Geophysical Research Abstracts
Vol. 16, EGU2014-10774, 2014
EGU General Assembly 2014

Using random forests to explore the effects of site attributes and soil properties on near-saturated and saturated hydraulic conductivity

Helena Jorda (1,2), John Koestel (1), and Nicholas Jarvis (1)

(1) Swedish University of Agricultural Sciences, Soil and Environment, Uppsala, Sweden (john.koestel@slu.se), (2) KU Leuven, Department of Earth and Environmental Sciences, Division of Soil and Water Management, Celestijnenlaan 200e, P. O. Box 2411, 3001 Leuven, BELGIUM

Knowledge of the near-saturated and saturated hydraulic conductivity of soil is fundamental for understanding important processes like groundwater contamination risks or runoff and soil erosion. Hydraulic conductivities are however difficult and time-consuming to determine by direct measurements, especially at the field scale or larger. So far, pedotransfer functions do not offer an especially reliable alternative since published approaches exhibit poor prediction performances. In our study we aimed at building pedotransfer functions by growing random forests (a statistical learning approach) on 486 datasets from the meta-database on tension-disk infiltrometer measurements collected from peer-reviewed literature and recently presented by Jarvis et al. (2013, Influence of soil, land use and climatic factors on the hydraulic conductivity of soil. Hydrol. Earth Syst. Sci. 17(12), 5185-5195).

When some data from a specific source publication were allowed to enter the training set whereas others were used for validation, the results of a 10-fold cross-validation showed reasonable coefficients of determination of 0.53 for hydraulic conductivity at 10 cm tension, K_{10}, and 0.41 for saturated conductivity, K_s. The estimated average annual temperature and precipitation at the site were the most important predictors for K_{10}, while bulk density and estimated average annual temperature were most important for K_s prediction. The soil organic carbon content and the diameter of the disk infiltrometer were also important for the prediction of both K_{10} and K_s.

However, coefficients of determination were around zero when all datasets of a specific source publication were excluded from the training set and exclusively used for validation. This may indicate experimenter bias, or that better predictors have to be found or that a larger dataset has to be used to infer meaningful pedotransfer functions for saturated and near-saturated hydraulic conductivities. More research is in progress to further elucidate this question.