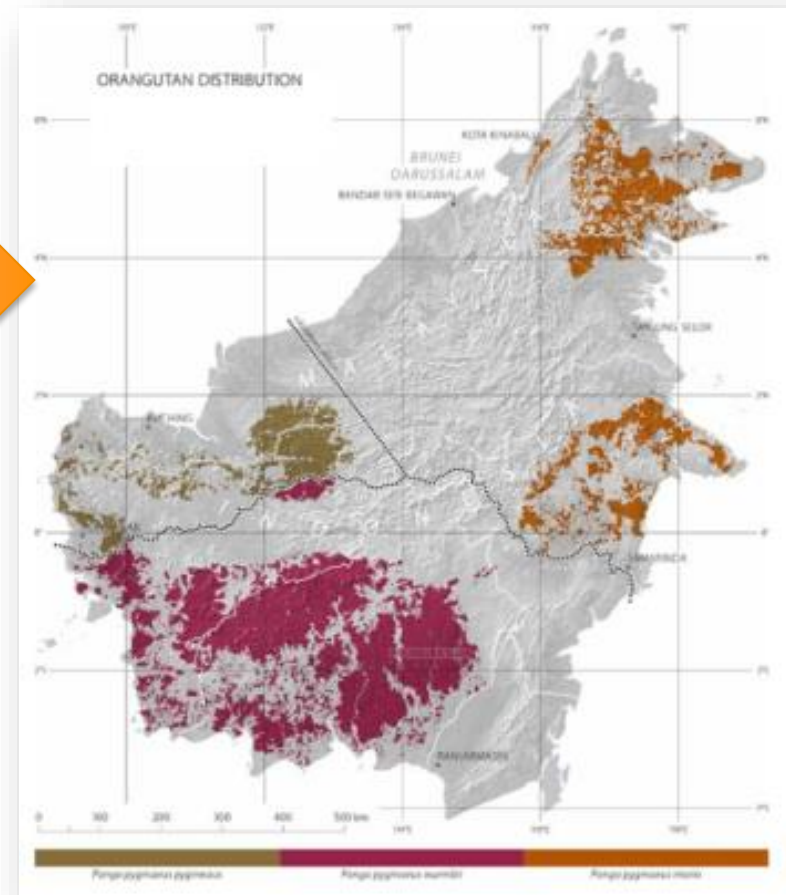
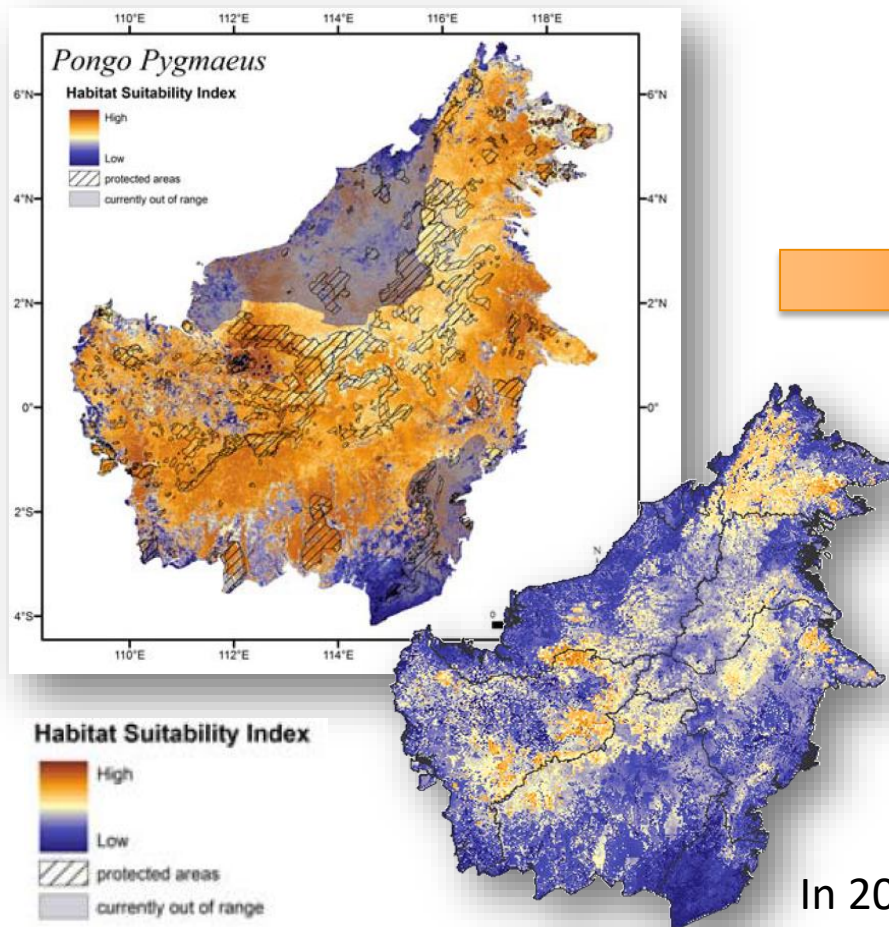


## 2 questions

- How good is our model at correctly predicting independent data
- What is the best threshold to separate the probability map into ,suitable‘ and ,unsuitable‘

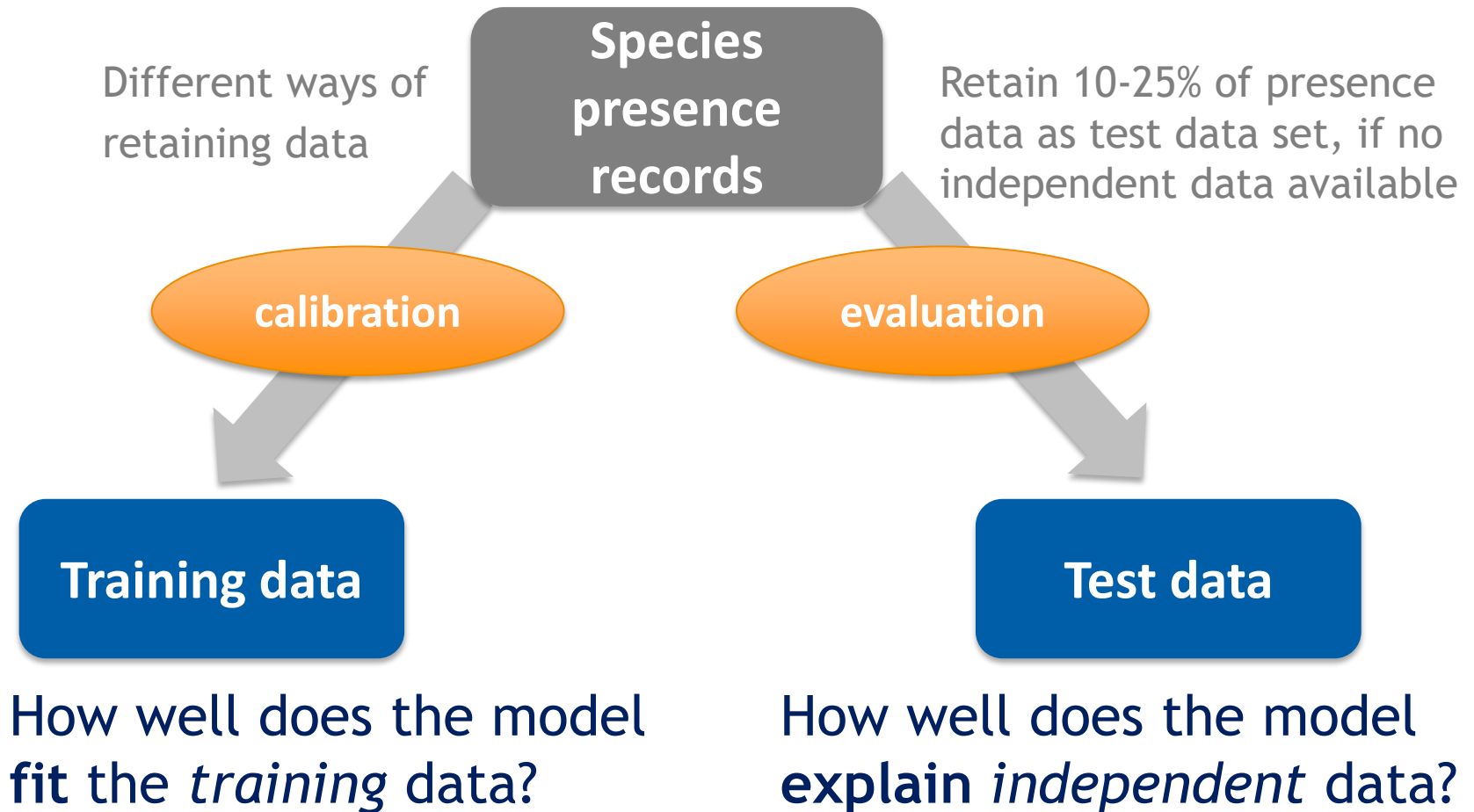
# Why is the correct threshold so important?

- Reserve design, Protected area borders!
- Planners need a binary map where to establish a protected area!



In 2080

# Independent data: Splitting the dataset



# Model calibration/ model fitting

Measuring the adequacy between model predictions and field observations.

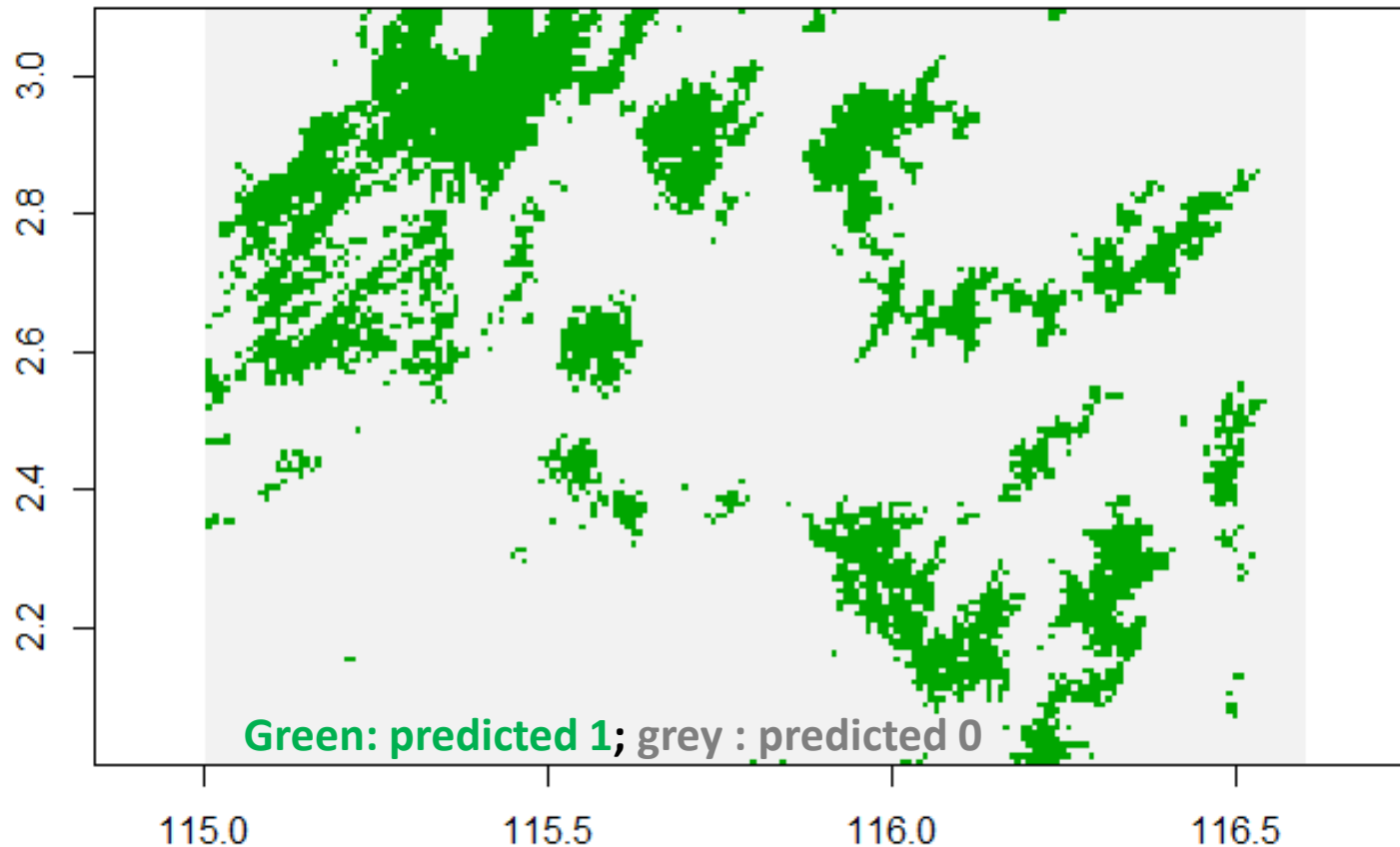
On the training data do...

- bootstrap (resampling with replacement)
- jackknife (,leave-one-out‘ = hold some points back)  
(do not mix this with the jackknife for predictors in MaxEnt!)
- or k-fold cross-validation (the training data is randomly partitioned into  $k$  equal size subsamples)

...and refit the model.

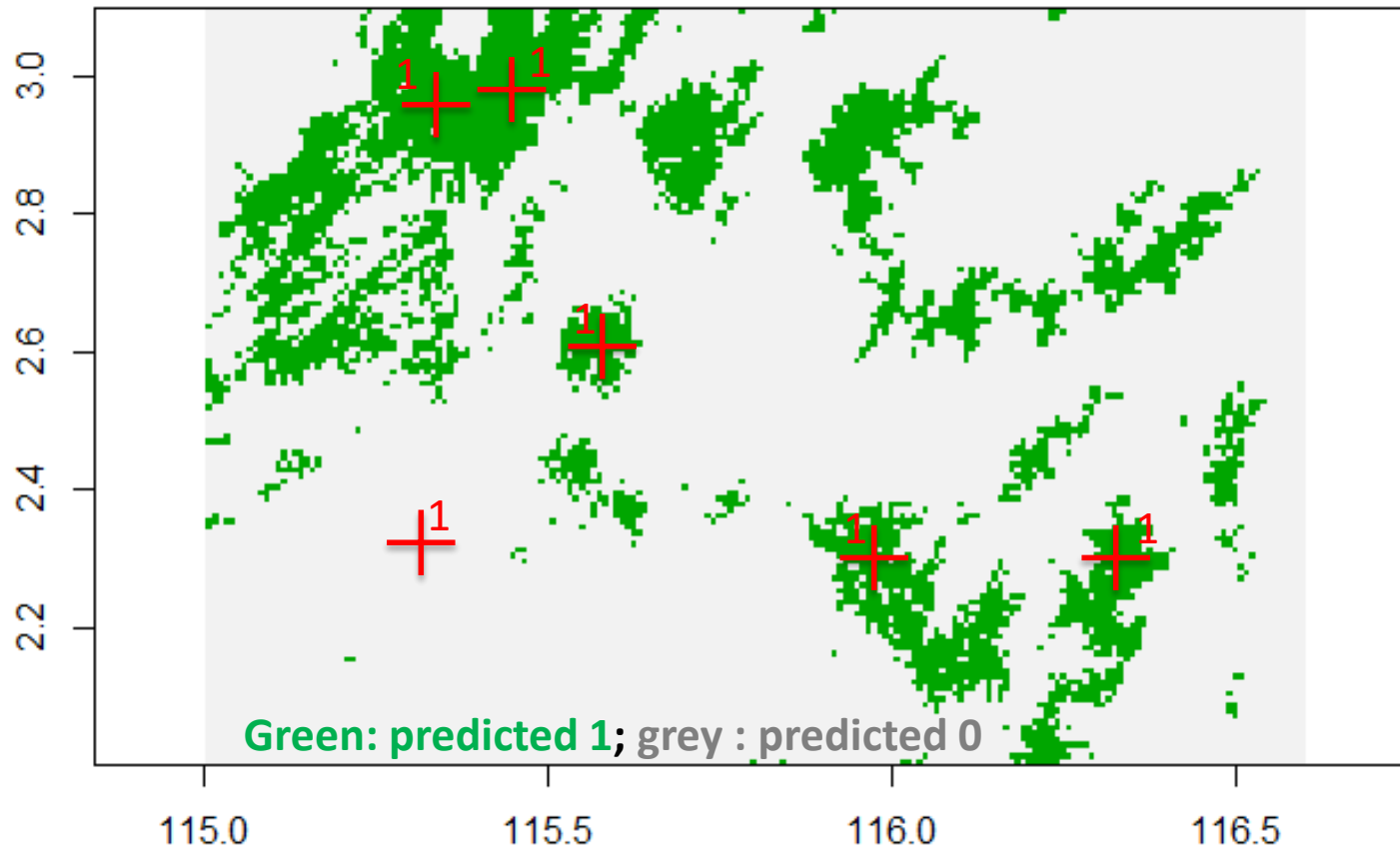
## Model evaluation diagnostics - test data

This is our model, say at  $p=0.5$  probability



## Model evaluation diagnostics - test data

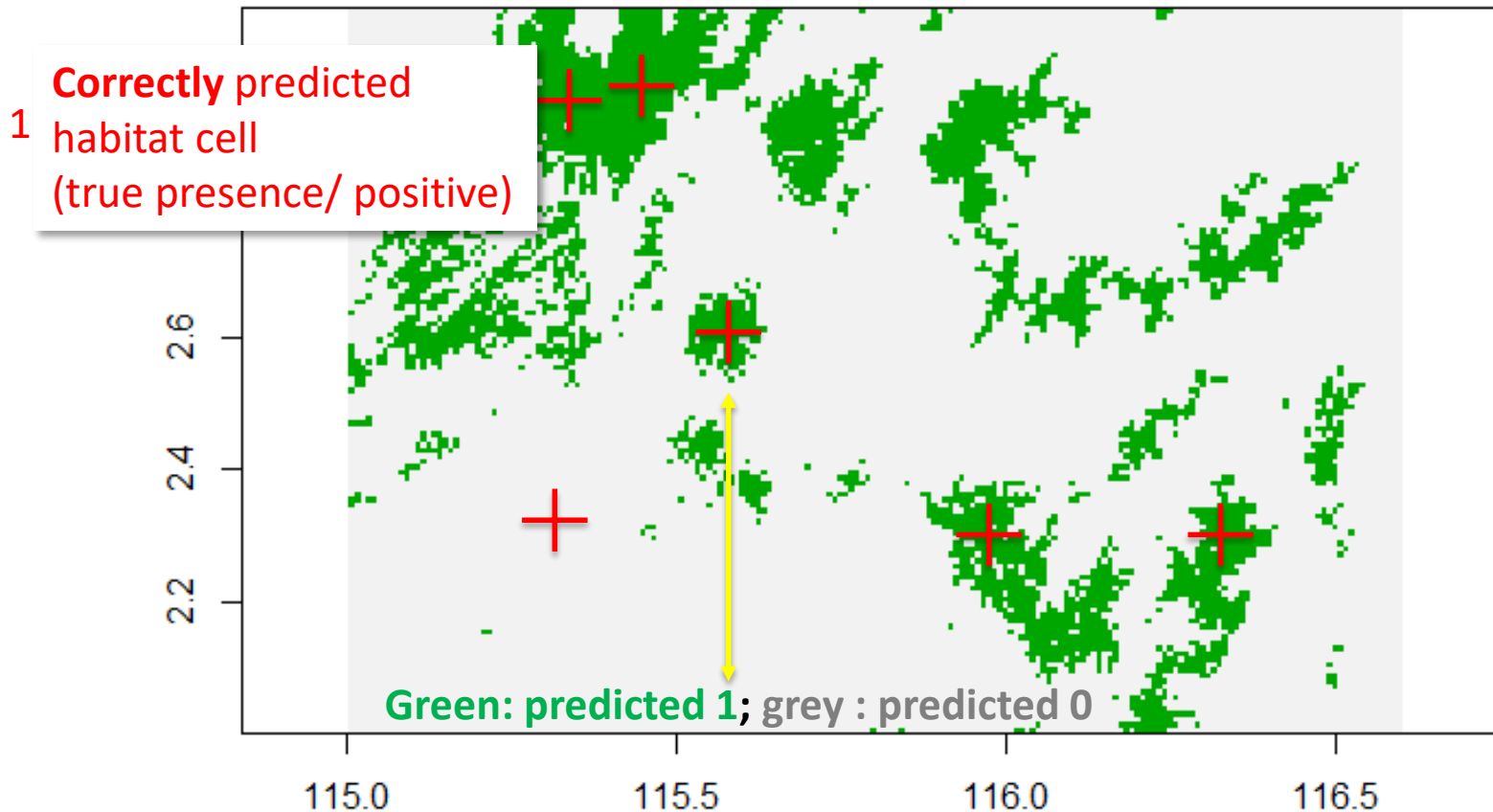
Using the test data set to assess model accuracy  
(= ability of model to correctly classify cells)





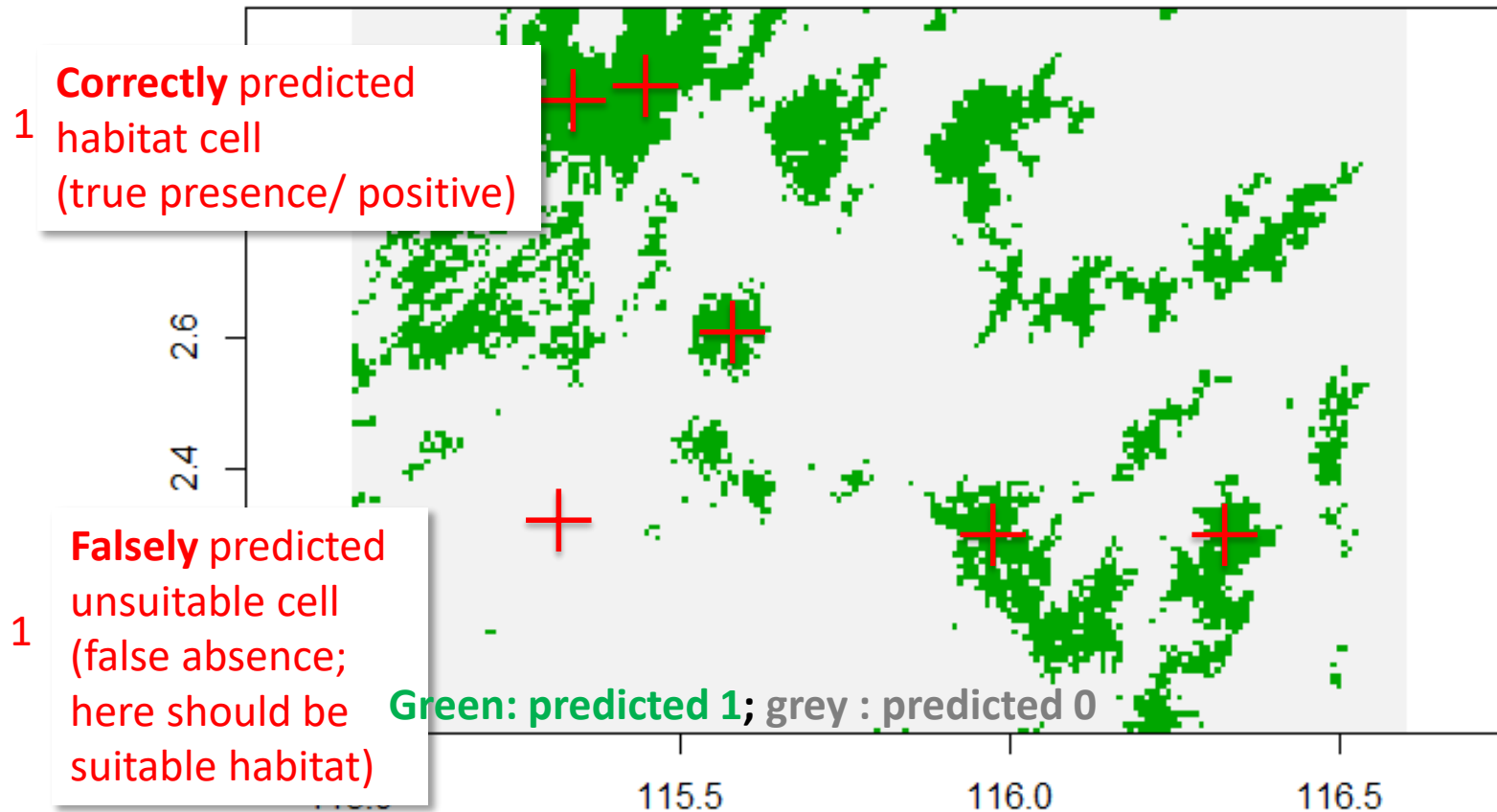
# Model evaluation diagnostics - test data

Using the test data set to assess model accuracy  
(= ability of model to correctly classify cells)



# Model evaluation diagnostics - test data

Using the test data set to assess model accuracy  
(= ability of model to correctly classify cells)



# Model evaluation diagnostics - test data

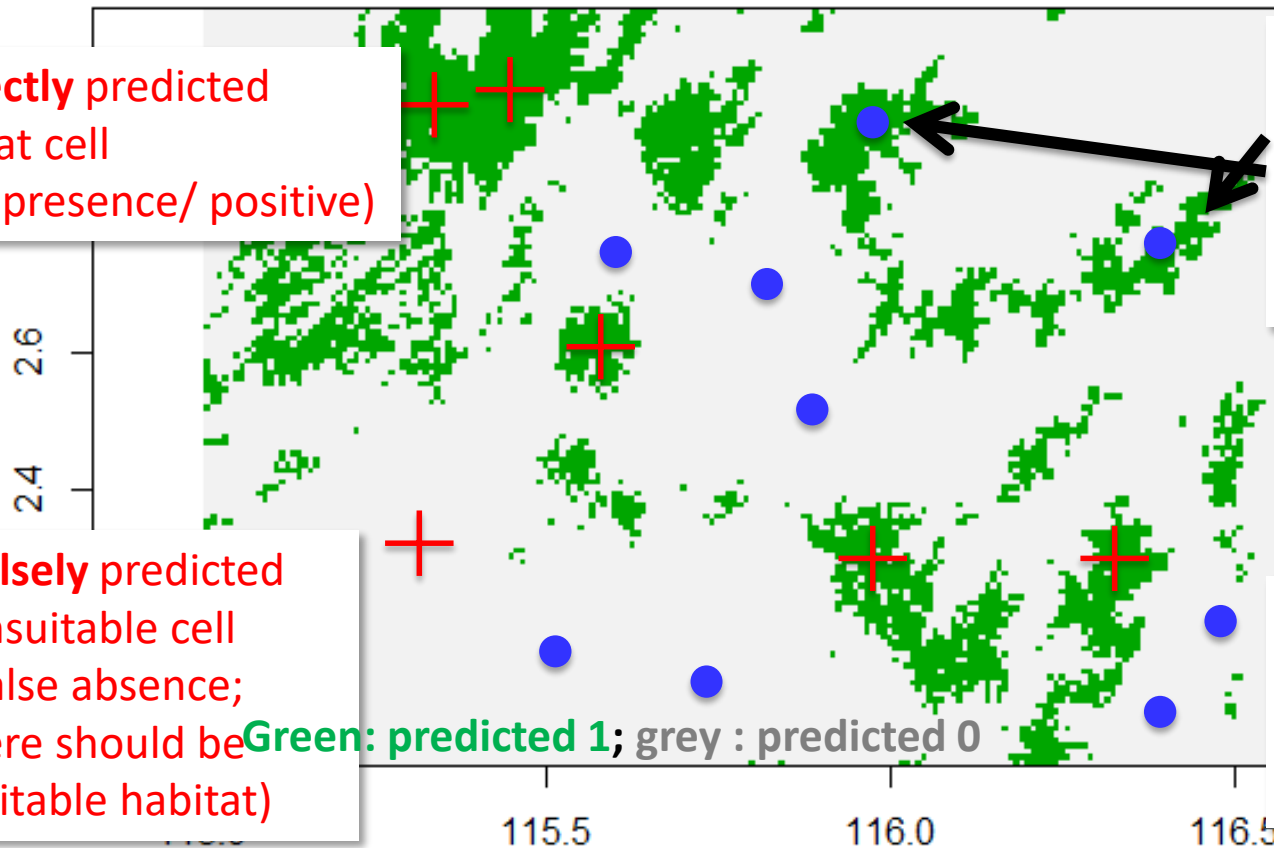
Using the test data set to assess model accuracy  
(= ability of model to correctly classify cells)

**Correctly predicted**  
habitat cell  
(true presence/ positive)

**Falsely predicted**  
habitat cell  
(false presence;  
here should not be  
Suitable habitat)





**Falsely predicted**  
unsuitable cell  
(false absence;  
here should be  
suitable habitat)

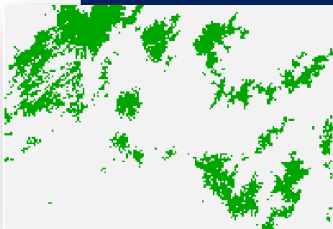
**Correctly predicted**  
unsuitable habitat  
(true absence/  
negative)



# ROC - receiver operating characteristics





- For binary classifiers
- Contingency table (also called: confusion matrix, error matrix)

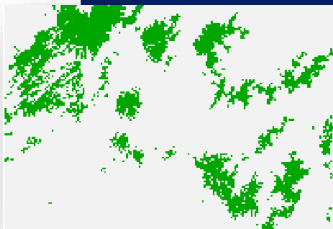
		Test data	
		Presence (positive)	Absence (negative)
Binary map; predicted model	Presence (pos.; suitable habitat)	<b>True presence (TP)</b>  	<b>False presence (FP)</b> (habitat overpredicted, commission)  
	Absence (neg.; unsuitable habitat)	<b>False absence (FN)</b> (no habitat detected, omission error)  	<b>True absence (TN)</b>   <i>What is more severe? Omission or commission?</i>



# ROC





- Sensitivity: true positive rate  $TPR = TP/P = TP/(TP+FN)$

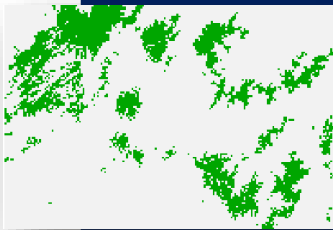
		Test data	
		Presence (positive)	Absence (negative)
Binary map; predicted model	Presence (pos.; suitable habitat)	<b>True presence (TP)</b>  	<b>False presence (FP)</b> (habitat overpredicted, commission)  
	Absence (neg.; unsuitable habitat)	<b>False absence (FN)</b> (no habitat detected, omission error)  	<b>True absence (TN)</b>  



# ROC





- Sensitivity: true positive rate  $TPR = TP/P = TP/(TP+FN)$
- Specificity: true negative rate  $TNR = TN/N = TN/(FP + TN)$

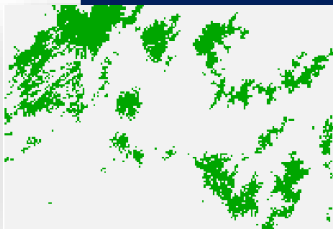
		Test data	
		Presence (positive)	Absence (negative)
Binary map; predicted model	Presence (pos.; suitable habitat)	<b>True presence (TP)</b> 	<b>False presence (FP)</b> (habitat overpredicted, commission) 
	Absence (neg.; unsuitable habitat)	<b>False absence (FN)</b> (no habitat detected, omission error) 	<b>True absence (TN)</b> 



# ROC - trade-off plot

- ROC is plotting  
Sensitivity by  $(1 - \text{Specificity}) = \text{Sensitivity by FPR} = \text{TPR by FPR}$   
 $(= \text{FP} / (\text{FP} + \text{TN}))$

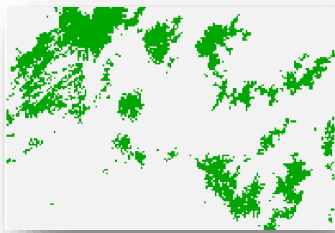
		Test data	
		Presence (positive) P	Absence (negative) N
Binary map; predicted model	Presence (pos.; suitable habitat)	<b>True presence (TP)</b> 	<b>False presence (FP)</b> (habitat overpredicted, commission) 
	Absence (neg.; unsuitable habitat)	<b>False absence (FN)</b> (no habitat detected, omission error) 	<b>True absence (TN)</b> 



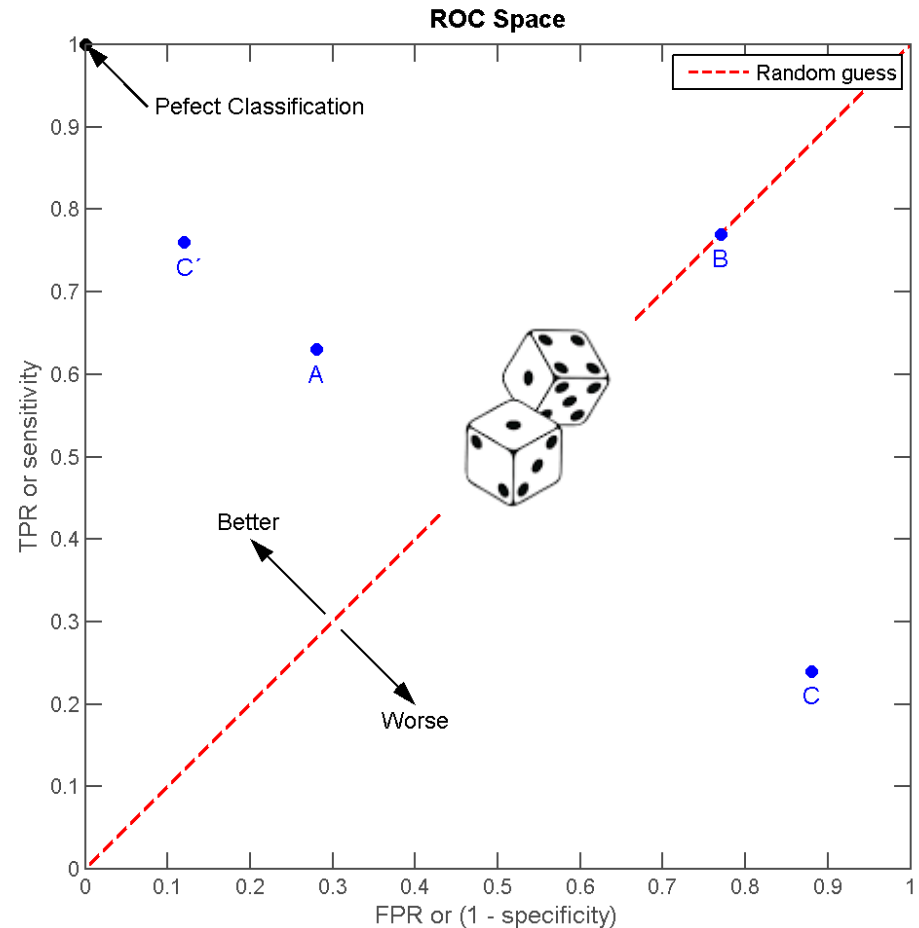
# ROC space

relative trade-off between true positive (benefits) and false positive (costs)

A-C' = **different** test data sets



A			B		
TP=63	FP=28	91	TP=77	FP=77	154
FN=37	TN=72	109	FN=23	TN=23	46
100	100	200	100	100	200
C			C'		
TP=24	FP=88	12	TP=76	FP=12	88
FN=76	TN=12	88	FN=24	TN=88	112
100	100	200	100	100	200

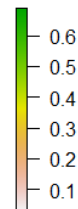
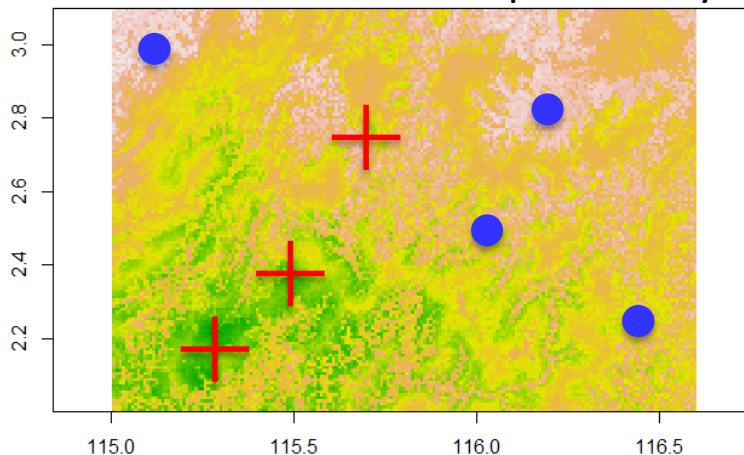






# ROC curve or AUC (area under curve)

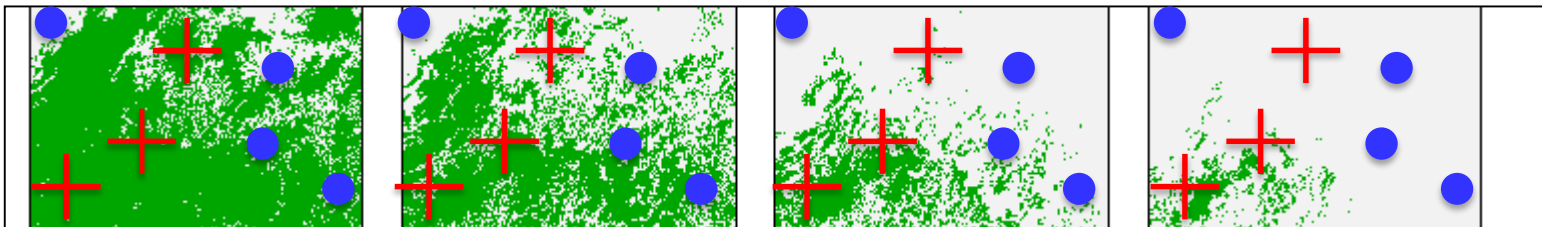
- ROC for different cut-off points of the model (data fixed, binary model changes!)
- Relative probabilities of map are continuous between 0 and 1
- Cut-off values (e.g. 0.1, 0.5, 0.7,...) separate map in binary classes (suitable vs unsuitable habitat)
- ROC can be defined across the whole value space of model

Relative occurrence probability



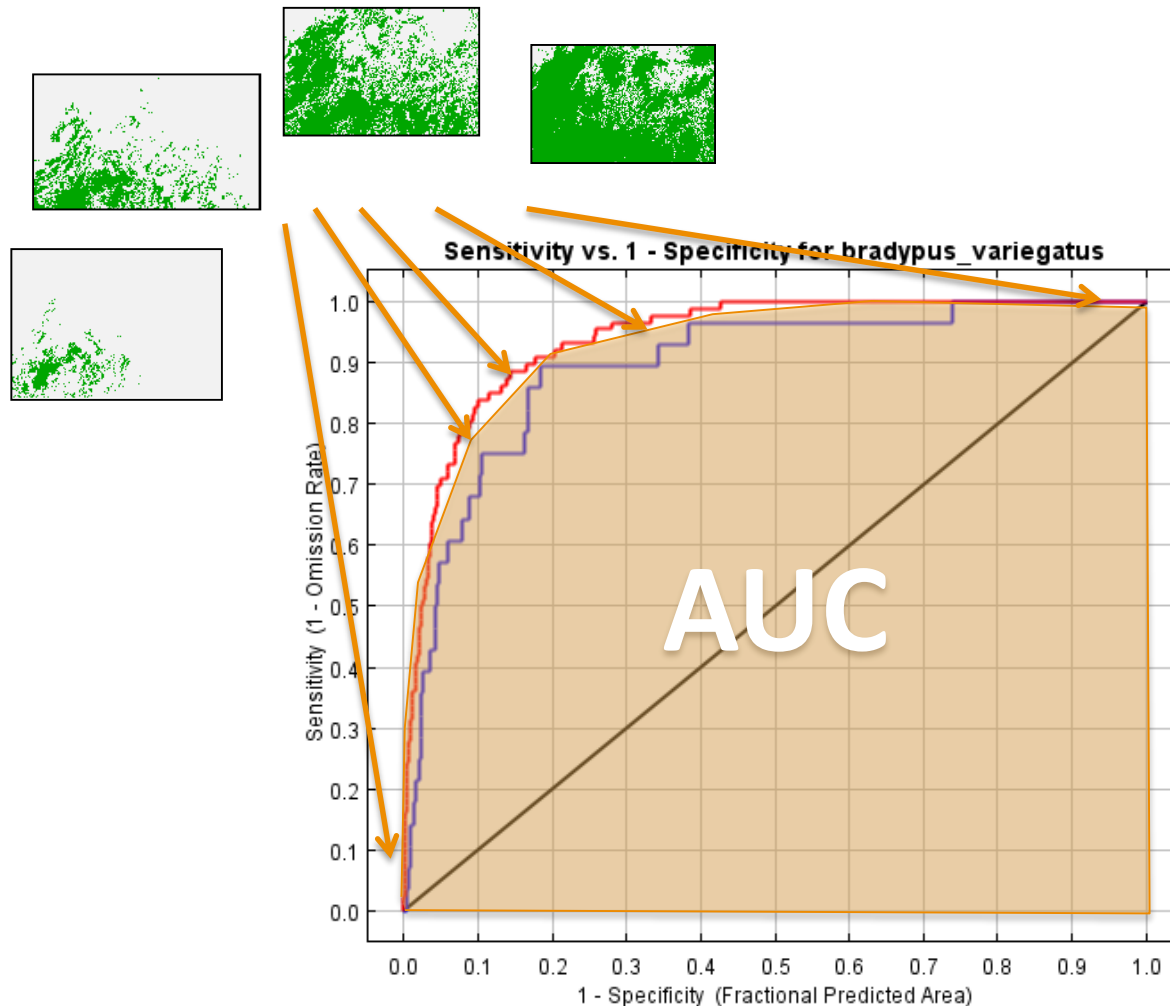
 Presence  
 absence  
 From same test data set

Binary maps



# AUC

Each point corresponds to a different cut-off value (threshold)

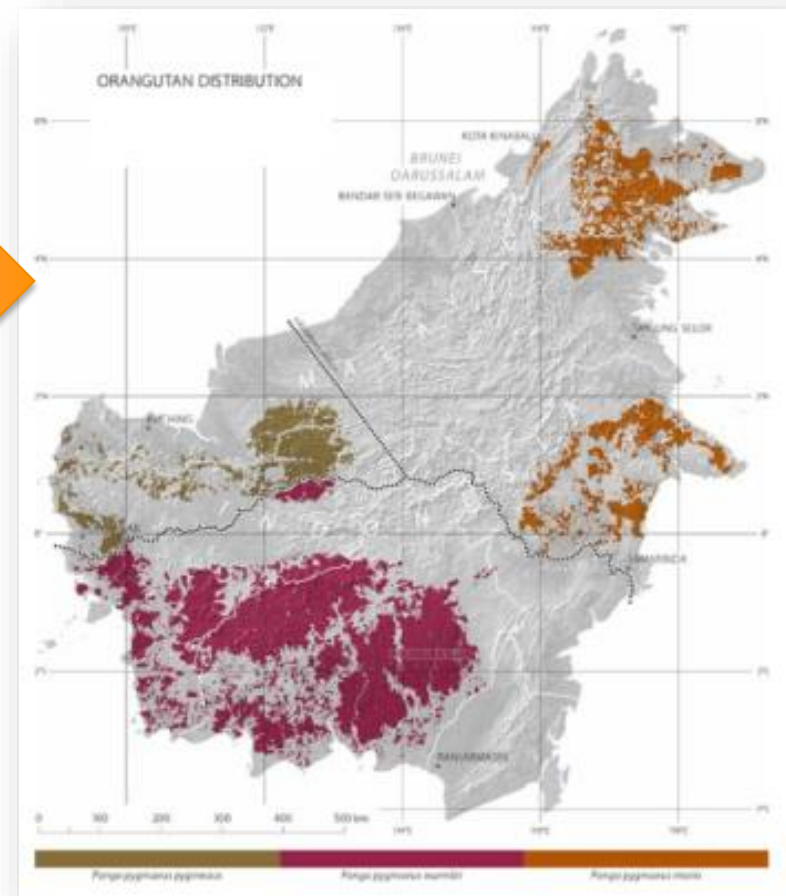
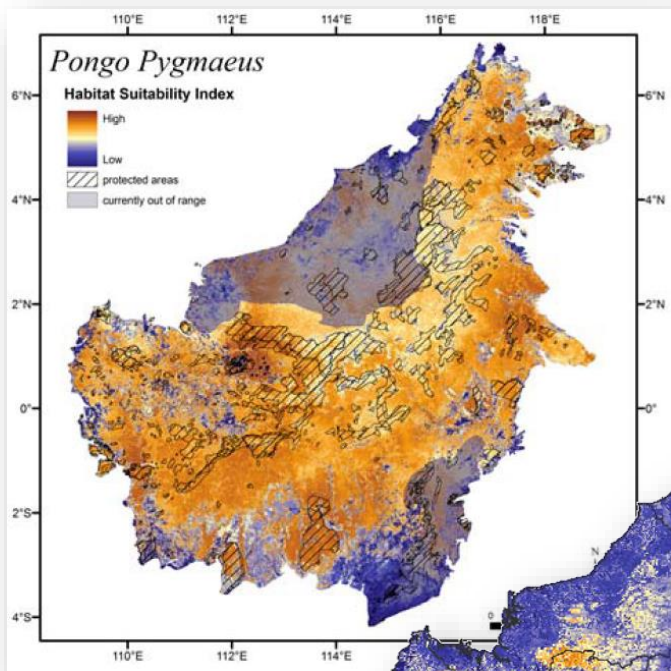
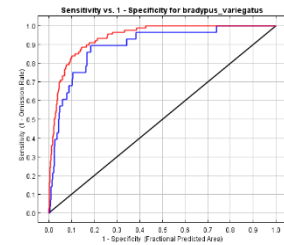


AUC  
> 0.7 good  
> 0.5 fair  
= 0.5 random  
< 0.5 low

discriminative ability  
of model.

# Why is the correct threshold so important?

- Reserve design, Protected area borders!
- Planners need a binary map where to establish a protected area!



In 2080

## Further reading

# Methods in Ecology and Evolution



*Methods in Ecology and Evolution* 2014, 5, 54–64

doi: 10.1111/2041-210X.12123

## Prevalence, thresholds and the performance of presence–absence models

Callum R. Lawson<sup>1,2\*</sup>, Jenny A. Hodgson<sup>3</sup>, Robert J. Wilson<sup>1</sup> and Shane A. Richards<sup>4</sup>

*Journal of Applied  
Ecology* 2006  
43, 1223–1232

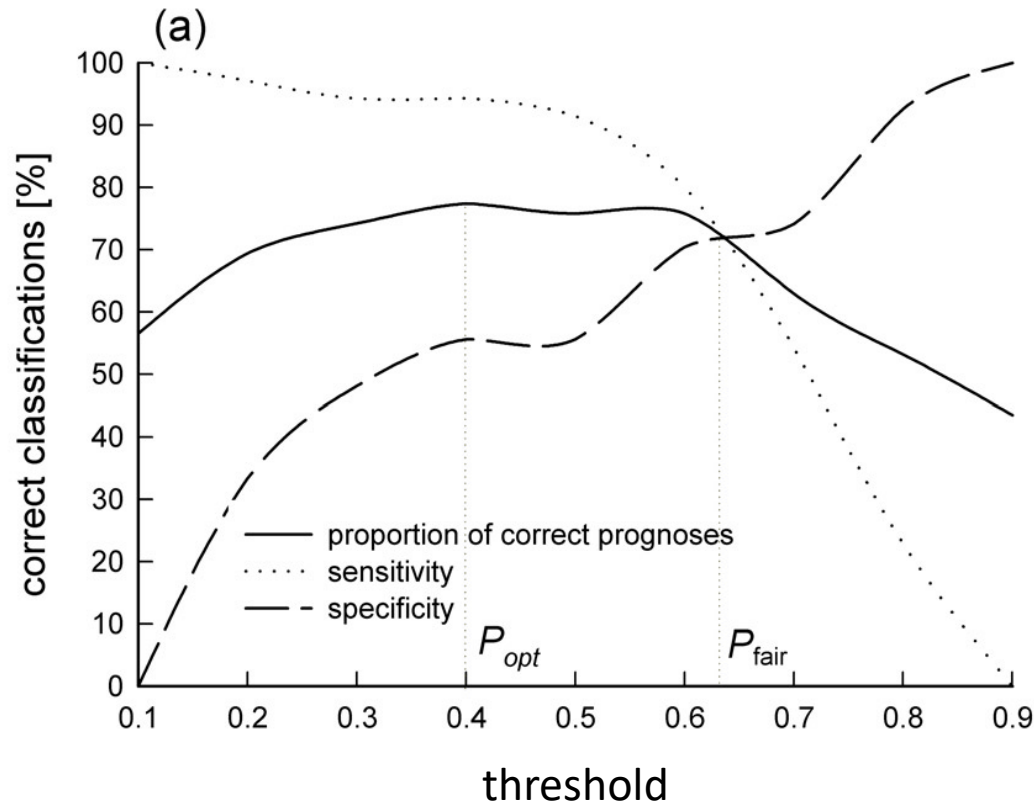
### METHODOLOGICAL INSIGHTS

## Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)

OMRI ALLOUCHE, ASAF TSOAR and RONEN KADMON

*Department of Evolution, Systematics and Ecology, Institute of Life Sciences, The Hebrew University, Givat-Ram, Jerusalem 91904, Israel*

# Threshold selection



- ROC-based (most widespread)
- Cohen's kappa (robust)
- 10th percentile (MaxEnt)
- fixed (usually at 0.5 - not recommended!!!)

Example for ROC-based threshold selection;  
Schadt et al. 2002 JAE



## Further reading

# Selecting thresholds of occurrence in the prediction of species distributions

Canran Liu, Pam M. Berry, Terence P. Dawson and Richard G. Pearson

Table 1. Indices for assessing the predictive performance of species distribution models, a is true positives (or presences), b is false positives (or presences), c is false negatives (or absences), d is true negatives (or absences), n (=a+b+c+d) is the total number of sites and  $\alpha$  is a parameter between 0 and 1 (inclusive).

Index	Formula
Sensitivity (or Recall, R)	$a/(a+c)$
Specificity	$d/(b+d)$
Precision (P)	$a/(a+b)$
Overall prediction success (OPS)	$(a+d)/n$
Kappa	$\frac{(a+d) - [(a+c)(a+b) + (b+d)(c+d)]/n}{n - [(a+c)(a+b) + (b+d)(c+d)]/n}$
Odds ratio	$(ad)/(cb)$