EGU25-8695, updated on 05 May 2025
https://doi.org/10.5194/egusphere-egu25-8695
EGU General Assembly 2025

Machine learning applications in hydrogeology, the case study of Lithuanian soils' hydraulic conductivity

Gintaras Žaržojus[1], Vytautas Samalavičius[1], Eveliina Kukka-Maaria Vanhala[2], Ieva Lekstutytė[1], Sonata Gadeikienė[1], and Saulius Gadeikis[1]

[1]Institute of Geosciences, Vilnius University, Vilnius, Lithuania (gintaras.zarzojus@gf.vu.lt)
[2]Independent researcher

The rapid advancement of artificial intelligence (AI) opens new opportunities across various scientific disciplines, including hydrogeology. AI-based methods, particularly machine learning (ML), are increasingly utilized to address complex, non-linear relationships in hydrogeological data, offering improved accuracy and efficiency over traditional approaches. This study is the first attempt to apply AI techniques to assess hydrogeological parameters (hydraulic conductivity (k)) in Lithuanian soils, aiming to compare the accuracy of traditional empirical formulas (EFs) and modern computational approaches.

Hydraulic conductivity (k) is a critical parameter for evaluating soil permeability and water movement in porous media, which is widely used in hydrogeological modelling, contaminant transport, and geotechnical design. This research compares the predictive performance of six ML algorithms (Elastic Net, Gradient Boosting Regressor, Huber Regressor, K-Neighbors Regressor, Multi-Layer Perceptron Regressor, Random Forest Regressor) with empirical formulas using a dataset of 282 unique soil samples. Grain size distributions and particle diameters were used as features (input parameters) for ML models to predict k values.

Statistical metrics reveal that ML models significantly outperformed EFs, achieving r-squared of 0.36–0.46, compared to 0.10–0.33 for EFs. However, some ML models displayed signs of overfitting, and performance varied depending on input feature combinations, with optimal models requiring 4–8 parameters. The study also highlights the limited size and diversity of the dataset as a key constraint, underscoring the need for a more extensive and diverse soil database for robust ML model development.

This pioneering effort demonstrates the potential of AI to enhance and revise geological research in Lithuania, suggesting that ML can provide more accurate and scalable solutions for other hydrogeological and engineering geology problems.