EGU22-8560, updated on 11 Aug 2022
https://doi.org/10.5194/egusphere-egu22-8560
EGU General Assembly 2022

Assessing machine-learning algorithms for digital soil mapping in an agricultural lowland area: a case study of Lombardy region.

Odunayo David Adeniyi[1,2], Alexander Brenning[2], and Michael Maerker[1]
[1]Department of Earth and Environmental sciences, University of Pavia, Pavia, Italy
(odunayodavid.adeniyi01@universitadipavia.it)
[2]Department of Geography, Friedrich Schiller University Jena, Jena, Germany

Sustainable agricultural landscape management needs reliable and accurate soil maps and updated geospatial soil information. The traditional process of soil surveying is time-consuming and limited in terms of accuracy and spatial distribution. This problem can be partly overcome by Geographical Information Systems (GIS) and the application of digital soil mapping (DSM) approaches. The DSM analyse the relations between soil properties and environmental variables derived from the Digital Elevation Models (DEM) as well as from remotely-sensed information. Moreover, DSM uses these relations and hence, allows for the regionalization of point observations of soil properties. Several machine-learning methods are used today for DSM. The main goal of this study is the evaluation of different supervised machine-learning techniques for mapping several topsoil properties in an agricultural lowland area of Lombardy region, Italy, and interpreting the modelled relationships. The methods analysed are Random Forest, Gradient Boosting Machine, Support Vector Machine and Generalized Additive Model. We applied the models to predict different correlated soil properties such as the soil organic carbon (SOC), texture (sand, silt, clay content) and topsoil depth. Cross validation performances of these models were determined, and diagnostic tools for the post-hoc interpretation of these black-box models were applied to assess their interpretability as well as similarities and differences in the modelled relationships, which reflect each model's abilities and biases. An important challenge is the interpretation of the effects of highly correlated predictors, which is achieved using a transformation-based post-hoc interpretation technique. The study helps to identify the best-performing predictive model for lowland area and to understand the robustness of the applied models. The selected models will be used to provide valuable information for facilitating a sustainable land use in an area with a unique soil water cycle as well as for assessing how future climate and socioeconomic changes may influence water content, soil pollution dynamics and food security.