Geophysical Research Abstracts
Vol. 18, EGU2016-9342, 2016
EGU General Assembly 2016

Understanding world soils: Machine Learning as a framework for analyzing global soil-landscape relationships

Tomislav Hengl and Jorge Mendes de Jesus
ISRIC — World Soil Information, Wageningen, Netherlands (tom.hengl@isric.org)

Soil information is an increasingly important input to global geochemical modelling, hydrological modelling, spatial planning and agricultural extension. Soil remains one of the least developed environmental layers globally with data available only at coarse resolutions and with limited accuracy. In 2013/2014 ISRIC — World Soil Information has released a Global Soil Information system (SoilGrids1km) and an app to serve 3D soil information globally in near real time (DOI: 10.1371/journal.pone.0105992)). At the time, this system was a proof of concept demonstrating that global compilations of soil profiles can be used in an automated framework to produce complete and consistent spatial predictions of soil properties and classes. It was primarily been based on linear statistical modelling, which resulted in a limited fitting success. Global models fit to large, noisy data, can often result in significant oversmoothing of the measured variation.

In year 2015, focus of the SoilGrids project has shifted towards improving data quality primarily considering of spatial detail and attribute accuracy. Initial testing using African soil data (DOI: 10.1371/journal.pone.0125814) has shown that the key to improving accuracy might lay in using Machine learning techniques such as random forests, neural networks and similar that are able to better represent complex, often non-linear soil-landscape relationships. In 2015 we have fitted machine learning using larger global compilations of soil profiles (about 150,000 points) and covariates at 250 m spatial resolution (about 150 covariates; mainly MODIS seasonal land products, SRTM DEM derivatives, climatic images, lithological and land cover and landform maps) and extracted more significant global soil-landscape relationships (R-square ranging from 0.42 to 0.83). Our results show that the key predictors for mapping soil classes are most commonly hydrological DEM parameters and climatic data; for soil texture fractions lithology and landform seem to be most important, while for chemical soil variables it is a combination of vegetation indices, DEM parameters and climatic images. This helps our understanding of the processes that drive soil variability and their respective scales. Building global models using such a diversity of covariates also helps us re-design soil mapping so that we can make it more efficient i.e. by putting more emphasis on remote sensing / DEM data that is the most efficient in explaining spatial patterns of soils.