Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.31138

Prediction of nitrate contamination prediction using machine learning with groundwater vulnerability assessment factors

YEONKYEONG CHOI[1], KYUNG-SEOK KO[2] AND HYOWON AN[1]

[1]Korea Institute of Geoscience and Mineral Resources, University of Science and Technology
[2]Korea Institute of Geoscience and Mineral Resources

In recent years, interest in the safe use of groundwater has been increasing, leading to a surge in research utilizing machine learning techniques to assess groundwater contamination levels and potential pollution risks. Traditionally, the DRASTIC model has been widely used for groundwater vulnerability assessment. However, groundwater contamination varies due to diverse environmental factors, limiting the DRASTIC model's ability to accurately represent actual contamination levels, such as nitrate pollution.

In contrast, AI-based data-driven models, including machine learning and deep learning, offer the advantage of effectively analyzing the nonlinear and complex characteristics of groundwater contamination. Therefore, this study aims to predict groundwater nitrate contamination by adding key factors from the existing DRASTIC model and environmental factors that can influence groundwater contamination distribution (e.g., land use, distance from rivers, distance to faults, and aquifer classification).

To predict groundwater nitrate contamination, this study employs the Random Forest model, an ensemble learning technique. To prevent overfitting and optimize model performance, cross-validation and Bayesian optimization were applied. Additionally, the model's predictive capability was evaluated using various metrics, including R^2, RMSE, and MAE. The importance of input variables was analyzed to identify key environmental factors significantly affecting groundwater contamination in the study area.

The study area, Hongseong-gun, South Korea, is a representative agricultural region where farmland accounts for approximately 37% of the total area, and the average groundwater nitrate concentration reaches 40.7 mg/L. Over the past 15 years, this region has experienced a sharp increase in facility-based agriculture (approximately 72-fold) and a high concentration of livestock facilities, both of which contribute to the rise in groundwater nitrate contamination.

The study results demonstrated that the Random Forest model outperformed the DRASTIC model in predicting groundwater nitrate contamination with higher accuracy. In particular, among the various input factors, we were able to confirm that land use and distance to faults, which are highly relevant to subsurface inflow of contaminants, are environmental factors that have a large impact on groundwater nitrate generation.