

EGU22-4032, updated on 13 Aug 2022

<https://doi.org/10.5194/egusphere-egu22-4032>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Building saturated hydraulic conductivity maps with machine learning and geostatistics

Héctor Aguilera, **Carolina Guardiola-Albert**, Luis Moreno Merino, Carlos Baquedano, Elisabeth Díaz-Losada, Pedro Agustín Robledo Ardila, Almudena de la Losa, and Juan José Durán Valsero
Geological Survey of Spain (IGME-CSIC), Madrid, Spain

Hydraulic conductivity (K_s) is one of the most challenging, time-consuming, and expensive soil hydraulic properties to estimate. Pedotransfer functions (PTFs) of general use for K_s estimation are often site and sample-scale specific and perform poorly when extrapolated to different regions and extents. The present work develops a stepwise methodology for topsoil K_s mapping at a catchment scale based on easy, fast, and inexpensive measurements and auxiliary data (Aguilera et al., 2022). It includes a double-scale sampling of the K_s to account for small-scale variability in the spatial geostatistical interpolation. A supervised selection of variables through correlation analysis and hierarchical clustering of variables precedes the development of site-specific PTFs with machine learning (ML) techniques. The ML model is then used to generate new K_s point data predictions to extend the spatial coverage for mapping. Finally, the consistency of the final K_s map is assessed in terms of a geomorphological base map.

The variable selection process filtered out four predictor variables from the initial pool of fourteen predictors. An artificial neural network (ANN) provided the best K_s prediction model with one hidden layer and six input variables (latitude, longitude, silt, clay, medium sand, and land use). Latitude and longitude coordinates and land use are surrogates for other physical and environmental (e.g., anthropic) factors. The relative importance of input variables in the ANN was determined as the sum of the product of raw input-hidden, hidden-output connection weights across all hidden neurons using Olden's algorithm. Longitude, percentage of clay, and percentage of medium sand presented a stronger positive relationship with K_s , while irrigated and dry land uses together with the percent of silt were the variables with a more significant negative influence on K_s . The fact that K_s was positively related to clay content is surprising, and it appears to be related to soil plowing before sampling.

The ANN was used to estimate new K_s values from a subsequent sampling of model covariates, which doubled the input information for spatial interpolation using ordinary kriging. Overall, the spatial distribution of K_s was consistent with the lithological variability and other superimposed anthropic factors, as the method adequately considers the spatial variability of K_s added by anthropization to the already high natural heterogeneity. The produced maps will help in the hydraulic planning and flood risk management in the study area where high and low K_s , respectively, are clearly outlined.

Reference:

- Aguilera, C. Guardiola-Albert, L. Moreno Merino, C. Baquedano, E. Díaz-Losada, P. Agustín Robledo Ardila, J.J. Durán Valsero, Building inexpensive topsoil saturated hydraulic conductivity maps for land planning based on machine learning and geostatistics, CATENA, Volume 208, 2022, 105788, <https://doi.org/10.1016/j.catena.2021.105788>.

This work is performed within the framework of the RESERVOIR project, part of the PRIMA Programme supported by the European Union. The PRIMA programme is supported under Horizon 2020 the European Union's Framework Programme for Research and Innovation. PRIMA RESERVOIR Grant Agreement number is 1924.