Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data

Patrick Schratz[1], Jannes Muenchow[1], Eugenia Iturritxa[2], Jakob Richter[3], Alexander Brenning[1]

September 27 2018, 10th International Conference on Ecological Informatics, Jena, Germany

[1] Department of Geography, GIScience group, University of Jena

[2] NEIKER, Vitoria-Gasteiz, Spain

[3] Department of Statistics, TU Dortmund

https://pjs-web.de @pjs_228 @pat-s @pjs_228

patrick.schratz@uni-jena.de Patrick Schratz

Crucial but often neglected: The important role of spatial autocorrelation in hyperparameter tuning and predictive performance of machine-learning algorithms for spatial data

Patrick Schratz[1], Jannes Muenchow[1], Eugenia Iturritxa[2], Jakob Richter[3], Alexander Brenning[1]

[1] Department of Geography, GIScience group, University of Jena

[2] NEIKER, Vitoria-Gasteiz, Spain

[3] Department of Statistics, TU Dortmund

Introduction

Introduction

LIFE Healthy Forest

Early detection and advanced management systems to reduce forest decline by invasive and pathogenic agents.

Main task: Spatial (modeling) analysis to support the early detection of various pathogens.

Pathogens

- Fusarium circinatum
- **Diplodia sapinea** (→ needle blight)
- Armillaria root disease
- Heterobasidion annosum

Fig. 1: Needle blight caused by **Diplodia pinea**

Introduction

Motivation

- Find the model with the **highest predictive performance**.
- Results are assumed to be representative for data sets with similar predictors and different pathogens (response).
- Be aware of **spatial autocorrelation** ⚠
- Analyze differences between **spatial and non-spatial hyperparameter tuning** (no research here yet!).
- Analyze differences in performance between algorithms and sampling schemes in CV (both performance estimation and hyperparameter tuning)

Data & Study Area

Data & Study Area

```
## Skim summary statistics
##  n obs: 926
##  n variables: 12
##
## Variable type: factor
##
##  variable    missing    n    n_unique                  top_counts
## ----------- --------- ----- ---------- --------------------------------------------
##   diplo01       0      926      2                 0: 703, 1: 223, NA: 0
##  lithology      0      926      5        clas: 602, chem: 143, biol: 136, surf: 32
##    soil         0      926      7         soil: 672, soil: 151, soil: 35, pron: 22
##    year         0      926      4        2009: 401, 2010: 261, 2012: 162, 2011: 102
##
## Variable type: numeric
##
##    variable      missing    n      mean       p0      p50      p100      hist
## --------------- --------- ----- ---------- ------- -------- -------- ----------
##       age           0      926    18.94       2       20       40
##    elevation        0      926    338.74     0.58   327.22   885.91
##    hail_prob        0      926     0.45     0.018    0.55      1
##      p_sum          0      926    234.17    124.4   224.55    496.6
##       ph            0      926     4.63      3.97     4.6      6.02
##      r_sum          0      926   -0.00004    -0.1   0.0086    0.082
##  slope_degrees      0      926    19.81      0.17    19.47    55.11
##      temp           0      926    15.13     12.59    15.23     16.8
```

Data & Study Area

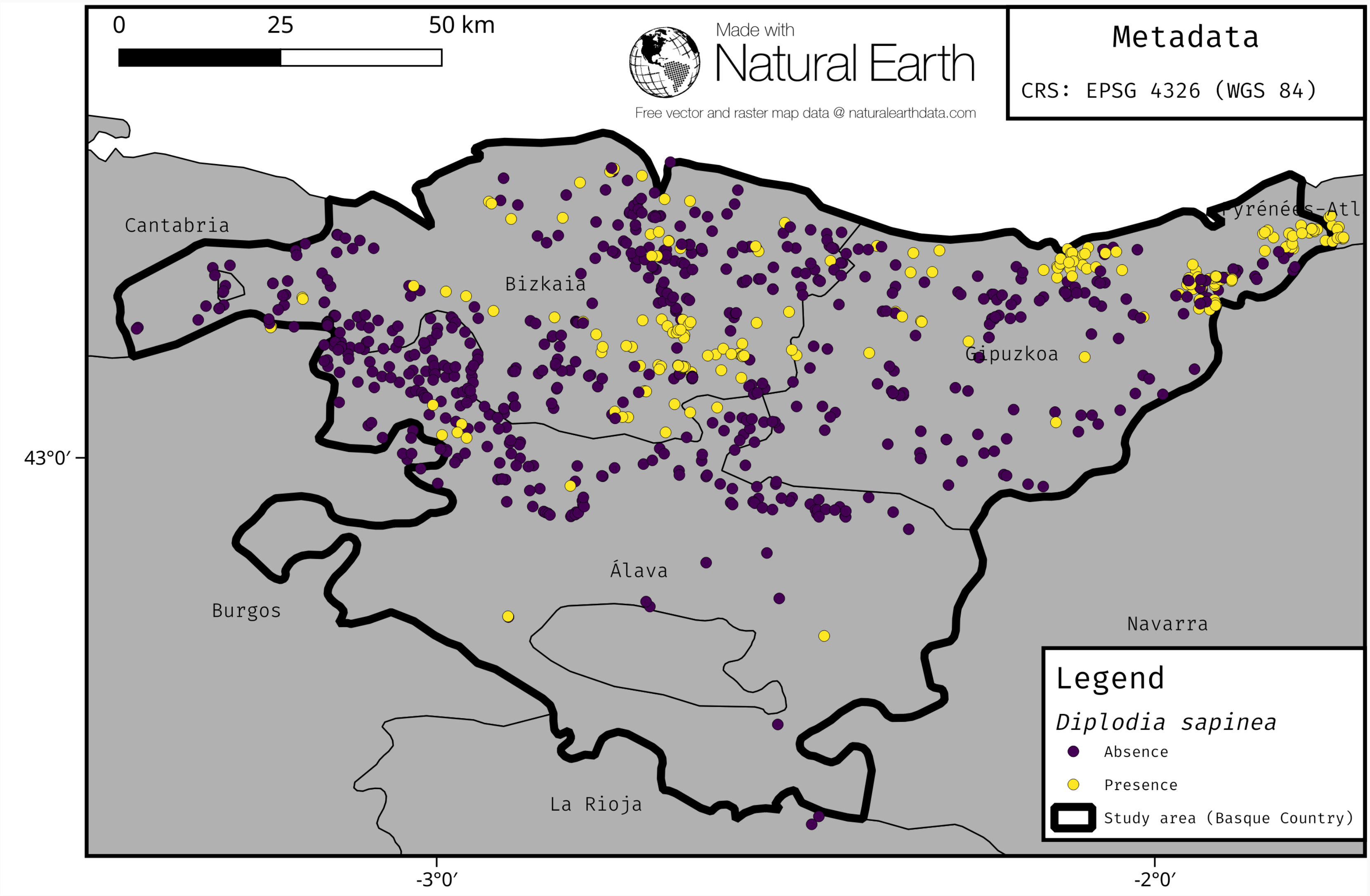


Fig. 2: Study area (Basque Country, Spain)

Methods

Machine-learning models

- Boosted Regression Trees (`BRT`)
- Random Forest (`RF`)
- Support Vector Machine (`SVM`)
- k-nearest Neighbor (`KNN`)

Parametric models

- Generalized Addtitive Model (`GAM`)
- Generalized Linear Model (`GLM`)

Performance Measure

Brier Score: Mean squared error of the probabilites, $\frac{1}{N}$ t t

Methods

Nested Cross-Validation

- Cross-validation for **performance estimation**
- Cross-validation for **hyperparameter tuning** (sequential model-based optimization (SMBO), Bischl, Richter, Bossek, et al. (2017))

Different sampling strategies (Performance estimation/Tuning):

- Non-Spatial/Non-Spatial
- Spatial/Non-Spatial
- Spatial/Spatial Brenning (2012)
- Non-Spatial/No Tuning
- Spatial/No Tuning

Methods

Nested (spatial) cross-validation

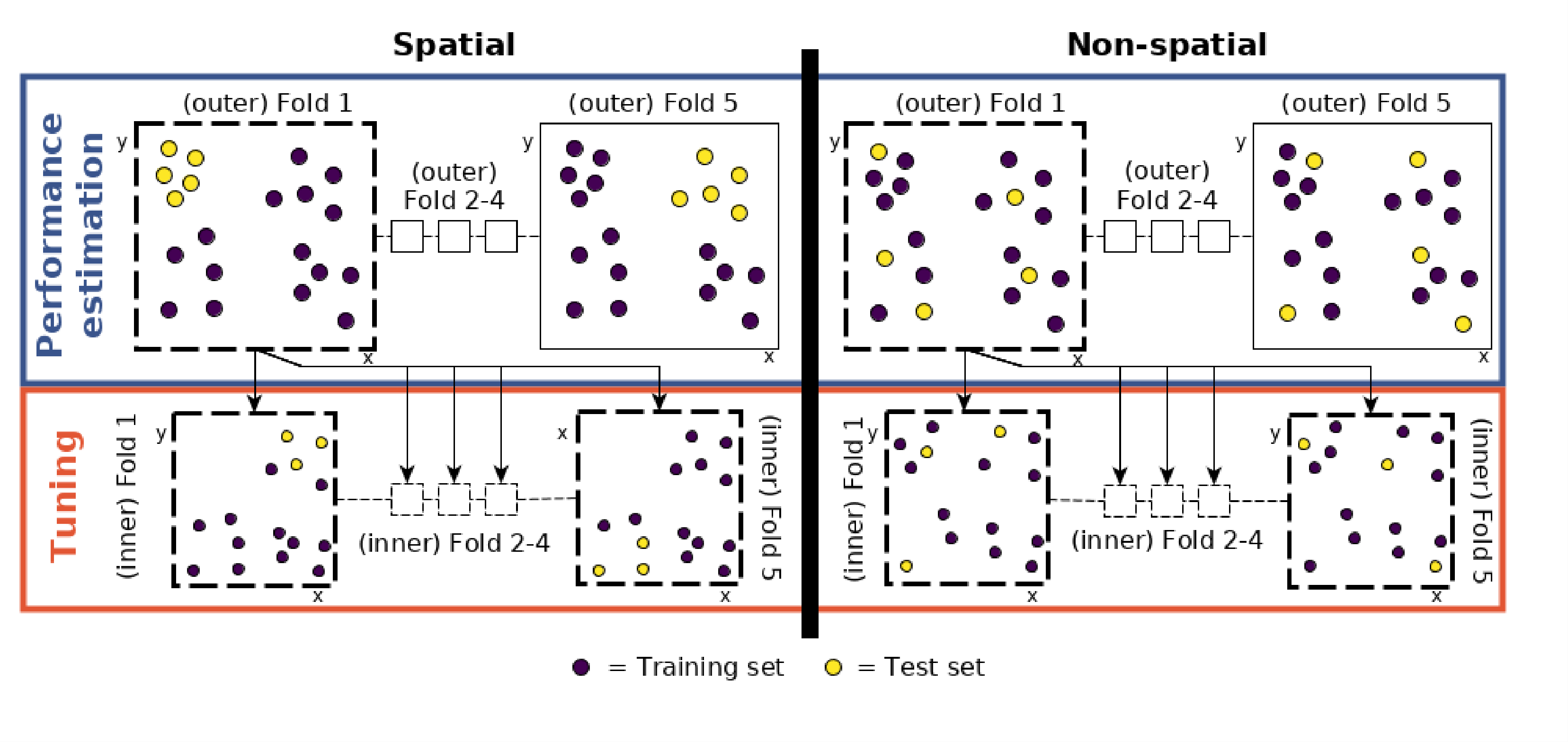


Fig. 3: Nested spatial/non-spatial cross-validation

Nested (spatial) cross-validation

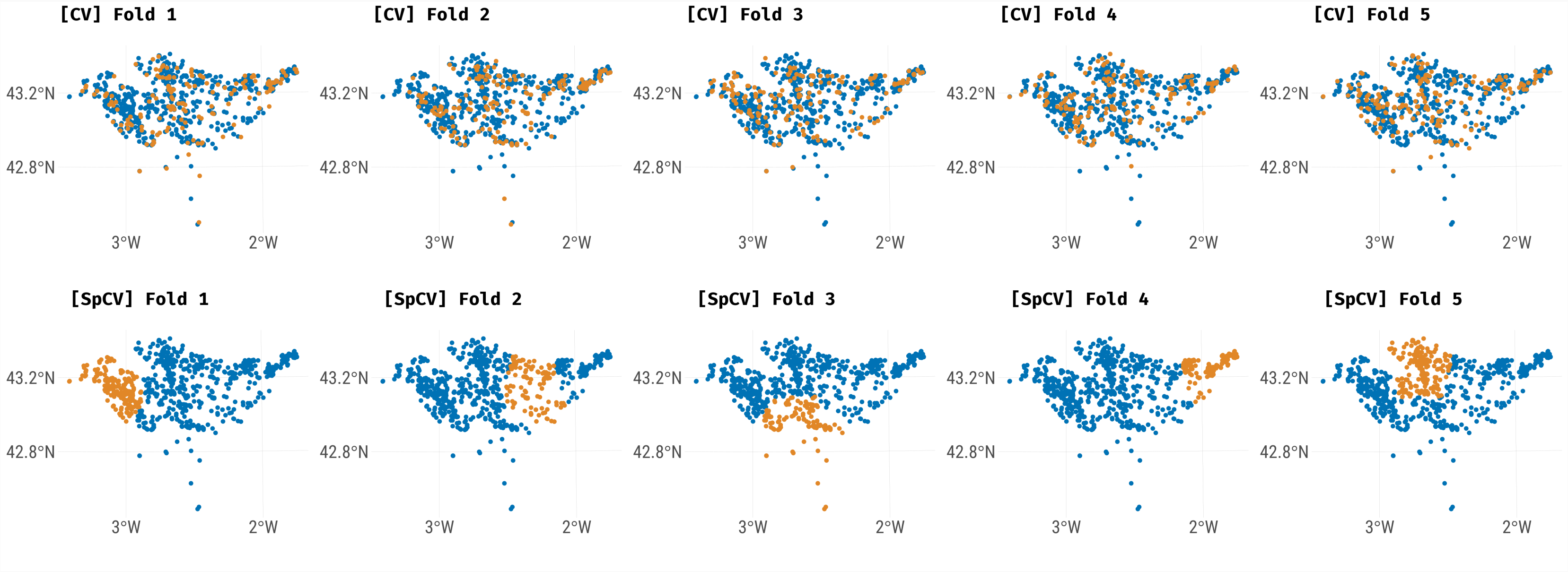


Fig. 4: Comparison of spatial and non-spatial partitioning of the data set.

Hyperparameter tuning search spaces

RF : Probst, Wright, and Boulesteix (2018)

BRT, SVM, KNN: R package *mlrHyperopt* Richter (2017)

Algorithm (package)	Hyperparameter	Type	Start	End	Default
BRT (gbm)	`n.tree`	integer	100	15000	100
	`shrinkage`	numeric	0	1.0	0.001
	`interaction.depth`	integer	1	20	1
KNN (kknn)	`k`	integer	1	250	7
	`distance`	integer	1	300	2
GAM (mgcv)	`sp`	numeric	0	10^6	-
RF (ranger)	`mtry`	integer	1	11	$\sqrt{p}$
	`min.node.size`	integer	1	10	1
	`sample.fraction`	numeric	0.2	0.9	1
SVM (kernlab)	`C`	numeric	2^{-15}	2^{15}	1
	σ	numeric	2^{-15}	2^{15}	1

Table 1: Hyperparameter limits and types of each model.
Notations of hyperparameters from the respective R packages were used.
p = Number of variables.

Results

Hyperparameter tuning

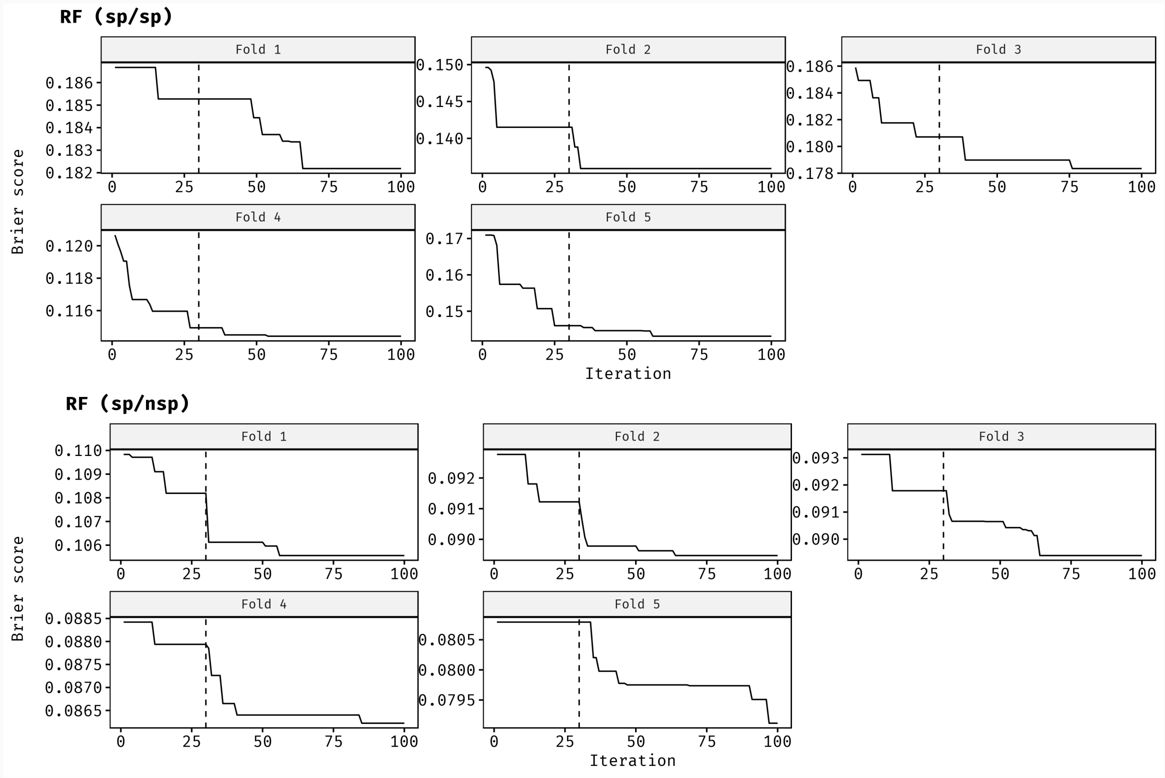


Fig 4: SMBO optimization paths of the first five folds of the **spatial/spatial** and **spatial/non-spatial** CV setting for RF. The dashed line marks the border between the initial design (30 randomly composed hyperparameter settings) and the sequential optimization part in which each setting was proposed using information from the prior evaluated settings.

Results

Hyperparameter tuning

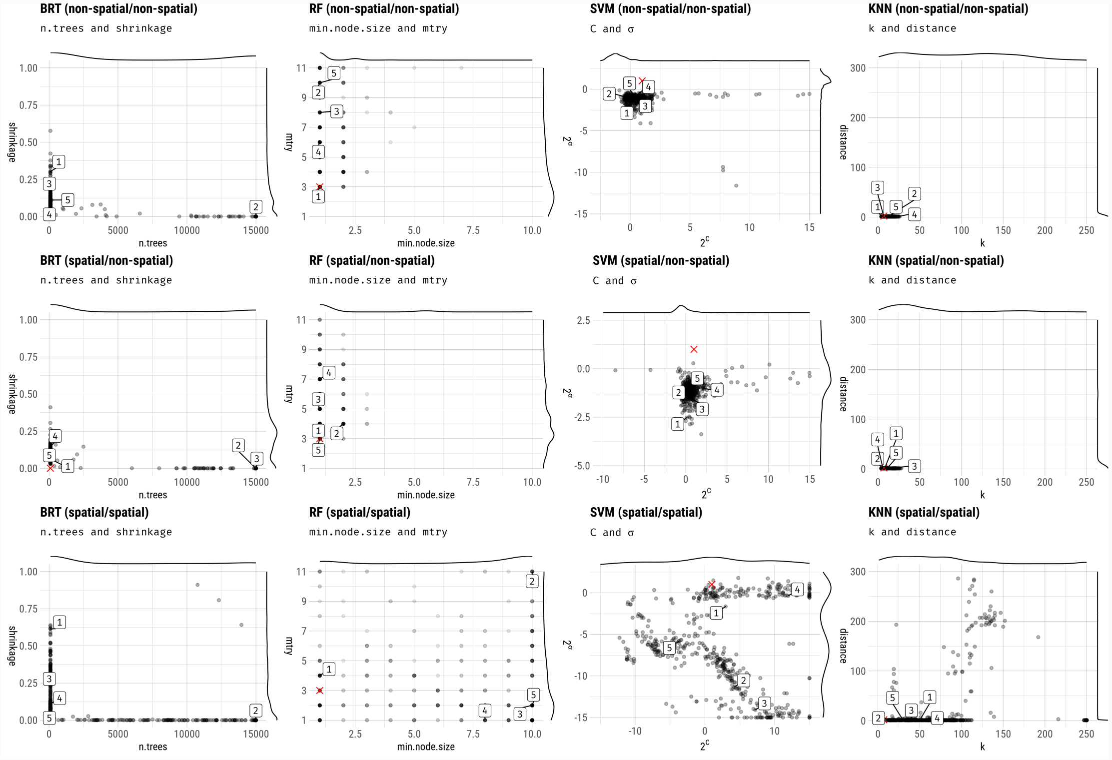
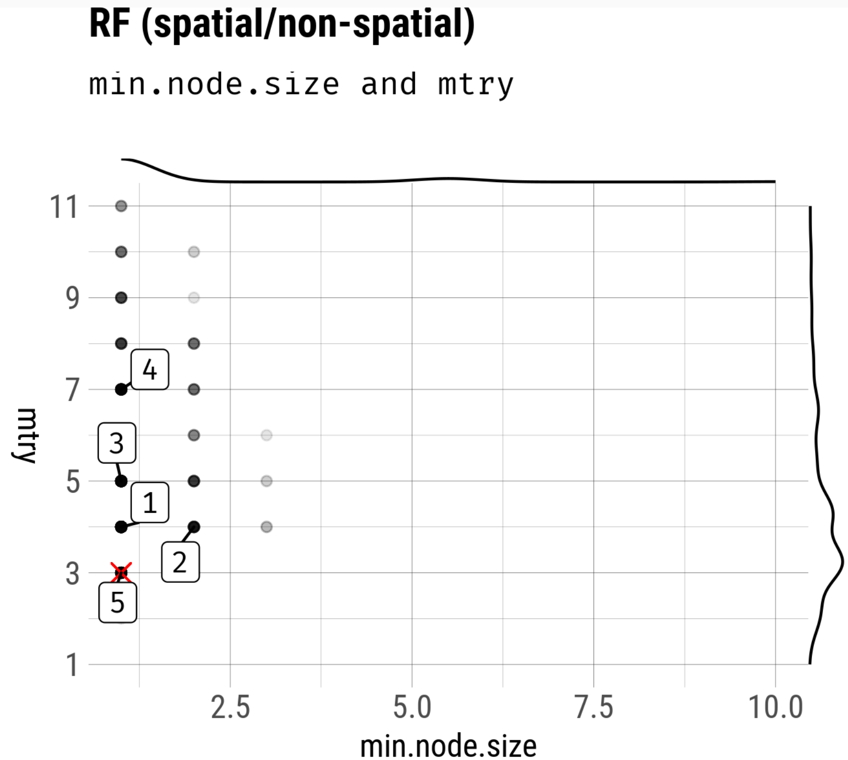
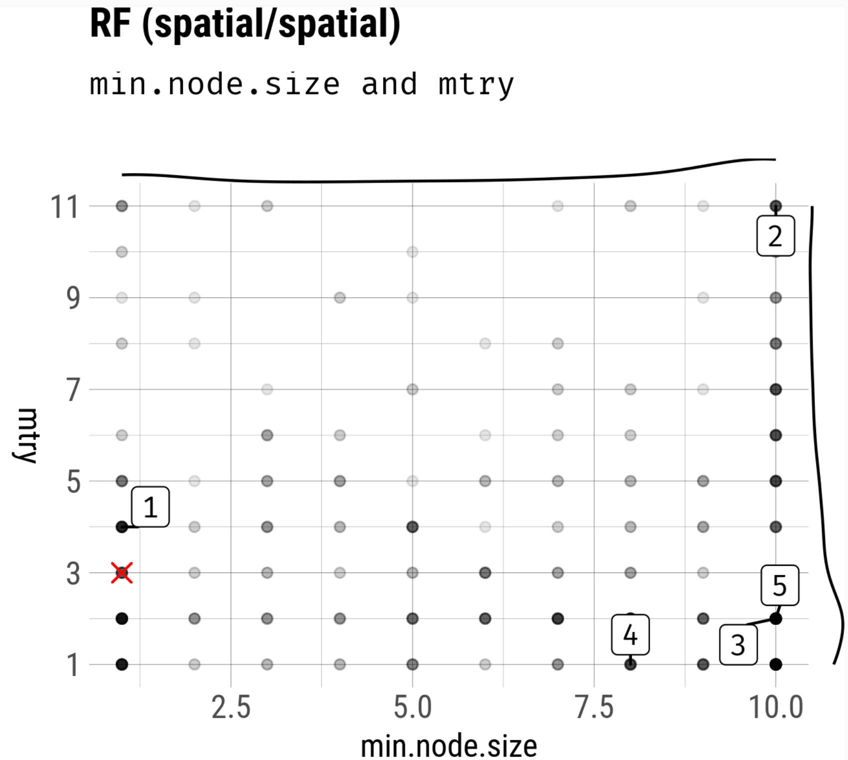


Fig 5: Best hyperparameter settings by fold (500 total) each estimated from 100 (30/70) SMBO tuning iterations per fold using five-fold cross-validation. Split by spatial and non-spatial partitioning setup and model type. Red crosses indicate the default hyperparameters of the respective model. Black dots represent the winning hyperparameter setting of each fold. The labels ranging from one to five show the winning hyperparameter settings of the first five folds.

Results

Hyperparameter tuning





Results

Predictive Performance

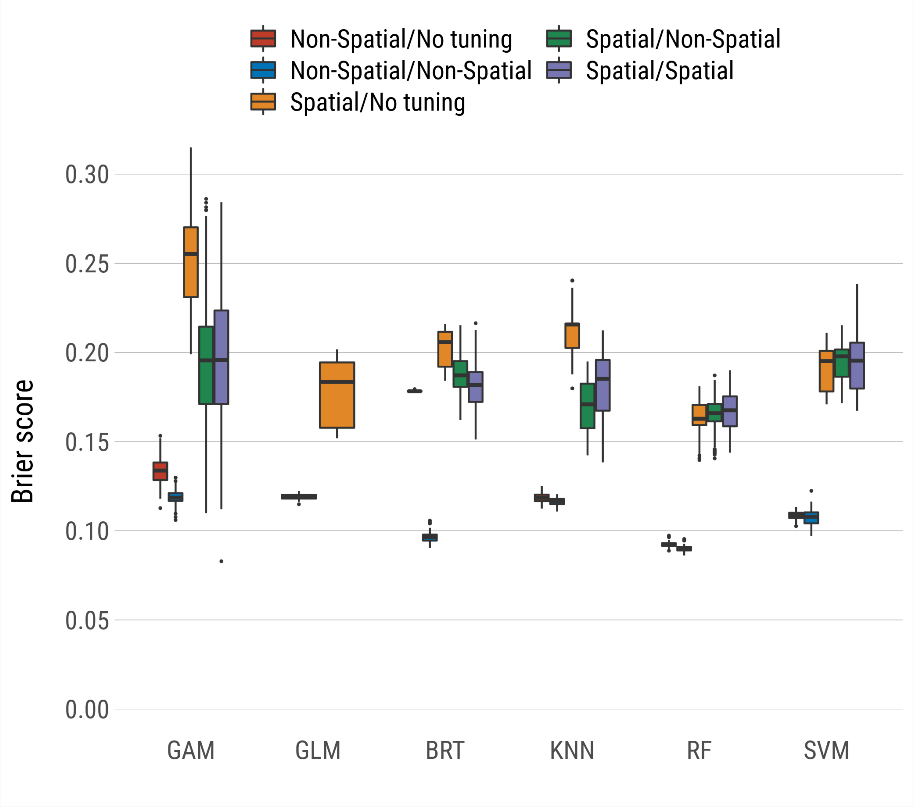
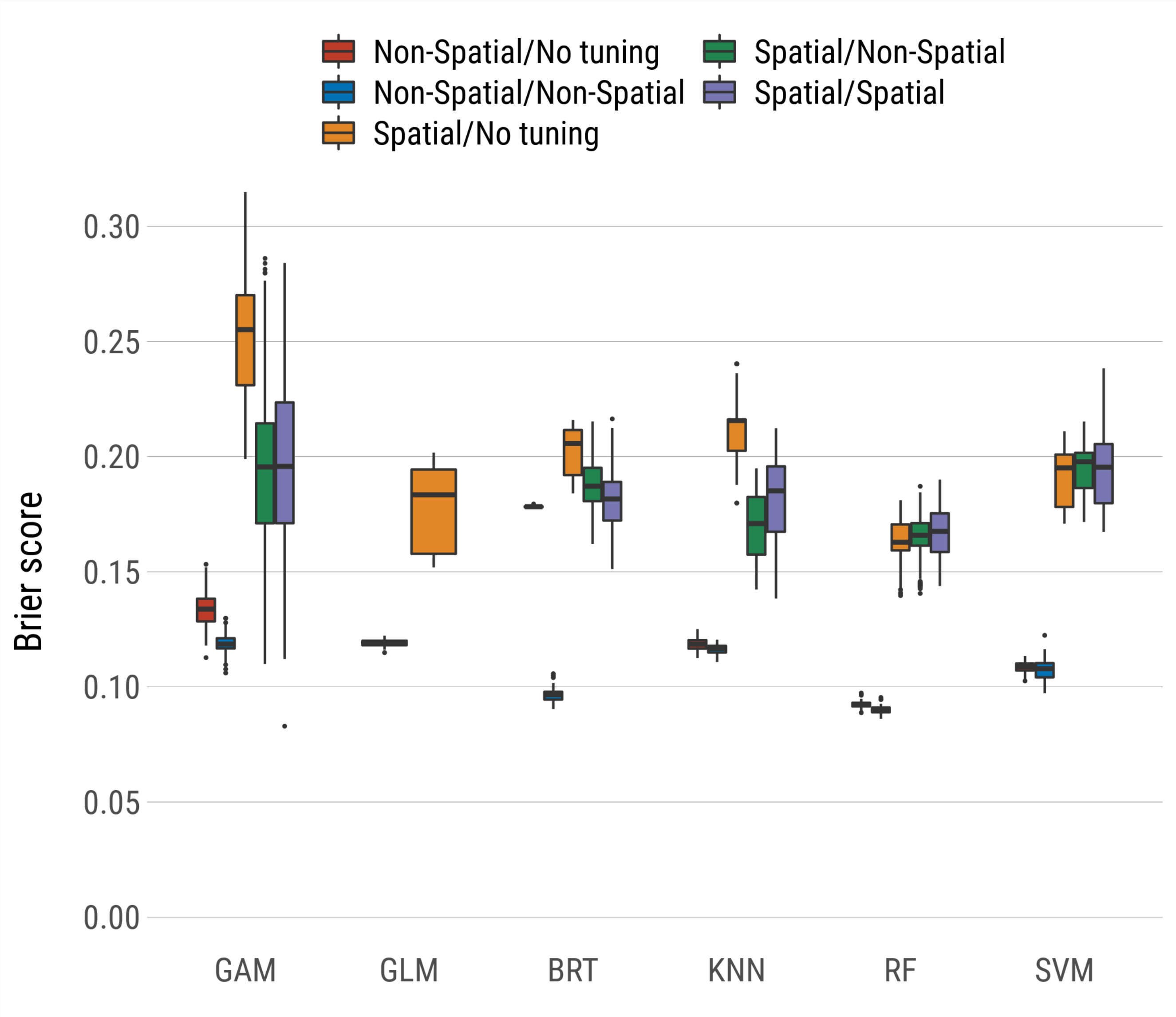


Fig 6: (Nested) CV estimates of model performance at the repetition level using 100 SMBO iterations for hyperparameter tuning. CV setting refers to performance estimation/hyperparameter tuning of the respective (nested) CV, e.g. "Spatial/Non-Spatial" means that spatial partitioning was used for performance estimation and non-spatial partitioning for hyperparameter tuning.

Results


Non-Spatial/No tuning
Non-Spatial/Non-Spatial
Spatial/No tuning
Spatial/Non-Spatial
Spatial/Spatial
Brier score
0.30
0.25
0.20
0.15
0.10
0.05
0.00
GAM
GLM
BRT
KNN
RF
SVM

Discussion

Discussion

Predictive performance

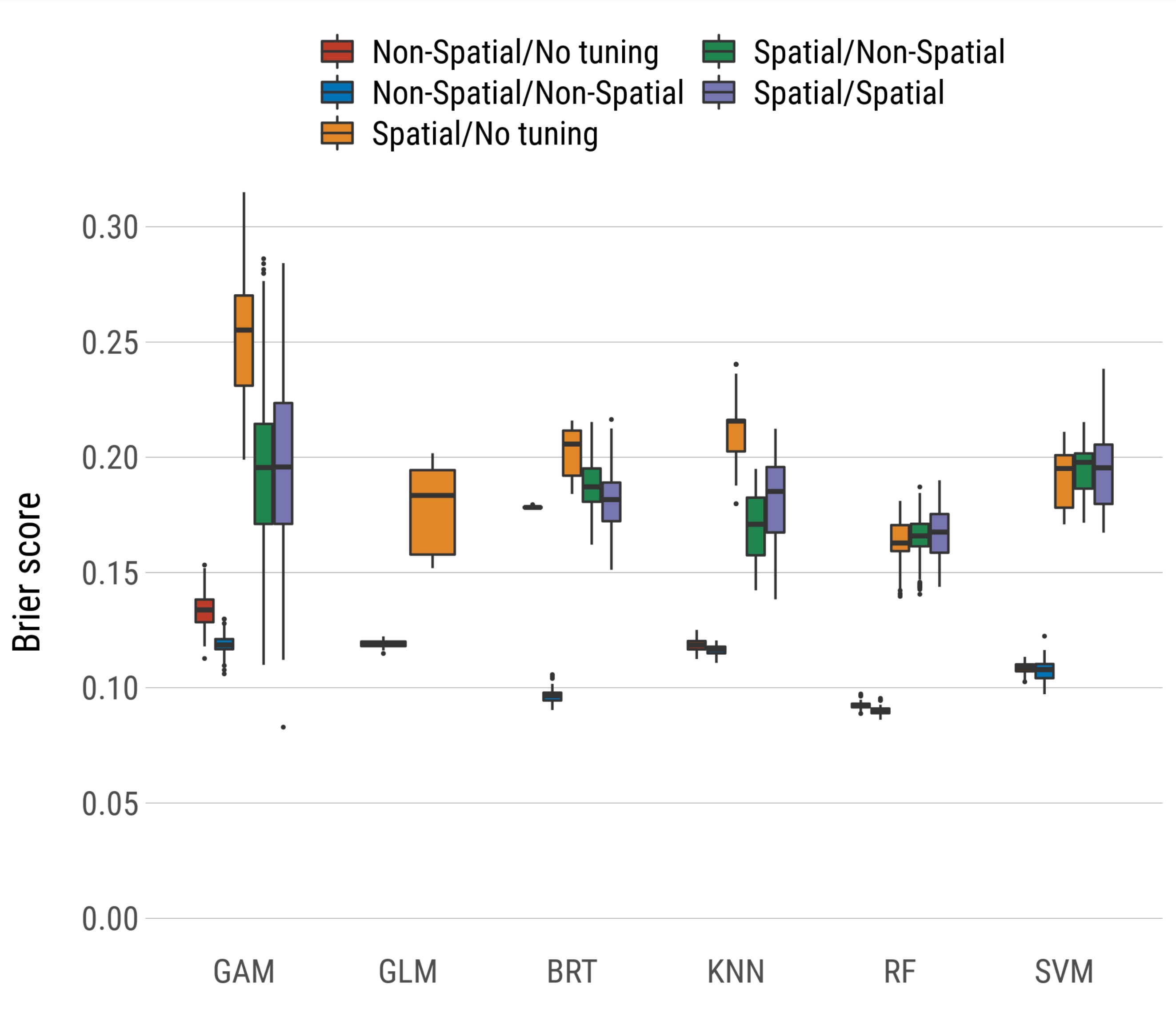
- `RF` showed the best predictive performance

Discussion

Predictive performance

- `RF` showed the best predictive performance
- High bias in performance when using non-spatial CV

Discussion



Non-Spatial/No tuning
Non-Spatial/Non-Spatial
Spatial/No tuning
Spatial/Non-Spatial
Spatial/Spatial
Brier score
0.30
0.25
0.20
0.15
0.10
0.05
0.00
GAM
GLM
BRT
KNN
RF
SVM

Discussion

Predictive Performance

- `RF` showed the best predictive performance
- High bias in performance when using non-spatial CV

Discussion

Predictive Performance

- `RF` showed the best predictive performance
- High bias in performance when using non-spatial CV
- The `GLM` shows an equally good performance as BRT, KNN and SVM

Discussion

Predictive Performance

- `RF` showed the best predictive performance
- High bias in performance when using non-spatial CV
- The `GLM` shows an equally good performance as BRT, KNN and SVM
- The `GAM` suffers from overfitting

Discussion

Hyperparameter tuning

- Almost no effect on predictive performance

Discussion

Hyperparameter tuning

- Almost no effect on predictive performance
- Differences between algorithms are higher than the effect of hyperparameter tuning

Discussion

Hyperparameter tuning

- Almost no effect on predictive performance
- Differences between algorithms are higher than the effect of hyperparameter tuning
- Spatial hyperparameter tuning has no substantial effect on predictive performance compared to non-spatial tuning

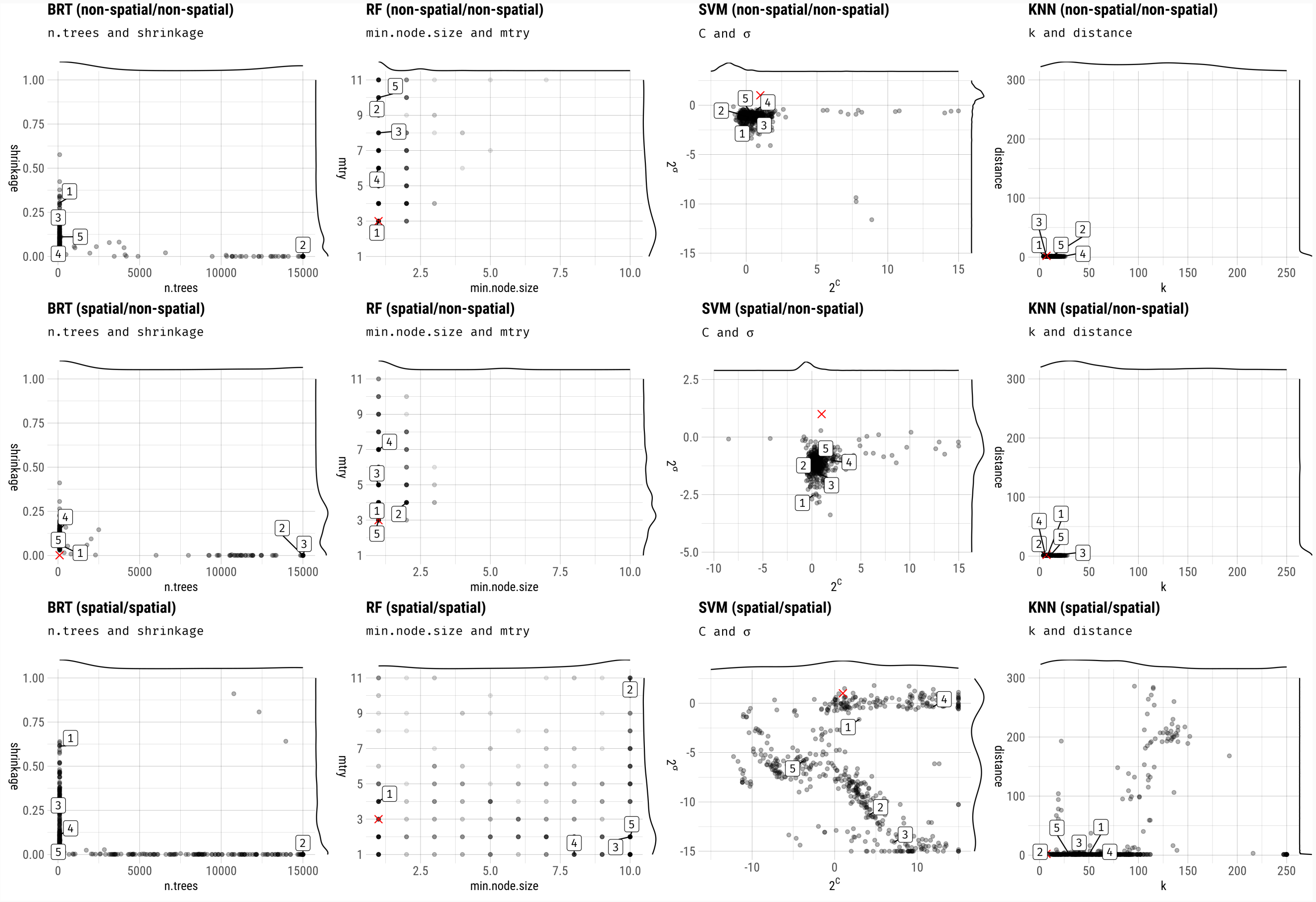
Discussion

Hyperparameter tuning

- Almost no effect on predictive performance
- Differences between algorithms are higher than the effect of hyperparameter tuning
- Spatial hyperparameter tuning has no substantial effect on predictive performance compared to non-spatial tuning
- Optimal parameters estimated from spatial hyperparameter tuning show a wide spread across the search space

Discussion

Tuning



Discussion

Hyperparameter tuning

- Almost no effect on predictive performance
- Differences between algorithms are higher than the effect of hyperparameter tuning
- Spatial hyperparameter tuning has no substantial effect on predictive performance compared to non-spatial tuning
- Optimal parameters estimated from spatial hyperparameter tuning show a wide spread across the search space

Spatial hyperparameter tuning should be used for spatial data sets to have a consistent resampling scheme

Thanks for listening!

Questions? Slides can be found here: https://bit.ly/2DsIEJg

Spatial modeling tutorial with *mlr*: http://mlr-org.github.io/mlr/articles/tutorial/handling_of_spatial_data.html
Spatial modeling tutorial with *sperrorest*: https://www.r-spatial.org/r/2017/03/13/sperrorest-update.html
arxiv preprint: https://arxiv.org/abs/1803.11266

References

Bischl, B, J. Richter, J. Bossek, et al. (2017). "mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions". In: *ArXiv e-prints*. arXiv: 1703.03373 [stat].

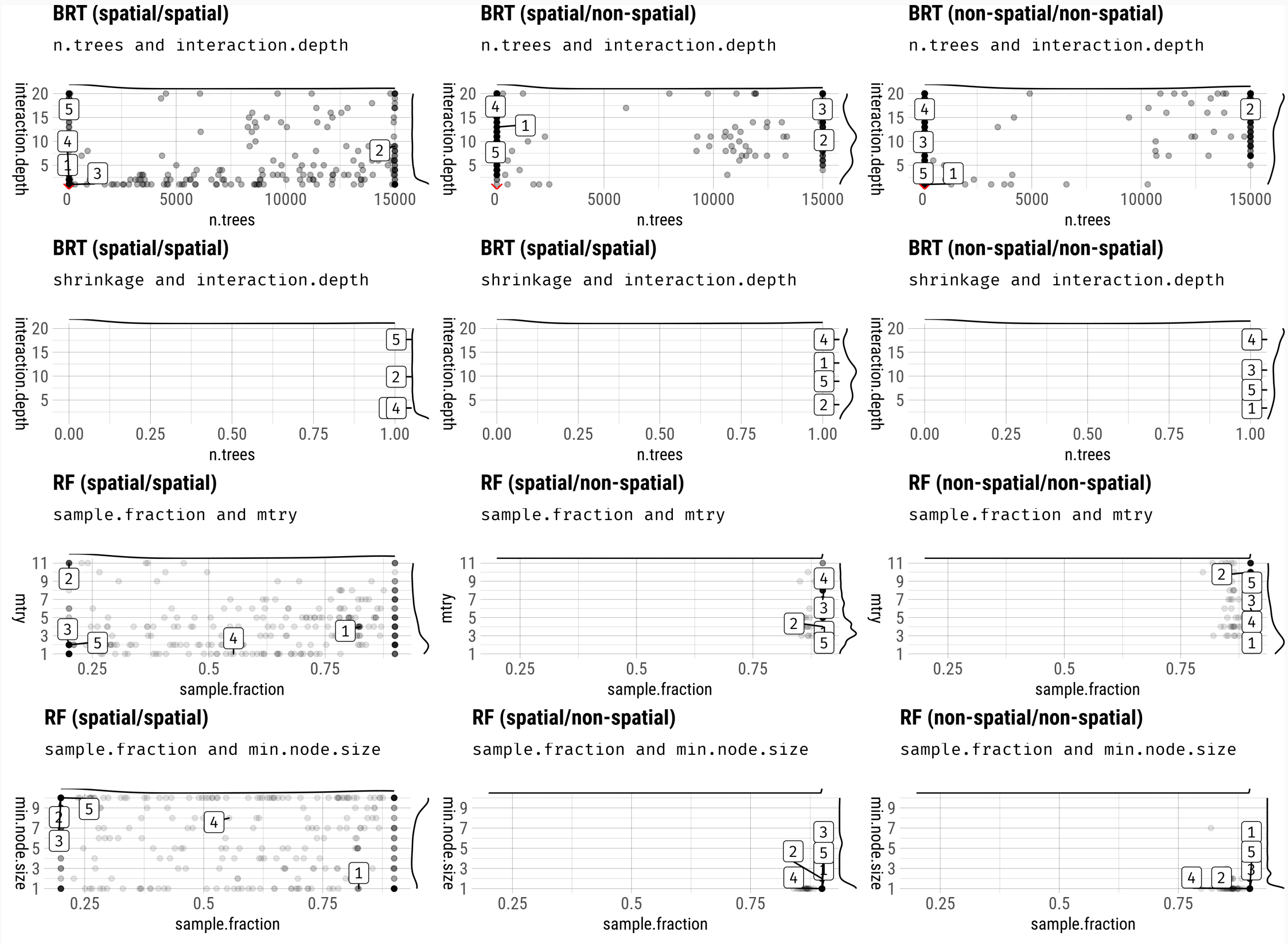
Brenning, A. (2012). "Spatial Cross-Validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The R Package Sperrorest". In: *2012 IEEE International Geoscience and Remote Sensing Symposium*. R package version 2.1.0. IEEE. DOI: 10.1109/igarss.2012.6352393.

Probst, P, M. Wright and A. Boulesteix (2018). "Hyperparameters and Tuning Strategies for Random Forest". In: *ArXiv e-prints*. arXiv: 1804.03515 [stat.ML].

Richter, J. (2017). "mlrHyperopt: Easy Hyperparameter Optimization with Mlr and mlrMBO". . R package version 0.1.1.

Backup ➕

Backup



BRT (spatial/spatial)
n.trees and interaction.depth
interaction.depth
n.trees
BRT (spatial/non-spatial)
n.trees and interaction.depth
interaction.depth
n.trees
BRT (non-spatial/non-spatial)
n.trees and interaction.depth
interaction.depth
n.trees
BRT (spatial/spatial)
shrinkage and interaction.depth
interaction.depth
n.trees
BRT (spatial/spatial)
shrinkage and interaction.depth
interaction.depth
n.trees
BRT (non-spatial/non-spatial)
shrinkage and interaction.depth
interaction.depth
n.trees
RF (spatial/spatial)
sample.fraction and mtry
mtry
sample.fraction
RF (spatial/non-spatial)
sample.fraction and mtry
mtry
sample.fraction
RF (non-spatial/non-spatial)
sample.fraction and mtry
mtry
sample.fraction
RF (spatial/spatial)
sample.fraction and min.node.size
min.node.size
sample.fraction
RF (spatial/non-spatial)
sample.fraction and min.node.size
min.node.size
sample.fraction
RF (non-spatial/non-spatial)
sample.fraction and min.node.size
min.node.size
sample.fraction