Commentary
HEI Energy Review Committee

HEI

Research Report 233, *Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania*, Jennifer Baka, Susan L. Brantley, and Lingzhou Xue* et al.

INTRODUCTION

The advent of high-volume and multistage hydraulic fracturing, combined with horizontal drilling in the early 2000s, has led to a substantial increase in onshore oil and natural gas development in the United States. This increase in the rate and scale of energy development now includes longer periods of development, with more materials being brought to and from larger well pads and more solid and liquid waste being generated on-site to be managed. This rapid expansion has created additional instances for potential human exposure to chemical and nonchemical stressors and has generated concerns about their effect on human health.

This newer form of oil and gas production, often referred to as unconventional oil and gas development (UOGD†), allows operators to access oil and gas in sandstone or shale formations. While UOGD has expanded into new areas, it is often co-located in areas with a history of extractive energy practices, such as conventional oil and gas development (COGD) or coal mining. A challenge to understanding exposure to UOGD, specifically in areas of legacy energy development, is that many of the potential contaminants of concern are the same.

As described in the Preface to this report, in 2020, HEI Energy issued a Request for Applications, *RFA E20-2*: Community Exposures Associated with Unconventional Oil and Natural Gas Development to fund research that would assess community exposures to chemicals in surface or groundwater originating from UOGD. This RFA solicited applications that would help determine the UOGD processes that have led or may lead to releases to water and identify and define the influence of various factors (e.g., varying geology) on exposures. The proposed research would engage local communities and should be able to maximize the generalizability of the research and inform decision-making, and, to the extent practicable, be able to distinguish potential UOGD exposures from other COGD and any other background sources.

In response to the RFA, Baka, Brantley, and Xue of The Pennsylvania State University submitted an application entitled "Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania." They proposed using a database of existing groundwater samples to investigate whether they could distinguish the chemical constituents associated with UOGD from legacy hydrocarbon development.

The investigators would use statistical analyses and machine learning to identify links between various potential sources and water contamination. Additionally, the team proposed hosting focus groups to elicit community concerns regarding UOGD, water contamination, and public health to inform their assessment of potential community exposures associated with UOGD. HEI's Energy Research Committee recommended funding their proposed study because it included a source apportionment approach that could be used in other regions, and they believed their community engagement plan brought value to the project and could inform decision-making.

This Commentary provides the HEI Energy Review Committee's independent evaluation of the study. It is intended to aid the sponsors of HEI Energy and the public by highlighting both the strengths and limitations of the study and by placing the results presented in the Investigators' Report into a broader scientific and regulatory context.

SCIENTIFIC AND REGULATORY BACKGROUND

The technological advances that have allowed for expanded oil and natural gas development from unconventional formations, such as low-permeability sandstone and shale, have caused the scale of UOGD to grow overall and at each process level. UOGD processes occur on and off the drilling site or *well pad*.

The life cycle of UOGD starts with field development and includes vertical and horizontal drilling and hydraulic fracturing. High-volume, multistage hydraulic fracturing along lengthy horizontal wells — often exceeding two or more miles — injects millions of gallons of water and chemicals that are sometimes proprietary mixtures, under high pressures, into

The 2-year study, "Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania," began in January 2022. Total expenditures were $636,753. The draft Investigators' Report was received for review in November 2023. A revised report was received in May 2024. A second revised report, received in September 2024, was accepted for publication in January 2025.

During the review process, the HEI Energy Review Committee and the investigators had the opportunity to exchange comments and clarify issues in both the Investigators' Report and the Review Committee's Commentary. This Commentary has not been reviewed by public or private party institutions, including those that support HEI Energy, and may not reflect the views of these parties; thus, no endorsements by them should be inferred.

*Co-principal investigators.

†A list of abbreviations and other terms appears at the end of this volume.

tight formations to release oil or natural gas. This mixture of water and anthropogenic chemicals, along with naturally occurring compounds (such as arsenic or petroleum hydrocarbons), is returned to the surface during flowback. Oil and gas wastewater, known as produced water, is generated at the onset of the production phase and continues to be generated over the life. Production involves the extraction, gathering, and processing of oil and gas, and the continued management of produced water and other materials. Post-production processes close the well and reclaim the land around the development.

UOGD*-related chemical releases to surface or groundwater can result from authorized or accidental discharges to the environment. Examples of authorized releases include permitted discharge of produced water to surface water or for applications outside the oil and gas production site (e.g., road treatments for dust or ice). While protections have been put in place to prevent unintentional chemical releases to groundwater and surface water, releases can occur due to wellbore failures, malfunctioning equipment, or other accidental conditions.[1]

Research on changes in surface water or groundwater quality from accidental UOGD releases is challenging, given the practical limits on knowing when and where such releases might occur. Once a release to surface water is known to have occurred, researchers may be able to detect associated impacts.[2] Accidental releases to the soil and groundwater pose a much greater challenge than releases to surface water. They typically require substantial effort and resources to determine the extent and severity of impacts on groundwater quality, as well as the conditions that influence the mobility of contaminants in the subsurface environment.

The rapid expansion of UOGD has given rise to concerns about effects on human health through the potential contamination of drinking water sources, particularly given the high volumes of water and chemicals used to fracture a well and the resulting wastewater. This concern is particularly acute in rural areas where many residents get their water from private wells. Private water wells are not monitored except by the owners. A review by the US EPA[3] did not identify extensive contamination of water related to UOGD but cited instances where contamination has occurred. Current evidence indicates that people can be exposed to chemicals in produced water released from UOGD processes, but there are important gaps in knowledge about these exposures that must be addressed to better understand potential impacts on health.

One of the major challenges to understanding or quantifying potential human exposures to UOGD through water contamination is distinguishing UOGD-related exposures from other geogenic or anthropogenic sources of the same chemicals in environmental media. In particular, many UOGD basins overlap with ongoing or historical COGD or coal extraction. Furthermore, many of the contaminants associated with leaks of oil and natural gas are either naturally occurring or have other potential sources. The chemicals used for drilling and in hydraulic fracturing fluid are often not naturally occurring; however, incidents and evidence of contamination by these chemicals are rare.[4] Because produced water is highly saline, contamination from produced water can be inferred through increased concentrations of salt species. For example, barium and strontium have been used previously as tracers for produced water released into the environment.[4] While a subset of water quality studies has used isotopic tracers and other markers associated with UOGD to isolate its effects from those of other sources, this type of fingerprinting is not typically conducted outside of specialized labs or for research purposes.[5]

The quantity of data on levels of UOGD-related chemicals in the environment continues to increase, along with efforts to use the data to quantify human exposure. The current study was designed to investigate trace contaminants from UOGD activity by using an ever-growing dataset of groundwater chemistry to find signatures of contamination from UOGD activities that distinguish the contamination from natural sources or other anthropogenic activities.

STUDY OBJECTIVES AND OVERALL STUDY DESIGN

The overarching objective of the study was to assess the impact of UOGD on groundwater resources and whether any such impact is exacerbated in shale gas basins where legacy forms of hydrocarbon extraction are prevalent. The investigators additionally used community perspectives to inform the geochemical analysis. They sought to explore five hypotheses:

1. The same potential drinking water contaminants from UOGD activities in a hydrocarbon-rich basin may also come from natural processes, COGD, coal extraction, or other land use activities.
2. Groundwater contamination from UOGD can be distinguished from other sources, even in areas with a heavy overlap of land use activities.
3. Co-located extractive processes, such as COGD, coal mining, and UOGD, are more likely to cause water contamination than when there is only one type of hydrocarbon exploitation.

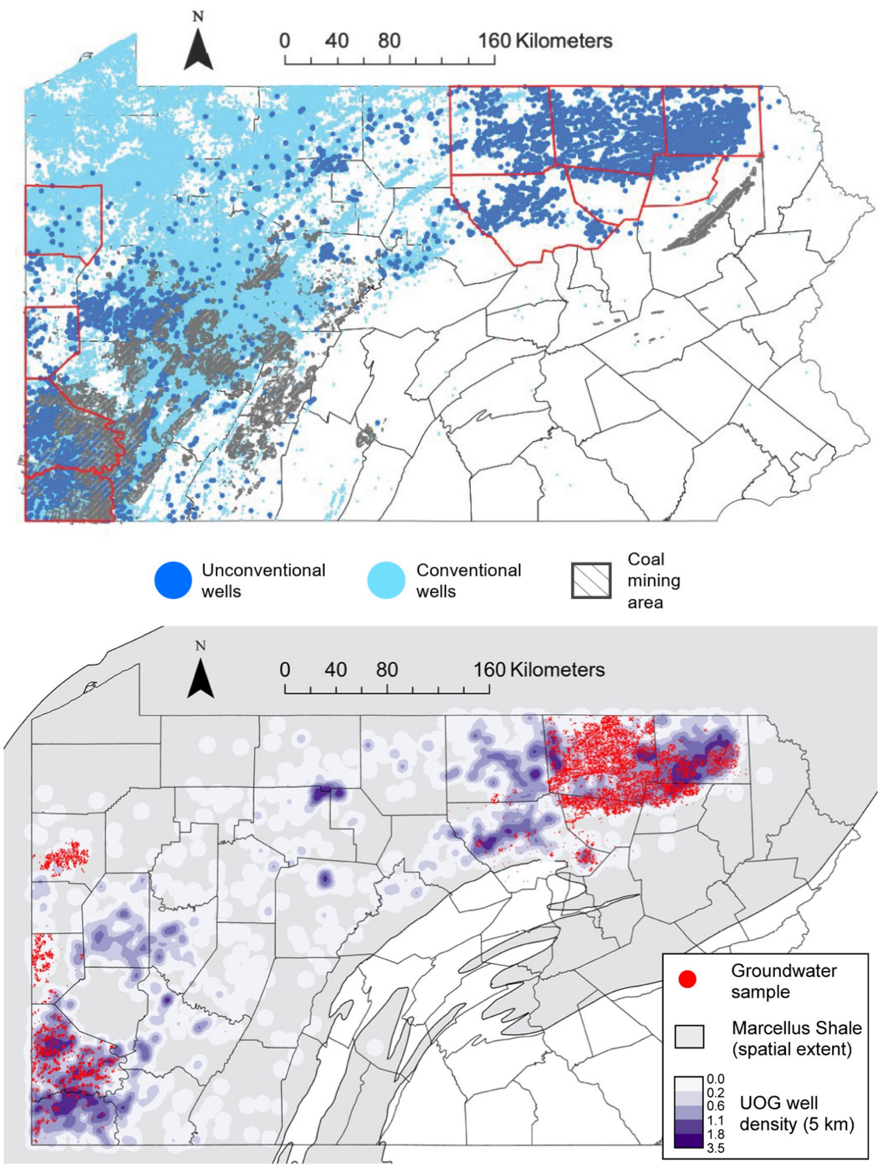
* UOGD refers to the development and production of oil and natural gas as practiced starting around the beginning of the 21st century through multistage hydraulic fracturing in horizontal wells. UOGD processes occur on and off the well pad and include the following:

- *Field development*: exploration, site preparation, vertical and horizontal drilling, well completion (casing and cementing, perforating, acidizing, hydraulic fracturing, flowback, and well testing) in preparation for production, and management of wastes
- *Production operations*: extraction, gathering, processing, and field compression of gas; extraction and processing of oil and natural gas condensates; management of produced water and wastes; and construction and operation of field production facilities
- *Post-production*: well closure and land reclamation

4. Communities may have different perceptions of risk related to UOGD water contamination than what may be identified through scientific analyses.
5. Gaps in community perceptions and scientific analyses can be reduced through facilitating multistakeholder dialogue and, when possible, through studies of contaminants emphasized by the public.

To explore these hypotheses, the investigators used a combination of quantitative and qualitative approaches that were broken into five specific aims. Briefly, they developed a technique to first estimate naturally occurring (i.e., background) groundwater chemistry in southwestern Pennsylvania in areas of multiple land uses, including UOGD, and second to assess groundwater signatures for different land uses and geologies (Aim 1) using data mining methods on a large dataset of groundwater data. The groundwater dataset had greater than 28,000 samples, which they used to isolate the influences of naturally occurring processes versus UOGD or legacy processes, including COGD and coal mining (Aim 2, **Commentary Figure 1**).

During the study, the team held six community focus groups in three counties of the study region to identify community concerns in areas of multiple types of hydrocarbon exploitation. They used these concerns to interpret preliminary findings and to determine the contaminants of most concern for both scientists and the public (Aim 3). The team then developed a methodology to map locations of potential UOGD water contamination in areas of overlapping land use (Aim 4). Finally, they used the focus group meetings to create a dialogue between communities and scientists and identified ways to reduce gaps between public perceptions and scientific findings (Aim 5).



Commentary Figure 1. Top: Map of UOG wells (dark blue), COG wells (light blue), and coal mining areas (gray) in PA. Counties included in the study are highlighted in red. Bottom: Locations of the 28,609 sampled groundwaters indicated on a map showing the average density of UOG wells within a 5-km radius in Pennsylvania. (Source: Reprinted from *Shaheen et al. 2024*; Creative Commons license *CC BY-NC-ND 4.0*.[6,7])

SUMMARY OF METHODS AND STUDY DESIGN

The team focused on southwestern Pennsylvania, as this part of the United States has one of the highest densities of unconventional wells in the world and over a century of energy extraction history, making it an area of co-located UOGD, conventional development, and coal mining. However, in exploring these methods and to better understand their findings, the team also considered northeastern Pennsylvania, which has a high incidence of UOGD but little history of hydrocarbon extraction (Commentary Figure 1).

GEOCHEMICAL DATA

The investigators applied quantitative geochemical and statistical analyses to an existing groundwater database to estimate background groundwater chemistry, look for evidence of groundwater impacts from different hydrocarbon extraction activities, and map locations of potential groundwater contamination from UOGD specifically.

The investigators used the Shale Network Database,[8] which includes groundwater analyses from the Marcellus Shale in Pennsylvania, as the groundwater chemistry database. These data were provided to co-PI Brantley based on an agreement with the Pennsylvania Department of Environmental Protection (PADEP), which is the regulator of UOGD in Pennsylvania. This dataset is continually growing, and at the time of the study, it included 28,609 groundwater analyses collected between April 2008 and April 2020. Approximately 7,000 of these analyses were from samples in the study area of southwestern Pennsylvania (Commentary Figure 1).

The database contains groundwater samples collected by professional hydrogeologists using standard practice, typically as established by the United States Geological Survey. The analyses are conducted by accredited commercial laboratories using EPA-approved techniques. By necessity, the number of chemicals that can be tested for in each sample is limited; therefore, the variable suite of hydraulic fracturing or drilling compounds is typically not included in chemical sample analyte lists.

The analytes in the database include volatile constituents of oil and natural gas (methane, ethane, propane), organic compounds found in petroleum hydrocarbons (benzene, toluene, ethylbenzene, oil and grease), indicators of hydraulic fracturing or drilling fluids (methylene blue active substances), and inorganic species (e.g., total dissolved solids, alkalinity, iron, and arsenic). The team focused on the six analytes that were most reported in the database for every site: sodium, calcium, chloride, barium, strontium, and sulfate.

Samples were collected from known or potential drinking water sources near proposed well sites and analyzed with funding from the oil and gas development companies to generate data to be used in potential legal disputes over water quality impacts. Given the reason for these sampling events, there was typically little to no repeat sampling of water sources.

In addition to the Shale Network database, the team collected relevant UOGD-related information, including when and where the wells were placed, any violation data (well casing or cementing violations, spills or leaks, or impoundment violations), location of wastewater impoundments,* and waste reports, as well as information on geological and surface water features. These data were used in conjunction with the Shale Network groundwater quality data, described above, to perform geochemical analyses.

AIM 1. GROUNDWATER CHEMISTRY

In Aim 1, the team first sought to develop a technique to estimate natural background groundwater chemistry in an area of multiple land uses. The team used the commercially available computer code, the Geochemist's Workbench, which is a geochemical equilibrium-solving model to understand background water chemistry, based on reactions of local rain with local bedrock systems.

Next, the team assessed groundwater signatures for different land uses and geologies. The land uses defined by the investigators were UOGD, COGD, coal mining, and combinations of one or more of these uses, using the groundwater database. Additionally, the team considered control sites, which had measured groundwater samples taken within 1 mile of areas with no prior hydrocarbon extraction.

They used nonparametric statistical tests (Wilcoxon-Mann-Whitney and Brunner-Munzel) to compare the median values of concentrations of sodium, calcium, chloride, barium, strontium, and sulfate in groundwater for the seven permutations of land uses or control sites. These inorganic chemicals were chosen because they were the most reported analytes in the database.

Both the Bruner-Munzel and Wilcox-Mann are based on assumptions of sample independence. To test for spatial correlation, the researchers used Moran's I, which is an assessment used to test if the data cluster. In this assessment, they used ArcGIS Pro's defaults to determine the distance threshold of 14,400 m.

Although all six anion and cation salt species (sodium, calcium, chloride, barium, strontium, and sulfate) were analyzed with respect to land use, a subset of analytes, barium and strontium, which have been used as UOGD tracers by others, were evaluated using regression and fixed-effects modeling to estimate average increases in their concentrations for a given increase in UOG well density, spill density, UOG well distance, and spill distance for different regions in Pennsylvania or statewide.

Using barium and strontium, the researchers examined the effect of other potential variables that may contribute to measurable impacts on water quality. These fixed-effects variables were proximity to conventional oil and gas wells, coal mining, highways, and streams; seasonality; and primary bedrock composition and geological features. Collectively, these variables describe other known saline impacts to groundwater through the presence or absence of legacy hydrocarbon development, road salting, natural pathways for deep fluids to migrate up into aquifers, or topographic lows where brines can also infiltrate into aquifers. Here, a "brine" refers to naturally occurring water that is brackish or more saline.

Finally, the team examined potential sources of UOGD-related contamination, including casing or cementing violations, impoundment violations, and reported spills. These violations were all reported to PADEP. Spill violations were only included in the analysis if they were associated with a well pad.

*In the context of UOGD development, an impoundment is a structure used to store wastes, including produced water, drilling fluids, or drilling muds.

Therefore, any spills that occurred off-site and were associated with transportation or disposal were not included. The locations of impoundments were identified from satellite imagery in a 2010 survey based on the United States Department of Agriculture aerial survey photography. Additionally, in 2014, eight impoundments were ordered by PADEP to be shut down to upgrade their liners and leak detection systems. Leaks from the impoundments were suspected due to increased chloride concentrations downgradient.

AIM 2. EVIDENCE OF UOGD AND OTHER LEGACY HYDROCARBON EXTRACTION CONTAMINATION

To further distinguish evidence of UOGD from legacy hydrocarbon development, the researchers built off the signatures for different land uses (Aim 1) and used a previously developed machine learning approach[6] of non-negative matrix factorization (NMF) to determine chemical signatures of groundwater in Pennsylvania associated with various sources of brine. NMF can be used to look for patterns in data. Here, the team used NMF to identify the proportions of each endmember, meaning the representative chemistry, in groundwater samples in the Shale Gas Network database and explore the sources of chloride based on the molar ratios of major cations and anions (barium, calcium, magnesium, sodium, and sulfate) as a line of evidence for UOGD.

AIM 3. IDENTIFY CONTAMINANTS OF BOTH SCIENTIFIC AND PUBLIC CONCERN

The investigators identified analytes associated with activities related to UOGD in different regions of Pennsylvania and across the entire state. They also hosted focus groups to engage with community groups in the study areas and learn their concerns regarding UOGD, water contamination, and public health. The contaminants identified by the focus groups were compared to the analytes identified in the groundwater database for any potential overlap. Community concerns were considered in completing the assessment of potential community exposures associated with UOGD. This aim is also discussed in the sections on Community Focus Groups (Aim 5).

AIM 4. HOTSPOT MAPPING

To develop a map of potential hotspots, the research team combined their statistical analysis on background chemistry with their findings from NMF and a sliding windows technique to isolate the influences of natural and anthropogenic processes on groundwater chemistry and to infer instances of specific UOGD activities on water contamination. Like their findings from NMF, the sliding windows technique was developed in previously published work and supported by alternative funding.[6] Briefly, the team used the sliding window geospatial technique[9] to identify 5×5-km grids or “windows” across the study area that were significantly correlated with UOGD. These grids were stepped across the area in 200-m increments. Within each grid, the team calculated the Kendall rank correlation between the concentration of chloride and either the distance to the nearest UOG well or the density of UOG wells within 1 km of the sample location. If a significant relationship was identified, the window was assigned a value of 1 or –1, indicating a positive or negative relationship.

The chloride concentrations measured in groundwater and their potential linkages with UOGD were further refined using the findings from NMF, which were able to identify the likely sources of chloride based on the molar ratios of the six major cations and anions.

AIMS 3 AND 5. COMMUNITY FOCUS GROUPS

The investigators conducted two meetings in three counties in southwestern Pennsylvania for a total of six meetings with the objectives to first gather community input on the relationship between UOGD, water contamination and public health (Aim 3), and second to share preliminary findings and to continue gathering information on any concerns that may have changed since the initial meeting (Aim 5). These participants were recruited by reaching out to different organizations within the study areas and asking them to recruit participants through their various networks via email and phone banking. The investigators make the important note that the community concerns, observations, and opinions included in their report reflect the views of the focus group attendees and are not necessarily representative of the greater community. Participants from the first meetings were invited to join in the second meetings.

During these meetings, the participants were asked questions related to observed changes in their water since the onset of UOGD in their community; any perceived health impacts; sources of information; obstacles to understanding the relationship between UOGD, drinking water, and health; and recommended changes improving community knowledge.

AIM 5. FACILITATE A DIALOGUE BETWEEN SCIENTISTS AND COMMUNITY

To facilitate dialogue between communities and scientists to reduce gaps between public perceptions and scientific findings, a second meeting was held to share preliminary findings from this study. Additionally, the community group members and focus group participants were invited to participate in the 2023 Shale Network Workshop, which was well attended by PADEP employees. From these conversations, the researchers developed policy recommendations for facilitating stakeholder dialogue on contentious environmental topics.

SUMMARY OF KEY RESULTS

AIM 1. BACKGROUND CHEMISTRY

Hydrocarbon-Related Land Use Patterns

A comparison of median water chemistry based on land use areas revealed significant differences (P value < 0.05) in analyte species relative to control sites, based on types of energy development and how those uses overlapped. While the study area was primarily southwestern Pennsylvania, the team additionally examined other areas in Pennsylvania to better understand the observed patterns and land-use correlations. The following relationships emerged.

Median strontium concentrations were higher for sites that included UOGD or coal mining. The team notes that acid mine drainage associated with coal mining can dissolve carbonate minerals and release strontium. Barium concentrations were higher for sites with UOGD and UOGD co-located with conventional development. This increase was not seen for sites that included coal mining, again due to acid mine drainage, which can lead to sulfate levels that can cause the precipitation of barium out of solution as barite ($BaSO_4$).

Produced waters are often highly saline, particularly in the Marcellus shale region. However, sodium concentrations are not significantly elevated within areas of hydrocarbon development. The researchers surmise that this is likely due to other geogenic and anthropogenic sources of sodium, including road salting, concluding that sodium is not a reliable indicator of groundwater impacts from UOGD in this area. Additionally, sulfate did not consistently show significant increases with UOGD. Median calcium concentrations were higher than the control sites for all permutations of land development, which did not allow the team to use calcium concentrations to differentiate between any of the land uses. Median chloride concentrations were significantly higher for all land uses that included UOGD. However, using chloride as a UOGD-indicator species suffers from the same challenge as using sodium in that chloride contamination can also arise through road salting.

Collectively, these patterns led the research team to conclude that the groundwater chemistry signatures for barium and strontium were the best and clearest indicators of potential UOGD impacts. Therefore, these two species became the focus of their analysis while also using chloride concentrations as an additional line of evidence, where appropriate. A table summarizing these correlations can be found in Appendix A, Table A2 (available on the HEI Energy website).

It should be noted that after applying Moran's I to the groundwater chemistry data to test for sample independence, the research team reported a small degree of positive spatial autocorrelation, meaning that similar concentrations of these elements are somewhat clustered together in space. They did not explore the effects of this autocorrelation.

Proximity to UOGD

Overall, the team noted small but statistically significant increases in both barium and strontium concentrations near UOGD, particularly in the regions of Pennsylvania where UOGD overlaps with legacy hydrocarbon extraction. Using a regression analysis, the team found that barium and strontium were significantly higher within 1 km of UOGD than beyond 1 km. Additionally, the concentrations of barium and strontium showed a small but significant increase with a higher density of wells within 1 km of the water sample location.

The team also examined the effects of elevation on the relationships between barium and strontium concentrations and UOGD well proximity and density, which were not considered in their initial calculations. Hydrological processes are usually topologically driven, meaning that water flows down gradient even in the subsurface. By considering only higher elevation UOGD wells in their regression analysis, the team observed a larger regression coefficient and higher confidence values, showing a strengthened positive correlation between barium and strontium concentrations and UOGD well proximity and density.

Subregions in Pennsylvania

While both the northeast and southwest of Pennsylvania have a high density of UOGD, they have differences in other land uses and geology. The northeastern area of the state has much less coal mining or COGD; it is also characterized by higher topographic relief and a higher prevalence of large faults. This makes it a useful area to compare to the primary study area of southwestern Pennsylvania to better understand some of the correlations that were observed by the team. For example, the team was able to identify an increase in barium and strontium concentrations associated with UOGD well density in southwestern but not northeastern Pennsylvania.

Some of the geological features that are more prevalent in northeastern Pennsylvania are associated with natural brine migration, and proximity to these features corresponded to higher concentrations of barium and strontium. Fixed-effects models that considered geological sources of salt species, as well as proximity to other anthropogenic sources (conventional development, coal mining, or road salt from highways), allowed the research team to account for some overlapping sources. They concluded that, overall, the two regions showed positive correlations between salt concentrations and UOGD, but the geological features of northeastern Pennsylvania may obscure some of these effects (Appendix A, Table A7).

AIM 2. EVIDENCE OF UOGD AND OTHER LEGACY HYDROCARBON EXTRACTION CONTAMINATION

To identify the characteristics of brine salt species that co-occur with chloride, the investigators used NMF to delineate the sources of different prominent chemistries. They were able to identify the signatures of Appalachian Basin brines (i.e., similar to Marcellus-produced water), meteoric recharge (precipitation and infiltration), and road salt. Based on their analysis, 29% of the chloride present in southwestern Pennsylvania groundwater sampled in this study is from Appalachian Basin brine, which they conclude reflects brines that are naturally present in groundwater. However, using this method, the team was able to identify potential contamination hotspots in areas proximate to UOGD, high-density UOGD areas, spills, and impoundments based on hydrocarbon-related land use patterns identified previously.

AIM 3. CONTAMINANTS OF BOTH SCIENTIFIC AND PUBLIC CONCERN

The focus groups indicated that communities in southwestern Pennsylvania were most concerned about potential radiation exposure from wastewater that spilled or leaked at the surface. They expressed frustration about not understanding if there was a relationship between UOGD and cancer. There has not been extensive testing for radioactive or other hazardous trace elements in the samples collected for the Shale Network database. Furthermore, in some cases, the method detection limits for some constituents of concern were above health-based risk guidance values, meaning it is impossible to know if those constituents were present at levels that may be harmful. To ascertain potential concentrations of these trace elements (e.g., thallium, arsenic, cadmium, or radium), the researchers performed a ratio analysis based on relationships of analytes in regionally produced water. While not directly observed, this analysis offers some insight into the potential presence of hazardous trace elements.

Assuming that the chloride to thallium ratio in produced water would be preserved regionally, the team identified certain hotspots where thallium might exceed EPA limits based on the high UOGD-attributable chloride concentrations. The other trace elements were calculated to not exceed EPA limits. However, the investigators noted that many variables could not be considered in this simple ratio analysis, including secondary changes in the groundwater chemistry from produced water released into the environment that could mobilize other toxic species.

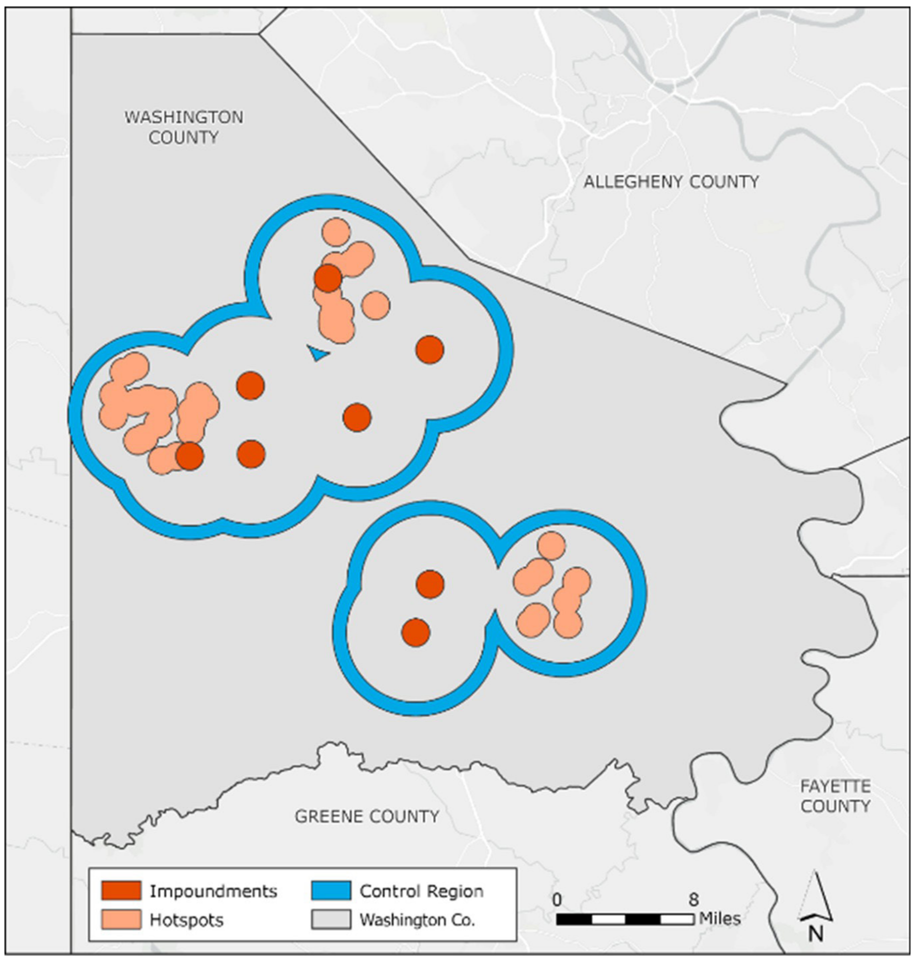
AIM 4. HOTSPOT MAPPING

Using the sliding window analysis, the team mapped hotspots that were areas of high chloride concentrations that were correlated with proximity to wells or dense UOGD. These areas were further refined using the representative chemistries revealed by NMF analysis in conjunction with land uses associated with UOGD-impacted groundwater.

Effects of Spills

The team examined barium and strontium concentrations in relation to the three identified spill mechanisms: (1) casing or cementing violations, (2) impoundment violations, and (3) pollution (i.e., spill) violations. Across the entire state, concentrations of barium and strontium increased with an increase in the number of spill violations within 1 km of the sampling site compared to sites farther away. There was no correlation with either impoundment or cementing or casing violations and increases in barium and strontium concentrations when considering statewide data. However, an additional spill increased the corresponding concentrations of these analytes more than the presence of an additional well, indicating that spills contribute more strongly to UOGD-related groundwater impacts than well density alone.

Similar to understanding the effects of well proximity and density on groundwater quality, when analyzing the impacts of spill violations on groundwater quality in the different subregions of the state, the team identified some differences



Commentary Figure 2. Locations of potential contamination, as suggested by the analysis of the groundwater dataset in Washington County, PA, within 1-km buffers around impoundments and UOG wells. Also shown are the UOGD control regions (in blue) defined as being greater than 3 km from UOG wells or impoundments. Minimal-UOGD (greater than 5 km from wells or impoundments) control sites are not shown but are all located just to the north of the Washington County-Allegheny County border. No spills were identified in this region. (Source: Investigators' Report Figure 3.)

in the salt species concentrations that they attribute to the geological features of northeastern Pennsylvania. Again, when accounting for the effect of geological features using a fixed-effect analysis, the differences were minimized. The team concluded that the geological features in northeastern Pennsylvania may have obscured the effects of UOGD spill-related impacts on groundwater and should be accounted for in analyses. The researchers also found that spills of produced water would need to be greater than 1,000 L to account for the increased barium concentrations, based on a mass balance calculation and their understanding of geological conditions.

Impoundments as a Source

In southwestern Pennsylvania, the team observed significantly higher concentrations of barium, but not strontium, in samples within 1 km of impoundments compared to those farther away. Median barium concentrations were 34% higher in samples near impoundments reprimanded by PADEP in 2014 than in those farther from such sites.

By combining regression analysis, NMF, and sliding windows analysis, the team created a map of locations of potential contamination in one of the counties in southwestern Pennsylvania, as suggested by their analysis of the groundwater dataset (**Commentary Figure 2**).

AIM 5. FACILITATE A DIALOGUE BETWEEN SCIENTISTS AND COMMUNITY

Obstacles to Community Knowledge

The focus group reported that the largest obstacle to learning more about UOGD-related water contamination and any potential public health impacts was government practices, noting a profound lack of trust by participants in state regulators and the PADEP, who are tasked with overseeing UOGD. The participants also cited industry practices, including the use of nondisclosure agreements, as impediments to knowledge. They also discussed conflicts within their communities, for example, tensions among neighbors regarding whether UOGD is helping (employment, financial gain for landowners) or hurting communities (health issues, environmental degradation), and the lack of resources to access external expertise. Finally, participants noted the logistical difficulties and immense resources required to conduct an environmental site investigation to understand UOGD-specific impacts in an area with a history of energy development and its legacy of pollution.

Policy Recommendations for Facilitating Stakeholder Dialogue

Based on the focus group participant feedback, the investigators developed the following policy recommendations for creating productive dialogue with communities on topics of contentious environmental issues:

- Engage the community early and throughout the process, such as permitting a facility, the life cycle of the project, and its decommissioning.
- Choose appropriate communication channels to improve engagement, including those appropriate for the area, such as local news outlets, nongovernment organizations, and citizen scientist efforts.
- Encourage regulators to improve transparency in their communication with the communities, including more information about the potential risks of the ongoing or proposed activities and the gaps and limits to scientific or regulatory understanding.
- Create a systems approach to permitting that allows for a better understanding of the cumulative effect of multiple UOGD-related facilities or infrastructure.
- Consider community concerns about the lack of environmental monitoring by making concerted efforts to address data gaps. For example, in the case of southwestern Pennsylvania, regular radiation testing in water should be a high priority.

HEI ENERGY REVIEW COMMITTEE'S EVALUATION

STUDY DESIGN, DATASETS, AND ANALYTICAL APPROACHES

In its independent evaluation of the study, the HEI Energy Review Committee agreed that the study scope and analysis that allowed for the separation of UOGD from legacy contamination were compelling and well executed. The Committee felt that the study design, methodology for data analysis, and modeling approach bring scientific rigor to understanding potential exposure to UOGD and advance the ability to investigate potential health effects from water contamination from UOGD. The following sections describe the strengths and limitations of the research.

Geoscientific Analysis The Committee recognized that a key strength of this work lies in identifying potential groundwater impacts from UOGD, specifically in areas with legacy contamination and multiple brine sources. The geostatistical analysis, which uses available groundwater data to isolate the effects of UOGD, represents a novel method and important contribution to understanding its influence on the quality of drinking water sources, particularly in regions with long histories of hydrocarbon extraction.

The Committee valued the integration of traditional geostatistical methods with machine learning techniques to correlate brine signatures with various land uses.

Focus Groups The investigators generated a rich new dataset of public opinion on the intersection of UOGD, groundwater contamination, and public health through a series of focus groups. The team demonstrated an effective approach for multidirectional communication, which brings value to and informs decision-making for both researchers and policymakers. The Committee appreciated the way the researchers framed the goals of the community conversations to reduce knowledge gaps rather than to change public perceptions. In particular, they commented that this approach is more likely to engender open and honest conversation and a receptive audience.

Furthermore, the Committee appreciated the integration of community concerns with quantitative geoscientific analysis, particularly in addressing concerns about potential radiation exposure from wastewater management spills as a key contamination pathway. Initially, the Committee struggled to understand the connection between the geochemical analyses and the focus group activities. Therefore, the Committee urges caution in describing this approach to ensure it accurately reflects that community concerns are used to guide the research design and interpret the results, which allows for the acknowledgment of these concerns and their potential resolution through scientific analysis.

Spatial Autocorrelation The Committee also noted some limitations to the research approach. The results from both the Bruner-Munzel and Wilcox-Mann tests to identify median concentrations of analytes in water samples across different hydrocarbon land use classifications are based on assumptions of sample independence. While the water samples included in the database do not have temporal dependence, they are not spatially independent, showing a small degree of spatial autocorrelation.

It is difficult to understand how any autocorrelation may influence the significance of the concentrations of barium or strontium relative to UOGD processes. Furthermore, the Committee noted that the distance of 14.4 km used in the test for spatial autocorrelation is much greater than the 1-km radius used in the geostatistical analysis. The team had defined 1 km as "hydrologically plausible" in Pennsylvania, meaning that the distance used in Moran's I test will likely not reflect the actual distance of groundwater migration. Importantly, however, the Committee noted that this potential autocorrelation does not invalidate the team's findings. Overall, the Committee notes that future studies may explore the potential impact of autocorrelation on geostatistical relationships.

Generalizability of Guidelines The research team developed helpful guidelines to generalize their approach, but the Committee noted the likely challenges to adapting this approach to areas beyond Pennsylvania. The second step of this approach is to "identify the largest dataset of water quality data available for the target area of UOGD," followed by geospatial regressions, fixed effects approaches, and geospatial analysis tools to identify hotspots. Much of the information about the types of incidents the investigators used in their regression analysis is publicly accessible in Pennsylvania. This large, publicly available groundwater database enabled them to perform a large-scale investigation of impacts on groundwater that highlighted relevant UOGD processes. These findings may be relevant for other major shale areas; however, many oil and gas regions lack the information and resources to assess UOGD impacts on groundwater.

FINDINGS AND INTERPRETATION

Effects of UODG Wastewater Spills and Leaks

The investigators concluded that the trends in salt species in southwestern Pennsylvania were explained by large spills or leaks of produced water from impoundments at a few specific locations. Overall, the Committee agreed and found that this report offered a comprehensive and thoughtful discussion and well-supported conclusions. Furthermore, the finding that a minor but observable salinization of groundwater in this area is likely from UOGD-produced water spills and leaks has important implications for best practices and policy development for waste management. The increase in salt species concentrations near decommissioned wastewater impoundments offers additional evidence that UOGD hydraulic fracturing (i.e., well simulation) or oil and gas well integrity issues are unlikely to be the cause of groundwater contamination in Pennsylvania. The management of waste and produced water at the surface should be an important component of risk mitigation practices nationwide, particularly where large volumes of produced water are stored or transported.

The geochemical analysis of a large existing database allowed the team to distinguish the influences of geogenic and anthropogenic processes, including legacy hydrocarbon development, on groundwater chemistry and thereby identify potential linkages between UOGD and water contamination. Small but significant increases in brine salt concentration may be attributed to UOGD, likely due to spills or leaks from impoundments of produced water. The geostatistical techniques used to identify and isolate where the salinization of groundwater is likely stemming from UOGD rather than other ongoing or historical land use practices are an insightful and novel data mining approach to leverage an existing groundwater dataset. It is unclear, however, the extent to which co-located extractive processes (conventional development or coal mining) exacerbated UOGD-related impacts. This was not explicitly explained in the report and could be explored further.

Identifying Contaminant Hotspots

The team described in general terms the types of contaminants that are associated with produced water that were contained in the dataset. Salt and brine-related chemicals are the most dominant species in produced water and, therefore, are the most likely to be detected in the environment if produced water is spilled or discharged. The limitations of the dataset mean that the researchers were restricted in what could be used for their geostatistical analysis. However, the team did not address the idea that the salt signatures that signify potential UOGD groundwater impacts may also be indicators for other anthropogenic organic chemicals found in produced water or from hydraulic fracturing fluids.

Community Concerns

The information gathered from the focus groups showed that participants in southwestern Pennsylvania are concerned about radiation exposure and cancer effects from produced water spills and leaks stemming from wastewater management, rather than leaks at the oil and gas wells. The team was not able to assess radioactivity in groundwater samples, as this information was largely absent from their dataset. However, they were able to extrapolate data to provide some insight into the potential for the presence of these and other UOGD-related contaminants of concern. These calculations were added to the scope of work of the research as a direct result of the concerns expressed by the community in the focus groups.

The Committee appreciated the analysis of potential metals and radium contamination based on the mass ratios expected with measurable salt species. This analysis may not put residents' minds completely at ease, but it is a good-faith scientific effort to present the pertinent findings. Furthermore, the research team was able to confirm participant concerns that UOGD wastewater management is the contamination pathway of the highest concern. A finding that was not underscored in the report's discussion, but one that bears recognition, is that the focus group(s) identified spills as a concern, and that the geochemical analyses also concluded that the most likely source of barium and strontium in drinking water samples from local wells was from spills. This supports the notion that local knowledge, in this case, of spills, is an important consideration in any exposure study.

CONCLUSIONS

In this study, the investigators used geostatistical techniques to distinguish the effects of UOGD on groundwater from those of other land uses and legacy contamination from historical energy production in southwestern Pennsylvania, a complex task due to the overlapping impacts of these activities. They found that there were statistically significant increases in barium and strontium primarily associated with UOGD wastewater spills or leaks from impoundments containing produced water in a small number of hotspot regions. This potential pathway was highlighted by community focus groups, who were most concerned about potential radiation exposure from spills and leaks through wastewater management. This study advances our understanding of UOGD-related exposure pathways through groundwater contamination, which is needed to identify potential health effects associated with UOGD processes. The findings of this analysis may inform other shale areas where contamination may have occurred through similar UOGD processes. However, more data related to groundwater compositions and UOGD-related violations in those places may be needed before similar analyses can be performed in regions across the country.

ACKNOWLEDGMENTS

The HEI Energy Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Donna Vorhees for oversight of the study, to Cloelle Danforth for assistance with review of the Investigators' Report and preparation of its Commentary, to Tara Hamilton for editing the Investigators' Report and its Commentary, and to Kristin Eckles and Hope Green for their roles in preparing this Research Report for publication.

REFERENCES

1. Maloney KO, Baruch-Mordo S, Patterson LA, Nicot JP, Entrekin SA, Fargione JE, et al. 2017. Unconventional oil and gas spills: materials, volumes, and risks to surface waters in four states of the US. Sci Total Environ 581–582:369–377, *https://doi.org/10.1016/j.scitotenv.2016.12.142*.

2. Cozzarelli IM, Skalak KJ, Kent DB, Engle MA, Benthem A, Mumford AC, et al. 2017. Environmental signatures and effects of an oil and gas wastewater spill in the Williston Basin, North Dakota. Sci Total Environ 579:1781–1793, *https://doi.org/10.1016/j.scitotenv.2016.11.157*.

3. US Environmental Protection Agency (US EPA). 2016. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report, EPA/600/R-16/236F). Washington, DC: US EPA. Available: *https://cfpub.epa.gov/ncea/hfstudy/recordisplay.cfm?deid=332990*.

4. Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, et al. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. Proc Natl Acad Sci USA 112:6325–6330, *https://doi.org/10.1073/pnas.1420279112*.

5. Ariana A, Cozzarelli I, Danforth C, McDevitt B, Rosofsky A, Vorhees D. 2025. Pathways for potential exposure to onshore oil and gas wastewater: what we need to know to protect human health. GeoHealth, 9:e2024GH001263, *https://doi.org/10.1029/2024GH001263*.

6. Shaheen SW, Wen T, Herman A, Brantley SL. 2022. Geochemical Evidence of potential groundwater contamination with human health risks where hydraulic fracturing overlaps with extensive legacy hydrocarbon extraction. Environ Sci Technol 56:10010–10019, *https://doi.org/10.1021/acs.est.2c00001*.

7. Shaheen SW, Wen T, Zheng Z, Xue L, Baka J, Brantley SL. 2024. Wastewaters coproduced with shale gas drive slight regional salinization of groundwater. Environ Science Technol 58:17862–17873, *https://doi.org/10.1021/acs.est.4c03371*.

8. Brantley SL. 2011. Shale Network Database, Consortium for Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI). Available: *https://doi.org/10.4211/his-data-shalenetwork*.

9. Li Z, You C, Gonzales M, Wendt AK, Wu F, Brantley SL. 2016. Searching for anomalous methane in shallow groundwater near shale gas wells. J Contam Hydrol 195:23–30, *https://doi.org/10.1016/j.jconhyd.2016.10.005*.