Soil Organic Carbon Prediction at National Scale (Germany)

Ali Sakhaee¹, Anika Gebauer², Mareike Ließ², and Axel Don¹

¹Thünen-Institut, Agrarklimaschutz, Braunschweig, Germany (ali.sakhaee@thuenen.de)
²Helmholtz Centre for Environmental Research - UFZ, Halle (Saale), Germany

Soil Organic Carbon (SOC) plays a crucial role in agricultural ecosystems. However, its abundance is spatially variable at different scales. In recent years, machine learning (ML) algorithms have become an important tool in the spatial prediction of SOC at regional to continental scales. Particularly in agricultural landscapes, the prediction of SOC is a challenging task.

In this study, our aim is to evaluate the capability of two ML algorithms (Random Forest and Boosted Regression Trees) for topsoil (0 to 30 cm) SOC prediction in soils under agricultural use at national scale for Germany. In order to build the models, 50 environmental covariates representing topography, climate factors, land use as well as soil properties were selected. The SOC data we used was from the German Agricultural Soil inventory (2947 sampling points). A nested 5-fold cross-validation was used for model tuning and evaluation. Hyperparameter tuning for both ML algorithms was done by differential evolution optimization.

This approach allows exploring an extensive set of field data in combination with state of the art pedometric tools. With a strict validation scheme, the geospatial-model performance was assessed. Current results indicate that the spatial SOC variation is to a minor extent predictable with the considered covariate data (<30% explained variance). This may partly be explained by a non-steady state of SOC content in agricultural soils with environmental drivers. We discuss the challenges of geo-spatial modelling and the value of ML algorithms in pedometrics.