SDMs algorithms & ensembles

Damaris Zurell

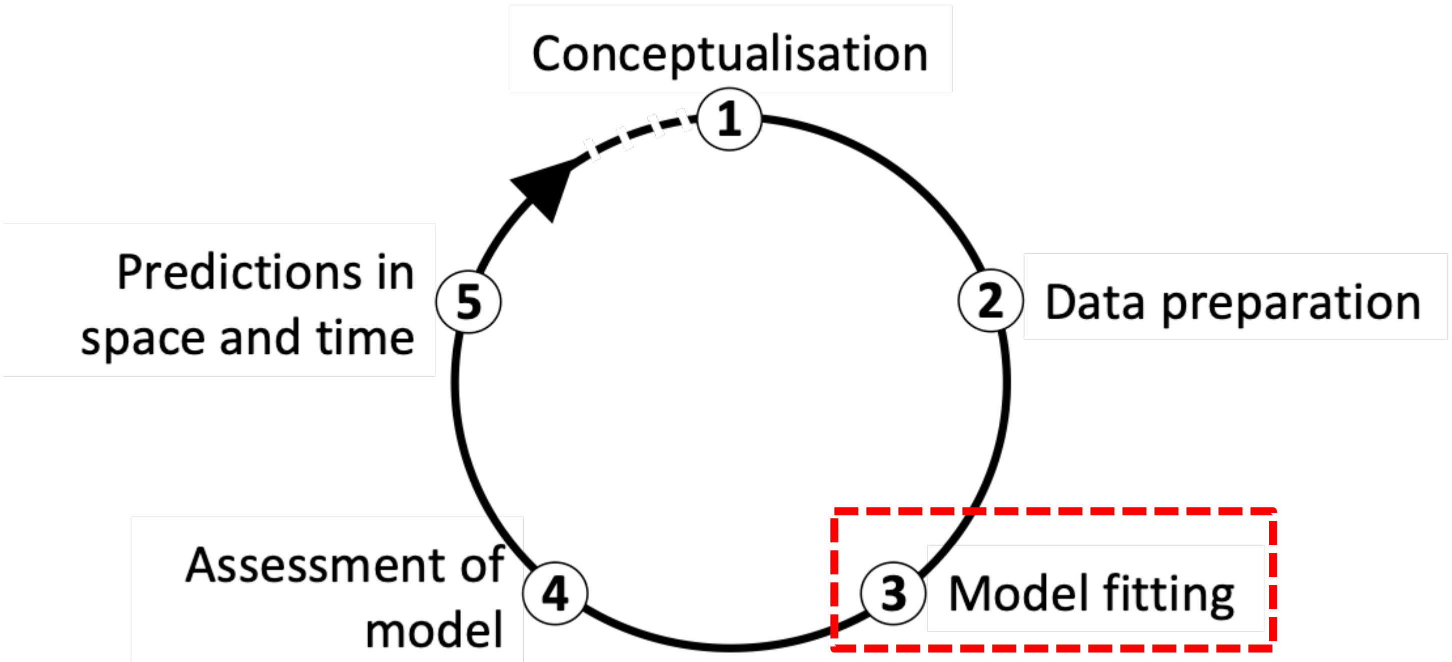
https://damariszurell.github.io

@ZurellLab

Universität Potsdam

Pictures from www.undekade-biologischevielfalt.de

SDM – model building steps



SDM algorithms

Many different algorithms available for SDMs:

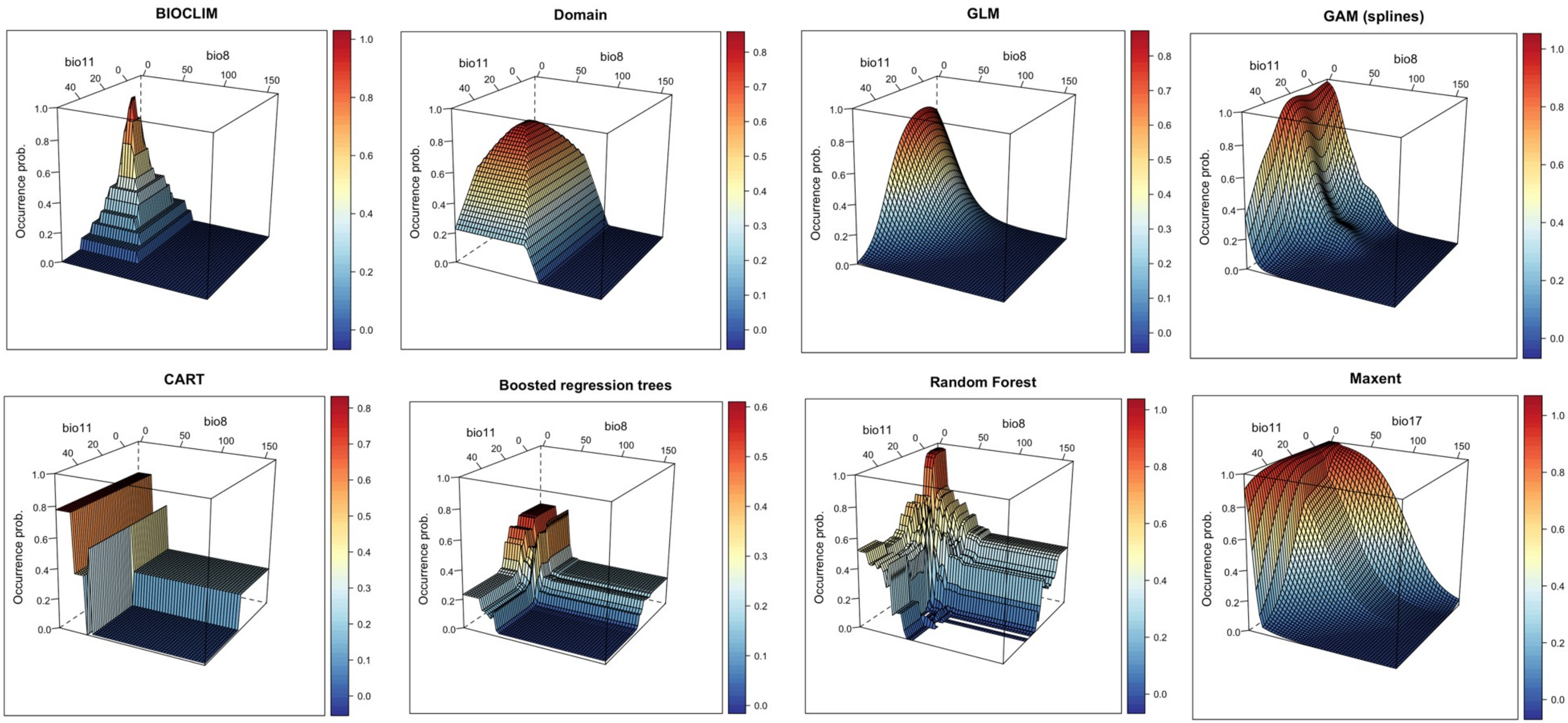
- Profile methods
- Regression
- Machine-learning

SDM algorithms

- **Profile methods** only consider species presences; use simple statistical techniques, e.g. environmental distance to known sites
  - e.g. BIOCLIM, DOMAIN, Mahalonobis distance

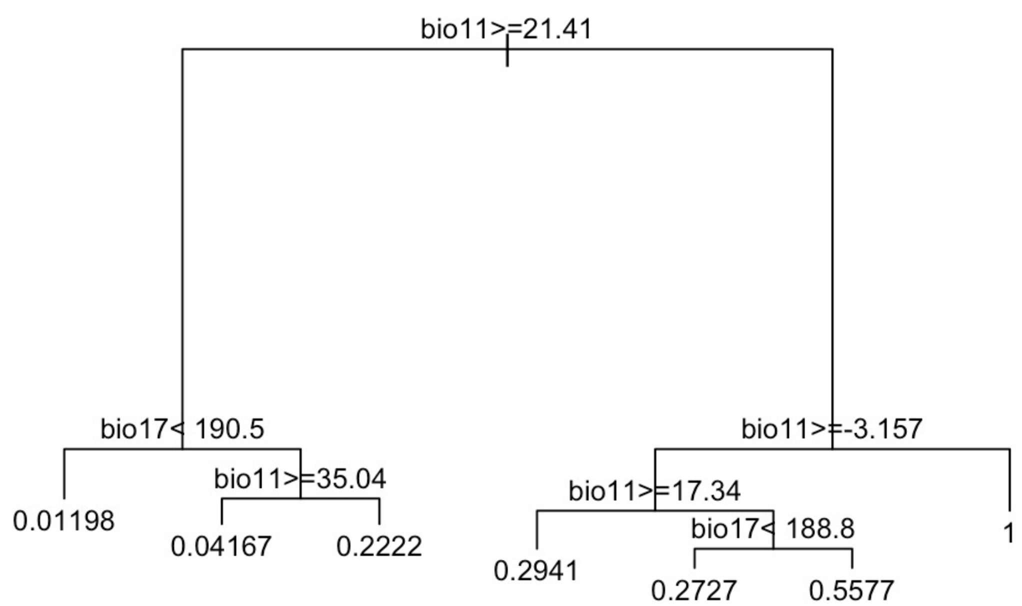
- **Regression-based** techniques and **machine-learning** algorithms use presence and absence (or background) data to contrast used and unused sites
  - **Regression**: e.g. generalised linear model (GLM), generalised additive model (GAM), multivariate adapative regression splines (MARS), ...
  - **Machine-learning**: e.g. classification and regression tree (CART), artificial neural network (ANN), generalised boosted model/boosted regression trees (GBM/BRT), random forest (RF), maximum entropy (Maxent), genetic algorithms, ...

SDM algorithms



Machine-learning: CART

- Classification and regression trees (CARTs)
- Recursive partitioning method to divide the data into homogeneous subgroups
- Find splits (nodes) that best separate the observations
- Interactions between variables fitted automatically



Machine-learning: CART extensions

- CARTs sensitive to noise: typically show low bias and high variance
- One solution: model averaging

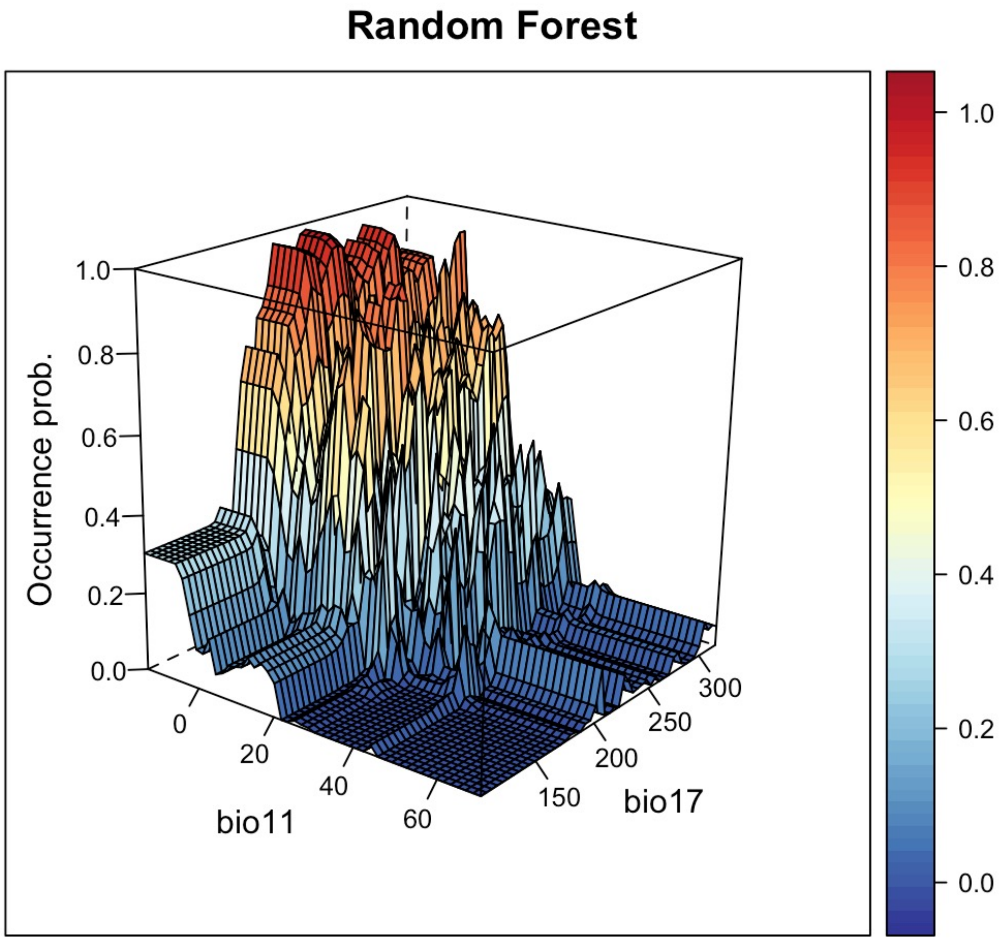
  - Bagging = bootstrap aggregation: fit many CARTs to bootstrapped samples of data and average results
    - → Random Forest

  - Boosting: fit relatively simple CARTs sequentially in adaptive way = each model depends on the previous ones
    - → Boosted regression trees

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

Machine-learning: random forest

- R package „randomForest“



Machine-learning: random forest

- R package „randomForest“

Data frame of predictors

Response

```
m_rf <- randomForest( x=sp_train[,my_preds], y=sp_train$Turdus_torquatus,
    ntree=1000, importance =T)
```

How many trees to grow?

Should variable importance be computed?

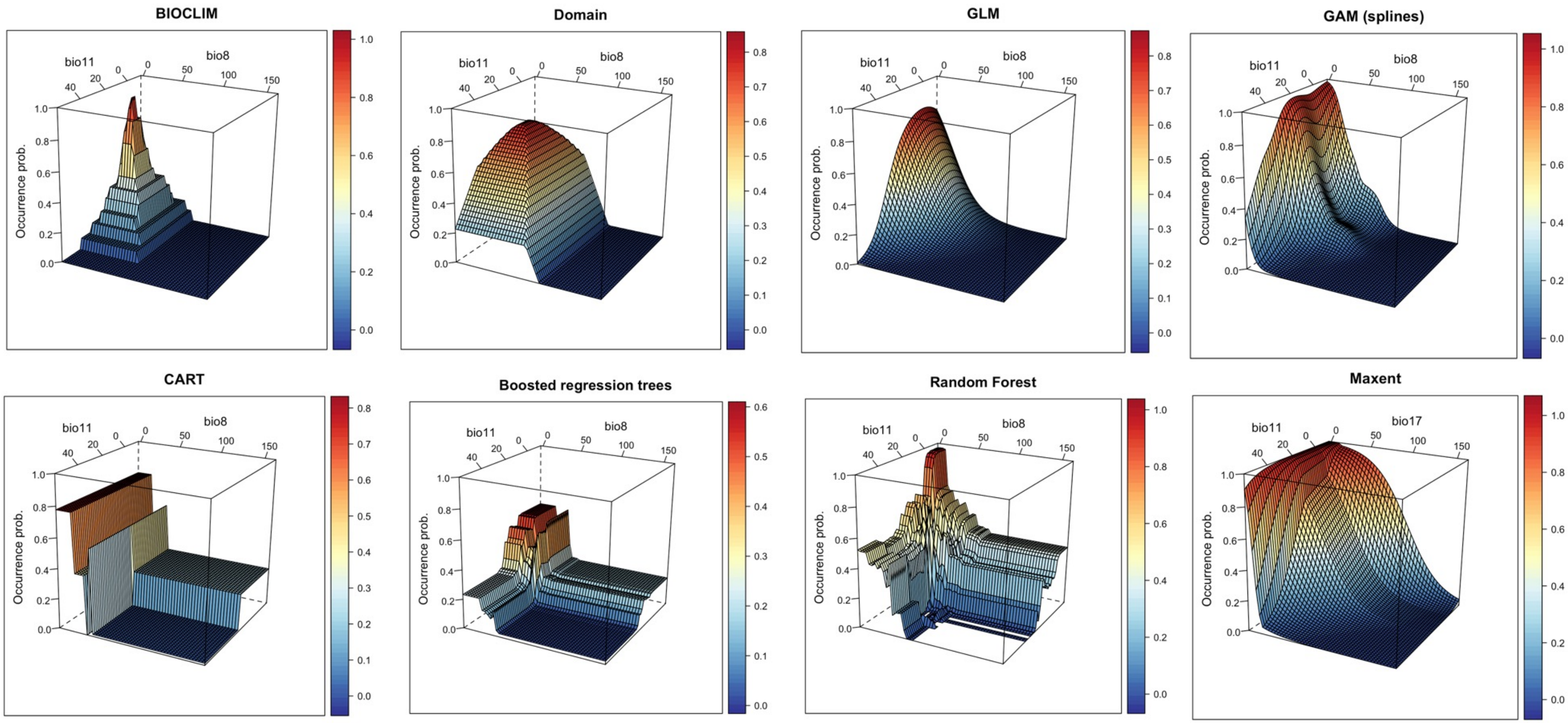
Response type: probabilities

Probability of presence

```
predict(m_rf, xyz, type='prob')[,2]
```

Data frame with predictor variables

SDM algorithms



SDM algorithms

- There is no single best approach for SDMs. (I have no favourite)

- Model choice should be guided by model purpose, available data, scale, ...

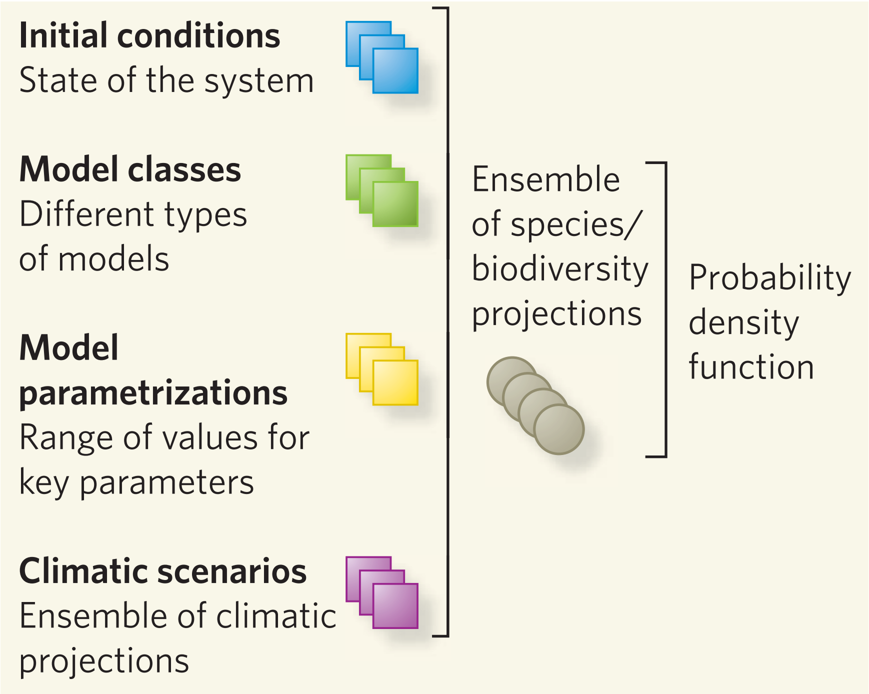
- More complex models tend to better fit current species-environment relationship. Yet, it is highly debated whether more complex models make better **predictions under global change**.

- For global change analyses, the IUCN recommends to **use at least three algorithms that are as independent as possible**.

Araujo & Rahbek (2006) Science 313: 1396-1397.
IUCN Standards and Petitions Committee (2019) http://www.iucnredlist.org/documents/RedListGuidelines.pdf

SDM ensembles

- Ensembles of forecasts are produced by making multiple simulations across more than one set of initial conditions (data), model classes, model parameterisations, and boundary conditions (scenarios)



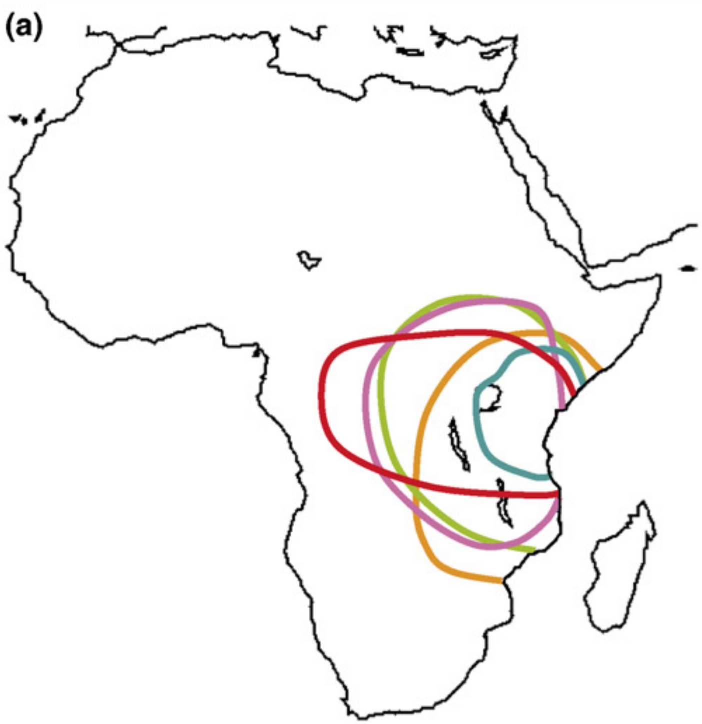
Araujo & New (2007) Trends in Ecology & Evolution 22: 42-47.

Thuiller (2007) Nature 448: 550-552.

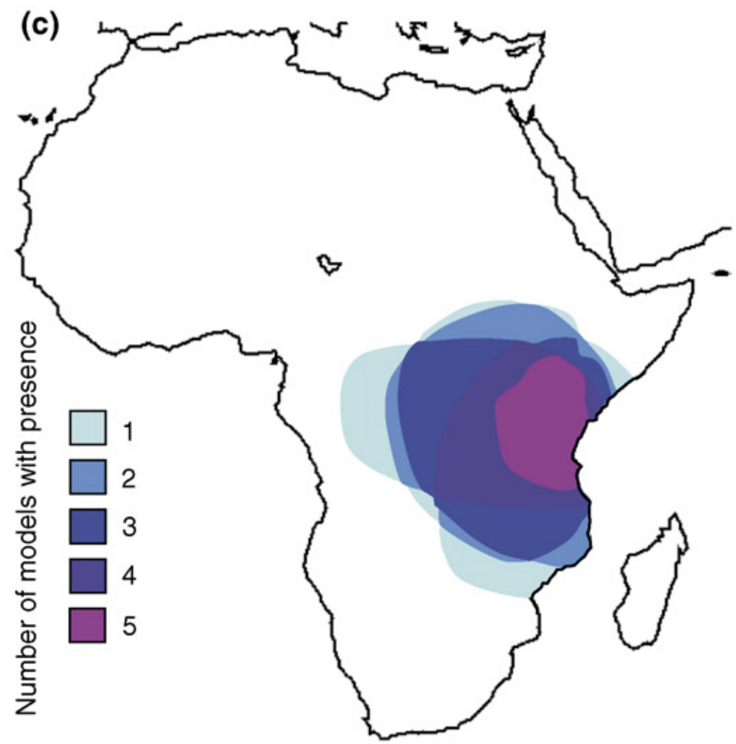
SDM ensembles

- The final predictions can be combined in different ways

Individual model predictions



Committee average of binary predictions

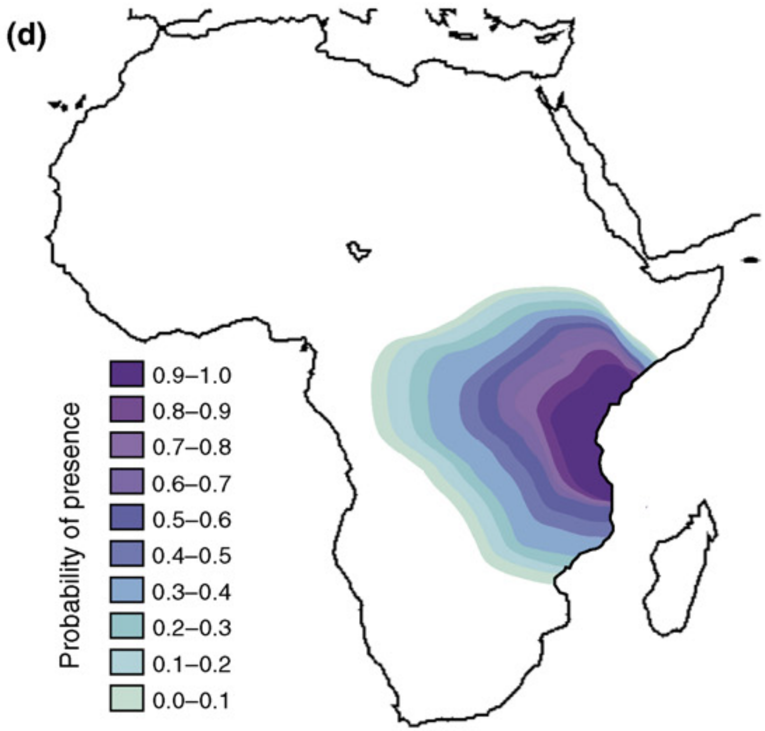


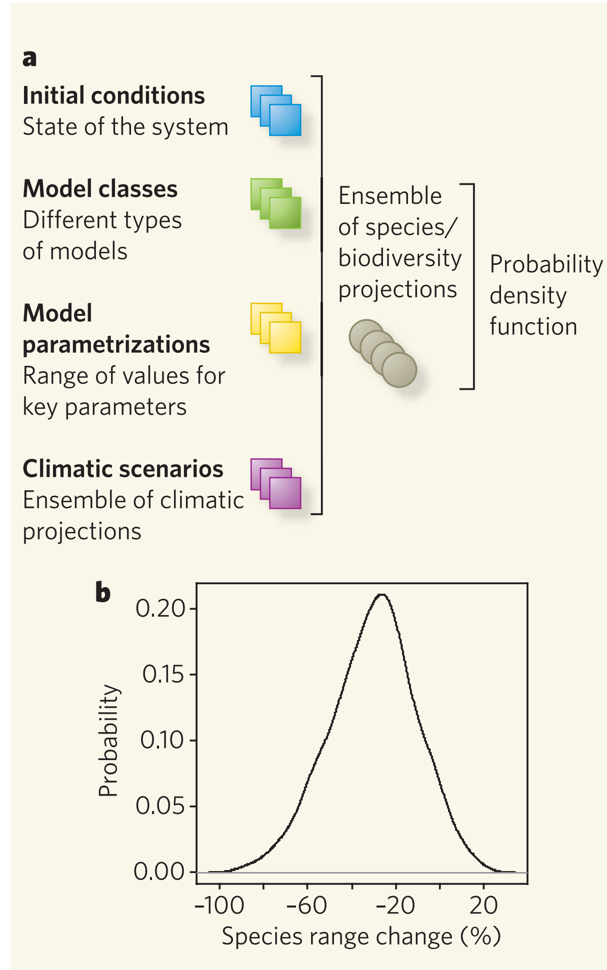
Araujo & New (2007) Trends in Ecology & Evolution 22: 42-47.

SDM ensembles

- The final predictions can be combined in different ways

Consensus – central tendency



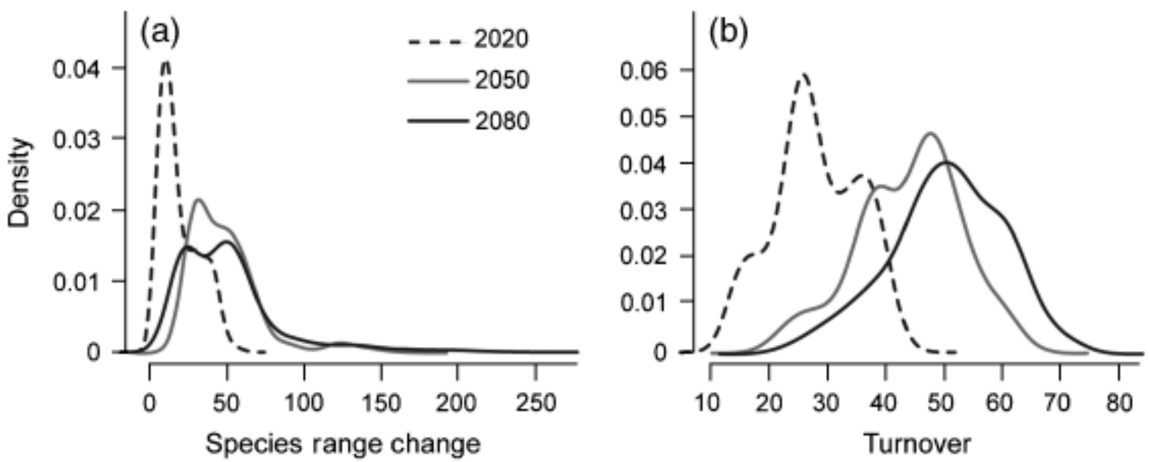


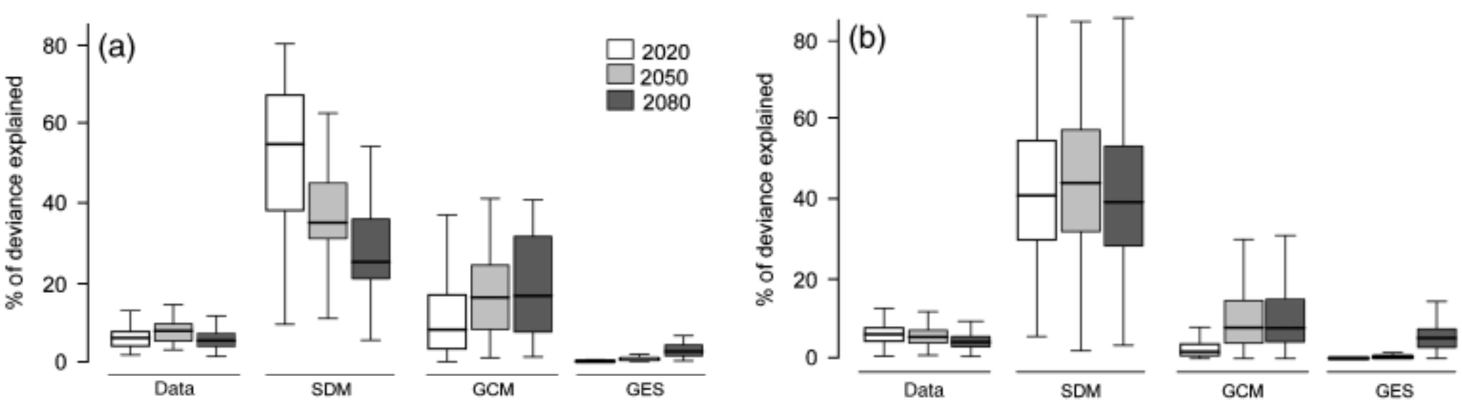
Araujo & New (2007) Trends in Ecology & Evolution 22: 42-47.

Thuiller (2007) Nature 448: 550-552.

SDM ensembles

- Purpose: accounting for sources of uncertainty

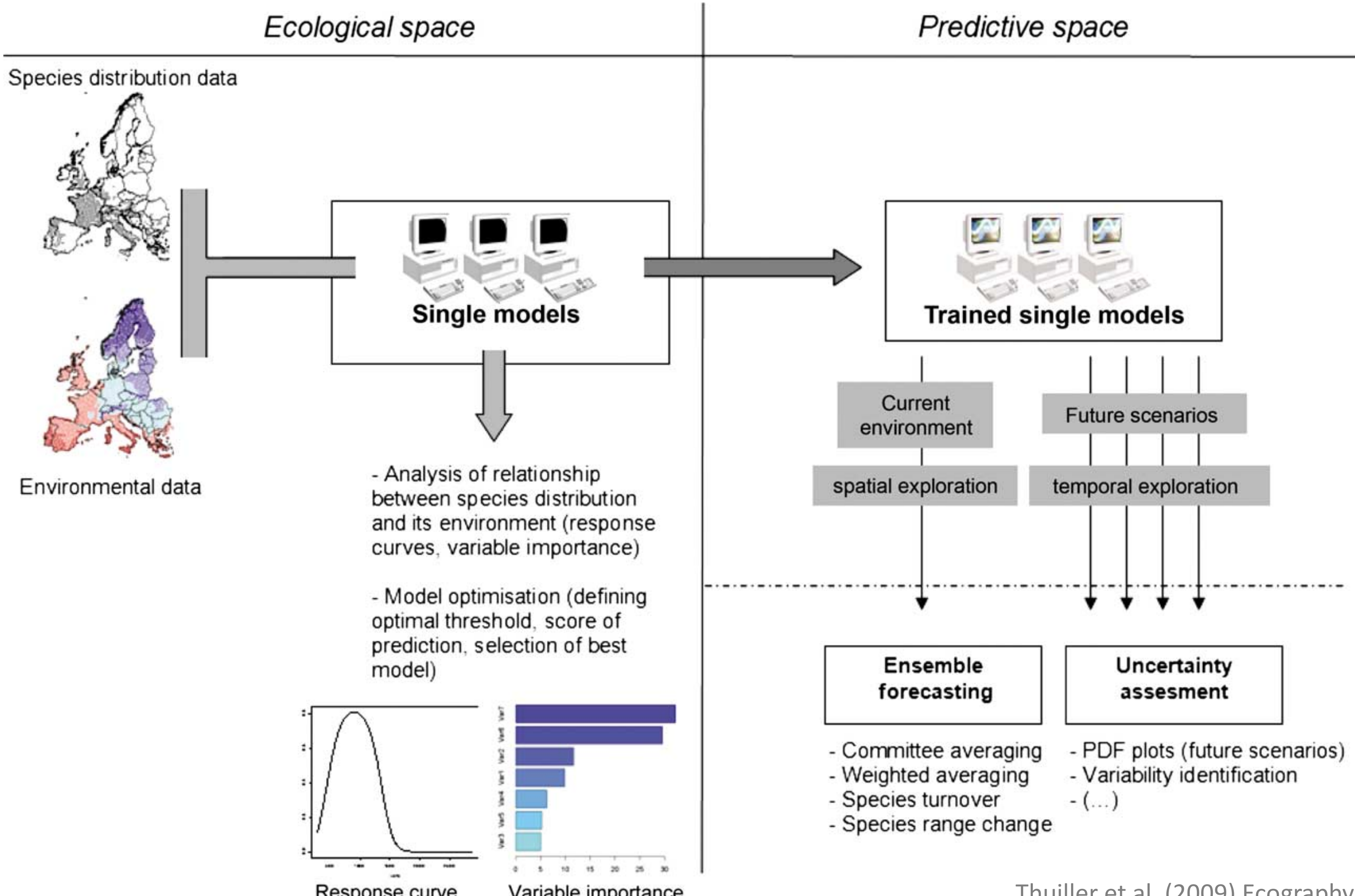




Buisson et al. (2010) Global Change Biology 16: 1145-1157.

SDM ensembles

- Dedicated R packages, e.g. *biomod2*



Thuiller et al. (2009) Ecography 32: 369-373.

Thank you for your interest

Contact:

Damaris Zurell

Ecology & Macroecology

University of Potsdam

https://damariszurell.github.io

Email: damaris.zurell@uni-potsdam.de

@ZurellLab