

## **Data mining-based approaches to investigate the sources and extent of groundwater contamination in areas of extensive oil, gas, and coal extraction**

SAMUEL W SHAHEEN<sup>1</sup>, TAO WEN<sup>2</sup>, ALISON HERMAN<sup>1</sup>  
AND SUSAN L BRANTLEY<sup>1</sup>

<sup>1</sup>Pennsylvania State University

<sup>2</sup>Syracuse University

Presenting Author: [sws41@psu.edu](mailto:sws41@psu.edu)

The water resource impacts of unconventional oil and gas development (UOGD) can be difficult to delineate from longstanding geogenic and anthropogenic impacts on groundwater chemistry, particularly where UOGD may overlap with decades to centuries of legacy hydrocarbon extraction (conventional oil and gas, coal mining). In this study, southwestern Pennsylvania, U.S.A. was selected as a testbed to investigate how the overlap of intensive recent UOGD and widespread, longstanding hydrocarbon extraction in the region may impact water chemistry and the ability of data mining methods to delineate the impacts of the wide array of geogenic and anthropogenic contaminant sources. Using a geospatial statistics-based data mining tool, locations where concentrations of a species (e.g. methane) significantly increase nearby features of interest (e.g. unconventional oil and gas wells) were mapped across the study area. Additionally, unsupervised machine learning (non-negative matrix factorization, NMF) was applied to delineate endmember sources of groundwater species and their contributions to groundwater chemistry in the region. A small number of subregions where UOGD may have contaminated groundwater with methane or brine components were identified. However, the overlap of UOGD with legacy hydrocarbon extraction does not appear to make UOGD more prone to contamination, and methane migration incidents appear less frequent than in other areas of the state with extensive UOGD. However, significant increases in sulfate concentrations are widely observed nearby coal mining areas, and coal mining has elevated concentrations of sulfate above “baseline” on a regional scale. Additionally, chloride concentrations are significantly elevated nearby highways across the study area, and road salting is a regionally significant source of chloride in groundwater.