

Does My Drinking Water Have Uranium?: Using machine learning to predict uranium levels in regions of untested well water in New Jersey

Kathryn Kaiser^{1,2}, Benjamin Bostick^{1,3}, Shams Azad³, Mason Stahl⁴

¹*Columbia University*, ²*Trinity College Dublin*, ³*Lamont-Doherty Earth Observatory of Columbia University*, ⁴*Union College*

Chronic exposure to uranium in drinking water poses serious health risks, especially in rural areas where private wells are unregulated and rarely tested. However, predicting the spatial distribution of uranium contamination is difficult due to sparse sampling data. This study applied a machine learning framework to estimate groundwater uranium concentrations in untested wells using limited chemical and geological data, with a focus on developing scalable methods for underserved communities. Over 100,000 groundwater quality samples collected in New Jersey between 2003 and 2023 were compiled and aggregated to 2-mile grid cells. Each record included uranium, pH, iron, manganese, nitrate, and gross alpha activity. After cleaning and aggregation, random forest regression models were trained to identify key geochemical predictors of uranium. pH and gross alpha emerged as dominant variables, consistent with uranium's redox-sensitive mobility. Predictions of uranium concentrations were found to be higher along borders of geological regions throughout New Jersey. To address data scarcity, a two-step modeling approach was used: first, a geochemical model predicted uranium from more commonly measured variables; second, a spatial model was trained using geological and hydrological covariates to predict uranium concentrations without relying on chemical inputs. The resulting model generated a 2-mile resolution predictive map of uranium distribution across New Jersey. This map identifies areas of potential contamination risk and provides a transferable framework for other regions with limited data, including the Great Plains. These methods contribute to environmental justice efforts by identifying communities at elevated risk of uranium exposure where direct testing is unavailable.