Manuscript_draft

Lorenzo Menichetti

2023-04-24

Title

Abstract

Introduction

Agroforestry and SOC

Introducing the problem: it is difficult to relate the SOC stocks with AFS treatments

Variability

Limited data availability over time and space

What we did to solve it

We collected more data

We integrated the data with more sources (SoilGrids)

We analyzed the data with a nonparametric ML technique

We decomposed the variance of the models

Materials and Methods

Data mining

This study was initially conducted with a literature review aiming studies related to AFS and SOC Stocks, including SOC Content and bulk density. Among the research engines, for this study it was used: ISI—Web of Knowledge, Google Scholar, Scopus, Research Gate, and Scielo.org. In order to increase the number of potential references, it was used the following keywords: soil AND ('carbon stock*' OR 'carbon pool*' OR 'carbon sequestration' OR 'carbon concentration') AND (agroforest* OR parkland* OR homegarden OR multistrata OR hedgerow OR windbreak OR shelter-belt OR 'live fence' OR 'tree intercrop*' OR silvo*arable OR silvo*pasture OR 'rotation*wood*' OR tree*fallow* OR fallow OR (tree* AND 'improve* fallow*') OR (tree* AND relay*crop*) OR (tree* AND alley*crop*)) (Cardinael et al., 2019). Them it was filled a table with several parameters found in the papers, including Location, Climate Classification according to Köppen, Physical Property of the Soil, AFS Description, Previous Land Use, Age of System, SOC Stock from the AFS and from the Control plot, following the IPCC guidelines (IPCC, 2022). Over 400 peer reviewed papers have been examined and 174 have been considered adequate for this study. The eligibility of the papers was the presence of SOC Stocks data, measured in the AFS and a control area (synchronic); or SOC data sampled before and after the AFS implementation (diachronic).

Figure 1: The dataset

Data treatment

Outliers detection

Integration with variables from the SoilGrids database

Random forest model development

Introducing the method briefly

The different models we tested

Model.description	Model.name
Stocks difference, without SoilGrids data	Model 1
Stocks difference, with SoilGrids data	Model 1.1
Stocks after AFS, without SoilGrids data	Model 2
Stocks after AFS, with SoilGrids data	Model 2.1
Stocks after AFS, with only clay from SoilGrids data	Model 3.1
Stocks after AFS, no previous land use but with SoilGrids data	Model 0
Stocks after AFS, no previous land use	Model 0.1

The variables we considered

Table 2: The variables

Variable	Abbreviation	Model.1	Model.1.1	Model.2	Model.2.1	Model.3.1	Model.0	Model.0.1
AFS management	Man	X	X	X	X	X	X	X
Climate (Köppen)	Clim	X	X	X	X	X	X	X

Variable	Abbreviation	Model.1	Model.1.1	Model.2	Model.2.1	Model.3.1	Model.0	Model.0.1
Previous land use	L	X	X	X	X	X		
AFS duration	Dur	X	X	X	X	X	X	X
Depth	Depth	X	X	X	X	X	X	X
Region	Reg	X	X	X	X	X	X	X
layer thickness	Thick	X	X	X	X	X	X	X
clay (SoilGrids)	Clay		X		X	X		X
sand (SoilGrids)	Sand		X		X			X
pH (SoilGrids)	pH		X		X			X
N (SoilGrids)	N		X		X			X

Predicting the stocks difference We tested how the differences in C stocks (after AFS treatment and before, assuming that the control treatment from each study was representative of the previous state of the system) can be predicted by a trained model.

Models were all in the form

$$\Delta_{stocks} \sim f(variable_1, variable_2, variable_{...})$$

Predicting the AFS stocks accumulation Models were all in the form

$$AFS_{stocks} \sim f(variable_1, variable_2, variable_{...})$$

Predicting the AFS stocks accumulation but without including information about previous land use

Validation approach: fully independent, stratified by cluster

Results

The performances of the models on the validation dataset

- The models did not manage to predict reasonably the Δ in stocks, but they did manage to predict the C stocks after a certain time of agroforestry treatment
- The models perform relatively well (better than random, there is some information) even on a fully independent dataset
- There was virtually no information added from the SoilGrids dataset
- Removing the previous land use information still retains a model with some value (although of course worse than with such information)

Table 3: A knitr kable.

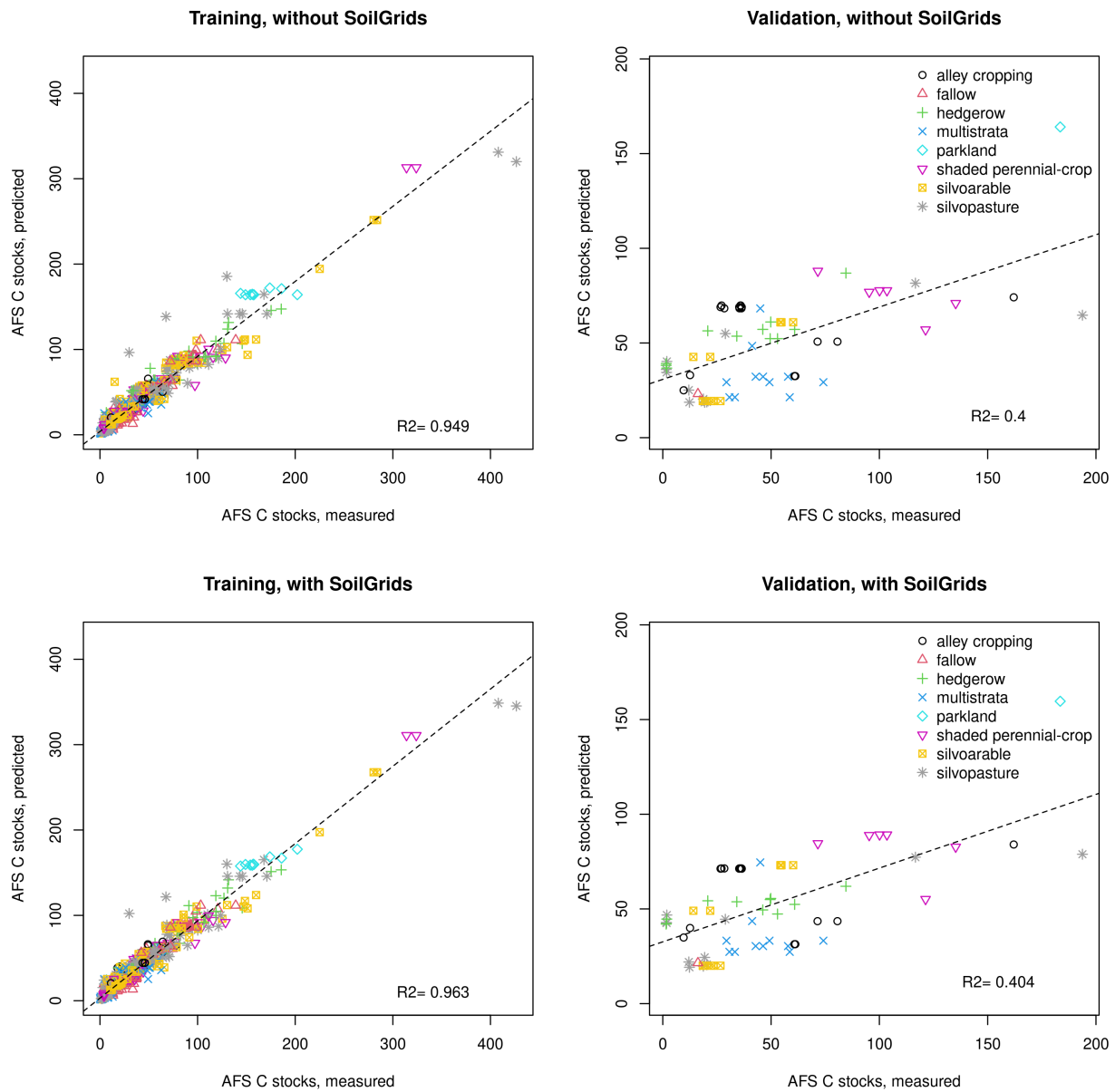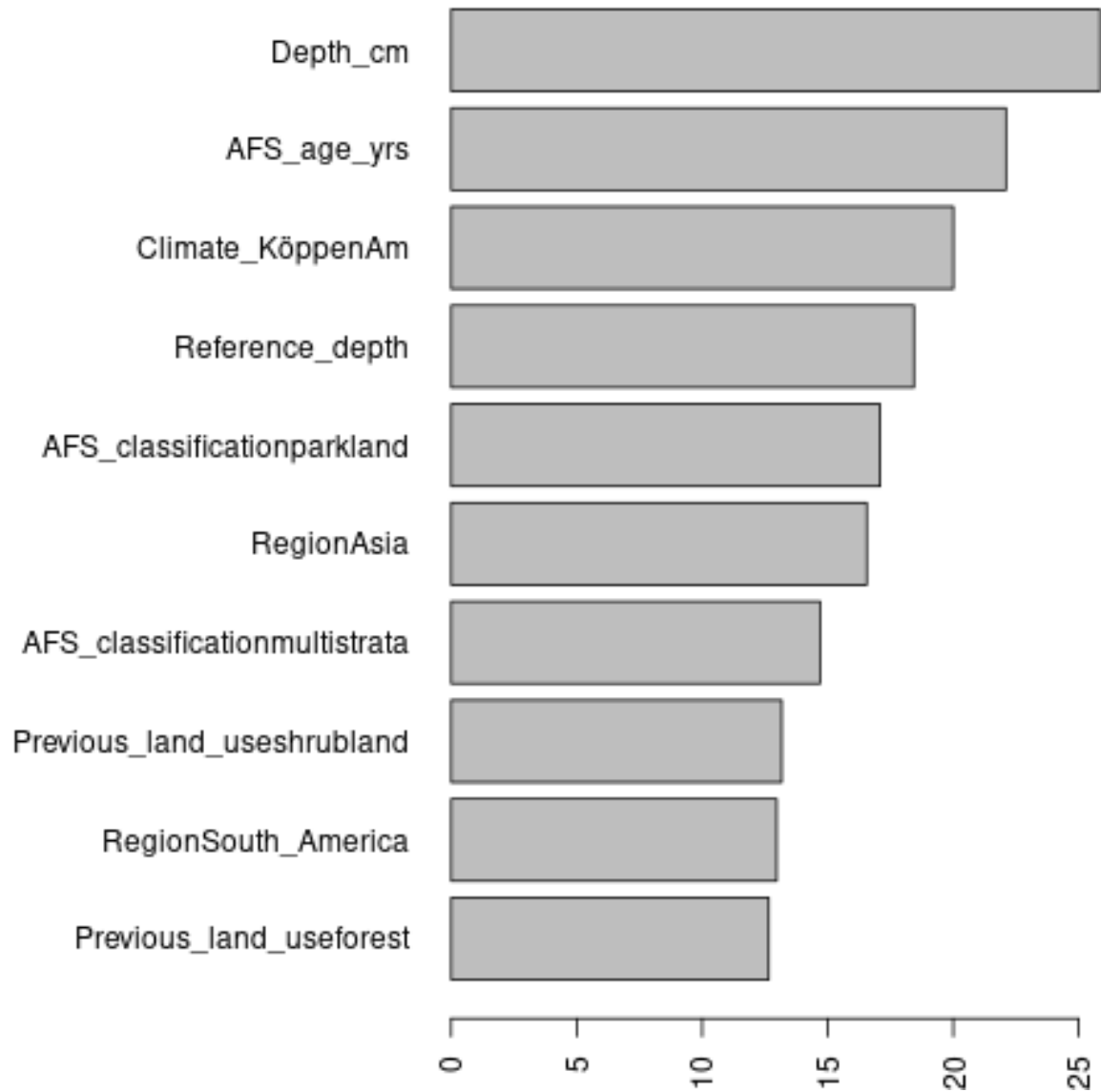| Model.type	R.2
Stocks difference, without SoilGrids data	0.2663974
Stocks difference, with SoilGrids data	0.2207099
Stocks after AFS, without SoilGrids data	0.3996744
Stocks after AFS, with SoilGrids data	0.4041894
Stocks after AFS, with only clay from SoilGrids data	0.3512694
Stocks after AFS, no previous land use but with SoilGrids data	0.3570584
Stocks after AFS, no previous land use	0.3086806

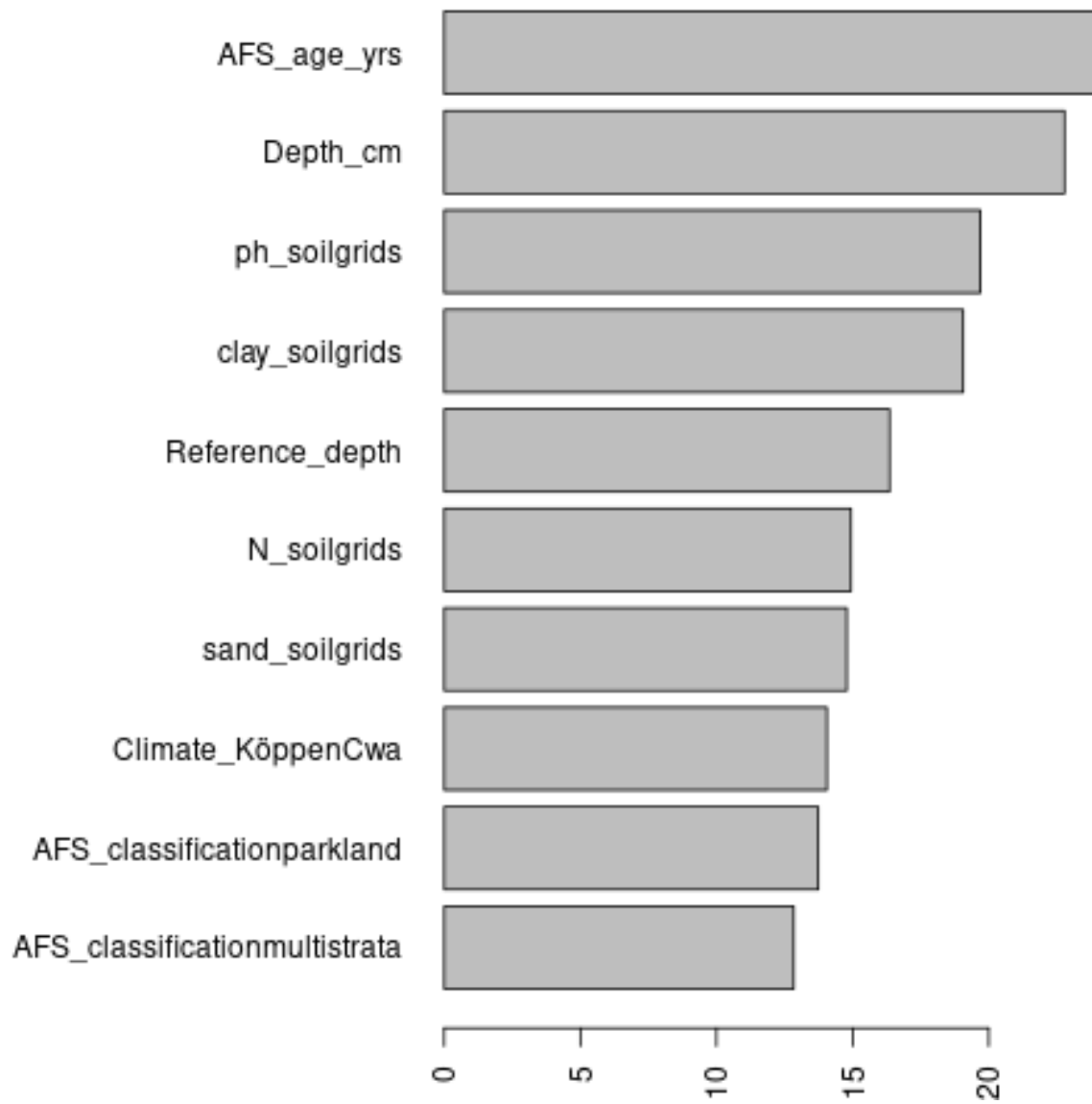Figure 2: Model evaluation (best models)

Decomposing the explained variance: the relative variable importance
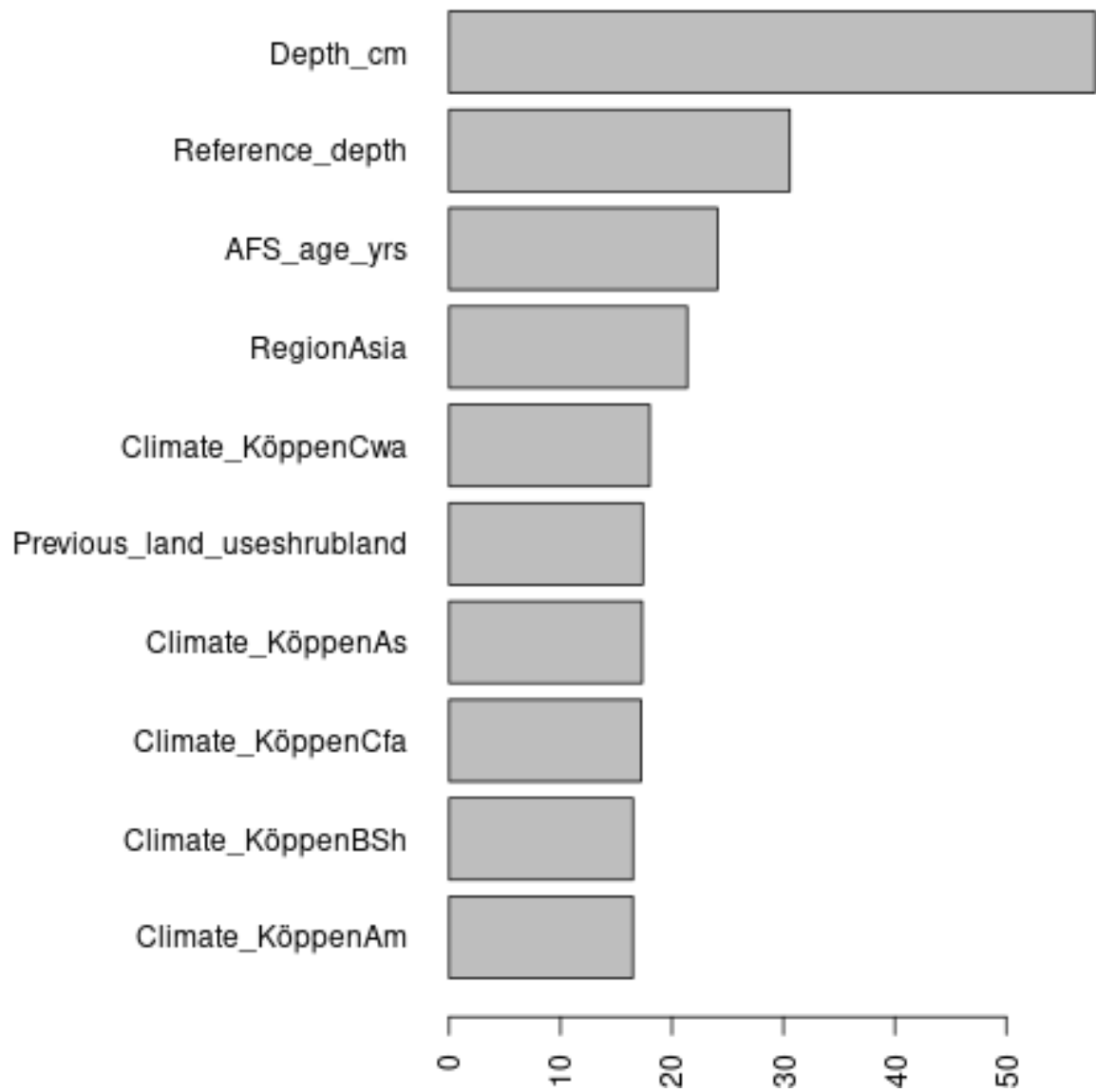
Predicting the stocks difference

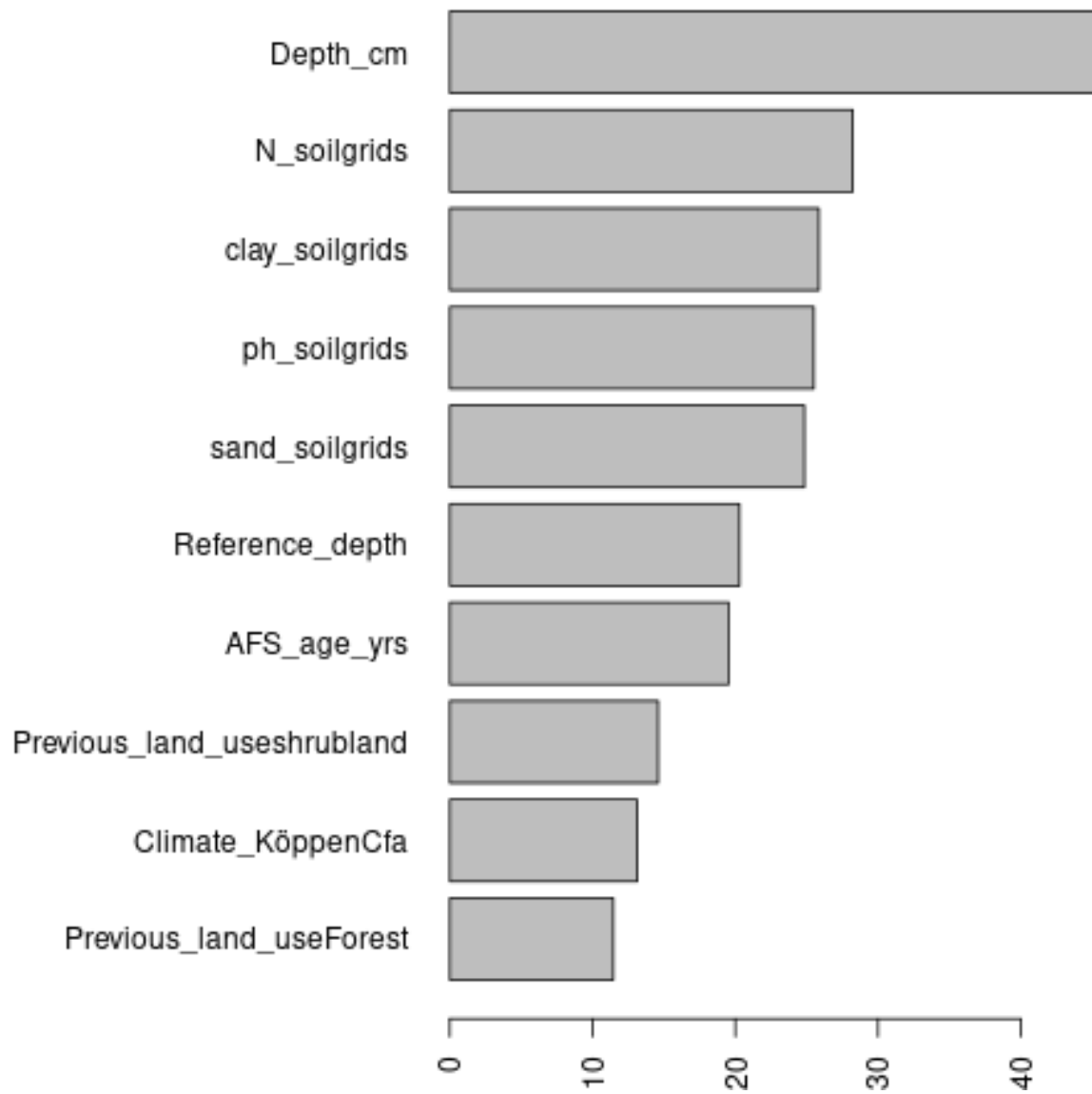Stocks difference, without SoilGrids data
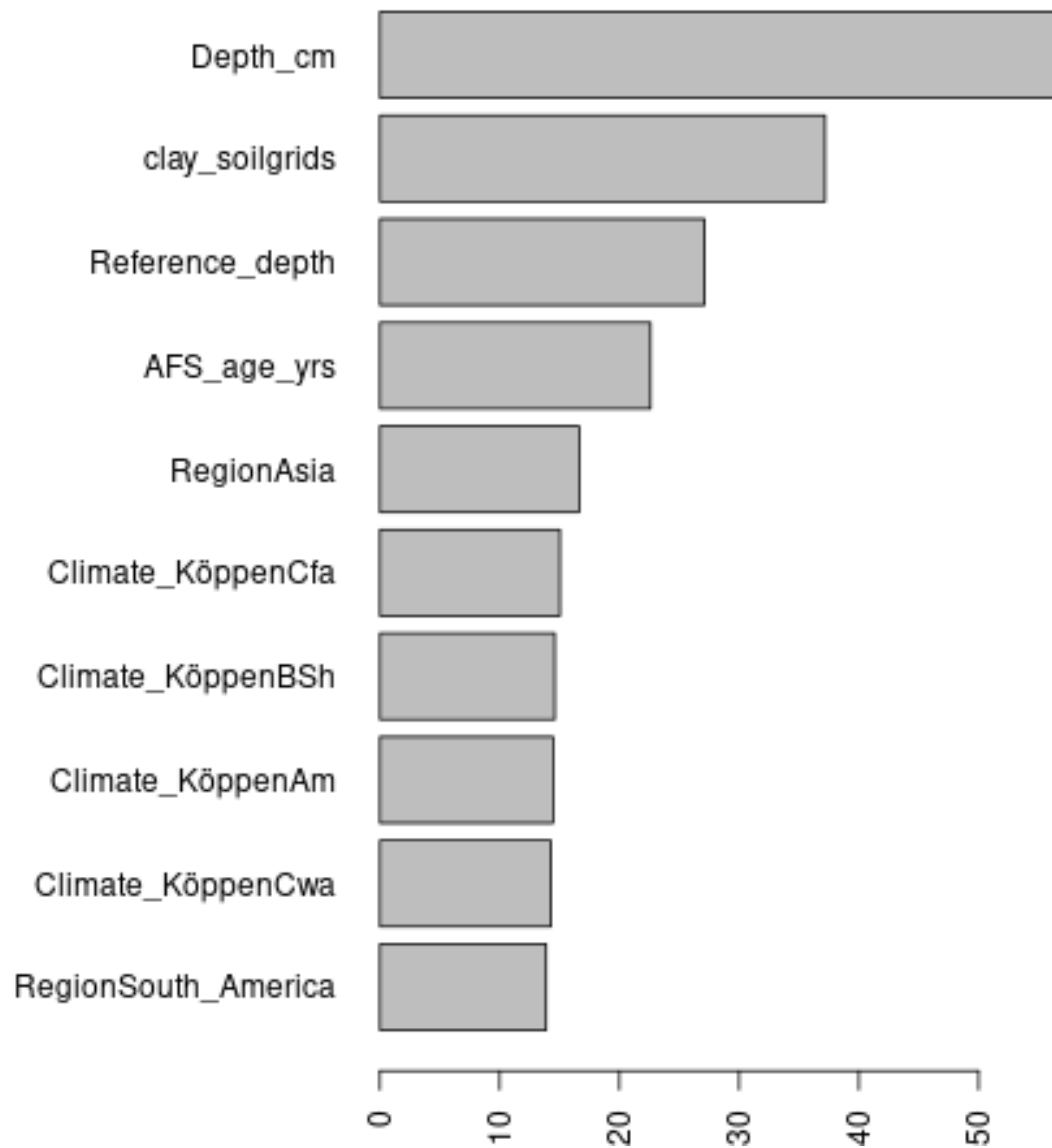

Stocks difference, with SoilGrids data

Predicting the AFS stocks (including information about previous land use) Stocks after AFS, without SoilGrids data

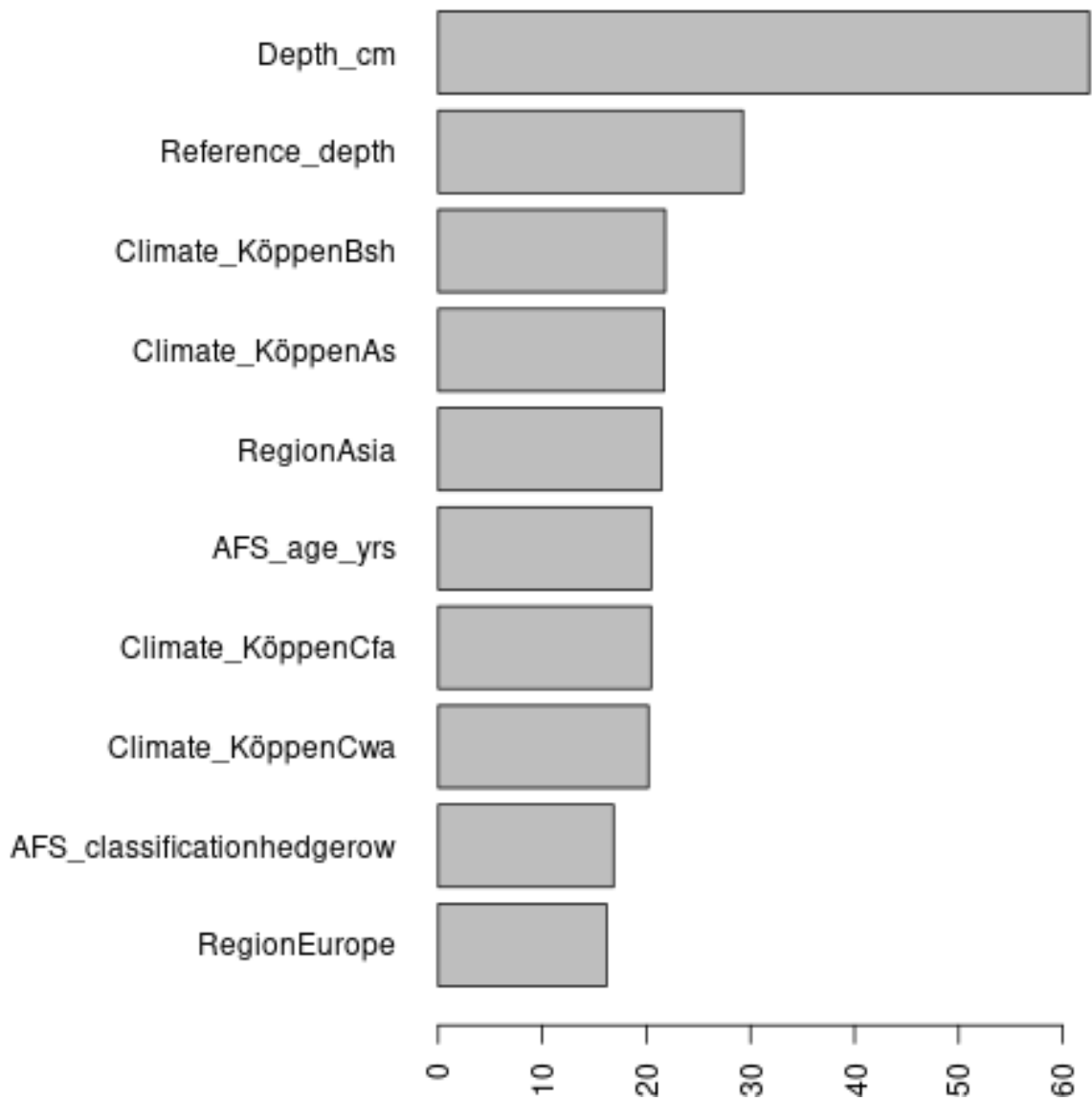Stocks after AFS, with SoilGrids data

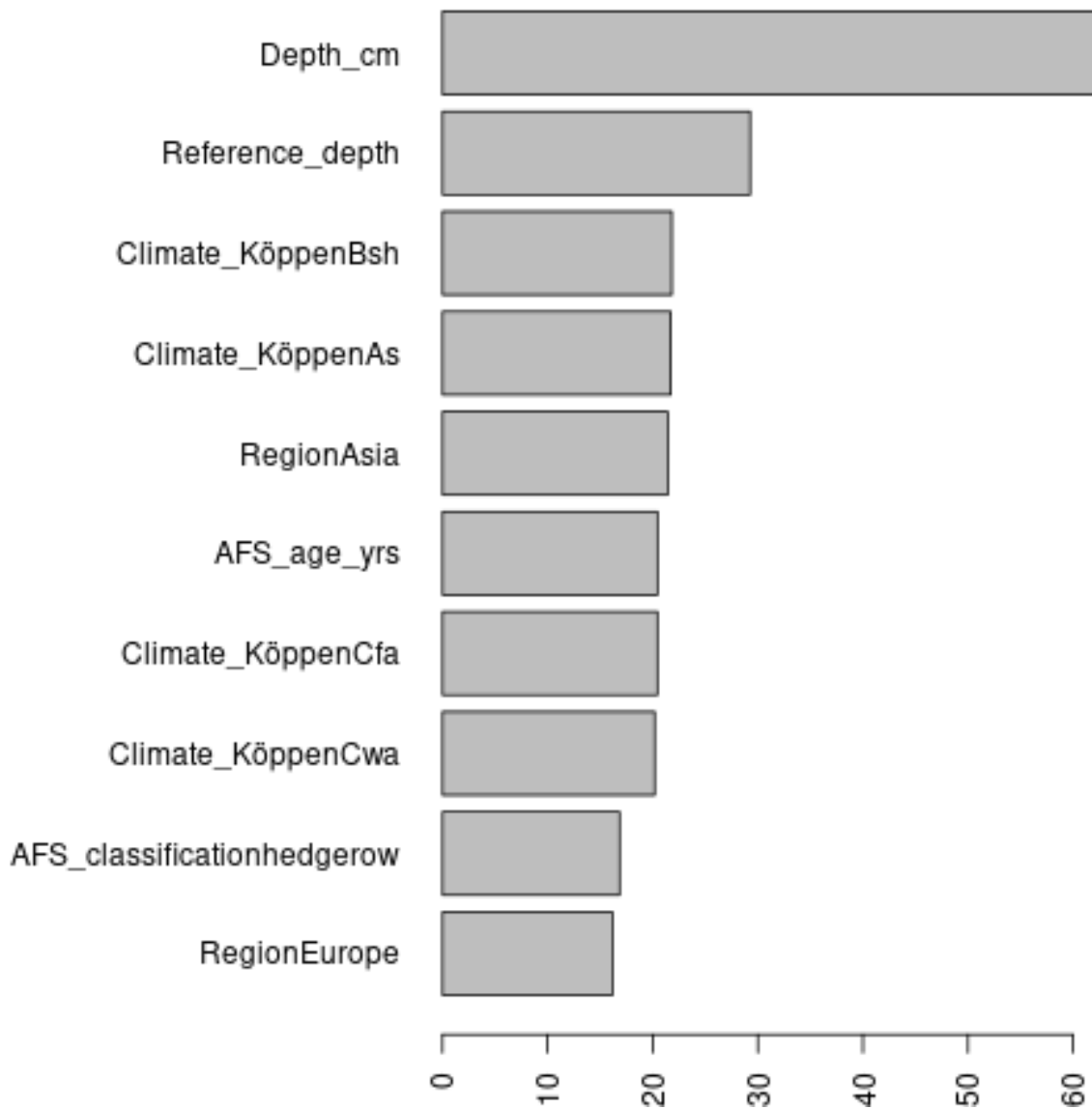Stocks after AFS, with only clay from SoilGrids data

Predicting the AFS stocks (but without information about previous land use)

Stocks after AFS, no previous land use but with SoilGrids data

Stocks after AFS, no previous land use

Discussion

The Δ in stocks

The information from the SoilGrids dataset

##The models as a predictive tool

Considering previous land use

Ignoring previous land use

Conclusions