Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.37777

Aqueous arsenic speciation across groundwater interfaces: a new perspective from measurements to interpretable models

ATHENA A. NGHIEM[1] AND LENNY H.E. WINKEL[2]

[1]University of Wisconsin-Madison
[2]ETH, Swiss Federal Institute of Technology

Naturally-occurring contamination of toxic arsenic (As) in groundwater is a pervasive issue worldwide, especially across South/Southeast Asia. Since its initial discovery, field campaigns and measurements of total As concentrations have identified a characteristic spatial heterogeneity in groundwater As distributions. In such settings, the commonly proposed mechanism of As release is reductive dissolution of As-sorbed iron (III) (oxyhydr)oxide minerals. Aqueous As is typically assumed to be an oxyanion species, despite only total concentrations being measured. However, it is also well-known from laboratory studies that aqueous As speciation itself – whether As(V) or As(III) oxyanions or more complex species – affect its sorption capacity onto iron minerals and other important surfaces (e.g., organic matter). Emerging aqueous As speciation measurements in the environment shows an undeniable impact on As mobility: in particular, a few non-oxyanion species such as thiolated, methylated and thio-methylated As species have been found to lead to faster As contamination in groundwater due their poor sorption onto varied iron minerals.

Despite its purported importance, however, these As species remain undercharacterized and verifiably difficult to measure in groundwater, due to complex stabilization protocols in the field. Meanwhile, thermodynamics governing As speciation have been noted not to match well with observations. To bridge this gap, we present novel machine learning (ML) predictions of aqueous As speciation, which are both interpretable and verified with subsequent measurements. Using typical geochemical measurements from As-affected groundwaters in Vietnam, we predict the presence of thiolated As species; interpretation of the ML model indicates that higher concentrations of sulfur increases the prediction of thioarsenic species, as expected. However, higher concentrations of methane, dissolved organic carbon, and ammonium surprisingly decrease the probability of thioarsenic species. Furthermore, investigating interpretable ML output such as Shapley values aggregated by geologic setting demonstrates that transitional interfaces such as (i) the shallowest aquifer and (ii) a likely transitional stratigraphic unit between oxidizing/reducing environments have higher predicted probability of thioarsenic species. Using ML predictions of these sparsely measured species, we can generalize and extend the presence of various As species to determine their impact on large-scale As biogeochemical cycling.