EGU21-7851, updated on 01 Dec 2022
https://doi.org/10.5194/egusphere-egu21-7851
EGU General Assembly 2021


# The interplay among analytical method, preprocessing, and modeling on soil organic carbon Vis-NIR-SWIR predictions

**Taciara Zborowski Horst-Heinen**[1], Ricardo Simão Diniz Dalmolin[1], Alessandro Samuel-Rosa[2], and Sabine Grunwald[3]

[1]Federal University of Santa Maria, Department of Soil Science, Santa Maria, Brazil (tacihorst@gmail.com)
[2]Federal University of Technology – Paraná, Department of Agronomy, Santa Helena, Paraná, Brazil
[3]University of Florida, Soil and Water Sciences Department, Gainesville, Florida, United States

The relationship between visible-near-infrared (Vis-NIR-SWIR) spectra and soil organic carbon (SOC) and the effects of preprocessing techniques on SOC predictive models have been shown in several studies. However, little attention has been given to the effect of analytical methods used to produce the SOC data used to calibrate those models. The predictive performance of Vis-NIR spectral models depends not only on the preprocessing technique and machine learning method but also on the analytical method employed to produce the SOC data. Our hypothesis is that some combinations of preprocessing and models may be more sensitive to laboratory (measurement) error than others. To test this hypothesis, we evaluated the leave-one-out cross-validation performance of three predictive models (Random Forest (RF), Cubist, and Partial Least Square Regression (PLSR)) calibrated using SOC data produced via three analytical methods (dry combustion (DC) and wet combustion with quantification by titration (WCt) and colorimetry (WCc)) and three Vis-NIR spectra preprocessing techniques (smoothing (SMO), continuum removal (CRR), and Savitzky-Golay first derivative (SGD)). The prediction performance varied among the models. DC and WCt provided a higher correlation between SOC and spectra than WCc, and thus, resulted in higher accuracy. The Cubist+CRR was ranked the best performing model, with an average of R2 = 0.81 and RMSE = 0.81% among analytical methods. Cubist+CRR also minimized the accuracy differences resulting from SOC analytical methods employed. The RF model had low accuracy and was unable to explain more than 46% of the variance. Overall, the analytical method significantly affects SOC predictions, and its impact may be larger than the preprocessing.