Prediction of topsoil organic carbon with Vis-NIR reflectance spectroscopy at the European scale: the LUCAS spectral library

A. Stevens (1), M. Nocita (1,2), P. Panagos (2), G. Tóth (2), B. van Wesemael (1), and L. Montanarella (2)

(1) Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, UCLouvain, Place Louis Pasteur, 3, 1348 Louvain-la-Neuve, Belgium (antoine.stevens@uclouvain.be), (2) Land Management and Natural Hazards Unit, Institute for Environment and Sustainability, Joint Research Centre of the European Commission, 21027 Ispra (VA), Italy

In the framework of an EU survey called *Land Use/Cover Area frame Statistical Survey* (LUCAS), $\sim 20,000$ soil samples have been collected in order to assess the state of the soil across Europe, under the supervision of the Joint Research Centre of the European Commission (JRC). Chemical (organic carbon, nitrogen, carbonate, etc), physical (particle size fraction) and spectral (visible and near-infrared diffuse reflectance - VNIRS) characteristics of the samples have been analyzed in the same laboratory. Soil spectra provide an integrative measure of soil properties and can be exploited to infer soil properties using multivariate statistical models developed on spectral libraries. While the ability of local scale spectral library to predict soil properties in homogeneous environments has been repeatedly demonstrated, there are still high uncertainties on the potential of VNIRS to produce useful soil predictions in large and diverse environments. Absorption features of several soil chromophores are not entirely unique, which can induce a mutual masking, shifting and distortion of absorption features when the soil composition is changing. Therefore, the relationship between soil properties and spectral data is likely to vary in space. Such problem is of high importance for the remote sensing of soils since it is used by its very nature for the monitoring/mapping of large areas. The LUCAS spectral library represents an exceptional opportunity to explore this issue and estimate the level of accuracy of VNIRS calibration models at large scales.

Our objective was to assess the performance of VNIRS for the prediction of soil organic carbon (OC) content at European scale. We tested several linear and non-linear models to relate spectral data with OC: partial least square regression, sparse partial least square regression, boosted regression tree, random forest, support vector machine, multivariate adaptive regression splines. The best model achieved a RMSE of $\sim 7\ g\ C\ kg^{-1}$ for mineral soils and $\sim 63\ g\ C\ kg^{-1}$ for organic soils. This level of accuracy is still 3-4 times larger than the error of repeatability of the laboratory method (SEL = $\sim 1.7\ g\ C\ kg^{-1}$ computed with 28 duplicates) for mineral soils and 6 times larger for organic soils (SEL = $\sim 10\ g\ C\ kg^{-1}$ computed with 4 duplicates). This quantitative analysis of LUCAS soil spectral library demonstrated the difficulty to produce spectroscopic models of OC content at the continental-scale that are sufficiently accurate to be useful for most applications at fine scales (e.g. spatio-temporal monitoring of individual locations, biogeochemical modeling). We show however some reasons for these difficulties and ways for improvement using additional covariates in the multivariate modeling.