Geophysical Research Abstracts
Vol. 21, EGU2019-7295, 2019
EGU General Assembly 2019

Soil sampling strategies based on multispectral Sentinel-2 and hyperspectral EnMAP satellite data

Bas van Wesemael (1), Fabio Castaldi (1), and Sabine Chabrillat (2)

(1) Université catholique de Louvain, Earth and Life Institute, Louvain-la-Neuve, Belgium (bas.vanwesemael@uclouvain.be), (2) Helmholtz-Zentrum Potsdam—Deutsches GeoForschungs Zentrum GFZ, 14473 Potsdam, Germany

The sampling strategy for mapping soil properties from remote sensing imagery entails making decisions about sampling pattern, size and location. The availability of a consistent number of ancillary data strongly related to the target variable allows applying sampling strategies that optimally cover the feature space. This study aims at evaluating the capability of multispectral (Sentinel-2) and hyperspectral (EnMAP) satellite data to select the sampling locations in order to collect a calibration dataset as basis for multivariate statistical modelling of the Soil Organic Carbon (SOC) content. Remote sensing spectra can be exploited first to set the sampling strategy and then as independent variables for the prediction models of the target variables. We tested different sampling strategies based on the feature space, where the ancillary data are the spectral bands of the Sentinel-2 and of simulated EnMAP satellite data acquired in Demmin (north-Est Germany). Some selection algorithms require to set the number of samples in advance (random, Kennard-Stones and conditioned Latin Hypercube algorithm) and others automatically provide the ideal number of sampling units (Puchwein, SELECT and Puchwein+SELECT algorithm). The SOC content and the spectra extracted at the sampling locations were used to build random forest (RF) models. The accuracy of the RF estimation models was evaluated on an independent dataset. The highest Sentinel-2 ratio of performance to deviation (RPD) for the validation set was obtained using Puchwein (RPD: 2.5), and Kennard-Stones (RPD: 2.4) algorithm. A strong positive correlation was detected between the standard deviation of the calibration dataset and the validation accuracy. The efficiency of the sampling strategies, as ratio between accuracy and number of samples per hectare, is highest using Puchwein with EnMAP and Puchwein+SELECT algorithm with Sentinel-2 data. The achieved results demonstrated that Sentinel-2 and EnMAP data can be exploited to build a reliable calibration dataset for SOC mapping; moreover the efficiency of the sampling strategy selection can be improved using algorithms that provide the number of sampling units. For EnMAP, the different selection algorithms provided very similar results. On the other hand, using Puchwein and Kennard-Stones algorithms, Sentinel-2 provided a more accurate estimation than the EnMAP. The calibration datasets provided by EnMAP data provided in this case lower SOC variability and lower prediction accuracy than compared to Sentinel-2. This was probably due to EnMAP coarser spatial resolution (30 m) less adequate for linkage to the sampling performed at 10 m scale.