Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.31978

Predictions of arsenic speciation in groundwater: using machine learning to optimize speciation sampling in low-sulfide environments

ATHENA A. NGHIEM[1], DUYEN VU[2], HUNG VIET PHAM[2] AND LENNY H.E. WINKEL[3]

[1]University of Wisconsin-Madison
[2]VNU University of Science, Hanoi
[3]ETH, Swiss Federal Institute of Technology

Naturally-occurring contamination of arsenic (As) in groundwater affects millions across South/Southeast Asia. Based on total groundwater As concentrations, Pleistocene aquifers are typically considered a low-As alternative to As-contaminated Holocene aquifers. Arsenic mobilization strongly depends on its chemical speciation; however, few studies have measured As speciation in groundwater, especially thiolated species that can lead to faster As mobility due to poor sorption onto oxidized iron minerals. Prof. Britta Planer-Friedrich was a world-renowned expert in measuring thioarsenate species, even in unexpected low-sulfide environments like groundwater. In initial work on Bangladesh groundwater, she discovered the presence of monothioarsenate [1]. Since then, Britta and her team have refined preservation techniques to measure methylated, thiolated, and methylated-thiolated As in diverse settings.

Meanwhile, we recently demonstrated that thioarsenic species were likely responsible for increasing As by simulating the infiltration of high-As Holocene groundwater into low-As Pleistocene sediment from Bangladesh by pumping [2]. In Vietnam, changes in As distributions related to groundwater pumping have also been suggested [3]. Here, we aim to test the generalizability of this biogeochemical mechanism to Vietnam and optimize site selection for As speciation measurements.

In an initial "untargeted" field campaign, we sampled 64 wells across Hanoi and preserved samples to enable As speciation analysis. In collaboration with Britta, we measured thioarsenic species, which were found in ~26% of wells and as a low fraction of total As. However, as wells were randomly chosen, to optimize sampling and target where thioarsenic species could play an important role, we predicted thioarsenic occurrence. Since thermodynamic predictions of thioarsenic species remain limited, we performed random forest modeling with the untargeted measurements using available geochemical parameters. We predicted targeted sampling locations and accurately confirmed thioarsenic predictions with measurements. For the first time, we demonstrate the strength of machine learning to predict As speciation and optimize site selection for As speciation, which is often tedious to preserve in remote environments. Finally, interpretation of our validated modeling results is valuable for determining the impact of As speciation in As-affected groundwaters.

[1] Planer-Friedrich et al. 2018, *ES&T*
[2] Nghiem et al. 2023, *Nature Water*
[3] Winkel et al. 2011, *PNAS*