Evaluating species distribution models and making predictions

Damaris Zurell

https://damariszurell.github.io

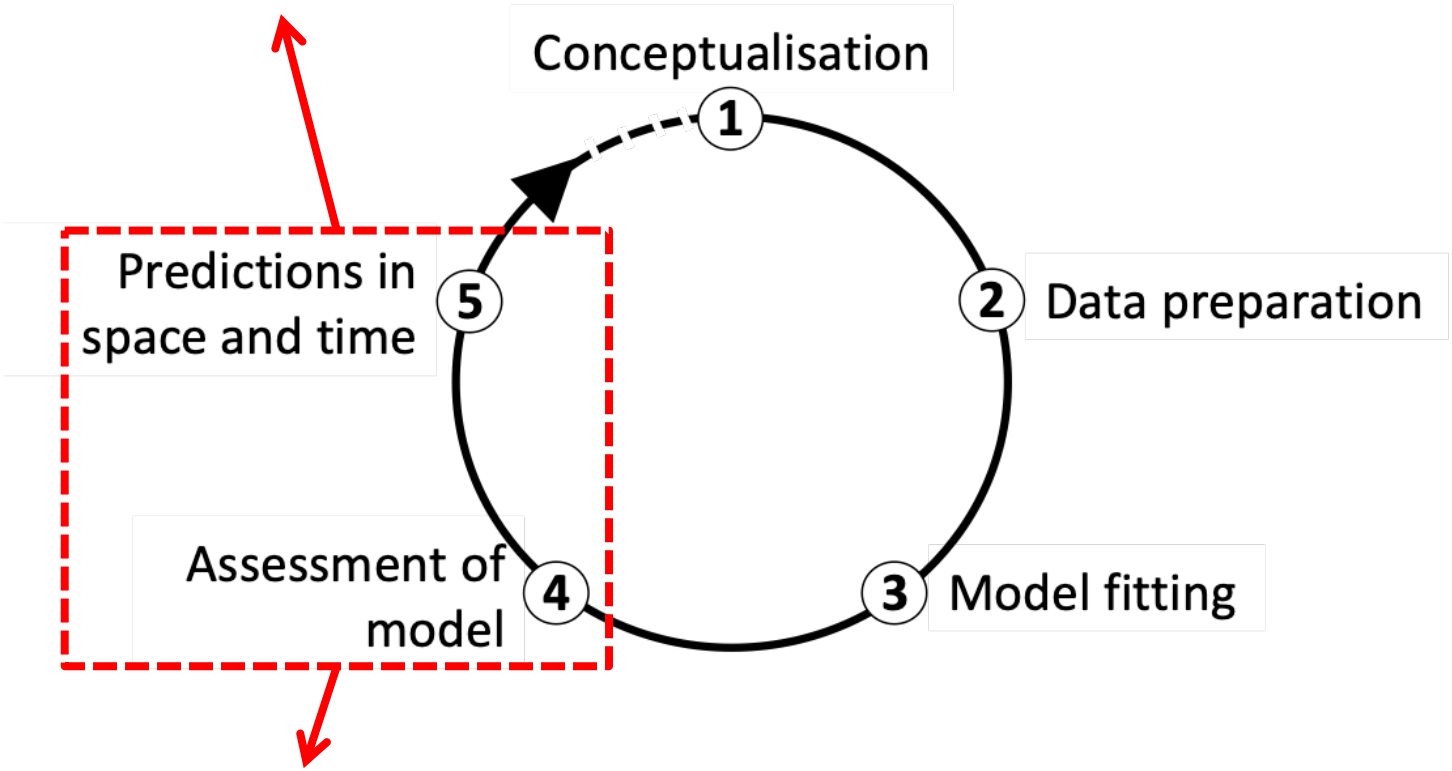
@ZurellLab

Universität Potsdam

Pictures from www.undekade-biologischevielfalt.de

SDM – model building steps

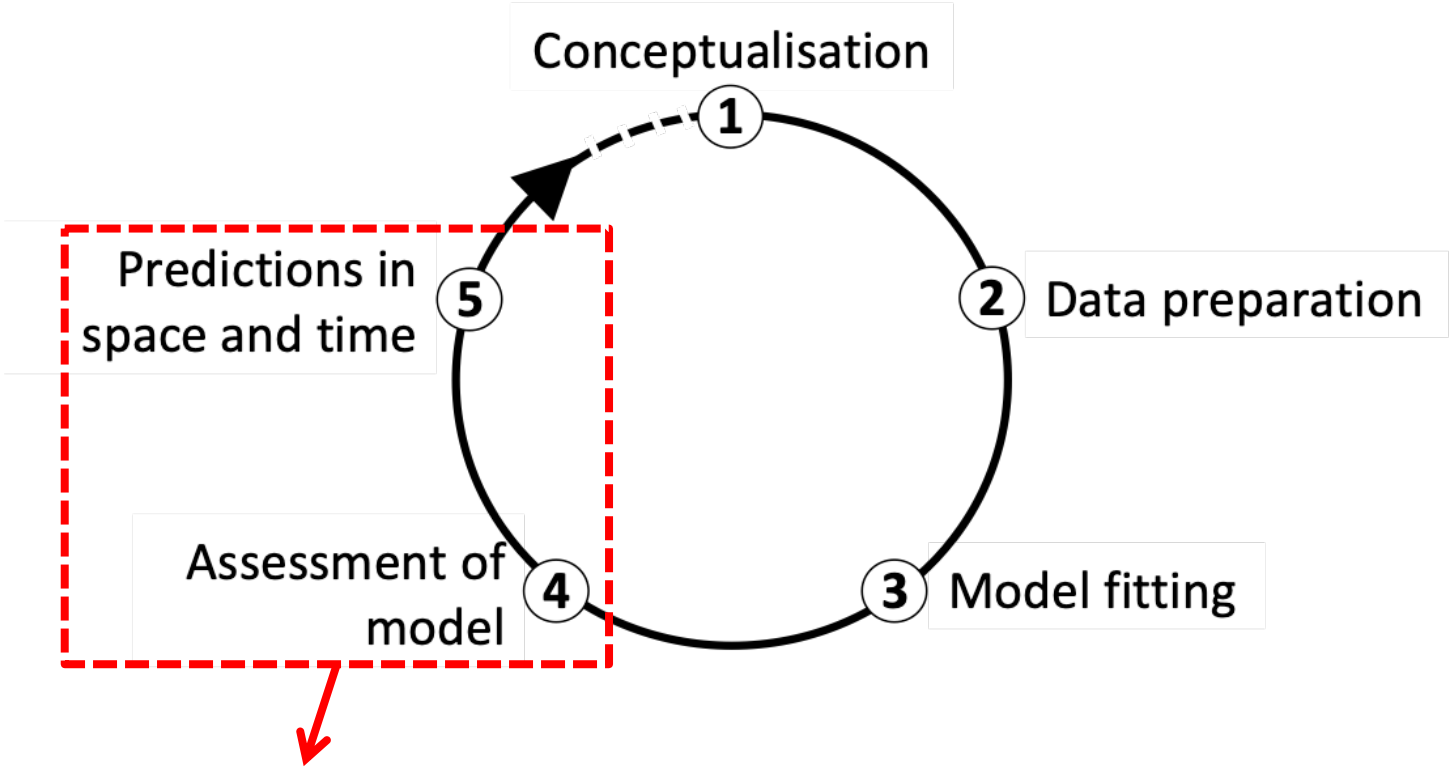
- What is the potential distribution of the species?
- How certain is this prediction?



- How does the species-environment relationship look like?
- How well is my model supported by data?
- How well does my model predict to independent data?

Zurell et al. (2020) Ecography 43: 1261-1277.

SDM – model building steps



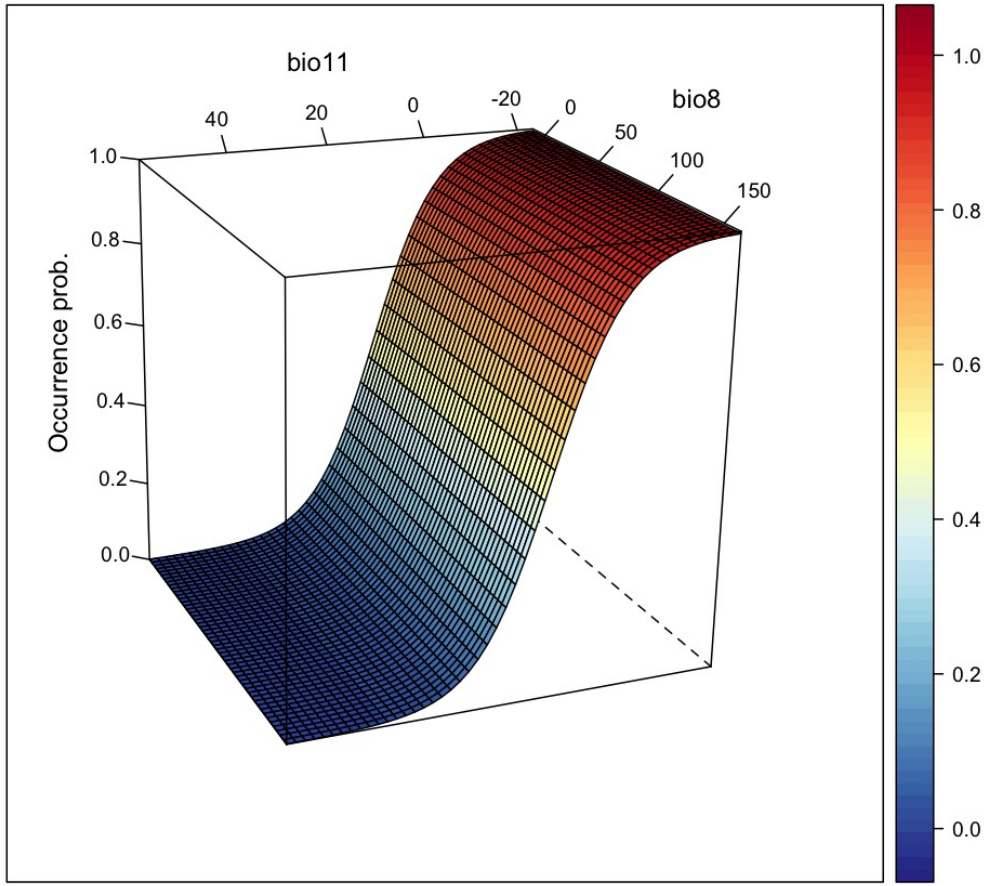
➢ How does the species-environment relationship look like?

Zurell et al. (2020) Ecography 43: 1261-1277.

SDM – visualise response

Response surfaces

- Purpose: visualise model predictions along two environmental gradients

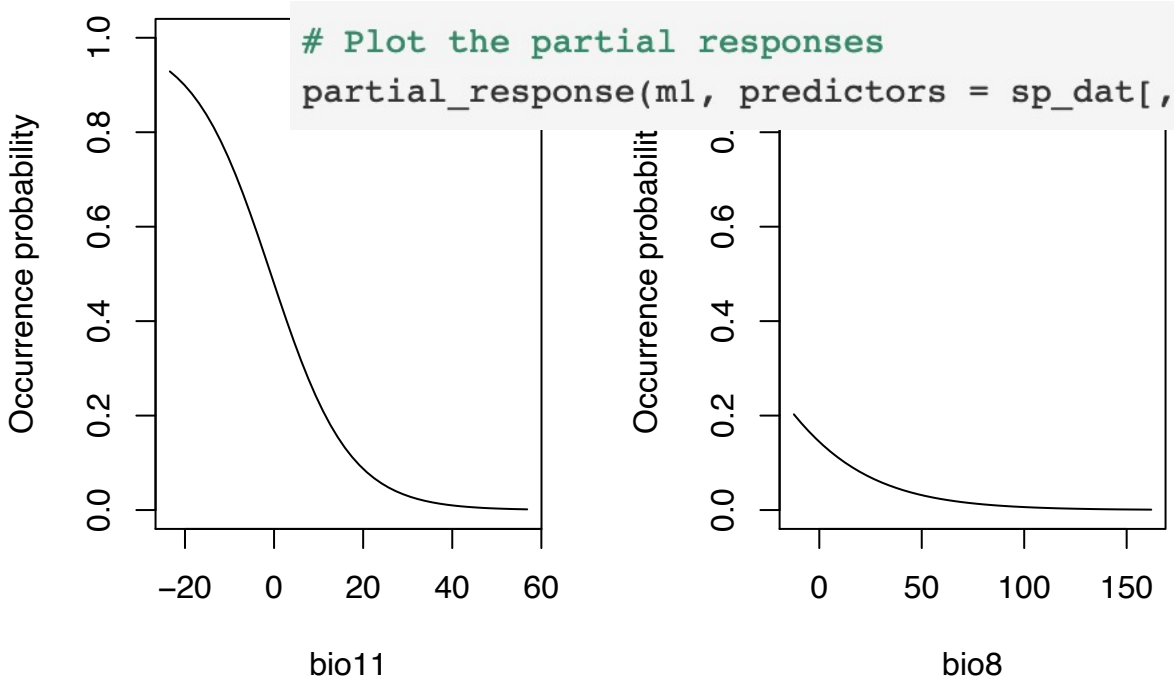


SDM – visualise response

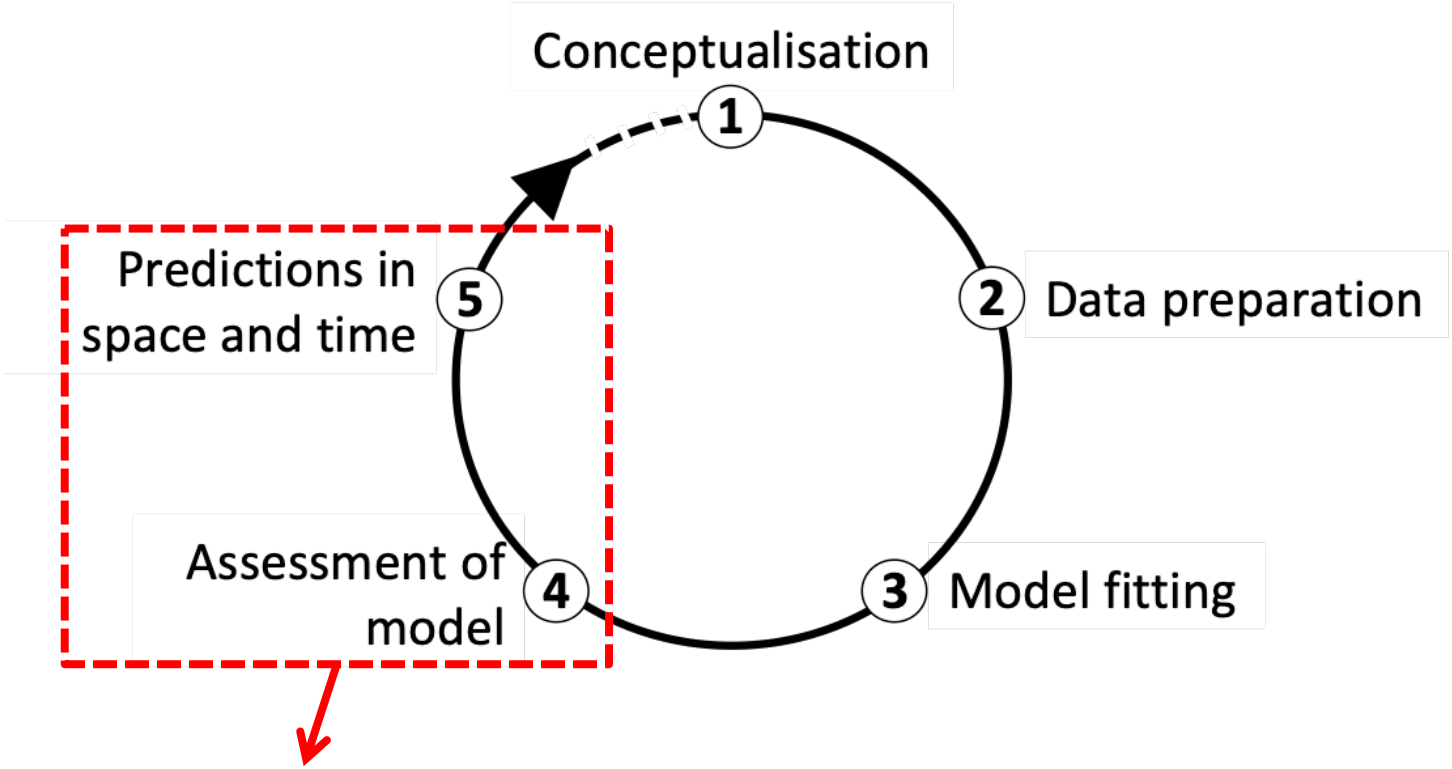
Partial response plots

- Purpose: abstract model behaviour to 2D
- Approach: plot response curve for each predictor separately while keeping the other predictors at their mean

```
# Plot the partial responses
partial_response(m1, predictors = sp_dat[,my_preds])
```



SDM – model building steps



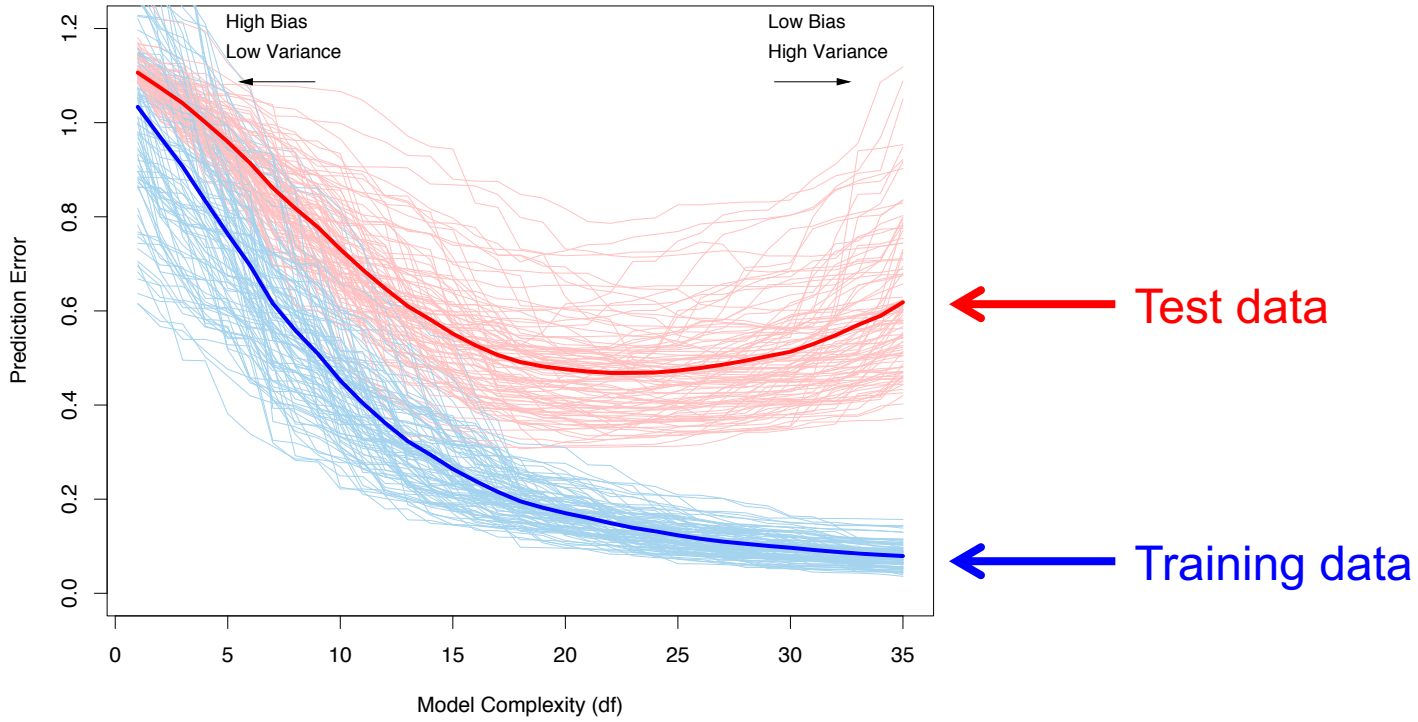
- How well is my model supported by data?
- How well does my model predict to independent data?

Zurell et al. (2020) Ecography 43: 1261-1277.

SDM – model assessment

The Bias-Variance trade-off:

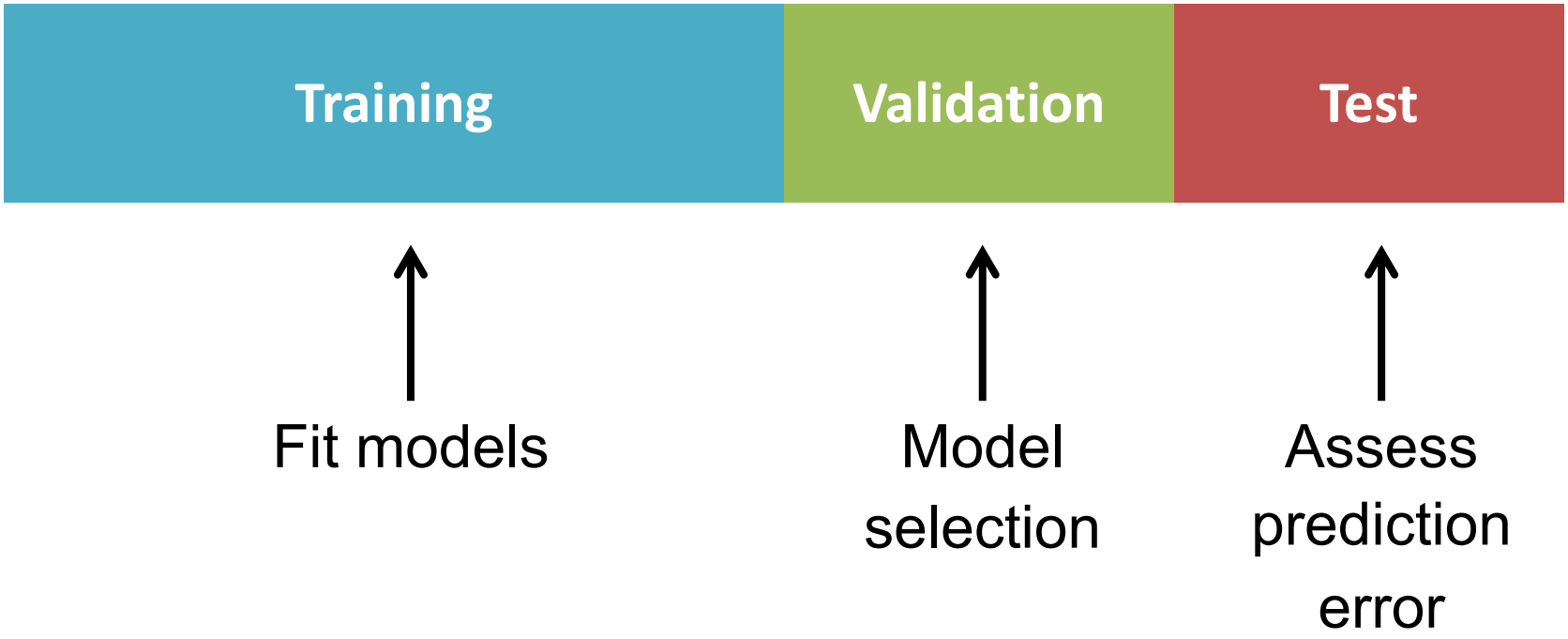
- Bias is caused by simplifying model assumptions
- Variance is caused by fluctuations in the data



Hastie et al. (2009) Elements of statistical learning. Springer

SDM – model assessment

Ideally, predictive performance should be validated on independent test data.



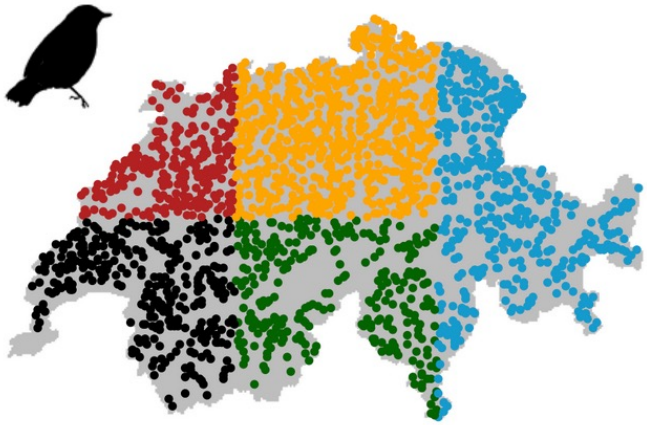
Hastie et al. (2009) Elements of statistical learning. Springer

SDM – model assessment

Ideally, predictive performance should be validated on independent test data (= external validation).

Typical approaches when independent test data missing:

- **Internal validation**: only on training data
- **Split-sample**, e.g. 70% training - 30% test
- **k-fold cross-validation**: (k-1)/k proportion training – 1/k proportion test, repeat k times
- **k-fold block cross-validation**:
  - Spatial blocks, or
  - Environmental blocks



SDM - performance measures

Goodness-of-fit: typically derived from the log-likelihood

- Example: explained deviance D^2

Log-likelihood:

$$L(\beta) = ln[l(\beta)] = \sum_{i=1}^{n} (y_i \times ln[\pi(x_i)] + (1 - y_i) \times ln[1 - \pi(x_i)])$$

Deviance:

$$D = -2 \times L$$

Explained deviance:

$$D^2 = 1 - \frac{D(model)}{D(Null.model)}$$

SDM - performance measures

Discrimination: in how far can model distinguish between presences and absences?

- Threshold-dependent measures:
  - Sensitivity (true positive rate)
  - Specificity (true negative rate)
  - TSS (true skill statistic)
  - Kappa

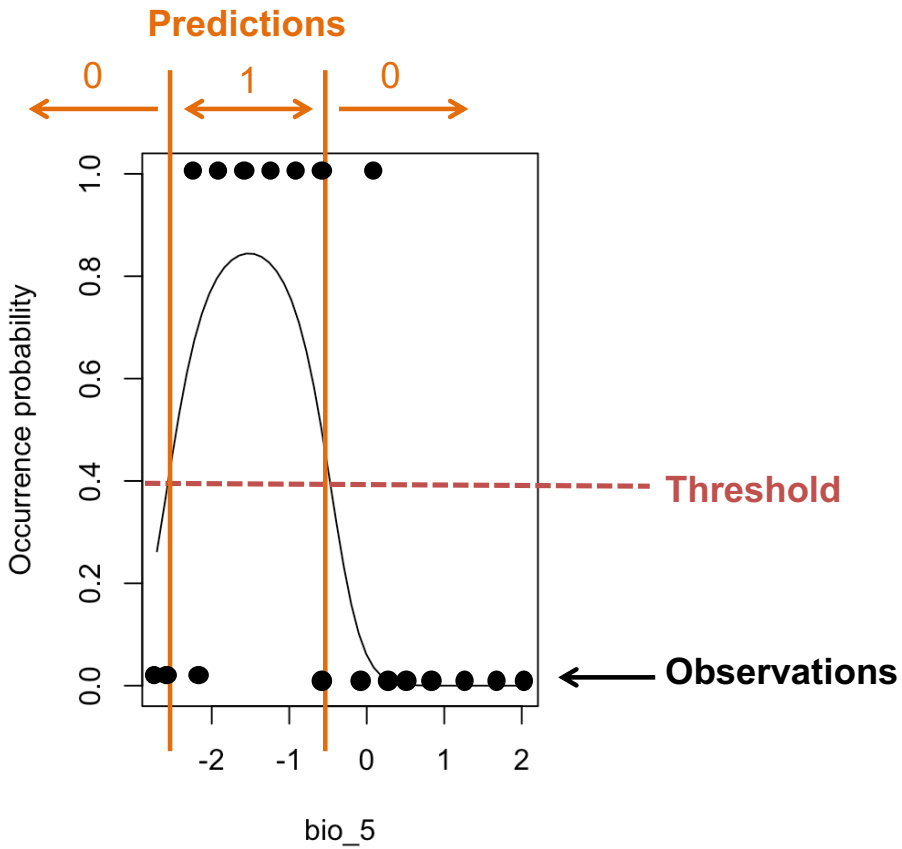
- Threshold-independent measures:
  - AUC (area under the receiver operating characteristic curve)

SDM - performance measures

Threshold-dependent measures:

Derived from confusion matrix

Predictions \ Observations	1	0
1	a	b
0	c	d

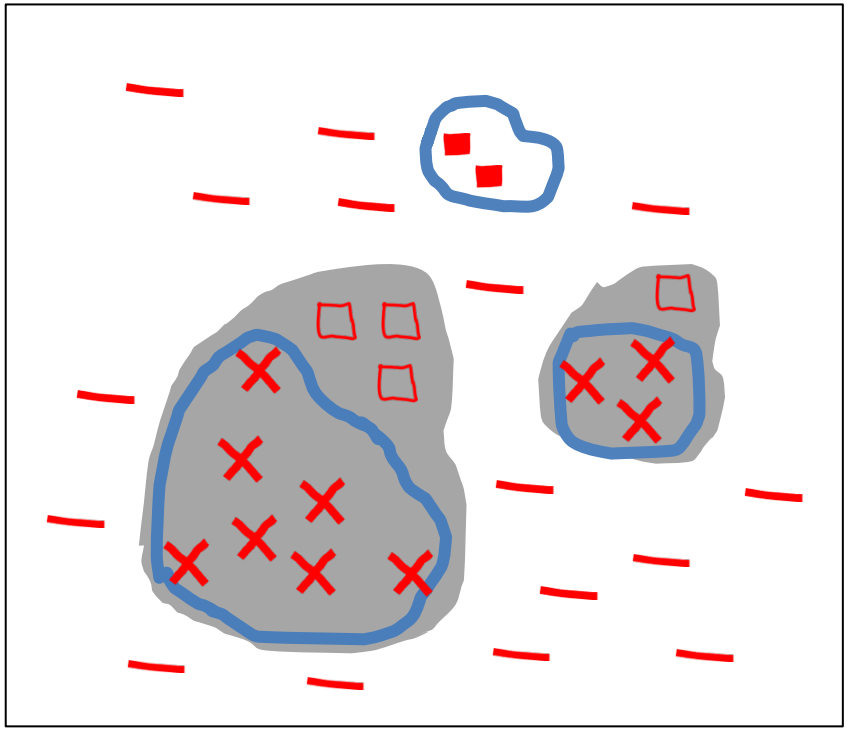


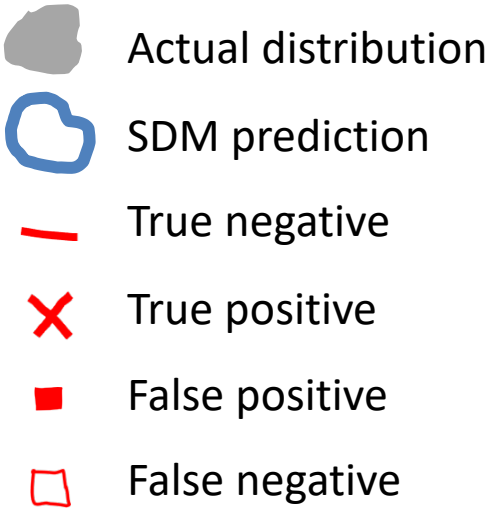
SDM - performance measures

Threshold-dependent measures:

Derived from confusion matrix

Predictions \ Observations	1	0
1	a	b
0	c	d





SDM - performance measures

Threshold-dependent measures:

Derived from confusion matrix

Predictions \ Observations	1	0
1	a	b
0	c	d

Measure	Formula	Fair prediction?
Overall accuracy	$\frac{a+d}{n}$	
Sensitivity	$\frac{a}{a+c}$	Sens > 0.75
Specificity	$\frac{d}{b+d}$	Spec > 0.75
Kappa statistic	$\frac{\left(\frac{a+d}{n}\right) - \frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}{1 - \frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}$	Kappa > 0.4
TSS	sensitivity + specificity – 1	TSS > 0.5

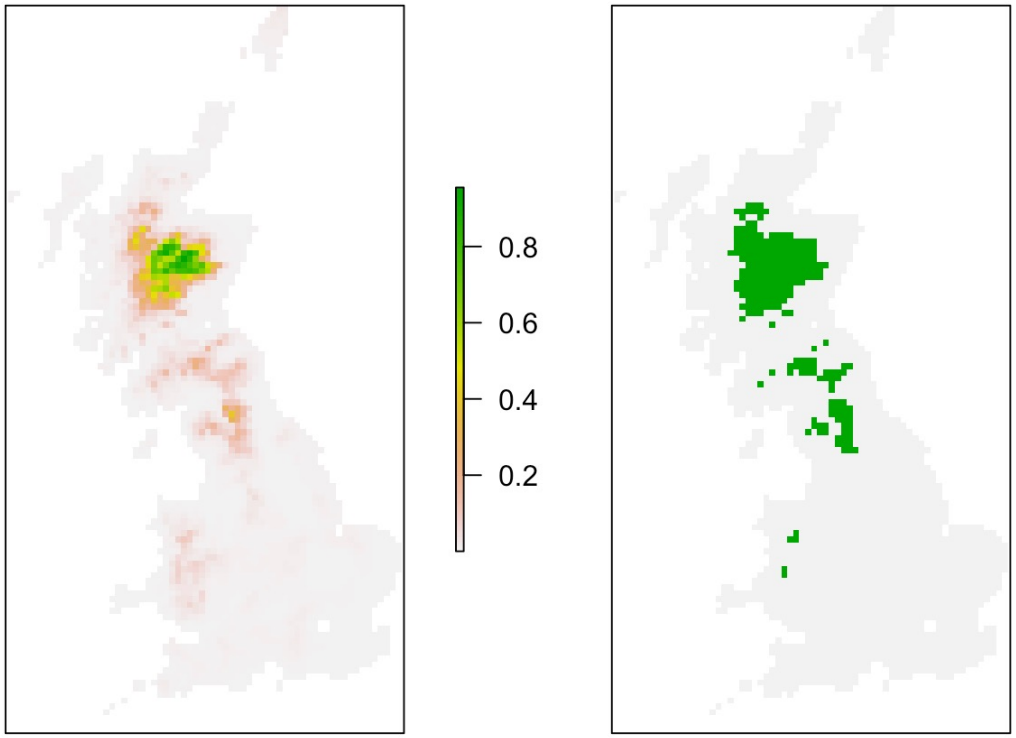
Allouche et al. (2006) J Applied Ecology 43: 1223-1232.

SDM - performance measures

Threshold-dependent measures:

Need to select a threshold!

Continuous SDM prediction => Presence/Absence prediction



SDM - performance measures

Threshold-dependent measures:

Need to select a threshold!

Table 2. Threshold-determining approaches studied in this paper.

Code	Approach	Definition	Reference
Subjective approach			
1	Fixed threshold approach	Taking a fixed value, usually 0.5, as the threshold	Manel et al. (1999), Bailey et al. (2002)
Objective approaches			
	Single index-based approaches:		
2	Kappa maximization approach	Kappa statistic is maximized	Huntley et al. (1995), Guisan et al. (1998)
3	OPS maximization approach	Overall prediction success (OPS) is maximized	
	Model-building data-only-based approach:		
4	Prevalence approach	Taking the prevalence of model-building data as the threshold	Cramer (2003)
	Predicted probability/suitability-based approaches:		
5	Average probability/suitability approach	Taking the average predicted probability/ suitability of the model-building data as the threshold	Cramer (2003)
6	Mid-point probability/suitability approach	Mid-point between the average probabilities of or suitabilities for the species' presence for occupied and unoccupied sites	Fielding and Haworth (1995)
	Sensitivity and specificity-combined approaches:		
7	Sensitivity-specificity sum maximization approach	The sum of sensitivity and specificity is maximized	Cantor et al. (1999), Manel et al. (2001)
8	Sensitivity-specificity equality approach	The absolute value of the difference between sensitivity and specificity is minimized	Cantor et al. (1999)
9	ROC plot-based approach	The threshold corresponds to the point on ROC curve (sensitivity against 1-specificity) which has the shortest distance to the top-left corner (0,1) in ROC plot	Cantor et al. (1999)
	Precision and recall-combined approaches:		
10	Precision-recall break-even point approach	The absolute value of the difference between precision and recall is minimized	Shapire et al. (1998)
11	P-R plot-based approach	The threshold corresponds to the point on P-R (Precision-Recall) curve which has the shortest distance to the top-right corner (1,1) in P-R plot	
12	F maximization approach	The index F is maximized. In this study, $\alpha = 0.5$ is used in F, i.e. there is no preference to precision and recall	Shapire et al. (1998)

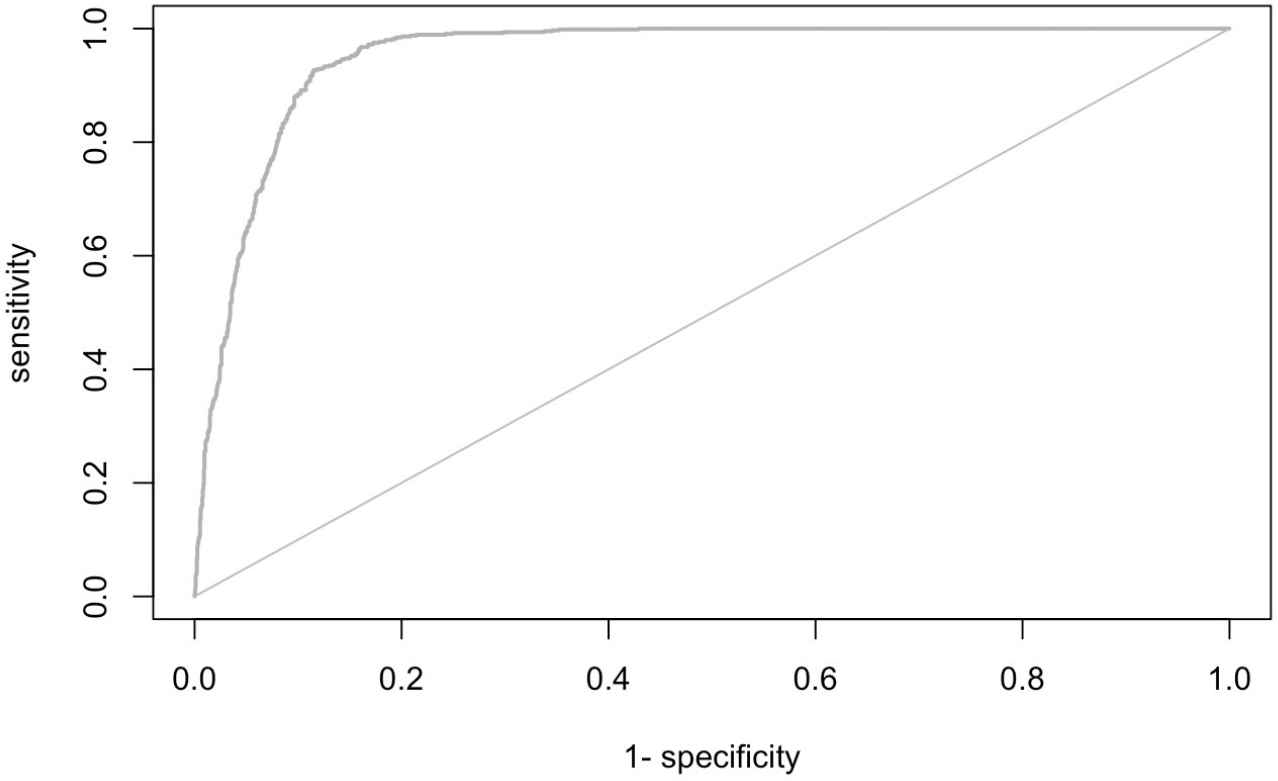
TSS

Liu et al. (2005) Ecography 28: 385-393.

SDM - performance measures

Threshold-independent measures:

- AUC (area under the receiver operating characteristic curve)

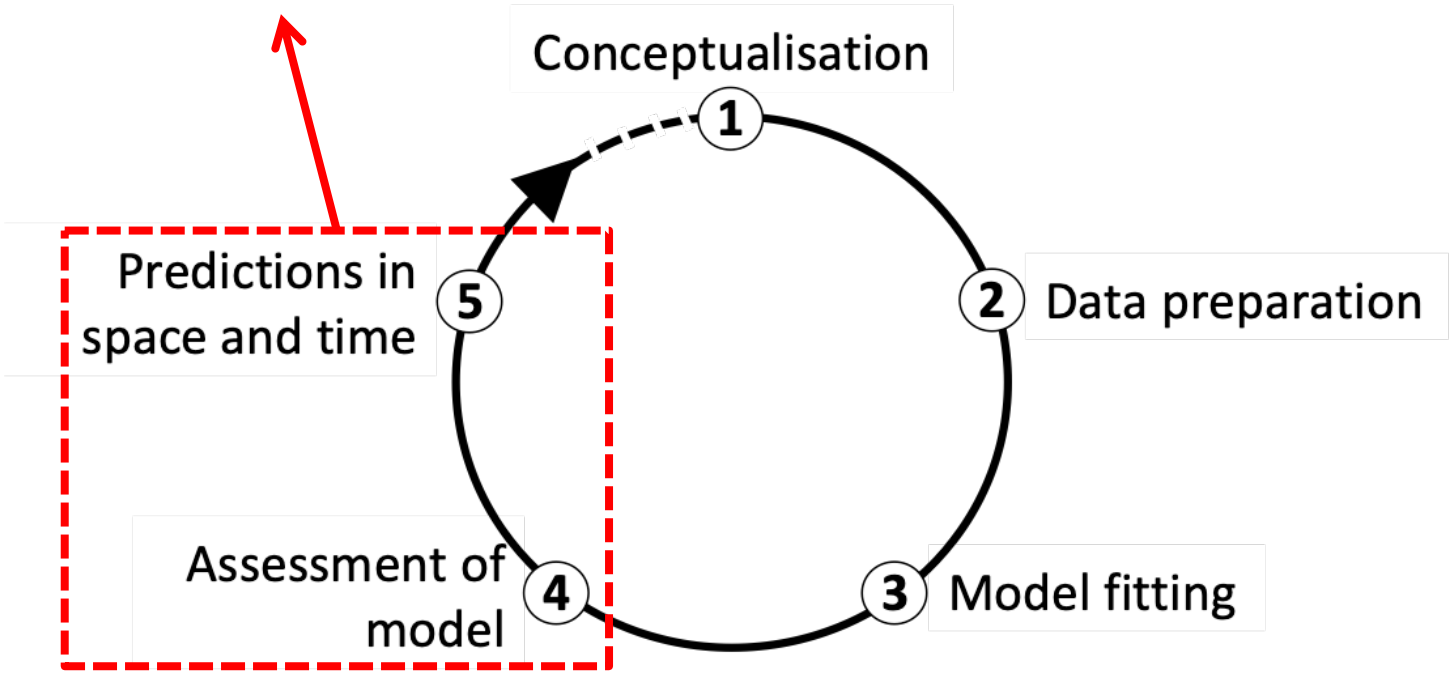


Fair prediction?

AUC > 0.7

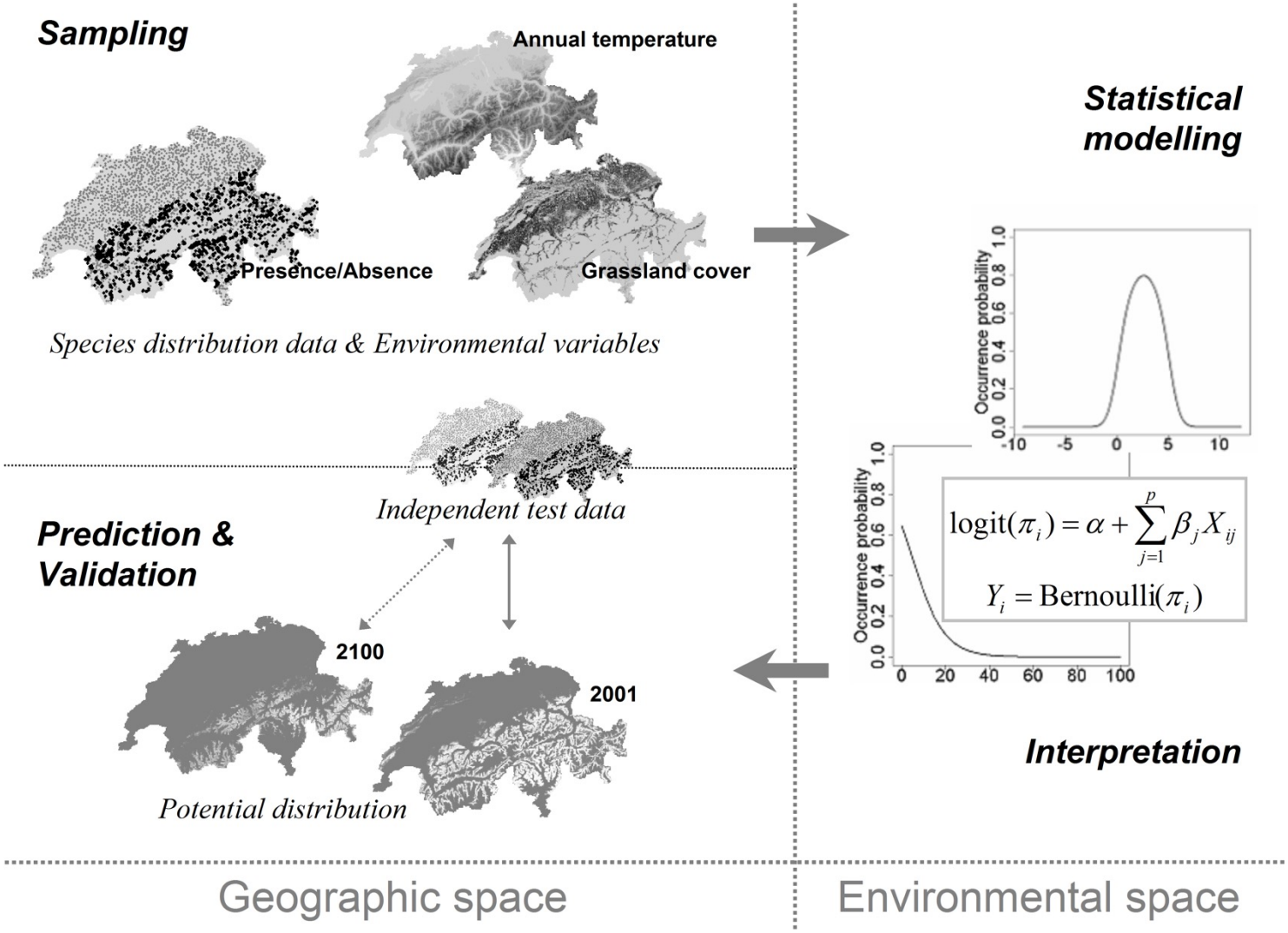
SDM – model building steps

- What is the potential distribution of the species?
- How certain is this prediction?



Zurell et al. (2020) Ecography 43: 1261-1277.

Species distribution models

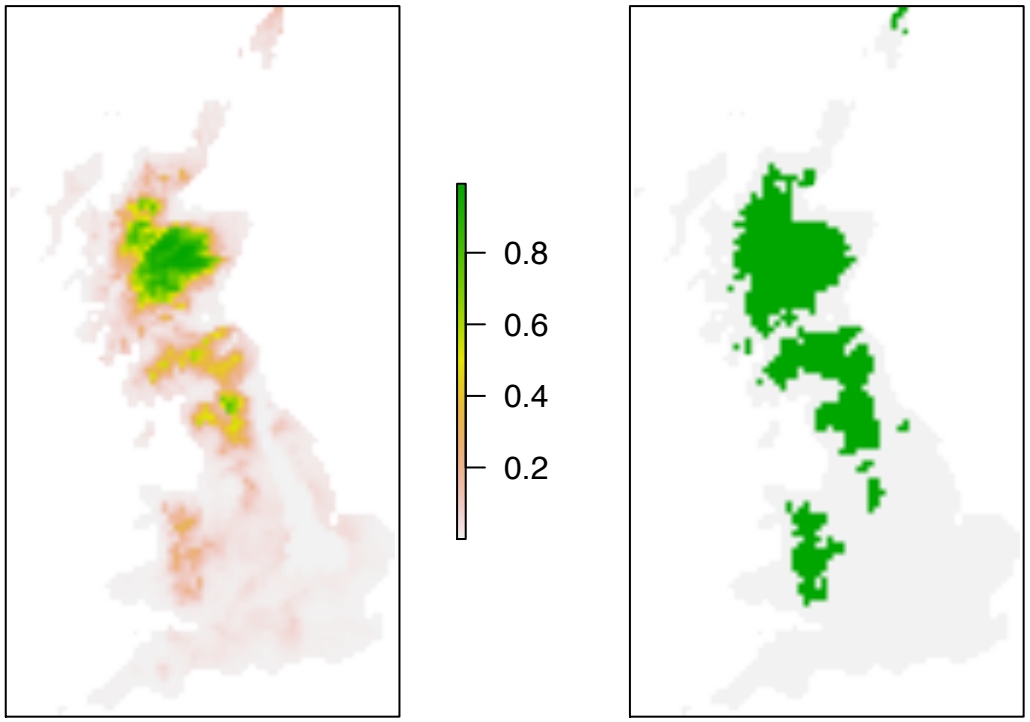


SDM predictions

Spatiotemporal predictions:

```
predict(m1, newdata= bio_fut_df, type="response")
```

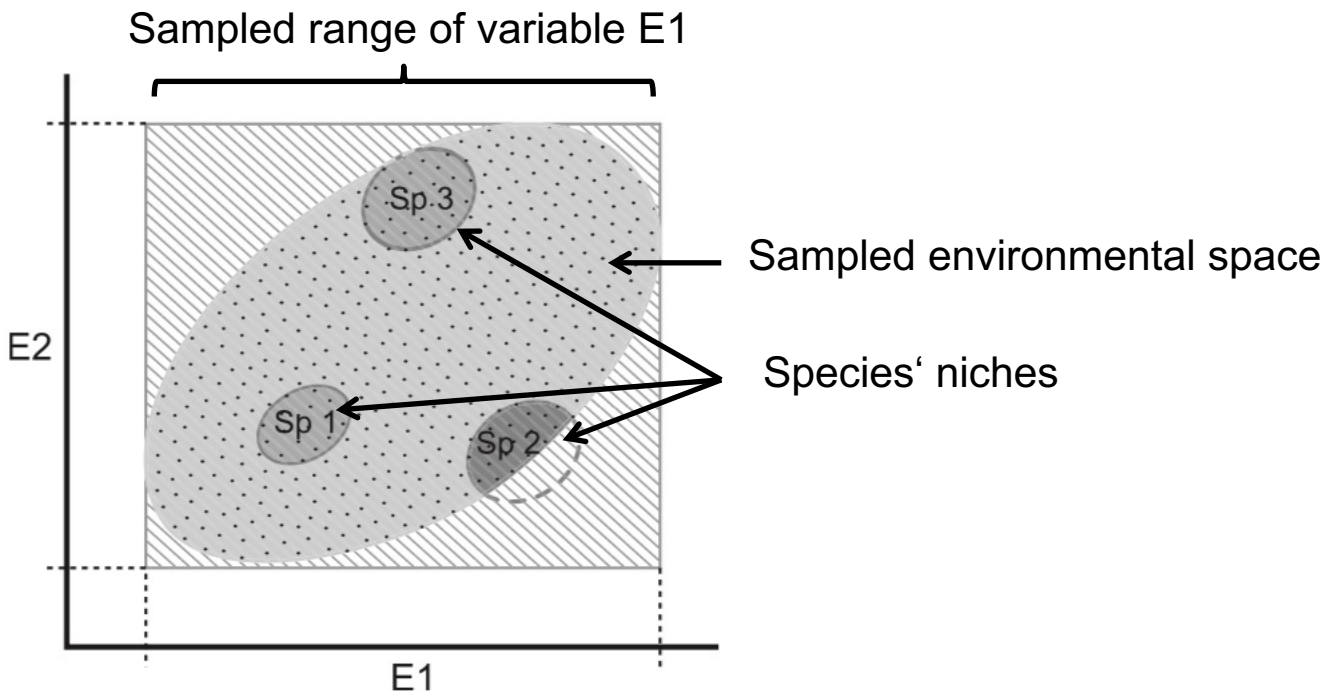
Predicted occurrence probabilities → (Threshold) → Binary predictions



SDM predictions

Spatiotemporal predictions:

- Interpolation: within sampled environmental space
- Extrapolation: beyond sampled environmental space



Zurell et al. (2012) Diversity and Distributions 18: 628-634.

Thank you for your interest

Contact:

Damaris Zurell

Ecology & Macroecology

University of Potsdam

https://damariszurell.github.io

Email: damaris.zurell@uni-potsdam.de

@ZurellLab