Geophysical Research Abstracts
Vol. 21, EGU2019-9045, 2019
EGU General Assembly 2019

Mapping Soil Organic Matter in the arable land of Europe with machine learning algorithms

Luís M. de Sousa (1), Gerard B.M. Heuvelink (1), Maria Ruiperez (1), Laura Poggio (1), Bas Kempen (1), and Erik van den Berg (2)

(1) ISRIC – World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands, (2) Wageningen Environmental Research, P.O. Box 47, 6700 AA Wageningen, The Netherlands

The evaluation of the environmental impacts of pesticide use is important for sustainable soil management. This includes the modelling of leaching of pesticides into the groundwater, pesticide drainage to surface waters and persistence of pesticides in top soils. PEARL and GeoPEARL are two models developed by Wageningen Environmental Research and its partners, enabling the evaluation of pesticide use policies and supporting pesticide registration procedures in Europe. This presentation reports the elaboration of a Soil Organic Matter map for the arable land of Europe, an important input to the PEARL and GeoPEARL models. Soil organic matter influences the sorption and transformation of pesticides in the soil and thus has a strong effect on pesticide leaching.

This Soil Organic Matter map was created with machine learning algorithms, with predictions at five different depth intervals. The algorithms random forests and gradient boosting were trained with the most recent available observations in Europe, gathered from more than 20 000 locations and harmonised into a single database. A set of over 200 global environmental covariates was also employed in training, primarily derived from remote sensing such as MODIS land products, but also including SRTM DEM derivatives, climatic variables and global landform and lithology maps. The prediction algorithms were assessed using a 10-fold spatially stratified cross-validation procedure. The spatial stratification was obtained by binning sampling locations according to a discrete global grid, allowing for a more realistic assessment of the prediction accuracy vis à vis the standard 10-fold cross-validation.

Implementation was done in the R language for statistical computing, with the resulting maps derived at a spatial resolution of 250 meters. This presentation further details how big data and numerical complexity issues were solved using high performance computing and tiling methods.