

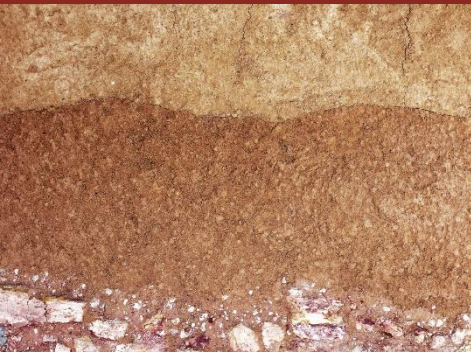


ISRIC
World Soil Information

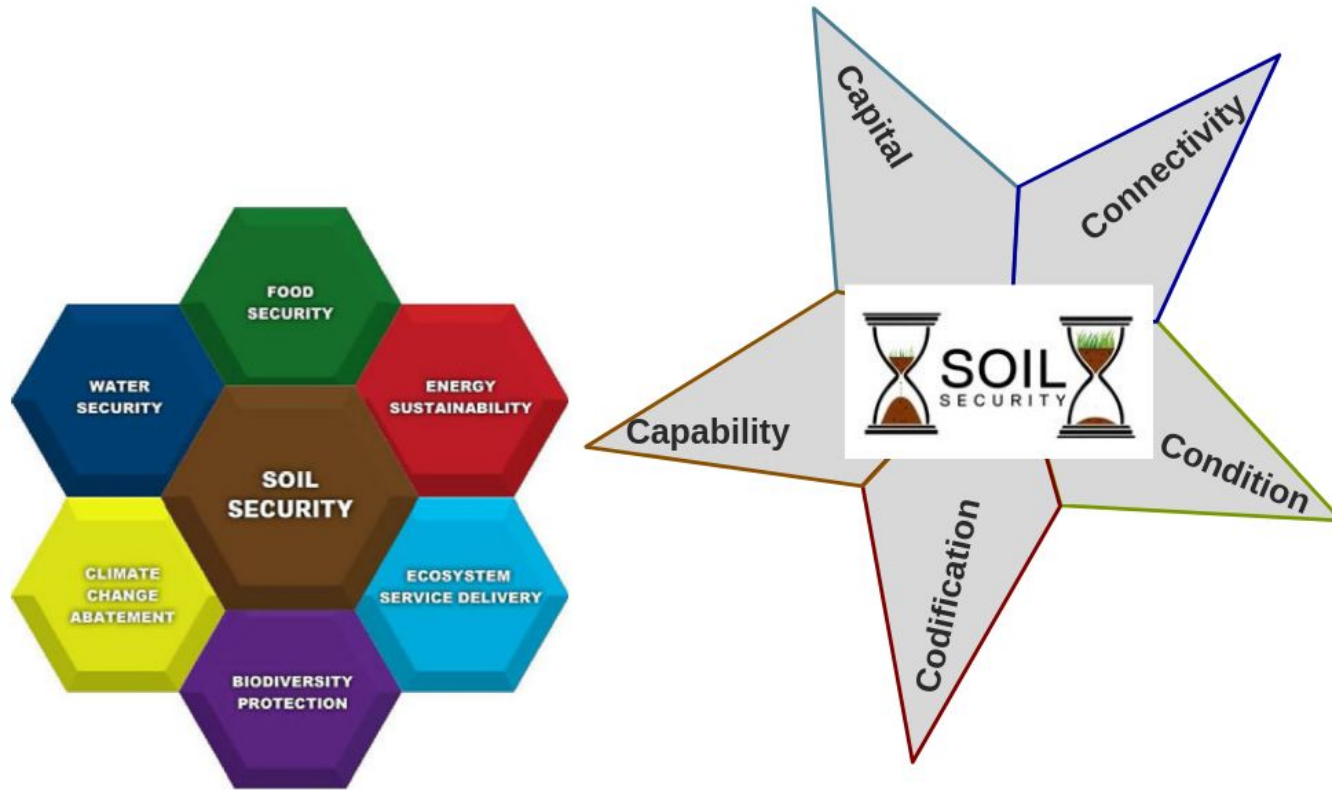


Modelling and mapping global soil information

Laura Poggio, Luis Moreira de Sousa, Gerard Heuvelink, Bas Kempen, Zhanquo Bai, Ulan Turdukulov, Maria Ruiperez-Gonzalez, Eloi Ribeiro, Niels Batjes, Rik van den Bosch

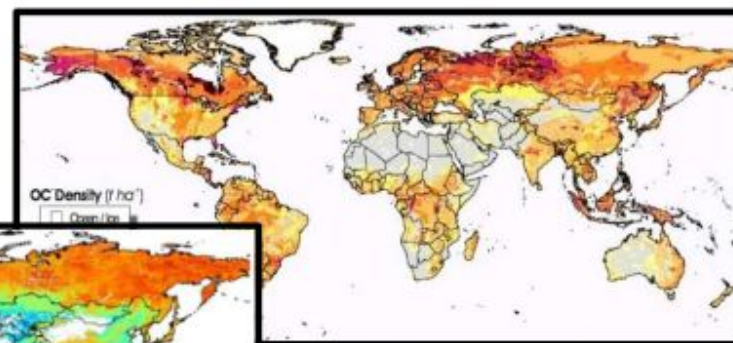


Why modelling at global scale?

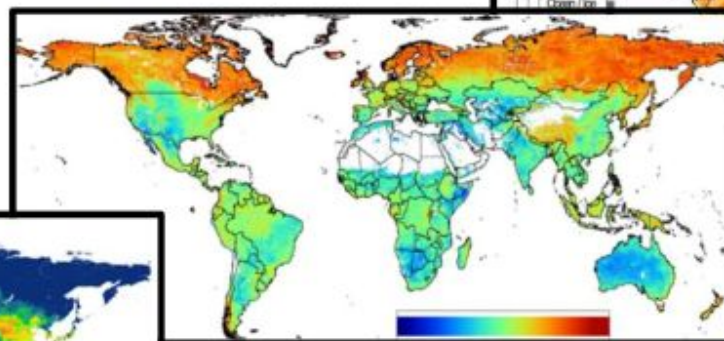


Koch et al, 2013, Global Policy;
McBratney et al, 2014, Geoderma

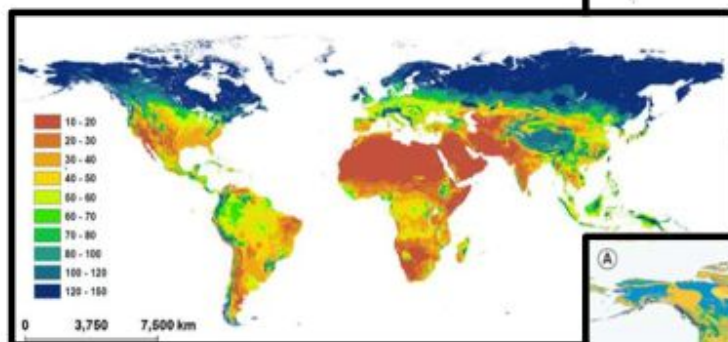
Examples of Global SOC maps



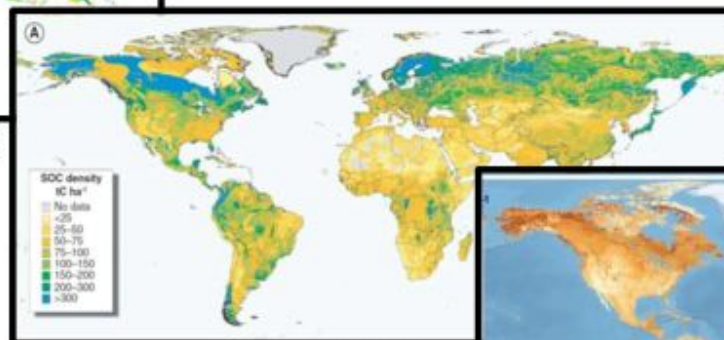
USDA



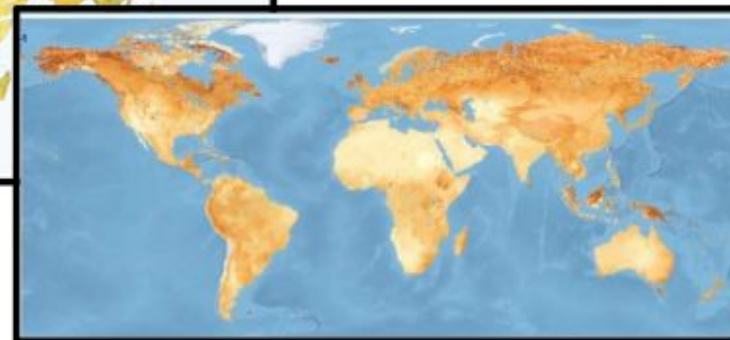
SoilGrids



Stockmann et al, 2013



HWSD

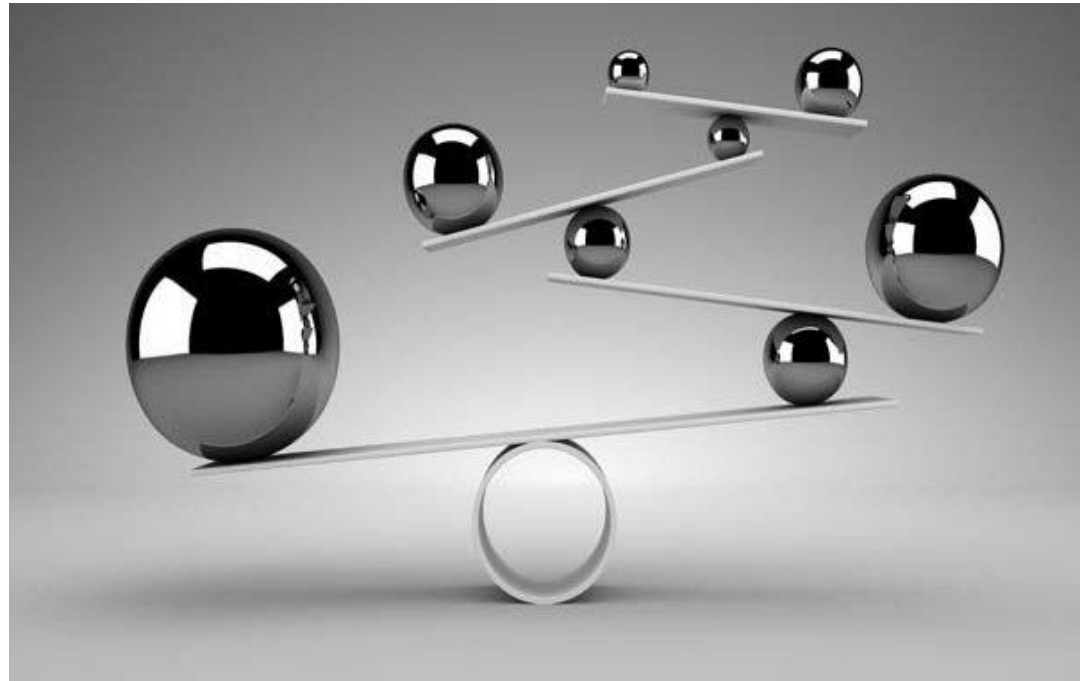


ITPS-FAO-GSP

Challenges



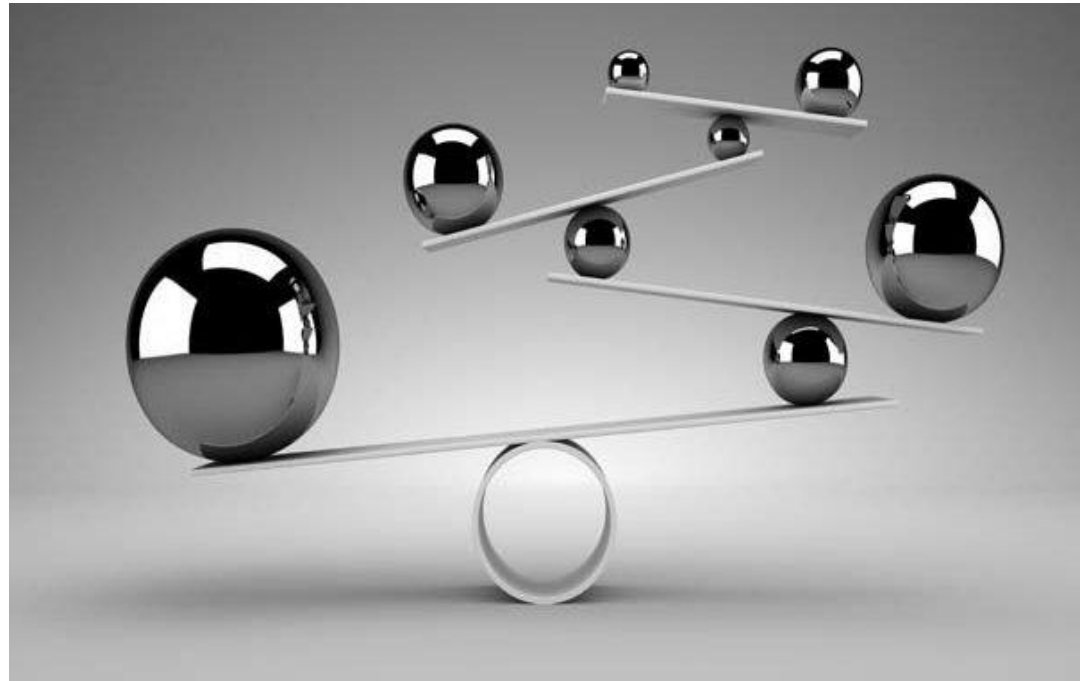
- Data
- Covariates
- Modelling



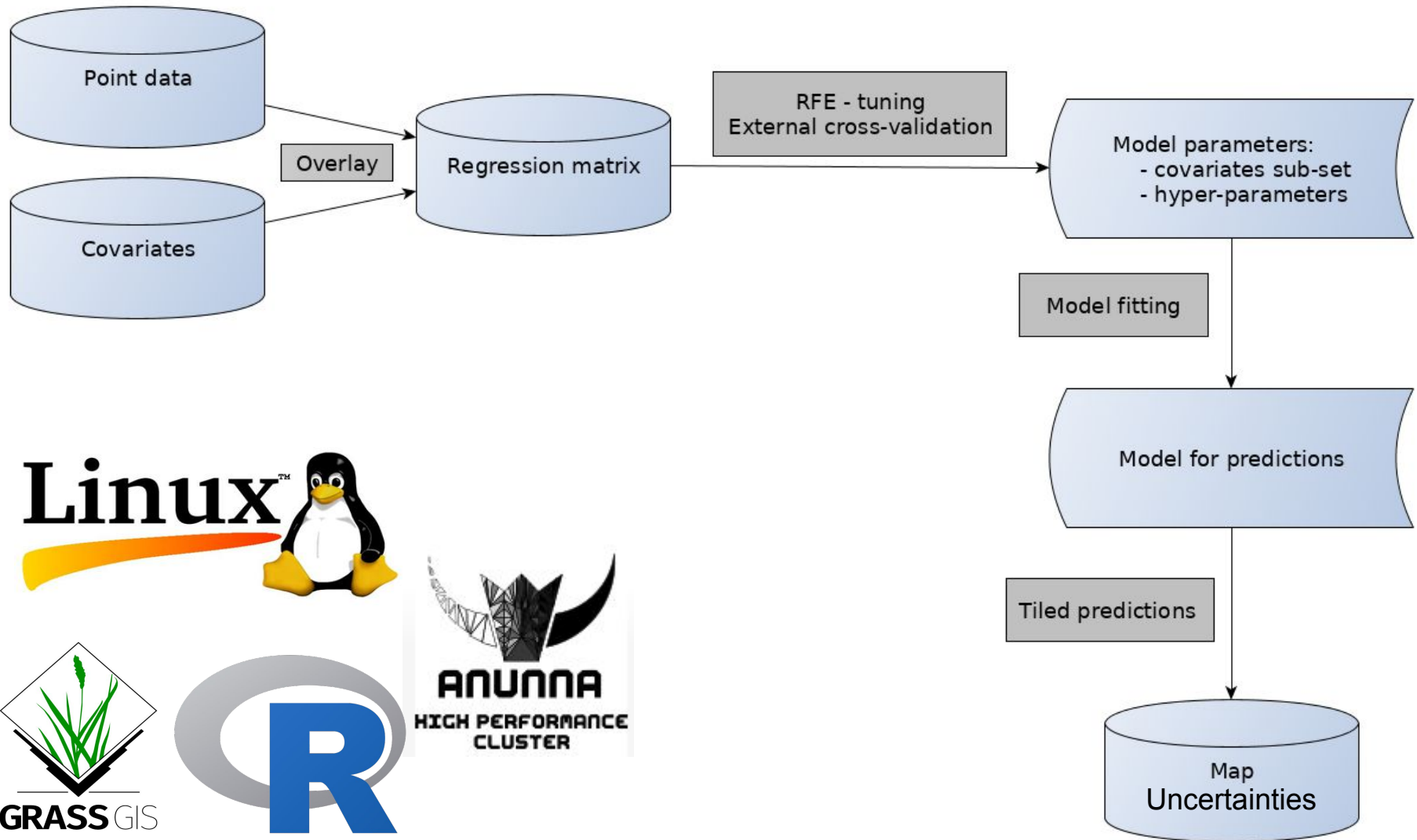
Challenges



- Quality control
- Computational resources



SoilGrids 2.0



Linux™



[illegible]

Model tuning for global modelling



- Spatially balanced 10 fold cross-validation
- Covariates selection for parsimonious model
- Model tuning: performances and uncertainty

Covariates



Tuning: parsimonious model

Covariates selection:

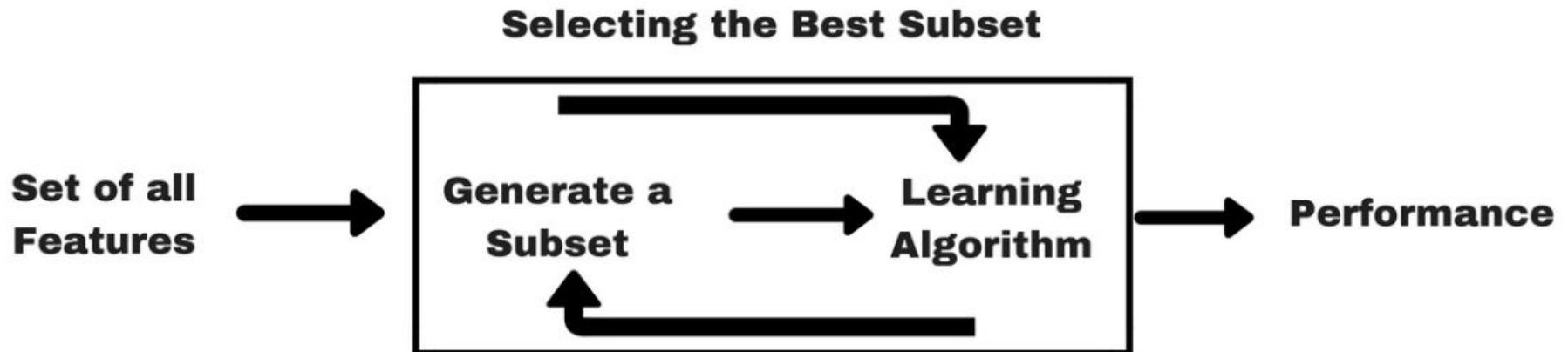
- De-correlation
- Successive removal by performances
maximisation
- Covariates grouping

Covariates RFE

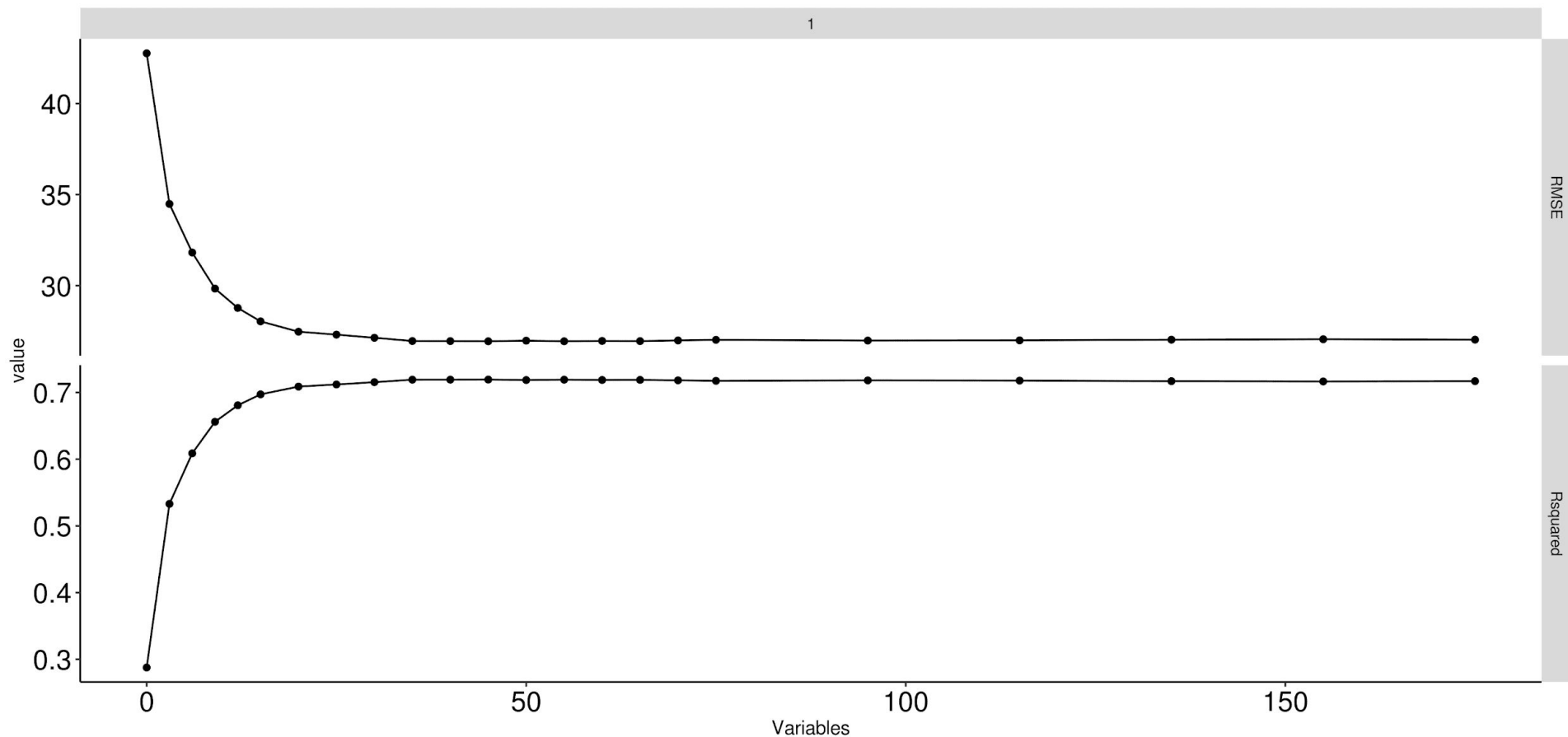


Recursive feature elimination:

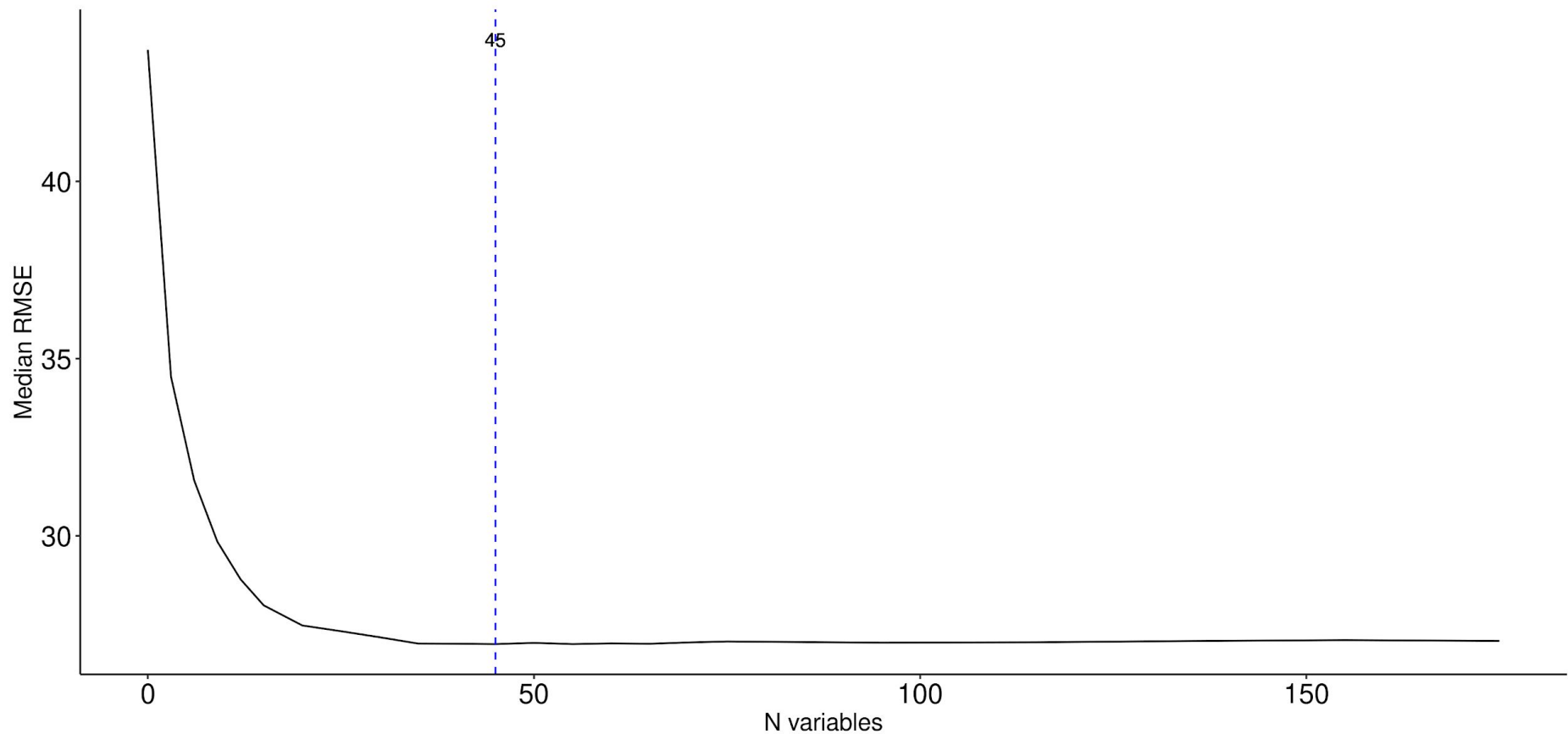
- RandomForest (machine learning) broad equivalent of backward stepwise regression
- Optimised for large sets of covariates



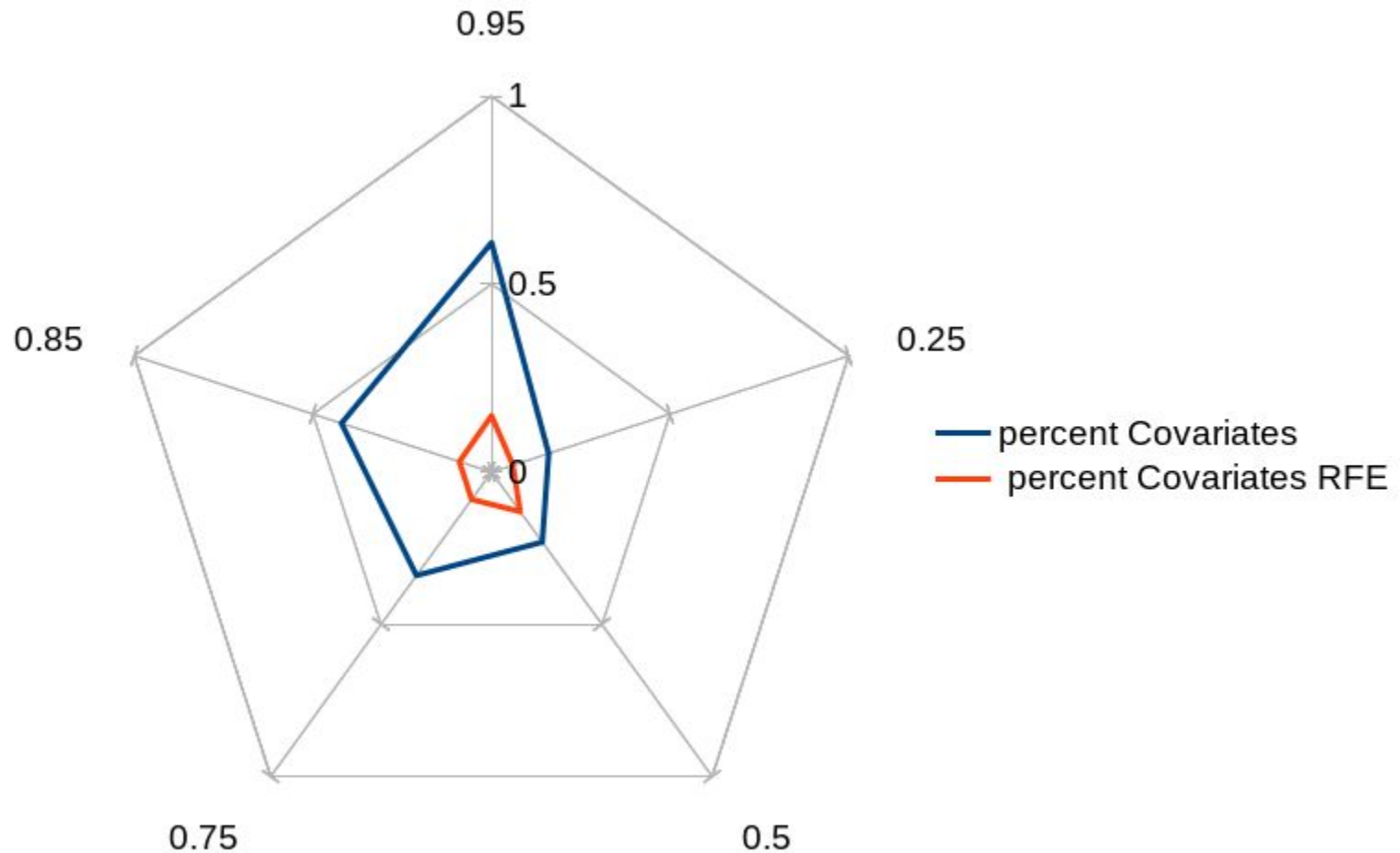
Covariates RFE



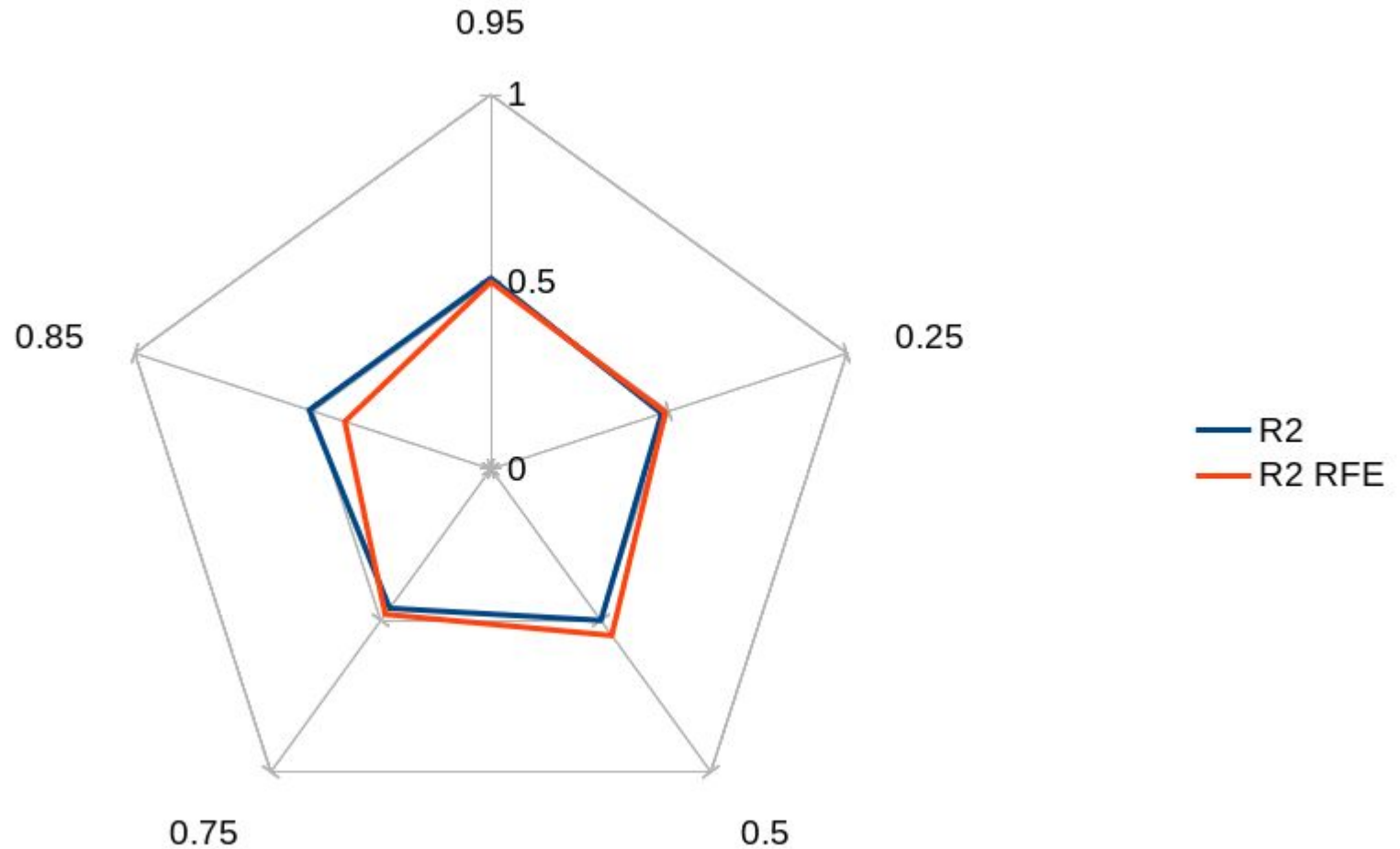
Covariates RFE



Covariates reduction



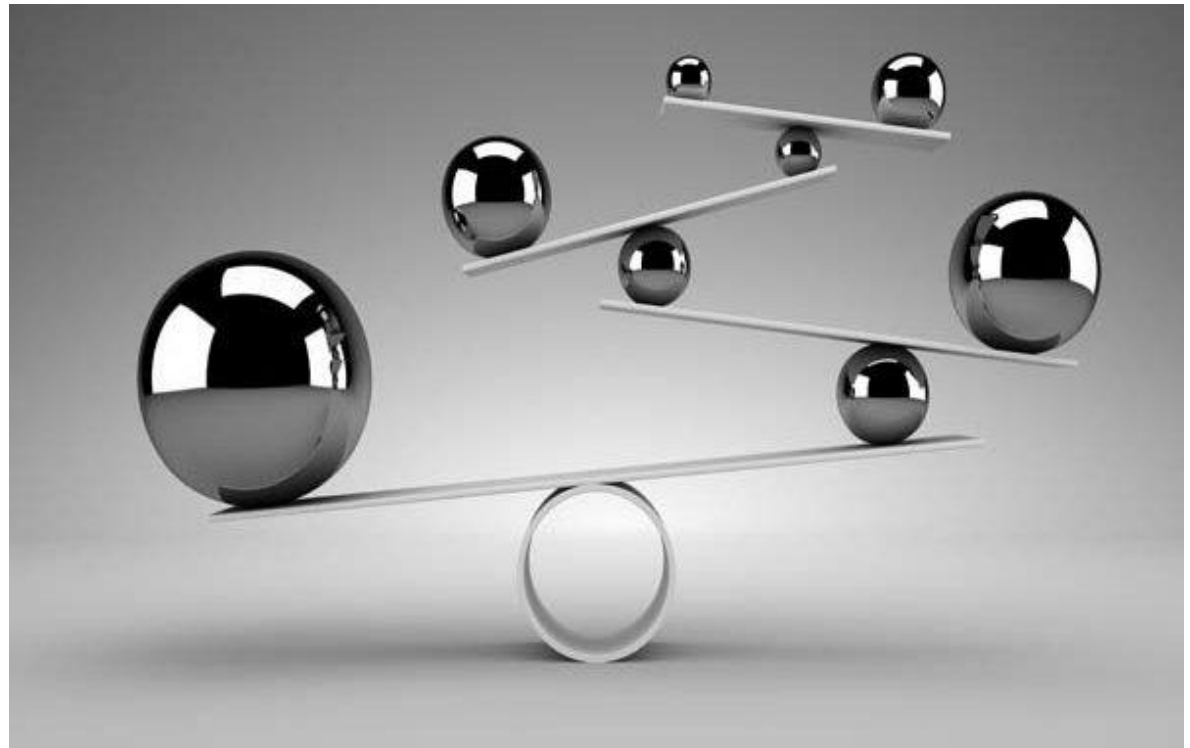
Covariates reduction



Parsimonious model



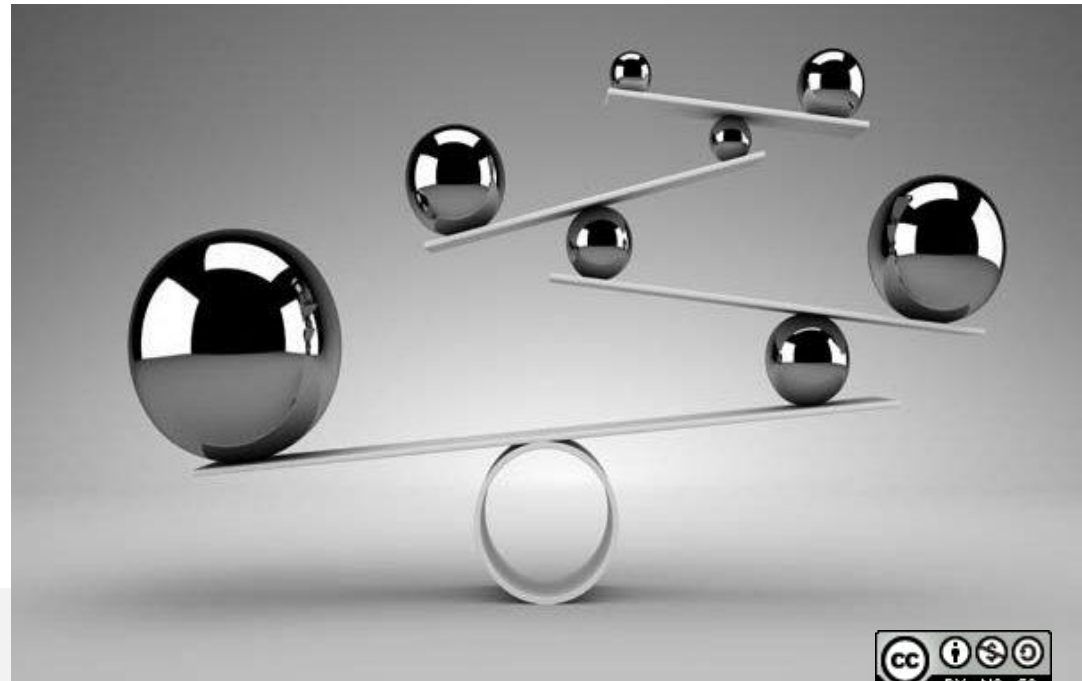
Less covariates similar performances



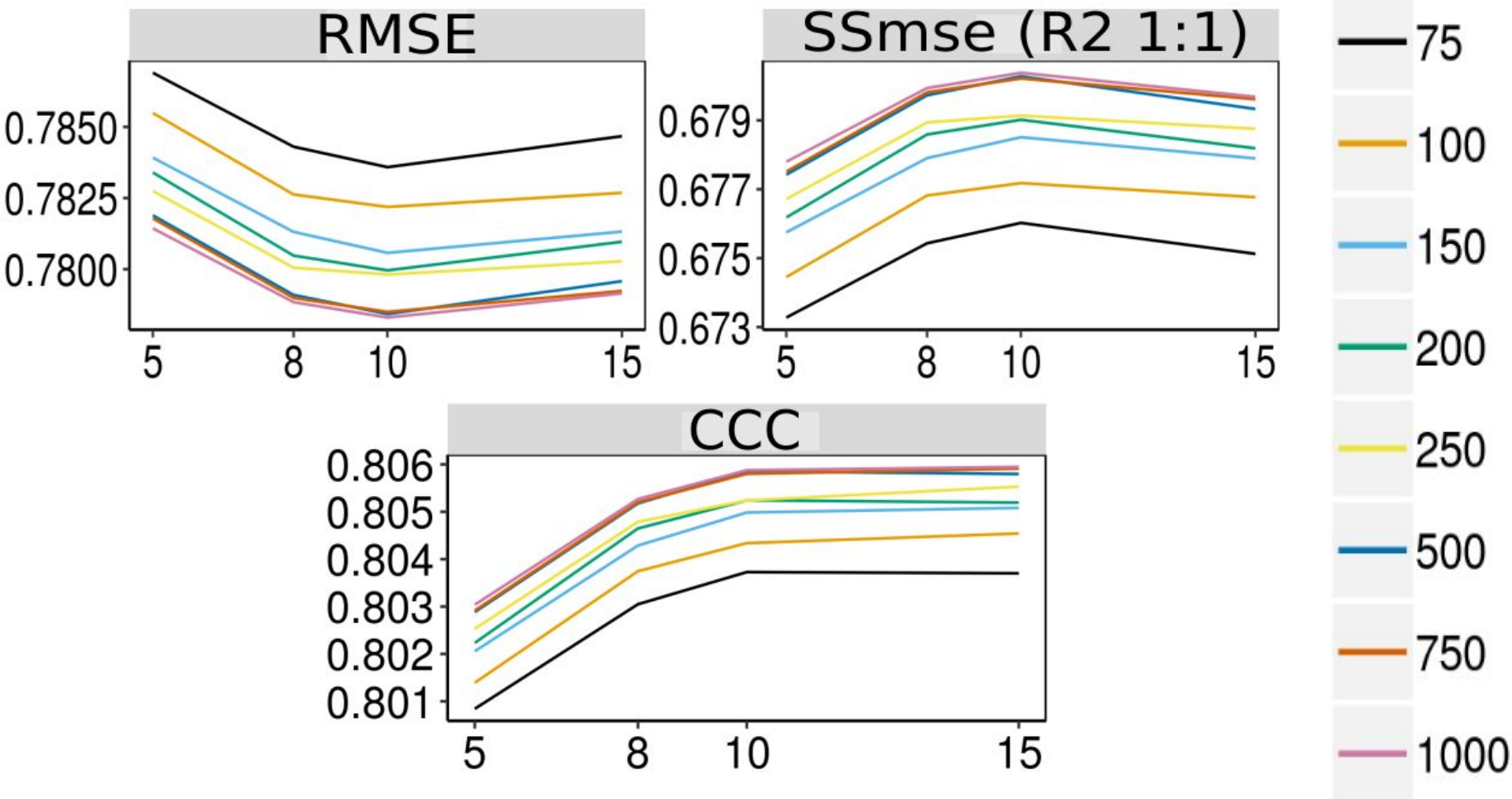
Tuning: hyperparameters



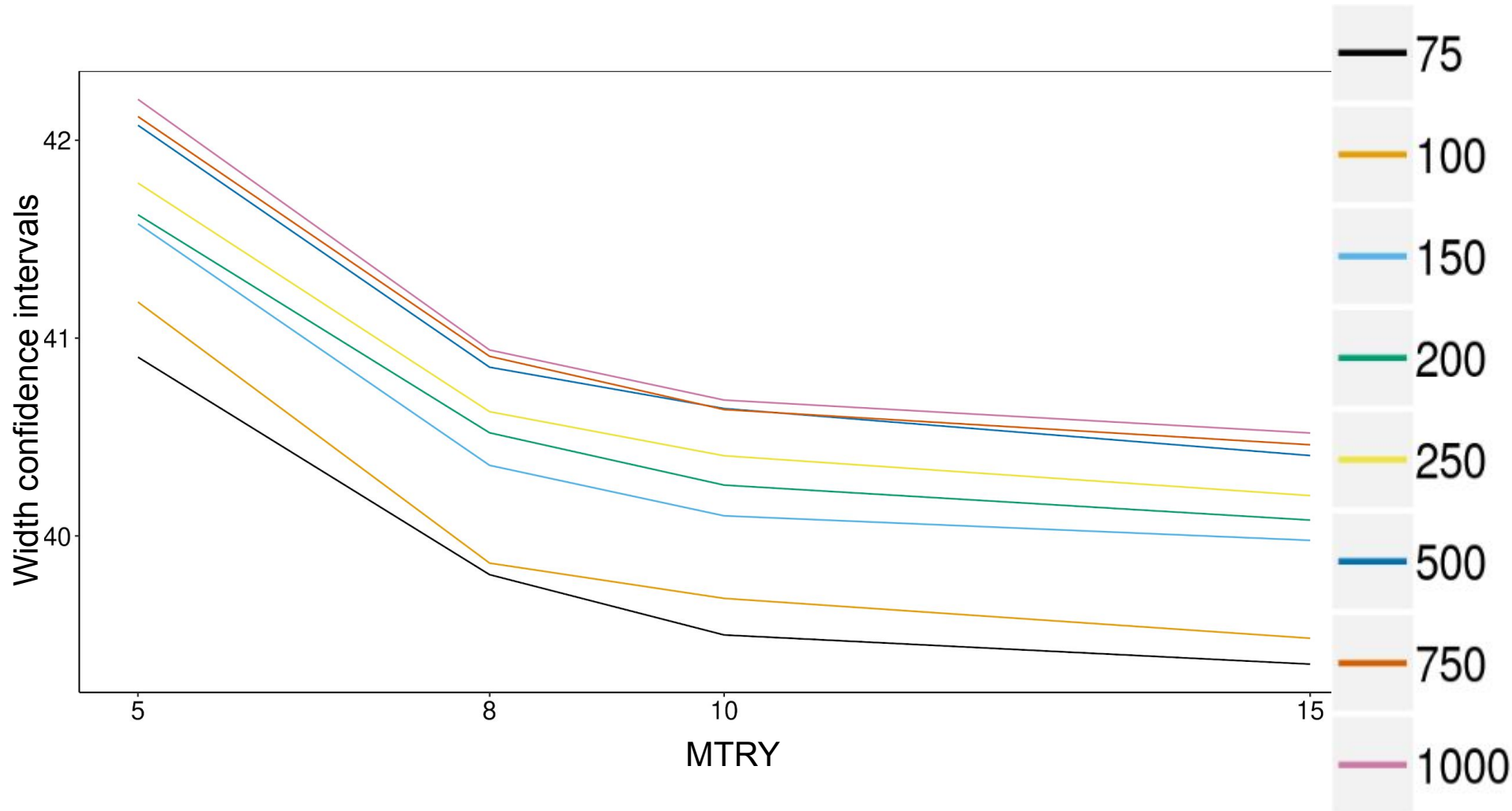
1. Covariates sub-set
2. Model depending ---> QuantilesRandomForest (in this example)



Tuning - hyperparameters



Tuning - Uncertainty and intervals



Tuning - Uncertainty and intervals



	Validation points included in 90% intervals
Correlation	0.928
Correlation + RFE	0.927
Correlation + RFE + Grouping	0.923

Open issues



- Depth:
 - 2D,2.5D,3D approach?
 - Spline?
 - Depth as covariate?
 - Something else?

Open issues

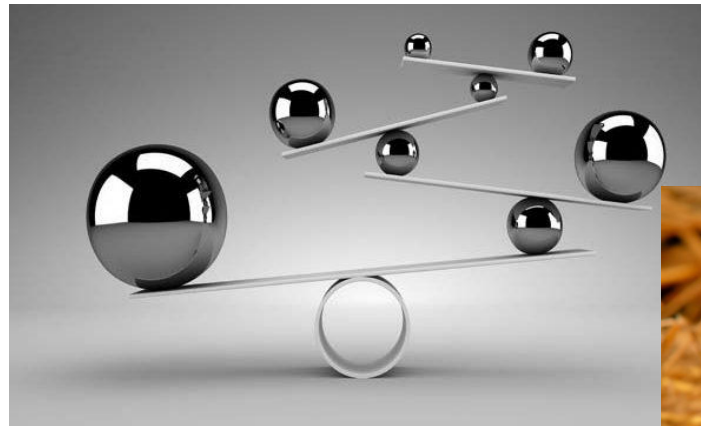


- Model:
 - Is ML (especially RandomForest) the best approach?
 - Are the predictions intervals created realistic?
 - Integration with national/regional models/results?

Concluding remarks



- Quality controlled and standardised input data
- Spatially aware cross-validation
- Parsimonious models
- Uncertainty assessment
- Open questions!





ISRIC
World Soil Information



Thank you for listening
laura.poggio@isric.org

www.isric.org

