Usage of visual and near-infrared spectroscopy to predict soil properties in forest stands

Felix Thomas, Ulrike Werban, Carina Becker, Rainer Petzold

Department Monitoring- und Erkundungstechnologien
Helmholtz-Zentrum für Umweltforschung GmbH - UFZ

FNR
Fachagentur Nachwachsende Rohstoffe e.V.

HELMHOLTZ
ZENTRUM FÜR
UMWELTFORSCHUNG
UFZ

Motivation

- high demand for up-to-date information about humus conditions is important for site-adapted management
- continuous impact through element input, management measures and changing climate
- need for inexpensive and fast methods

Image: Thomas Frey/dpa

Background

- Vis-NIR spectroscopy as known method to estimate physical and chemical soil properties (see Kooistra et. al. (2001), Gubler (2011), Riedel et. al. (2018))
- Application so far mainly on agricultural sites and mineral soils
- First examples for analysis of forest soils (Ludwig et. Ak. 2017)
- Usage of organic layers (Cardelli et. al. 2017)

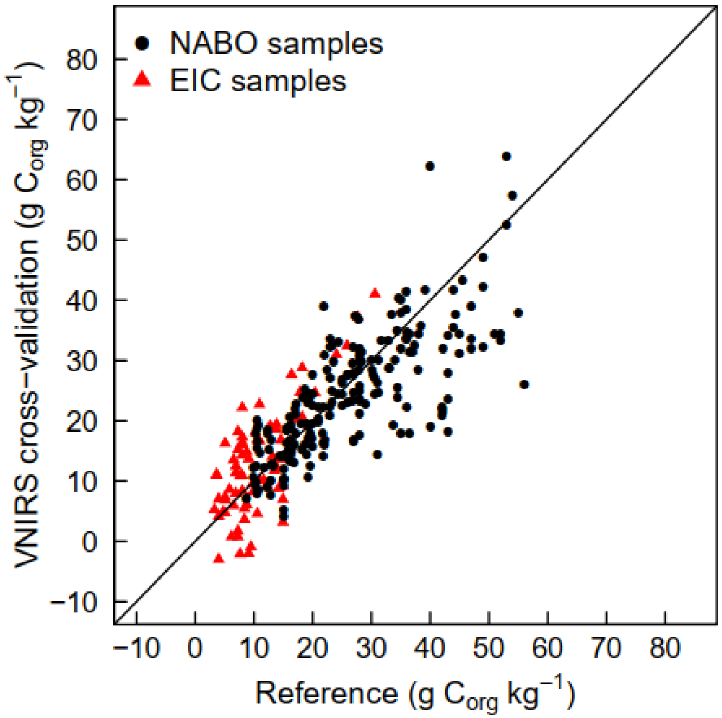


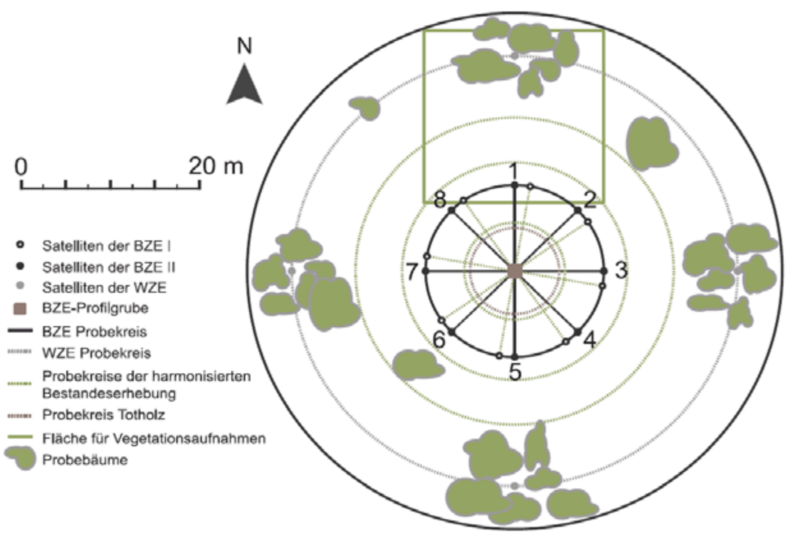
Image:Andreas Gubler

Aim

- Proximal soil sensing as support for forest monitoring
- Cooperation: Public Enterprise Sachsenforst and UFZ
- Analysis of forest soil samples with vis-NIR spectroscopy and setup of a spectral library
- Modelling of soil parameters
- Development of a fast, cost-efficient and applicable procedure for periodic forest soil mapping as supplement for current methods

Data Collection

- 378 retained samples (National Forest Soil Inventory and samples from test sites) from Saxony
- 109 additional samples from field campaign 2019
- Satellite sampling scheme (8 satellites per point, Ah/Oh horizon)



Images: Staatsbetrieb Sachsenforst

Data Collection

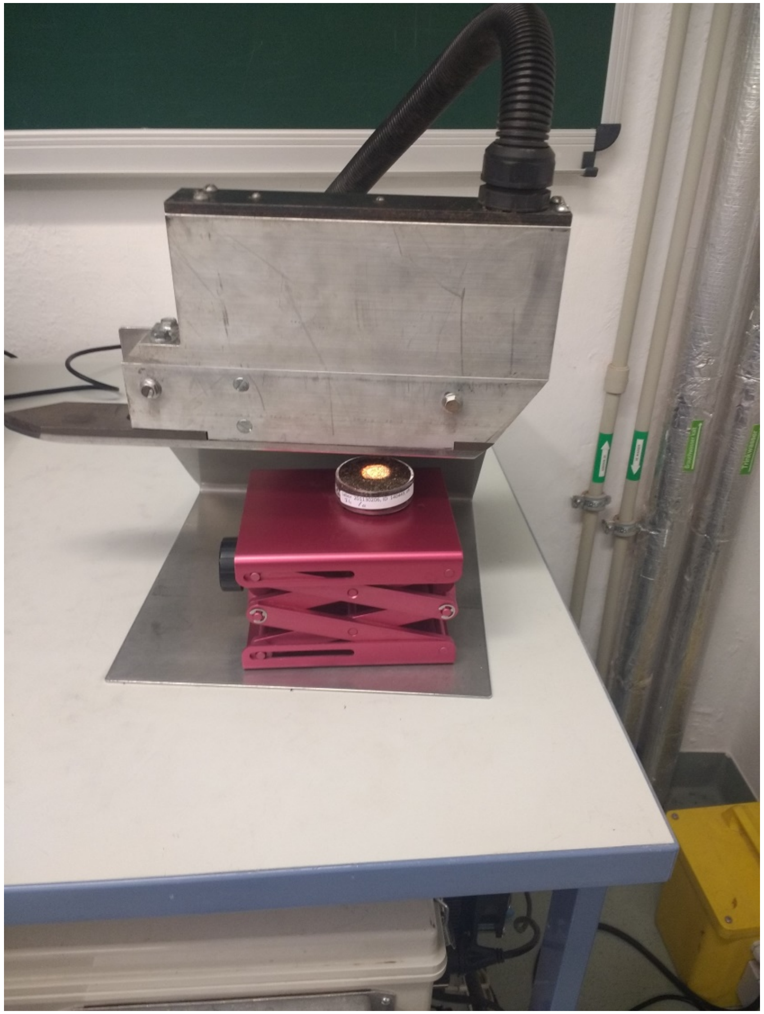
Spectral data

Protocol: Capture variability of ech sample

- Veris VIS-NIR Spectrophotometer
- Per samples five measurements from two petri dishes (ten in total)
- turn/shift der sample after each run
- Taking external references before each sample

Preprocessing

- Removing Wavelengths
- Savitzky-Golay Filter
- Standard Normal Variate
- Outlier detection



UFZ

Regression Analysis

Algorithms

- Partial Least Square Regression (PLSR)
- Suport Vector Machine (SVM)

Target Variables

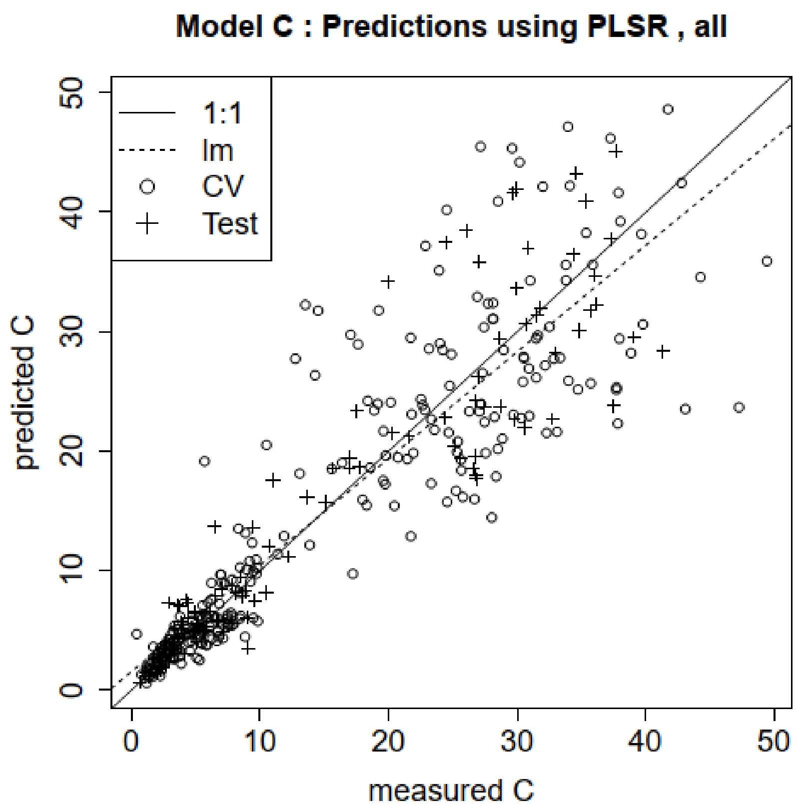
- C, N, pH-value, base saturation
- Usage of logarithmic values
- Separation of the data according to soil horizons and forest stand

Validation

- Independent test set (train/test split 70/30)
- 10-fold cross validation (cv)
- Plots (predicted/observed), RMSE, R^2
- New data from 2019 as independet test data set

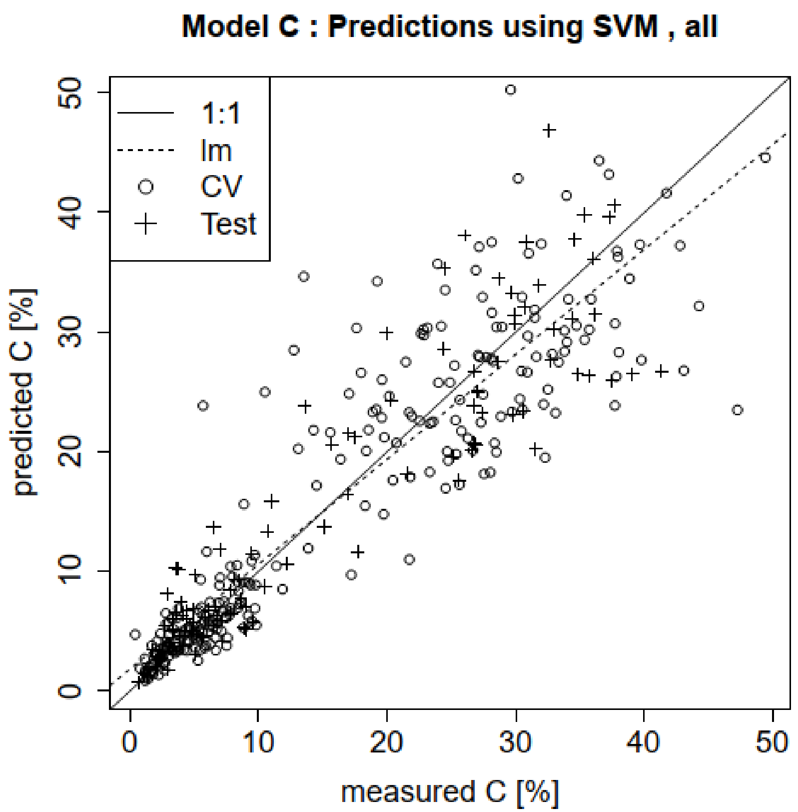
Selected Results: Carbon

Prediction results for carbon content based on retained samples for both algorithms



R^2: 0.8 (cv), 0.8 (test)

RMSE: 5.8 (cv), 5.9 (test)

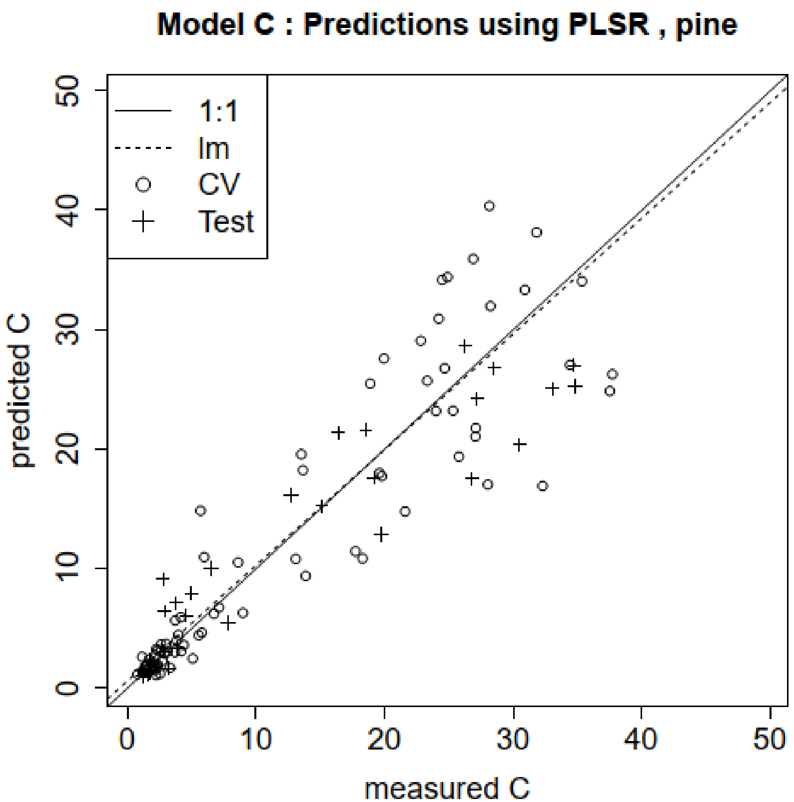


R^2: 0.82 (cv), 0.85 (test)

RMSE: 5.4 (cv), 4.9 (test)

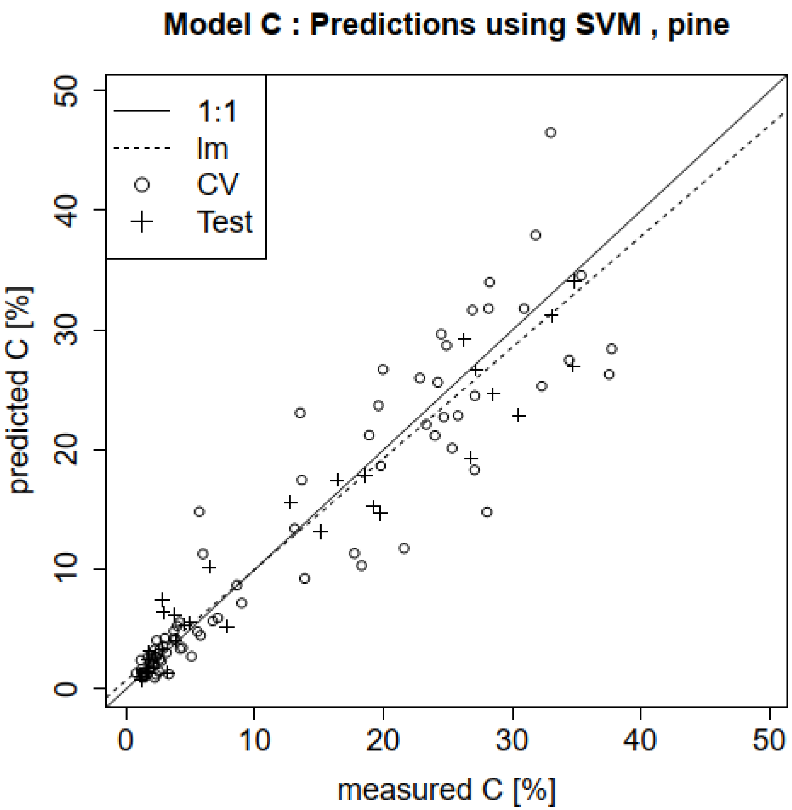
Selected Results: Carbon

Prediction results for carbon content based on retained samples for both algorithms in **pine stands**



R^2: 0.84 (cv), 0.87 (test)

RMSE: 4.9 (cv), 4.4 (test)

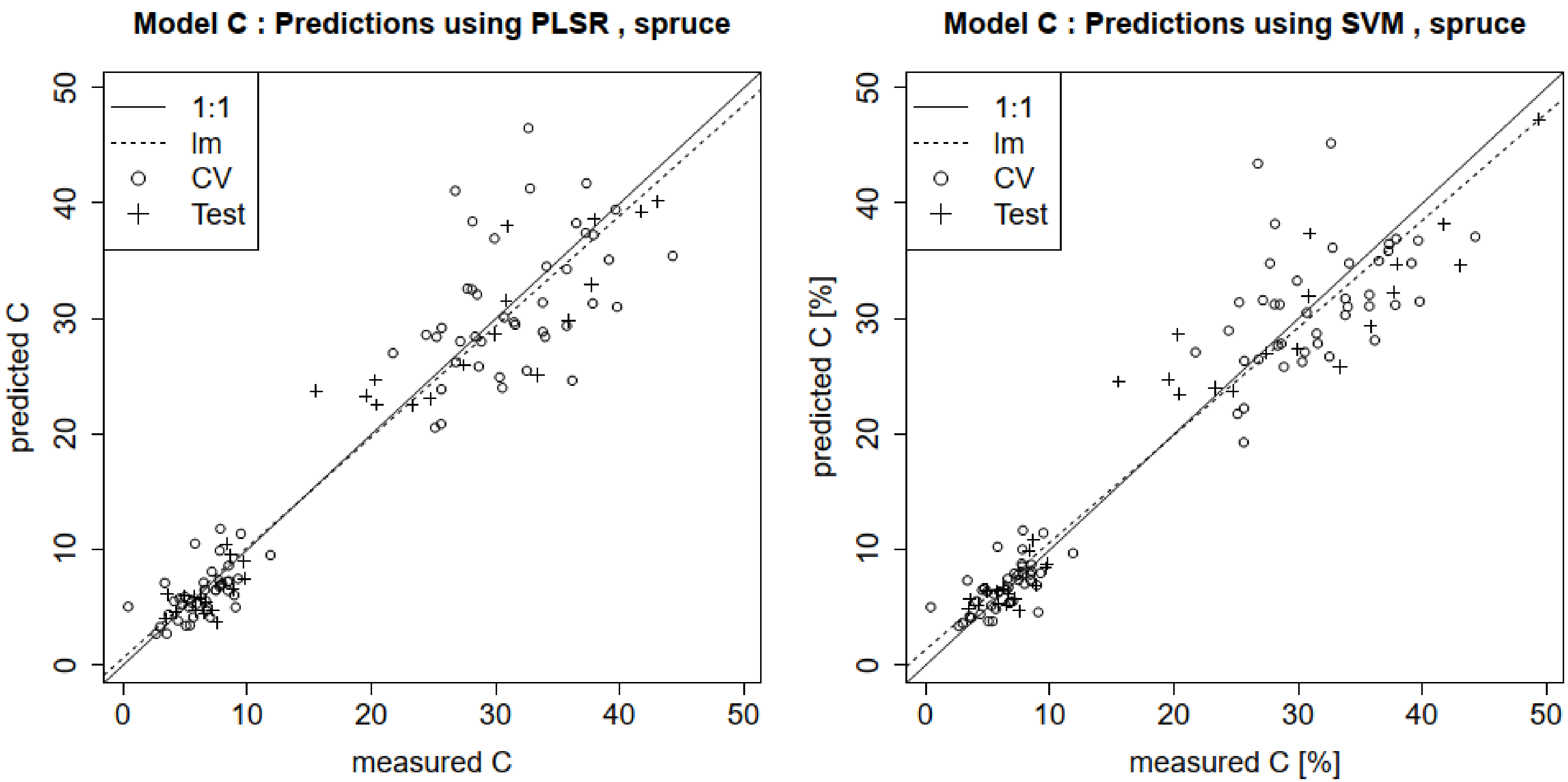


R^2: 0.86 (cv), 0.94 (test)

RMSE: 4.4 (cv), 3.2 (test)

Selected Results: Carbon

Prediction results for carbon content based on retained samples for both algorithms in **spruce stands**

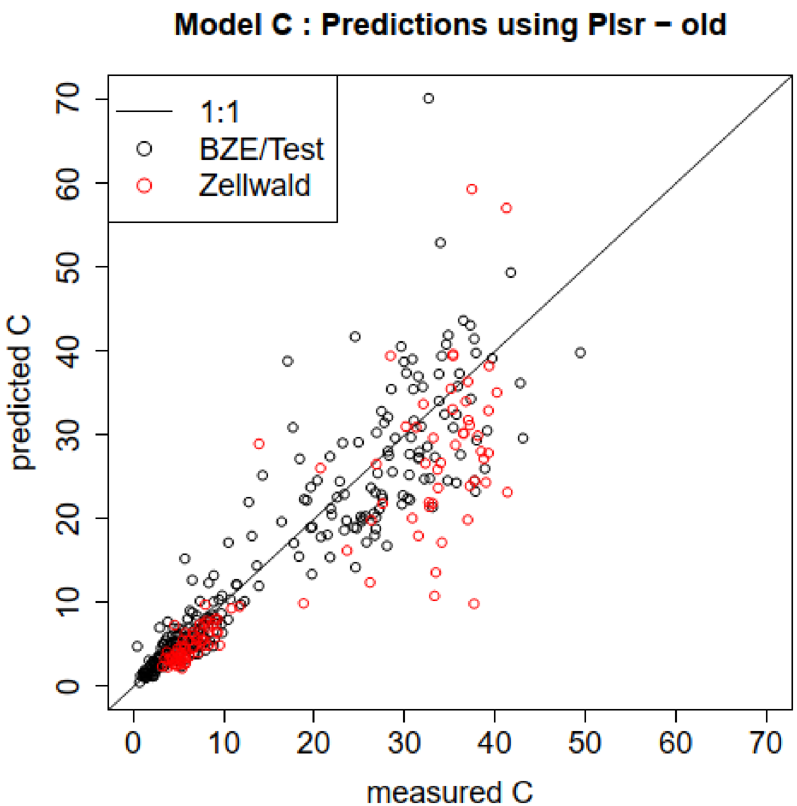


R^2: 0.91 (cv), 0.9 (test)
RMSE: 3.8 (cv), 4.8 (test)

R^2: 0.91 (cv), 0.92 (test)
RMSE: 4 (cv), 3.9 (test)

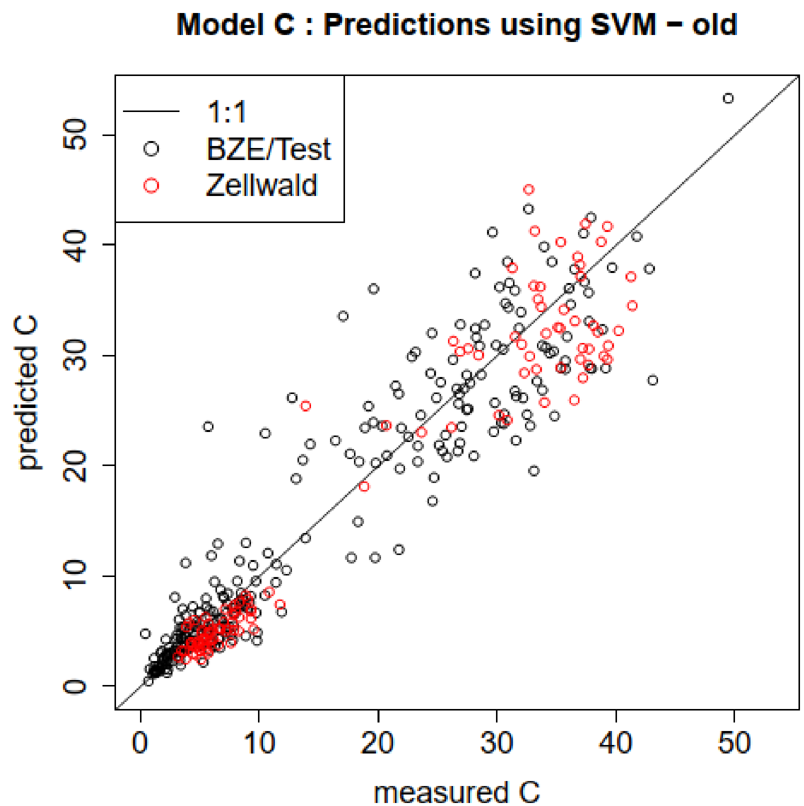
Selected Results: Carbon

Prediction results for carbon content based on retained samples with independent predictions for new data (red), for both algorithms



R²: 0.77

RMSE: 7.7

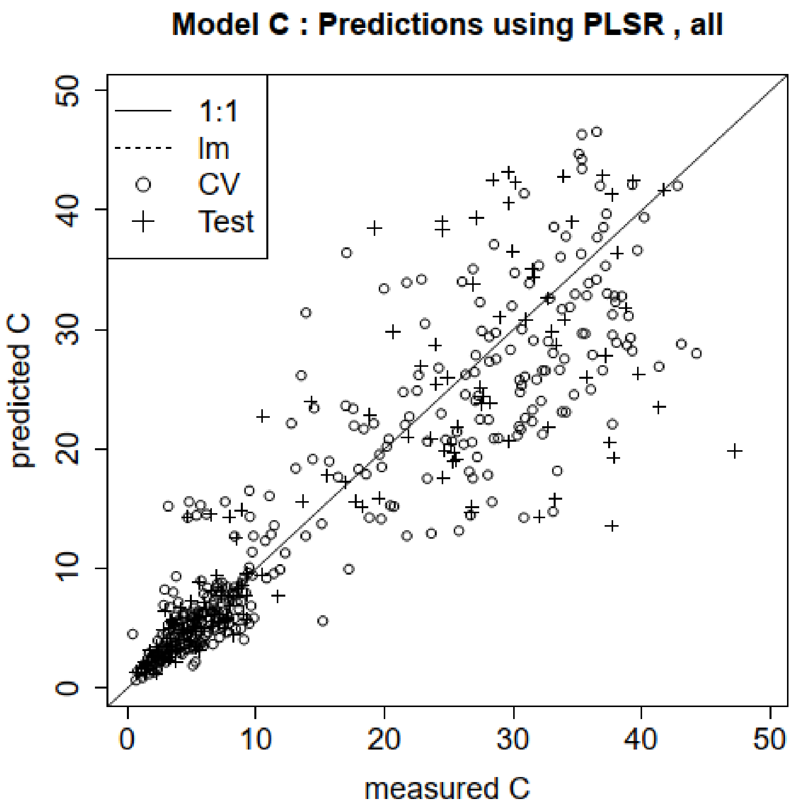


R²: 0.92

RMSE: 4.1

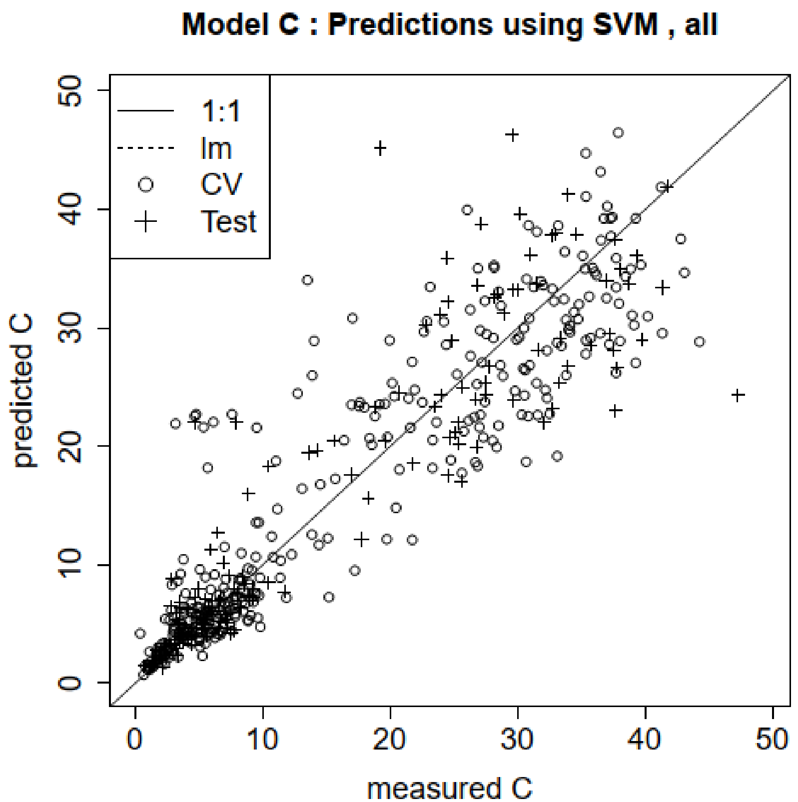
Selected Results: Carbon

Prediction results for carbon content based on retained and new samples for both algorithms



R²: 0.78

RMSE: 6.25



R²: 0.84

RMSE: 5.23

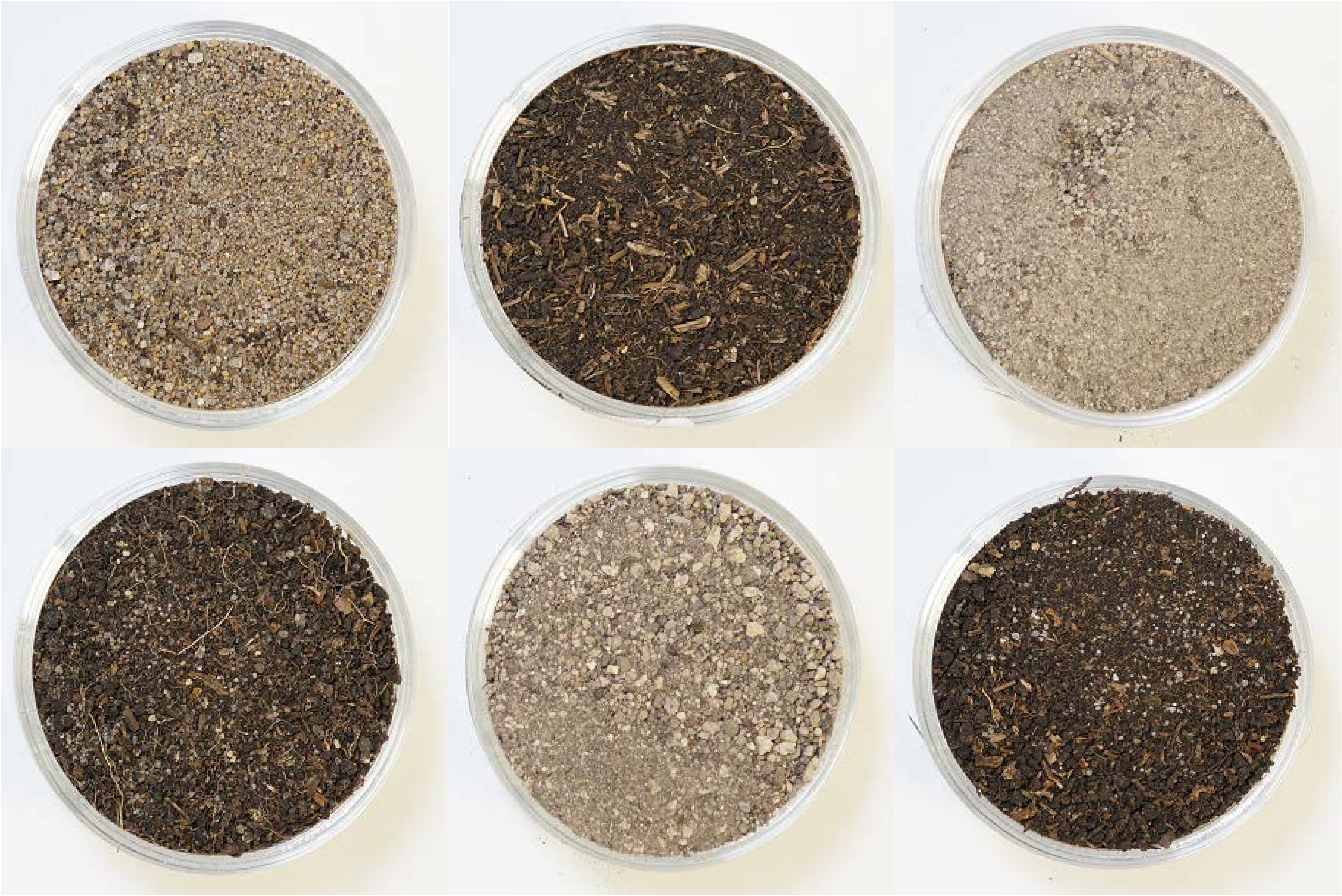
UFZ

Selected Results: Carbon

- R^2 values range from 0.25 to 0.94
- RMSE ranges from 1.78 to 7.69
- SVM produces more accurate predictions
- Separating horizons results in lower performance
- Increase in prediction accuracy for spruce and pine stands
- Models calibrated on retained samples are able to predict on new data from 2019
- Combining retained samples and new data produces results comparable to modelling based on retained samples only
- Model are robust to expansion of data basis

UFZ

Thank you for your attention!



UFZ

References

Wold, S.; Martens, H. & Wold, H.
The multivariate calibration problem in chemistry solved by the PLS method
Matrix pencils, Springer, 1983, 286-293

Gubler, A. (2011), Quantitative Estimations of Soil Properties by Visible and Near Infrared Spectroscopy: Applications for Laboratory and Field Measurements, PhD thesis.

Kooistra, L., Wehrens, R., Leuven, R. & Buydens, L. (2001), 'Possibilities of visible{near-infrared spectroscopy for the assessment of soil contamination in river floodplains', *Analytica Chimica Acta* 446(1-2), 97-105.

Ludwig, B., Vormstein, S., Niebuhr, J., Heinze, S., Marschner, B. & Vohland, M. (2017), 'Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects', *Geoderma* 288, 37-46.

Cardelli, V.; Weindorf, D. C.; Chakraborty, S.; Li, B.; De Feudis, M.; Cocco, S.; Agnelli, A.; Choudhury, A.; Ray, D. P. & Corti, G.
Non-saturated soil organic horizon characterization via advanced proximal sensors
Geoderma, Elsevier, 2017*, 288*, 130-142

Riedel, F., Denk, M., Müller, I., Barth, N. & Gläßer, C. (2018), 'Prediction of soil parameters using the spectral range between 350 and 15,000 nm: A case study based on the permanent soil monitoring program in saxony, germany', *Geoderma* 315, 188-198.

Müller, A. C.; Guido, S. & others
Introduction to machine learning with Python: a guide for data scientists
" O'Reilly Media, Inc.", 2016

Appendix

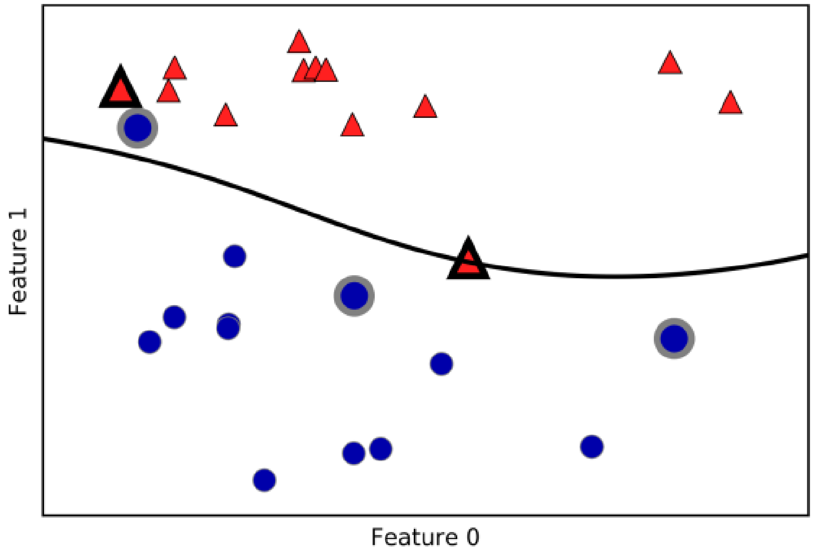
Partial Least Sqaure Regression

- Introduced by Wold et. Al. to solve multivariate calibration problems
- Extraction of components that describe the greates correlation between predictors and response variables
- Implementation of regressions based on these components instead of original data
- Highly reduces amount of data
- Is used when many collinear predictors are in use
- Reliable method to predict soil parameters

Appendix

Support Vector Machines (SVM)

- Representation of every object of the training data as a vector
- SVM creates a hyperplane that sepearates the training objects in two classes while maximising the distance to the adjacent vectors
- For the creation of the plane, only the adjacent vectors are required (support vectors, see picture)



Müller, et. al. (2016)

Appendix

Value Distribution for retained (red) and new samples (blue)