Download from CRAN

Documentation & References

Training and assessment of spatial prediction models: challenges, conceptual frameworks and implemented strategies in the R package CAST

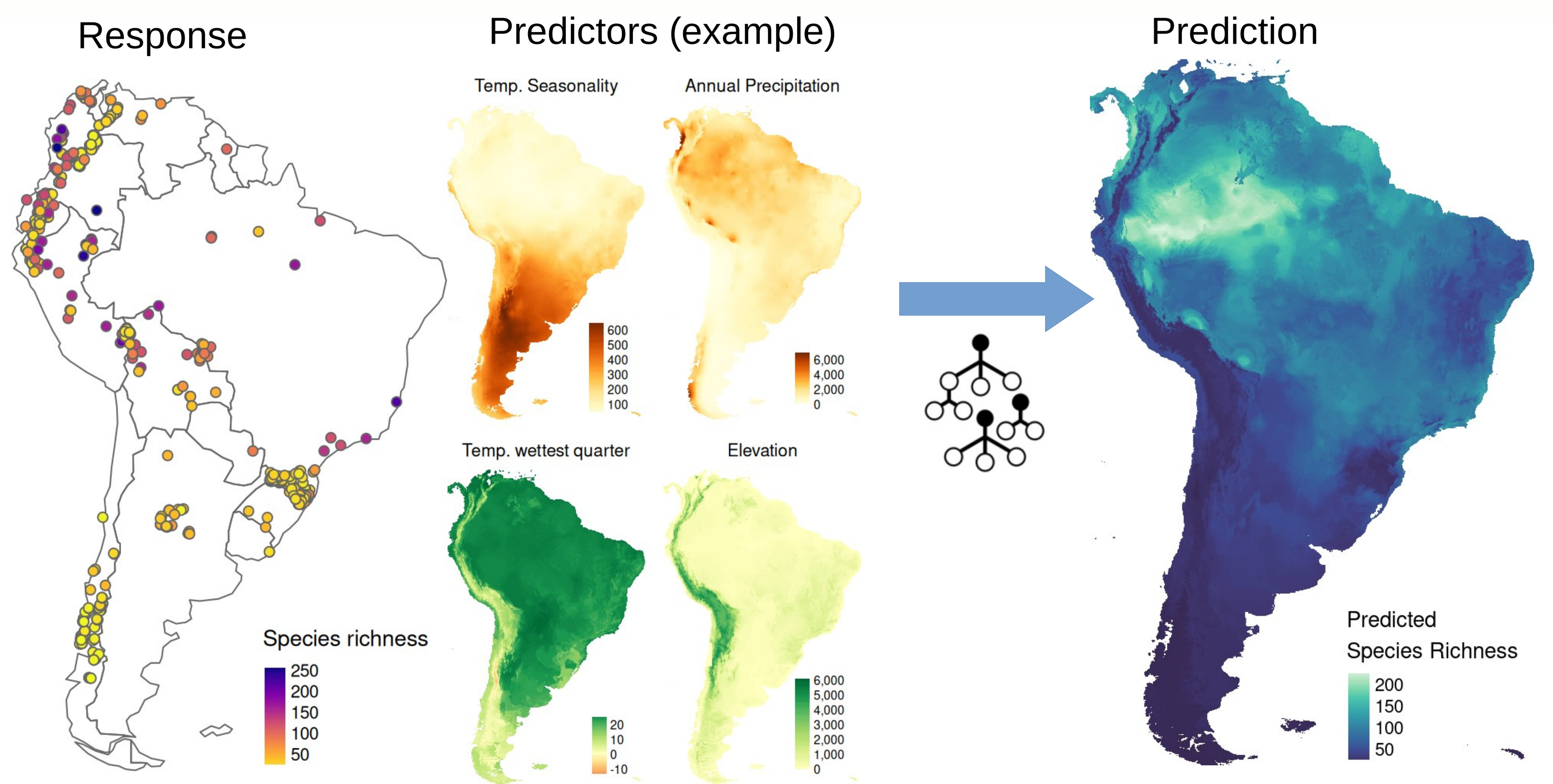
Hanna Meyer, Jan Linnenbrink, Marvin Ludwig, Carles Milà, Jakub Nowosad, Fabian Schumacher

Motivation

Aim: Derive spatial continuous data ("maps") of the environment based on limited field observations and remote sensing data
Problem: Consideration of the characteristics of geoscientific data, e.g. reference data are not IID, spatial and temporal autocorrelation.
Challenges: Assessment of models, avoiding spatial overfitting, model transferability.

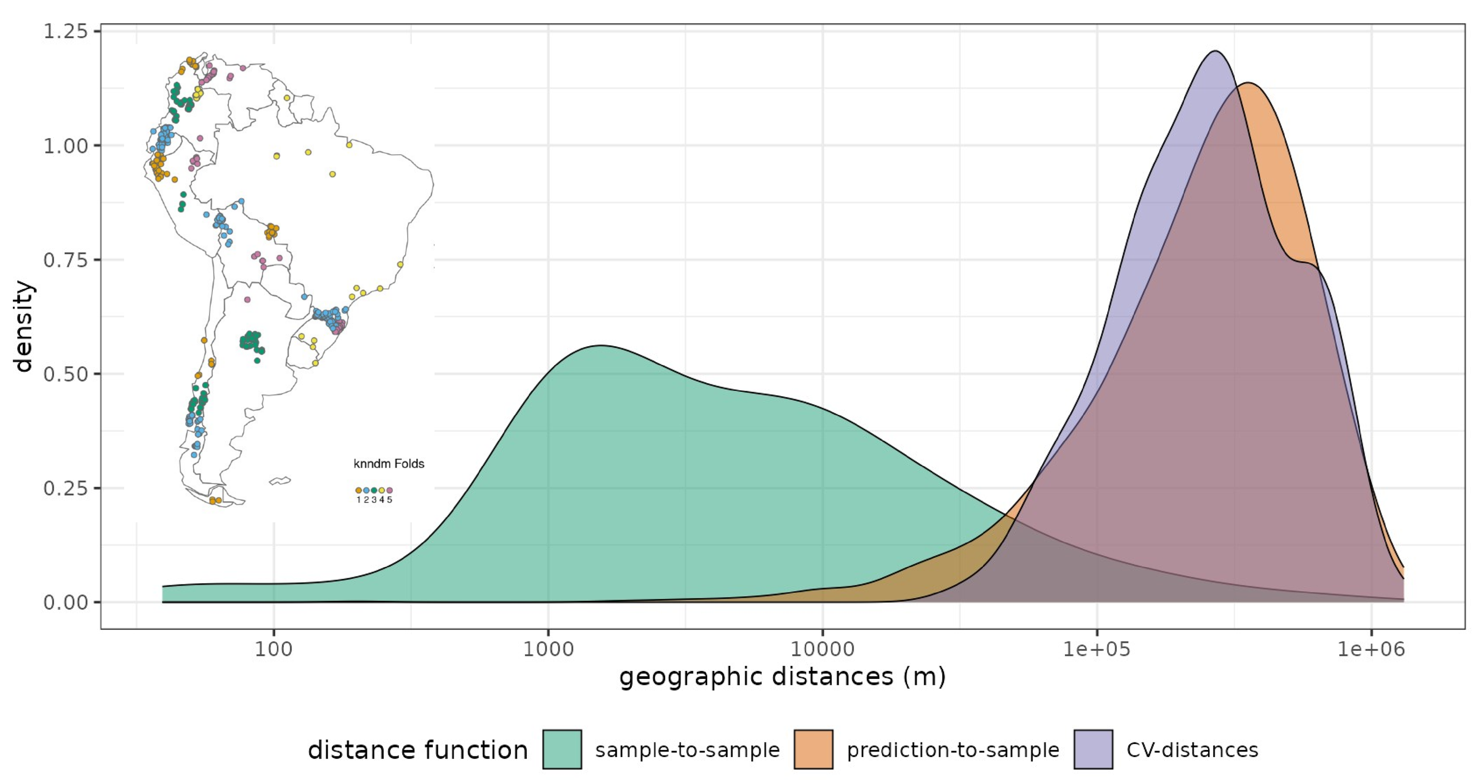
In the past few years, we developed a number of methods to support the application of machine learning for spatial data. We implemented these methods in the R package CAST for their easy integration into conceptually strong modelling workflows and strategies for predictive mapping.

Example: Mapping plant Species richness of South America based on limited field observations



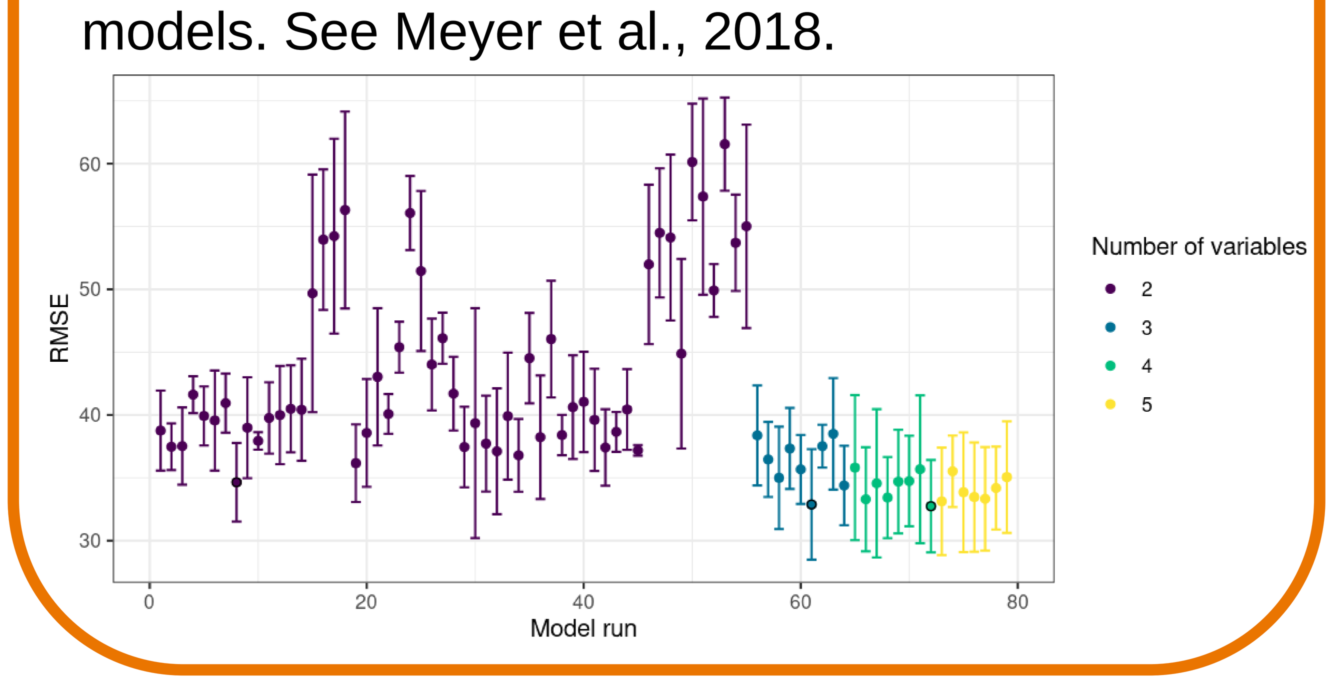
Prediction domain adaptive cross-validation

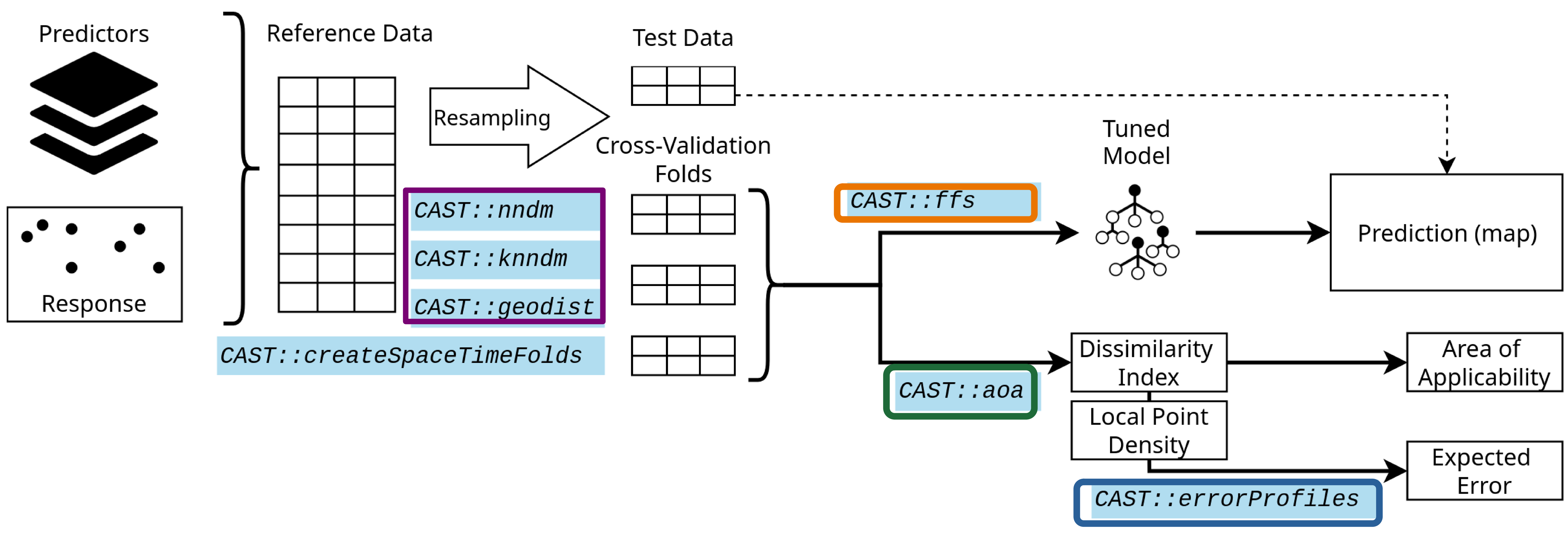
Prediction domain adaptive cross-validation takes the prediciton area into account to estimate the spatial prediction perfomance during model tuning, selection and assessment. This results in more realistic estimates of map accuracy. See Mila et al., 2022 and Linnenbrink et al., 2024.



Spatial feature selection

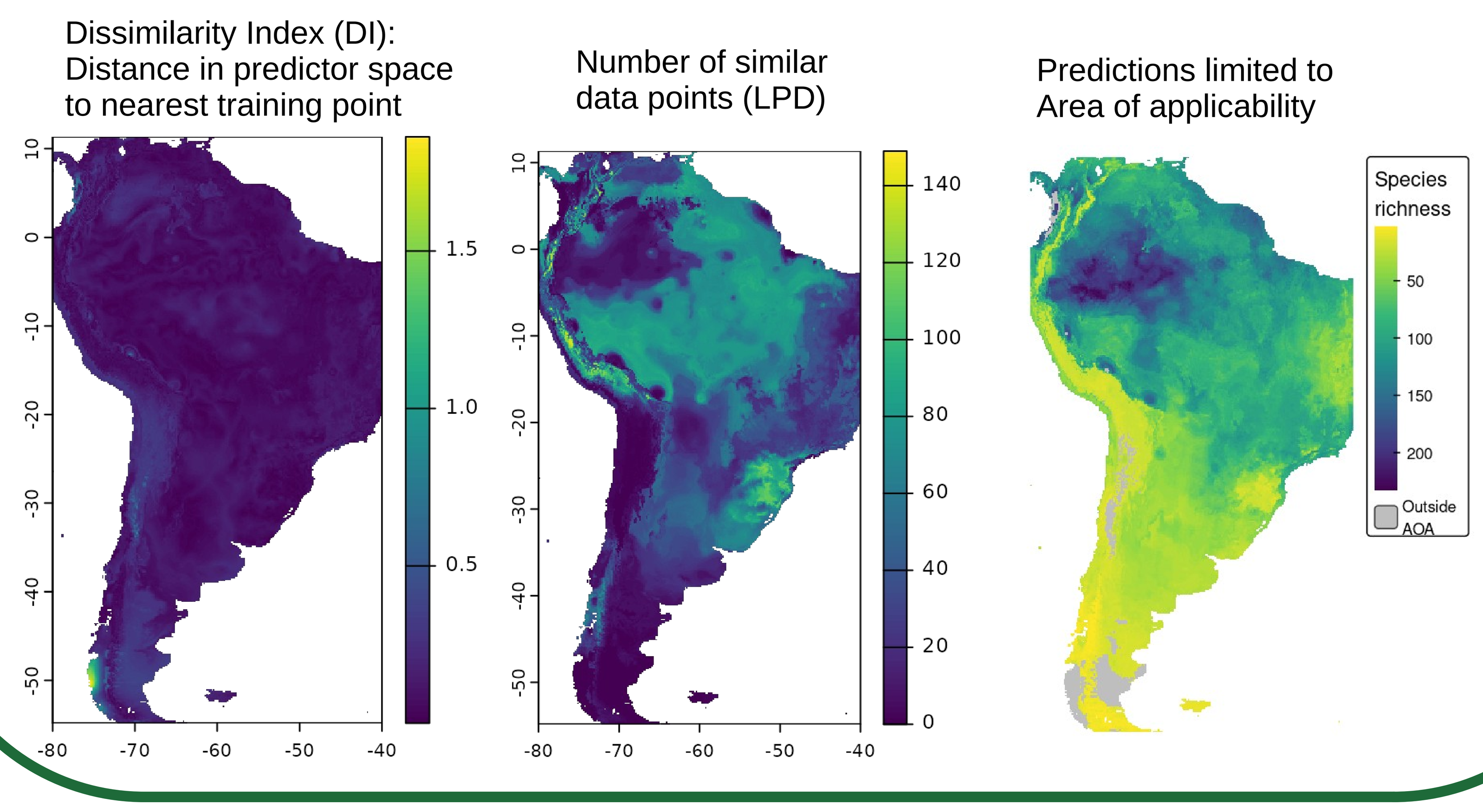
Spatial forward feature selection identifies which and how many predictor variables are most suitable for spatial prediction, leading to more generalizable models. See Meyer et al., 2018.





Area of applicability

The area of applicability (AOA) delineates where spatial prediction models can be reliably applied by accounting for dissimilarities (DI) and local training point densities (LPD) in feature space. This helps identify regions where models are supported by training data and avoids unreliable extrapolation (see Meyer & Pebesma, 2021; Schumacher et al., 2025).



Error profiles

Spatial error profiles relate the dissimilarity index and the local point density to prediction errors using shape-constrained additive models. The models are used to generate spatially explicit uncertainty estimates. This enables mapping of prediction errors (see Meyer et al., 2026).