



## Analyzing the prediction error of large scale Vis-NIR spectroscopic models

Antoine Stevens (1), Marco Nocita (1,2), Luca Montanarella (2), and Bas van Wesemael (1)

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

Based on the LUCAS soil spectral library ( $\sim 20,000$  samples distributed over 23 EU countries), we developed multivariate calibration models (model trees) for estimating the SOC content from the visible and near infrared reflectance (Vis-NIR) spectra. The root mean square error of validation of these models ranged from 4 to 15 g C kg<sup>-1</sup>. The prediction accuracy is usually negatively related to samples heterogeneity in a given library, so that large scale databases typically demonstrate low prediction accuracy compared to local scale studies. This is inherent to the empirical nature of the approach that cannot accommodate well the changing and scale-dependent relationship between Vis-NIR spectra and soil properties. In our study, we analyzed the effect of key soil properties and environmental covariates (land cover) on the SOC prediction accuracy of the spectroscopic models. It is shown that mineralogy as well as soil texture have large impacts on prediction accuracy and that pedogenetic factors that are easily obtainable if the samples are geo-referenced can be used as input in the spectroscopic models to improve model accuracies.