

## **Using machine learning to detect anomalous methane in groundwater within shale gas production areas**

T. WEN<sup>1\*</sup>, M. LIU<sup>2</sup>, G. ZHENG<sup>2</sup>, S.L. BRANTLEY<sup>1</sup>, Z. LI<sup>2</sup>

<sup>1</sup>Earth and Environmental Systems Institute, Penn State Univ., University Park, PA 16802, USA  
(\*correspondence: [tzw138@psu.edu](mailto:tzw138@psu.edu))

<sup>2</sup>College of Information Sciences and Technology, Penn State Univ., University Park, PA 16802, USA

The rapid development of shale gas in the U.S. has occasionally caused environmental issues including the leakage of methane out of gas wells. Elevated methane (CH<sub>4</sub>) in groundwater is the most common identified concern due to potential explosion hazards and its contribution to global warming.

Many environmental forensic studies using geochemical tracers (e.g., organic compounds, stable carbon/strontium/lithium/boron isotopes, halogen elements, and noble gas isotopes) have been conducted to investigate the source and transportation pathway of dissolved methane in groundwater. These studies have reached conflicting conclusions due partly to variable dataset size (small vs. large). In addition, the natural variability of these geochemical tracers can make definitive assessments of the provenance of dissolved methane uncertain. Furthermore, in some shale gas production areas, e.g., in Pennsylvania (PA) in the U.S.A., the presence of natural methane is not uncommon and the incidents of methane contamination are relatively rare compared to the large number of drilled shale gas wells that are distributed across wide regions. This poses more challenges to the task of discrimination of natural vs. anthropogenic methane in groundwater. Other difficulties in these environmental forensic studies also arise from the lack of pre-drill water chemistry data, especially the dissolved methane concentrations.

This study develops a geochemical fingerprinting tool to distinguish between methane from natural and anthropogenic sources. We develop this tool from published data (~11,000 water samples) on groundwater quality in northeastern Pennsylvania. We apply a machine learning model to predict methane concentration based on other water chemistry parameters and to subsequently study the importance of features in predicting methane concentration. In turn, this leads to the finding of helpful chemical analyte(s) to distinguish between natural and anomalous methane dissolved in groundwater. Learned models have also been tested on compiled data from many previously published studies in U.S.