

< Supporting Information >

Explainable machine learning for groundwater contamination by arsenic remobilization from a vadose zone

Tho Huu Huynh Tran^{a,b,†}, Sang Hyun Kim^{a,b,†}, Quynh Hoang Ngan Nguyen^{c,d}, Man Jae Kwon^e,
Jaeshik Chung^{a,b,*}, Seunghak Lee^{a,b,f*}

^a Water Cycle Research Center, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea

^b Division of Energy and Environment Technology, KIST School, Korea University of Science and Technology, Seoul 02792, Republic of Korea

^c Intelligence and Interaction Research Center, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea

^d Department of AI-Robotics, KIST school, Korea University of Science and Technology, Seoul 02792, Republic of Korea

^e Department of Earth and Environmental Science, Korea University, Seoul 02841, Republic of Korea

^f Graduate School of Energy and Environment (KU-KIST Green School), Korea University, Seoul 02841, Republic of Korea

[†]These authors contributed equally.

*Corresponding author.

Phone: +82-2-958-5816; E-mail: jschung@kist.re.kr; (J. Chung)

Phone: +82-2-958-5817; E-mail: seungleee@kist.re.kr; (S. Lee)

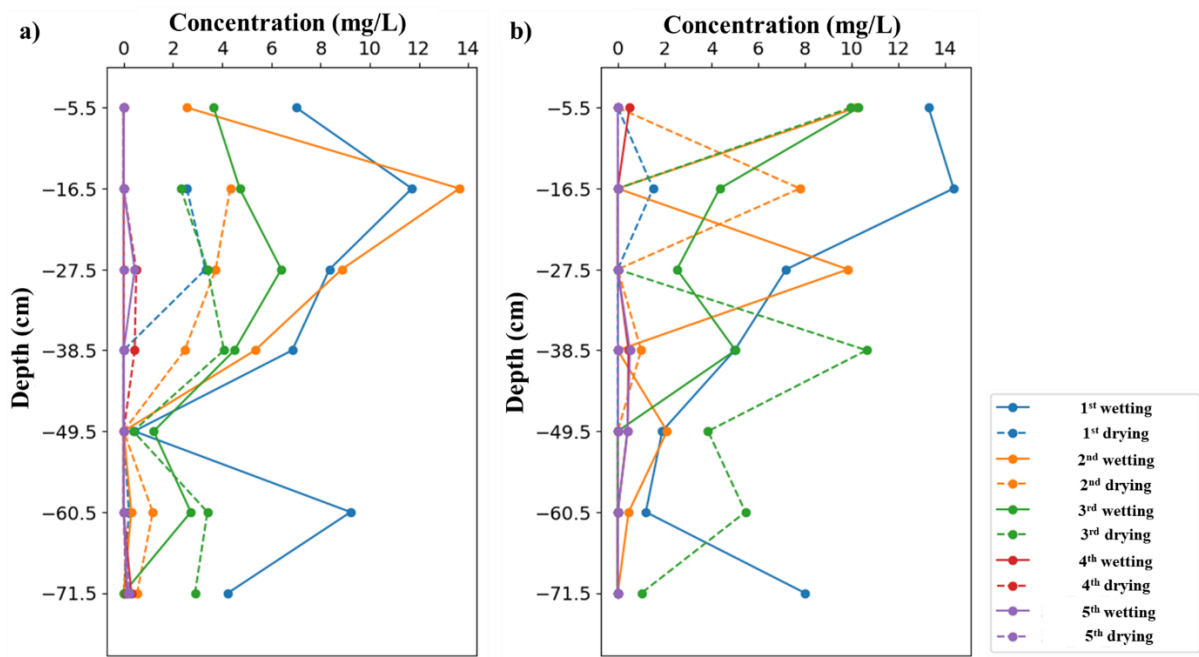


Figure S1: Pore water As concentrations along the depth in two soil columns from the preliminary study during the 1st to 5th cycle of wetting-drying.

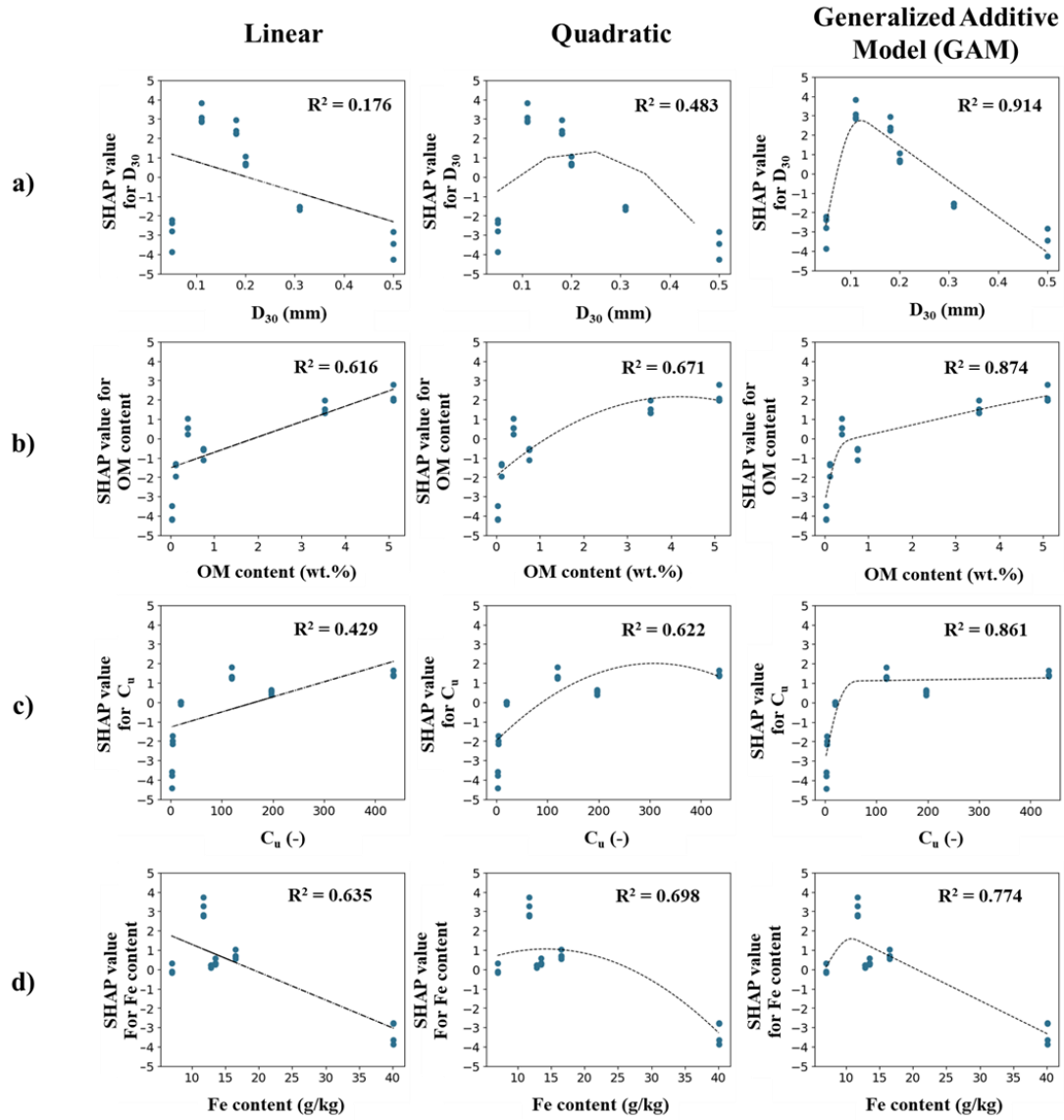


Figure S2: Fitting curves for Linear, Quadratic, and Generalized Additive Models (GAM) illustrating the relationship between Shapley Additive explanations (SHAP) values and soil properties. a) D_{30} (particle size finer than 30%), b) organic matter (OM) content, c) coefficient of uniformity (C_u), and d) iron (Fe) content. The fitting models are arranged in order: Linear, Quadratic, and GAM for each soil property. R^2 in the graphs refer to coefficient of determination of the fitting curves. SHAP values for bulk density did not show discernible trends, and thus were not presented here.

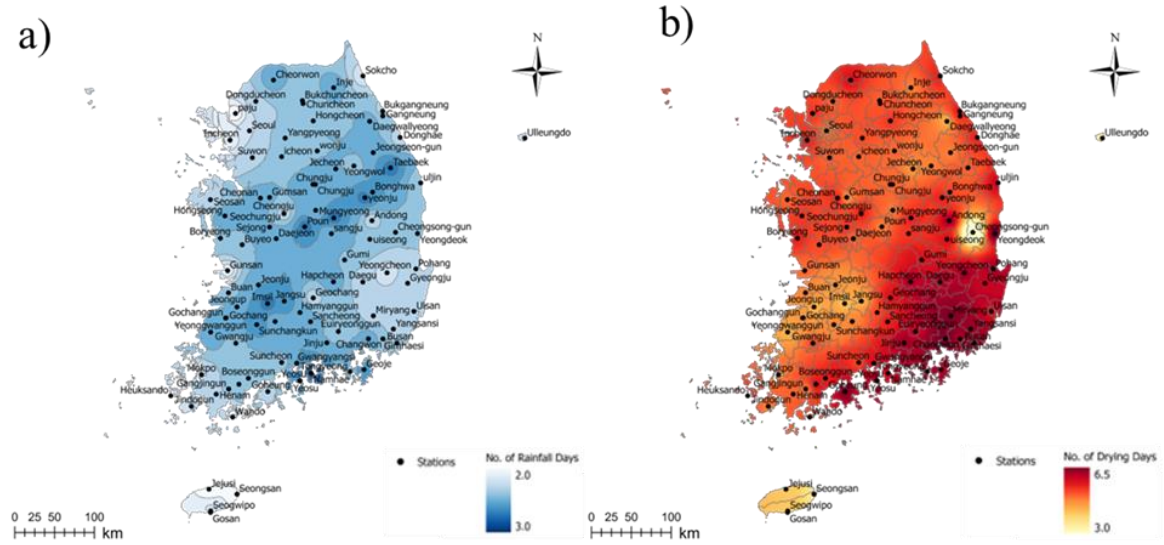


Figure S3: Average number of continuous a) rainfall, and b) drying days in South Korea.

Notes: data was collected from Korean Meteorological Administration.