A comparision of statistical models of arsenic in groundwater at the national scale

MELISSA A. LOMBARD[1]

[1]United States Geological Survey
New England Water Science Center
331 Commerce Way, Suite #2
Pembroke, NH 03275-3718
mlombard@usgs.gov

Arsenic occurs naturally in many groundwater systems. The U.S. Environmental Protection Agency has established a maximum contaminant level of 10 μg/L for public drinking-water supplies, however domestic supplies are largely unregulated and do not require testing. Therefore, domestic well users are more likely than those on public supplies to unknowingly consume water containing arsenic concentrations greater than 10 μg/L. In cooperation with the Centers for Disease Control and Prevention (CDC), the U.S. Geological Survey recently developed a statistical model to predict the probability of elevated arsenic levels (>10 μg/L) in water from domestic wells located in the continental United States. The logistic regression model uses geologic, geochemical, hydrologic, and physical landscape variables. It was applied to estimate the population exposed to elevated arsenic from domestic drinking water wells nationally.

An updated predictive model is currently being developed using machine learning methods (boosted regression trees, BRT) in an effort to improve the predictive capability. Although similar in overall accuracy, a preliminary BRT model more accurately predicts the occurrence of elevated arsenic levels than the original LR model. The preliminary BRT model variables with the most influence are average annual precipitation and the arsenic concentration in the C2 soil-horizon. Based on the standardized coefficients in the logistic regression model, the average bismuth concentrations in the C soil horizon and bedrock geology are important variables. A comparison of the two models will be presented and insights into large-scale geochemical processes that impact the mobility of arsenic in groundwater will be discussed.