EGU22-1945, updated on 19 Aug 2022
https://doi.org/10.5194/egusphere-egu22-1945
EGU General Assembly 2022

Predicting the risk of groundwater nitrate contamination using machine learning tools

Xin Huang[1], Menggui Jin[1], Xing Liang[1], Jingwen Su[2], and Bin Ma[1]
[1]China University of Geoscience, School of Environmental Studies, Hydrogeology, Wuhan, China (cughx2014@163.com)
[2]Nanjing Geological Survey Center, China Geological Survey, Nanjing 210016, China

Nitrate contamination in groundwater is affected by both anthropogenic activities and natural conditions, becoming one of the most prevalent problems worldwide. In this study, several machine learning methods including decision tree (DT), k nearest neighbors (KNN), logistic regression (LR), support vector machine (SVM), and extreme-gradient-boosted trees (Xgboost) were applied to predict the risk of groundwater nitrate contamination (NO3- > 50 mg L^{-1}) in the riverside areas of lower reaches of Yangtze River, east China. The developed model included 13 hydrochemical parameters (K^+, Na^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, NH_4^+, NO_2^-, Fe, Mn, As, Sr, pH) and well depth as explanatory variables, and a total of 1089 groundwater samples. The results showed the hydrochemical dataset could effectively predict the risk of nitrate contamination, with a minimum accuracy of 82.7% in LR and maximal accuracy of 91.7% in SVM and Xgboost. However, only the Xgboost model under a cutoff probability of 0.3 had the best performance with the highest sensitivity of 80.3% and AUC 0.95, whereas other models had sensitivity lower than 60% with insufficient capability of identifying contaminated groundwater samples. The results showed that the ensemble learning method had a strong, robust prediction capability. In addition, the relative importance of K^+, SO_4^{2-}, and Cl^- exceeded 0.65, indicating the dominant influence of domestic or industrial sewage in the study area due to widespread urbanization. Finally, we examined the relationship among nitrate contamination risk, land use type, the intensity of anthropogenic activities, and redox conditions and obtained the risk map of nitrate contamination in the study area. This study successfully proved the validity of predicting the risk of groundwater nitrate contamination using machine learning tools, which favors regional groundwater management and protection.

Keywords: groundwater; nitrate contamination; risk prediction; machine learning