September 13,2024

Representative Sample Size for Estimating Saturated Hydraulic Conductivity via Machine Learning

Behzad Ghanbarian

Associate Professor, Department of Earth and Environmental Sciences, University of Texas at Arlington

Seminar Details

Friday, September 13, 2024 2:30pm – 4:00pm

UH Campus Classroom & Business Building Room CBB 104

Online via Teams https://www.cive.uh.edu/research/beyer-distinguished-lecture

ABSTRACT: Although machine learning (ML) has been recently applied widely, not much attention has been paid to data heterogeneity and the number of samples required to train ML models in hydrology. In this study, we selected 17,990 soil samples from the USKSAT database and created random subsets N=2000, 4000, 6000, 8000, 10000, 12000, 14000, 16000, and 17990, 80% of which were algorithm to estimate saturated hydraulic conductivity (Ks) from bulk density, soil depth, texture, organic content, etc. For each subset, we conducted the learning curve analysis on the training and cross-validation datasets. Results showed that all training sample sizes the number of samples was not enough for the training and cross-validation curves to reach a plateau. We also applied that concept of representative elementary volume by plotting the average coefficient of determination (R2) and root mean square log-transformed error (RMSLE) against the training sample size. For testing dataset, as the number of training sample size increased from 1600 to 14392 the average R2 value in creased from 0.74 to 0.90, while the average RMSLE value decrease from 1.08 to 0.69. Either the learning curve or representative sample size analysis demonstrated that the number of samples was enough!

BIOGRAPHY: Dr. Behzad Ghanbarian is an Associate Professor at the Department of Earth and Environmental Sciences, University of Texas at Arlington. He is the author of 120 peer-reviewed journal articles and three books. His research interests center around a wide range of multidisciplinary topics, such as climate change, unconventional reservoirs, upscaling techniques, and fluid flow and contaminant transport in heterogeneous porous media. He is a member of AGI, SSSA, and SPE and received the 2015 Donald L. Turcotte Award in nonlinear geophysics from the American Geophysical Union as well as the 2020 Soil Physics and Hydrology Division Early Career Award from the Soil Science Society of America. Behzad was also listed among the top 2% of scientists in the world in 20221, 2022, and 2023. He also received the TWISS Graduate Teaching and Mentoring Award from Kansas State University in 2021 and 2022.