

EGU22-5507

<https://doi.org/10.5194/egusphere-egu22-5507>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Comparing prediction algorithms in FTIR-based chemometric analysis to predict soil variables within centuries-old charcoal rich Technosols

Shaghayegh Ramezany, **Alexander Bonhage**, and Thomas Raab

BTU Cottbus-Senftenberg, BTU Cottbus-Senftenberg, Geopedology and landscape development, Cottbus, Germany
(ramezsha@b-tu.de)

As an alternative to the costly wet chemical analysis of soil, Machine Learning (ML) algorithms can be applied for the quantification of soil properties through prediction models. In this study, we evaluate the performances of the Random Forest (RF) algorithm and Partial Least Square Regression (PLSR), in the prediction of soil variables, including CEC, pH, total contents of C and N as wells as other elements (Al, Fe, Ca, Mn, Mg, K, and Na) based on FTIR-spectra of Relict Charcoal Hearth (RCH) soils and reference forest soils (Non-RCH). We investigate the effect of high quantities of charcoal in the soil on the prediction models. Preliminary results suggest that there is no significant difference in the results of prediction models for total N, C, and Fe contents, while the accuracy of PLSR in the prediction of pH, Mg, and Ca, and RF performance in prediction of pH and C decreased for RCH soils. Both algorithms demonstrate higher accuracies in the prediction of Al within the RCH soils. Prediction of CEC, Na, K, Mg (RF) within the RCH soils, Al (PLSR) for Non-RCH soils, and Mn and Ca (RF) for both soil types resulted in lower quality of predictions.

It can be inferred from the results that the performance of FTIR-based prediction models can be affected by the presence of charcoal in soil due to the nature of spectral features reflecting the soil composition. The presence of charcoal in soil likely alters the absorption interference and peak overlaps, which can result in lower accuracy of the prediction models. In addition to the accuracy of the prediction models, we evaluate the reliability of the weighted wavenumbers (as important variables) in each prediction, which provides information about the correlation of spectral features and chemical properties. It can be studied through the Variable Importance Plot in RF and Variable Importance on Projection through PLSR (VIPs), which show high potential for studying soil composition and metal distribution in mineral and organic soil fractions despite the observed weaknesses in weighing wavenumbers in predictions. Therefore, we assess and compare the quality of information gained regarding soil chemical properties from the algorithms besides a sole quantification of soil parameters. Furthermore, we applied the developed prediction models on a large number ($n > 600$) of FTIR-spectra of RCH soils to investigate the practical application of the models and thereby compare spectral derived chemical properties of the studied Technosols and reference forest soils.

