SoilGrids: using big data solutions and machine learning algorithms for global soil mapping

De Sousa, L.M.[1], G.B.M. Heuvelink[1], N.H. Batjes[1], B. Kempen[1]

[1] *International Soil Reference and Information Centre, Wageningen University & Research, Droevendaalsesteeg 3, 6708PB Wageningen, the Netherlands*
Corresponding author's e-mail: luis.desousa@wur.nl

The SoilGrids system (www.soilgrids.org) uses machine learning algorithms to predict soil type and basic soil properties at seven depths on global extent. These algorithms (i.e., random forests, gradient boosting) are trained with soil observations assembled from 150 000 locations across the globe as stored in WoSIS (ISRIC's World Soil Information System including licences stipulated by the data providers) and datasets made available by data owners for Soilgrids specifically. These are used together with 200 global environmental covariates, mainly derived from MODIS land products, SRTM DEM derivatives, climatic images and global landform and lithology maps. The software implementation is done in the R language for statistical computing. Numerical complexity issues in Soilgrids were addressed by employing cloud computing and tiling. Currently the resulting maps are derived at a spatial resolution of 250 m and published under an Open Database License. The prediction accuracy of the maps is reported using cross-validation statistics. Prediction accuracy varies between soil properties and depths, with usually between 55 and 85 per cent of the total variance explained. Since ISRIC aims to continue applying state-of-the-art knowledge in data science and data management and incorporate newly available data, Soilgrids is continuously improved. Future developments of SoilGrids include the extension of the set of predicted soil properties, refining its spatial resolution, further improve overall prediction accuracy, as well as quantifying the spatial prediction accuracy by mapping prediction interval boundaries. Main uses of SoilGrids have been in Green Life Sciences such as Environmental Sciences, Ecology, Agriculture, Geology, Water Resources and Science Technology. The SoilGrids team is eager to connect with users to maximise the use of the product and tailor the product to specific user needs.