



NUMBER 243
OCTOBER 2026

RESEARCH REPORT

A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development

James E. Saiers, Mario Soriano Jr., Joshua Warren, Reed Maxwell, Shuqi Lin, Yingcan Cathy Wang, David van Velden, and Katie Baker

INCLUDES A COMMENTARY BY THE HEI ENERGY REVIEW COMMITTEE

www.healtheffects.org
www.HEIenergy.org



A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development

James E. Saiers, Mario Soriano Jr., Joshua Warren, Reed Maxwell,
Shuqi Lin, Yingcan Cathy Wang, David van Velden, and Katie Baker

with a Commentary by the HEI Energy Review Committee

Research Report 243
Health Effects Institute
Boston, Massachusetts

Trusted Science · Cleaner Air · Better Health

Publishing history: This report was posted at www.healtheffects.org in October 2026.

Citation for report:

Saiers JE, Soriano Jr. M, Warren J, Maxwell R, Lin S, Wang YC, et al. 2026. A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development. Research Report 243. Boston, MA: Health Effects Institute.

© 2026 Health Effects Institute, Boston, MA, USA. Compositor: David Wade, Virginia Beach, VA. Cover photos by A.W. Eustes III, Colorado School of Mines. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

Contents of this report may not be used without prior permission from the Health Effects Institute. Please send requests to pubs@healtheffects.org.

Health Effects Institute and HEI are service marks registered in the US Patent and Trademark Office.

ISSN 2688-6855 (online)

CONTENTS

About HEI Energy	v
About This Report	vi
HEI ENERGY STATEMENT	1
INVESTIGATORS' REPORT <i>by Saiers et al.</i>	5
Abstract	5
Introduction	5
Specific Aims	6
Modeling Approaches and Data Analysis	7
Vulnerability Modeling	7
<i>Groundwater Flow Model and Ensemble Calibration</i>	7
<i>Particle Tracking and Vulnerability Assessment</i>	8
<i>UOGD Spill Data</i>	9
Associations Between Vulnerability and Well Water Chemistry	10
<i>Water Quality Data: Chemical Fingerprints of UOGD Impacts</i>	10
<i>Statistical Modeling</i>	10
Results	11
Drinking Water Vulnerability to UOGD Spills	11
<i>Groundwater Flow Simulation</i>	11
<i>Particle Tracking</i>	11
<i>Vulnerability Assessment</i>	11
Association Between Vulnerability and Fingerprints of UOGD Spills	13
Discussion	14
Contaminant Transport Trajectories in Context of Setback Distances	14
Vulnerable Populations and Vulnerability Hot Spots	14
Vulnerability Is Associated with Chemical Fingerprints of UOGD Spills	14
Limitations	15
Conclusions	15
Data Availability Statement	16
Acknowledgments	16
References	16
HEI Quality Assurance Statement	19
Supplementary Appendices	19
About the Authors	20
Other Publications Resulting from This Research	21

COMMENTARY	23
INTRODUCTION	23
SCIENTIFIC AND REGULATORY BACKGROUND	23
UOGD Overview	23
UOGD and Water Quality	24
STUDY OBJECTIVES	24
SUMMARY OF METHODS AND STUDY DESIGN	24
Aim 1. Estimating Well Water Vulnerability to Potential Contamination from UOGD in the Marcellus Region	24
Aim 2. Evaluating Associations Between Well Water Vulnerability and UOGD Chemical Fingerprints in Well Water Measurements	25
SUMMARY OF KEY RESULTS	26
Aim 1. Well Water Vulnerability to Potential Contamination from UOGD in the Marcellus Region	26
Aim 2. Associations Between UOGD Chemical Fingerprints in Well Water Measurements and Well Water Vulnerability	26
HEI ENERGY REVIEW COMMITTEE'S EVALUATION	27
Study Design, Datasets, and Analytical Approaches	28
Findings and Interpretation	28
Conclusions	28
ACKNOWLEDGMENTS	29
REFERENCES	29
Abbreviations and Other Terms	30
HEI Board, Energy Committees, and Staff	31

ABOUT HEI ENERGY

The Health Effects Institute's Energy Research program (HEI Energy) was formed to identify and conduct high-priority research on potential population exposures and health effects from the development of oil and natural gas in the United States. Since 2022, HEI Energy has supported population-level exposure research in multiple oil and gas regions. This research followed an extensive planning process that included preparing reviews of the scientific literature, hosting multisector workshops to learn about research priorities, and developing an online curated database and spatial bibliography to advance both public and scientific understanding. The research scope of HEI Energy is expanding beyond oil and gas to other forms of energy development, with an overarching goal of providing impartial knowledge about the benefits and drawbacks associated with various technologies.

The peer-reviewed scientific research provided by HEI Energy contributes high-quality and credible science to the public debate about unconventional oil and natural gas development and provides support needed for decisions about how best to protect public health. To achieve this goal, HEI Energy has put into place a governance structure that mirrors the one successfully employed for nearly 40 years by its parent organization, the Health Effects Institute (HEI), with several critical features:

- Balanced funding from the US Environmental Protection Agency under a contract that funds HEI Energy exclusively and from the oil and natural gas industry, with other public and private organizations periodically providing support
- An independent Board of Directors consisting of leaders in science and policy who are committed to fostering the public-private partnership that is central to the organization
- A research program governed independently by individuals having no direct ties to or interests in sponsor organizations
- An HEI Energy Research Committee, consisting of internationally recognized experts in one or more subject areas relevant to the Committee's work, that conducts and reviews scientific research impartially and has been vetted to avoid conflicts of interest
- Research that undergoes rigorous peer review by the HEI Energy Review Committee, which is not involved in the selection and oversight of HEI Energy studies
- Staff and committees that participate in open and extensive stakeholder engagement before, during, and after research and communicate all results in the context of other relevant research

In addition, HEI Energy publicly shares all literature reviews and original research that it funds, along with summaries written for a general audience. Without advocating policy positions, it provides impartial science targeted to make better-informed decisions.

HEI Energy is funded separately from the Health Effects Institute's other research programs (www.healtheffects.org), with financial support from the US Environmental Protection Agency, the oil and gas industry, and private foundations.

ABOUT THIS REPORT

HEI Energy Research Report 243, *A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development*, presents a research project funded by HEI Energy and conducted by James E. Saiers at the Yale School of the Environment and colleagues. The report contains three main sections:

The **HEI Statement**, prepared by staff at HEI Energy, is a brief, nontechnical summary of the study and its findings; it also briefly describes the HEI Energy Review Committee's comments on the study.

The **Investigators' Report**, prepared by Saiers and his colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The **Commentary**, prepared by members of the HEI Energy Review Committee with the assistance of HEI staff, places the study in a broader scientific context, points out its strengths and limitations, and discusses the remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI Energy's rigorous review process. When an HEI Energy-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. Outside technical reviewers first examine this draft report. The report and the reviewers' comments are then evaluated by members of the Review Committee, an independent panel of distinguished scientists who are not involved in selecting or overseeing HEI Energy studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. This report also has not been reviewed by private party institutions, including those that support HEI Energy, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

HEI STATEMENT

Synopsis of Research Report 243

Potential Groundwater Vulnerability to Spills Associated with Unconventional Oil and Gas Development Across the Marcellus Region

BACKGROUND

Over the past two decades, unconventional oil and gas development (UOGD)* has allowed for a significant increase in oil and gas development across the United States.

UOGD-related releases to groundwater and surface water may occur under accidental conditions (e.g., spills). Thus, their potential for groundwater contamination has garnered much public concern, especially in the Marcellus region where there is considerable UOGD alongside populations that rely heavily on private wells for domestic water. However, understanding remains limited regarding how UOGD releases to groundwater and surface water could lead to population exposures through contaminated water.

Dr. James E. Saiers at the Yale School of the Environment and colleagues — Drs. Mario Soriano Jr., Joshua Warren, and Reed Maxwell — developed a groundwater modeling approach to improve our understanding of the potential for documented UOGD spills in the Marcellus region to adversely affect water quality that may lead to population exposures. Their study was funded through HEI Energy's Request for Qualifications [RFQ E23-2: Regional Groundwater Modeling to Understand Community Exposures Associated with Oil and Gas Development](#).

APPROACH

Saiers and colleagues developed a modeling approach to evaluate the vulnerability of groundwater wells to potential contamination from spills associated with UOGD well pads in the Marcellus region states of Ohio, Pennsylvania, and West Virginia. They also sought to evaluate potential associations between their estimates of well water vulnerability and UOGD-related chemical fingerprints, indicating potential contamination from UOGD.

What This Study Adds

- This study provides a modeling approach to estimate groundwater vulnerability of well water to potential contamination from spills associated with UOGD in the Marcellus region.
- The team found that up to 122,000 people who rely on domestic water wells live in locations where groundwater is vulnerable to such contamination.
- The team collected data on documented UOGD spills in the region and well water chemistry. They evaluated associations between their estimates of vulnerability of well water and the occurrence of chemical fingerprints resembling UOGD-produced waters in groundwater-supplied well water samples.
- The team found higher odds of detecting chemical fingerprints of UOGD in well water samples in areas that were classified as vulnerable to documented spills from well pads. This finding suggests that domestic water wells might be vulnerable to potential chemical contamination from UOGD.
- This study provides an important contribution to our knowledge of the locations of groundwater-supplied wells most vulnerable to UOGD spills. The modeling framework and datasets can inform decision-making related to water quality and UOGD, including potential setback distances from drinking water sources and spill investigation.

For the modeling approach, the investigators combined a two-dimensional physics-based groundwater flow model with a particle-tracking algorithm to simulate the transport trajectories of contaminants through groundwater in the region following their release from a UOGD well pad. They performed multiple model simulations and integrated the results to generate estimates of well water vulnerability

This Statement, prepared by HEI Energy, summarizes a research project funded by HEI Energy and conducted by James E. Saiers at the Yale School of the Environment and his colleagues. Research Report 243 contains the detailed Investigators' Report and a Commentary on the study prepared by the HEI Energy Review Committee.

*A list of abbreviations and other terms appears at the end of this report.

throughout Ohio, Pennsylvania, and West Virginia. The estimates consisted of the probability that a potential contaminant associated with a UOGD spill would reach a groundwater-supplied domestic well across model simulations. The investigators defined “high vulnerability” as a vulnerability estimate greater than 0.5 (the level of agreement among the model simulations that the area lies in the path of contaminants released from a UOGD well pad).

Well water vulnerability estimates were computed to evaluate two scenarios: (1) all UOGD well pads in the region ($N = 25,617$), and (2) a subset of UOGD well pads in the region with documented spills ($N = 1,761$), using data collected from state repositories. Subsequently, the investigators calculated the potentially vulnerable land area and population for both scenarios. They also aggregated their estimates to various geographic scales to explore potential spatial clustering in the region.

Next, Saiers and colleagues used publicly available water quality data and statistical models to evaluate associations between chemical fingerprints of UOGD spills and their groundwater vulnerability estimates. They collected information on over 27,000 groundwater-supplied well water samples across the three states, of which over 16,000 could be included in their statistical models. They adjusted their models to control other factors potentially associated with UOGD-like chemical fingerprints, such as geologic features (e.g., folds, faults), topography, and the presence of nearby conventional oil and gas well pads.

KEY RESULTS

Saiers and colleagues found that the average distance of a potential contaminant’s trajectory from a well pad was 671 m, and that most estimated travel distances were not greater than 2 km. Nonetheless, estimated distances generally exceeded regulatory setback distances between UOGD and wells supplying drinking water in Pennsylvania, Ohio, and West Virginia.

They estimated that the total land area vulnerable to potential contamination from spills from all UOGD well pads (scenario 1) in the region was around 3,600 km². The investigators were able to find and compile 3,853 documented spills from UOGD wells across the three states; however, due to missing information, only 1,761 spills were included in the analysis (scenario 2). For this second scenario, the estimated total land area vulnerable to groundwater contamination was approximately 600 km².

The investigators found that as many as 122,000 people who rely on groundwater wells for domestic water live in locations that are potentially vulnerable to contamination under scenario 1, which considers all

UOGD well pads. This estimate amounts to about 3% of the 3.9 million people who use domestic wells within the Marcellus region. For areas with documented UOGD spills, 22,000 people live in locations that were estimated to be highly vulnerable to contamination of their water wells. When aggregated to areas of about 30 km, the investigators observed spatial clusters of high vulnerability to groundwater contamination throughout Northeastern Pennsylvania, Southern West Virginia, and the area where those two states meet with Ohio (**Statement Figure**).

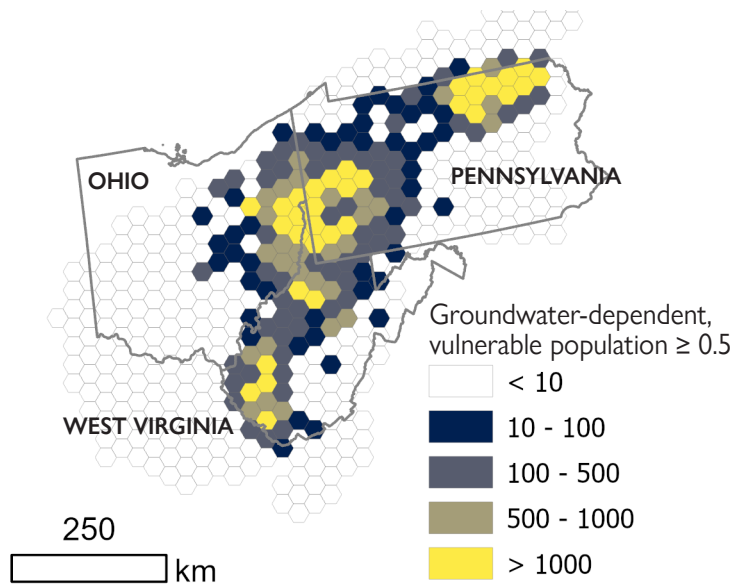
In areas with documented UOGD spills, Saiers and colleagues found positive associations between estimated well water vulnerability and the odds of detecting a UOGD-related chemical fingerprint in the groundwater samples. For example, the likelihood of detecting UOGD chemical fingerprints in water samples increased 13% (odds ratio 1.13) for every 0.1 increase in well water vulnerability. When modeled as a binary variable for high well water vulnerability, high vulnerability was associated with a 224% increase in the likelihood of detecting UOGD chemical fingerprints in water samples.

INTERPRETATION AND CONCLUSIONS

Saiers and colleagues developed a modeling approach to estimate groundwater vulnerability to potential contamination from UOGD spills in the Marcellus region states of Ohio, Pennsylvania, and West Virginia, and explored potential associations between estimated well water vulnerability and chemical fingerprints of UOGD in water samples.

In its independent review of the study, the HEI Energy Review Committee concluded that this study provided an effective modeling approach for examining groundwater vulnerability to potential contamination from UOGD spills that may be useful to inform decision-making. It also appreciated that this approach could be extended to other regions in the United States.

The Committee noted that the study was unavoidably limited by missing information for a large fraction of documented spills that could not be included in the analysis. This missing information may have resulted in an underestimation of well water vulnerability in areas with documented spills. They also noted some uncertainties associated with some simplifying assumptions used in the groundwater modeling (such as a two-dimensional rather than three-dimensional representation). Nonetheless, the Committee appreciated that the study’s findings, and its limitations, were thoughtfully and prudently presented. Moreover, they found that these limitations did not meaningfully change their conclusions about the utility of the modeling framework.



Statement Figure. Estimated number of people who depend on private wells and live in locations estimated to be highly vulnerable to potential groundwater contamination from UOGD spills in the Marcellus region. Source: Adapted from Investigators' Report Figure 7.

Overall, this study found that up to 120,000 people who rely on groundwater-supplied domestic water wells in the Marcellus region live in areas with “high vulnerability” to potential groundwater contamination associated with UOGD spills. Well waters determined to be vulnerable to documented spills from UOGD well pads were more likely to exhibit chemical fingerprints of UOGD produced water, although these fingerprints are not equivalent to exceedances of either health-based or aesthetics-based standards. The study provides a strong modeling framework and valuable datasets that can be used to inform decision-making by helping to predict the location of drinking water wells that are most vulnerable to UOGD spills.

A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development*

James E. Saiers^{1†}, Mario Soriano Jr.^{2‡}, Joshua Warren^{1†}, Reed Maxwell^{2‡}, Shuqi Lin¹, Yingcan Cathy Wang², David van Velden², Katie Baker¹

¹*Yale School of the Environment, New Haven, Connecticut, USA;* ²*Princeton University, Princeton, New Jersey, USA*

ABSTRACT

Unconventional oil and gas development (UOGD),^{††} the extraction of hydrocarbons from shale and other low-permeability formations, was made possible by technological advances in horizontal drilling and hydraulic fracturing. With favorable policy and market drivers, the industry's dramatic expansion throughout the mid-2000s and 2010s has bolstered domestic energy production and the economy in the United States. Nevertheless, public concerns about the water- and chemical-intensive industry's potential impacts on drinking water resources, the environment, and human health remain unabated. In this report, we use high-resolution physically based groundwater flow-and-transport modeling to quantify the vulnerability of domestic well waters to contamination from wastewater spills at UOGD well pads. We applied this modeling framework to a 305,000 km² region of the Appalachian Basin where more than 20,000 UOGD wells have been drilled to extract hydrocarbons from the Marcellus and Utica-Point Pleasant shales. We estimate that 3% of the 3.9 million residents served by domestic wells reside in locations where groundwater is vulnerable to contamination from contemporary UOGD activities. Using water quality data from 16,000 groundwater samples, our analysis indicates that high vulnerability to contamination is associated with a three-

fold increase in the odds of detecting produced water-like signatures in domestic well waters, independent of other contributing factors. Our framework is transferable to oil- and gas-producing regions beyond the Appalachian Basin and has applications to a range of scientific, management, and policy activities, including exposure assessment in health studies, optimization of water quality monitoring, and the establishment of regulatory setback distances.

INTRODUCTION

More than 20,000 unconventional oil and gas wells have been drilled and hydraulically fractured in the Marcellus region since 2004.¹ While unconventional oil and gas development (UOGD) has made natural gas plentiful, created jobs, and increased economic activity, it has been accompanied by persistent concerns that it contaminates drinking water. Near the onset of the Marcellus drilling boom, groundwater contamination by methane released from improperly constructed gas wells gained national attention.^{2–6} As UOGD expanded through the Marcellus region, public apprehension remained unabated, partly reflecting growing recognition of the potential for UOGD to pollute groundwaters with methane and other chemicals that may imperil human health.^{7–9}

UOGD involves the management of large volumes of freshwater, hydraulic fracturing fluids, and wastewater in the form of flowback, produced water, and spent drilling fluids.^{10–15} The hydraulic fracturing fluids and wastewaters, which can contain known and suspected carcinogens, endocrine disruptors, and reproductive toxicants, have been released into the environment, primarily at UOGD well pads but also into surface waters and along transport corridors between well pads and wastewater disposal and treatment facilities.^{16–20} Releases at well pads have resulted from spills during fluid transfer and storage, seepage from impoundments, and leaks from flowlines, tanks, and other UOGD infrastructure.^{21,22}

Spills and other fluid releases on and beneath UOGD well pads have been implicated as sources of water quality impairment within residential wells of the Marcellus region.^{23–27} Nevertheless, contaminant-source attribution is often equivocal owing to the occurrence of alternative anthropogenic and natural sources. For example, elevated salinity in

This Investigators' Report is one part of Health Effects Institute Research Report 243, which also includes a Commentary by the HEI Energy Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. James E. Saiers, Yale School of the Environment, 195 Prospect Street, New Haven, CT 06511; email: james.saiers@yale.edu. No potential conflicts of interest were reported by the authors.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. This report has also not been reviewed by private-party institutions, including those that support HEI Energy, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

* Portions of this report are reproduced or adapted from *Soriano et al. 2026*, published in *Nature Sustainability* under Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

† Co-principal investigators.

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

drinking water wells could stem from a produced water spill or could reflect natural upwelling of deeper basin brines that occurs beneath valleys throughout the Appalachian Basin.^{28,29} Moreover, organic compounds added to drilling and hydraulic fracturing fluids and found in UOGD wastewaters have other industrial uses and may also be sourced from coal mining operations and conventional oil and gas development (COGD), which co-occur with UOGD in parts of the Marcellus region.³⁰ Isotopic compositions, elemental ratios, and noble gases have been used as chemical fingerprints to distinguish UOGD from non-UOGD contamination,^{8,30–35} but chemical fingerprinting is retrospective and cannot be employed in a predictive capacity, which is essential if a goal is to inform actions to prevent exposure.

Groundwater flow-and-transport modeling can support inferences on contamination risks, but, with few exceptions,^{26,36,37} has not been applied to the UOGD problem. In the most basic sense, these models can support causal inference by illuminating whether contaminant transport from a source (e.g., UOGD well pad) to a drinking water receptor (e.g., household well) is hydrologically and geochemically plausible. Before this project, our team engaged in groundwater modeling for the purpose of understanding linkages between groundwater quality and UOGD in the Marcellus region. We have created hydrologic-based models that simulate groundwater flow patterns and solute transport for the purposes of quantifying the vulnerability of domestic wells to water quality impairments from UOGD.^{36,37} We have employed these models, in conjunction with our own water quality measurements, to distinguish between chemical impairments caused by UOGD and those related to other sources, including landfills, COGD, coal mining, and industrial activities,^{23,26,30} and we have gone on to explore associations between hydrologic vulnerability to UOGD contamination and sociodemographic characteristics, including income, race, and educational status.³⁸

We have extended this past work in substantial ways to support the objectives of this HEI-funded project. In particular, we have (1) expanded the geographic coverage of our vulnerability assessment to the entire Marcellus Shale region; (2) quantified the distribution of vulnerability to UOGD-related spills throughout this region at high spatial resolution; and (3) evaluated associations between our estimates of vulnerability and 16,000 water quality measurements from domestic wells that tap shallow aquifers above the Marcellus Shale. Going forward, we believe our modeling framework could be used to support exposure assessment for health studies, prioritize sampling locations for water quality monitoring programs, inform decisions on regulatory setback distances, and illuminate relationships between sociodemographic factors and areas at high risk from UOGD-sourced contamination. Our approach has been intentionally designed to be computationally tractable, amenable to updating as new spills occur or new information becomes available, and adaptable to other oil- and gas-producing regions. We hope our contributions will spur new collaborations with other researchers working

at the nexus of extractive-industry impacts, water quality, and human health.

SPECIFIC AIMS

Our work was guided by two aims:

1. **Advance a physics-based model for well water vulnerability to contamination from UOGD spills that is suitable for application across the entire Marcellus region.** Vulnerability quantifies the probability of a contaminant reaching a drinking water receptor upon release from a source within or above the aquifer. It depends on local groundwater flow direction and the spatial relationships between potential sources and receptors. As a basis for estimating vulnerability, we constructed and calibrated a high-resolution groundwater flow model to simulate aquifer flow patterns across portions of Ohio, Pennsylvania, and West Virginia that are underlain by the Marcellus Shale. We used the flow fields estimated from the groundwater model in a particle-tracking algorithm to predict the transport trajectories (path lines) of contaminants following their release from UOGD well pads. Through analysis of path lines generated from multiple model realizations, we inferred and mapped the spatial distribution of hydrologic vulnerability. The results illuminate the susceptibility of locales to water quality impacts arising from UOGD-related spills.
2. **Evaluate associations between UOGD fingerprints measured in well waters and vulnerability, adjusting for other features of the study area.** We explored associations between modeled-predicted vulnerability and chemical fingerprints of UOGD spills in domestic well waters using logistic regression analysis. We expected that, if we reasonably approximated hydrologic conditions that could lead to spill-associated impairments, then we should observe associations between spatially explicit estimates of vulnerability and the occurrence of spill-related chemical fingerprints within well water samples. Chemical fingerprints of UOGD spills were identified based on ion ratios. In accordance with other research groups,^{8,31,32} we used bivariate combinations of ion ratios to distinguish unimpacted meteoric waters from those affected by mixing with oil- and gas-produced waters that comprise a substantial fraction of fluids that are spilled at UOGD well pads. Additionally, sensitivity analyses were conducted to evaluate the robustness of the findings.

The project aims involve the development and application of physics-based and statistical models that draw from various kinds of data and lead to specific outcomes (**Figure 1**). Collectively, these outcomes are intended to elucidate linkages (if any) between UOGD-related spills and the chemical quality of residential well waters of the Marcellus Shale region. A detailed description of the modeling approaches and data analyses is presented in the following section, organized according to the Research Roadmap (**Table 1**).

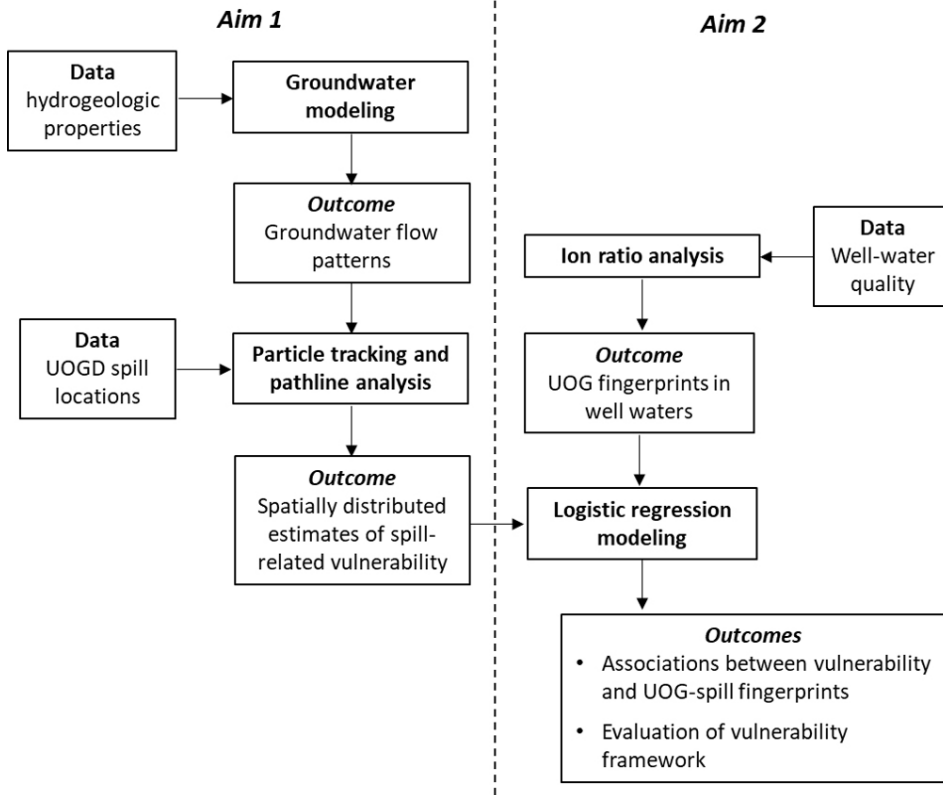


Figure 1. Interdependencies among project aims, their respective outcomes, and data used in support of these aims.

Table 1. Research Roadmap

Aims and Research Conducted	Methods Description
Aim 1 — Vulnerability Modeling	Section 3.1
Groundwater flow model construction and ensemble calibration	Section 3.1.1
Particle tracking	Section 3.1.2
UOGD spills data	Section 3.1.3
Aim 2 — Statistical Model for Vulnerability/UOGD Fingerprint Associations	Section 3.2
Water quality data	Section 3.2.1
Logistic regression model	Section 3.2.2

UOGD = unconventional oil and gas development.

MODELING APPROACHES AND DATA ANALYSIS

VULNERABILITY MODELING

Groundwater Flow Model and Ensemble Calibration

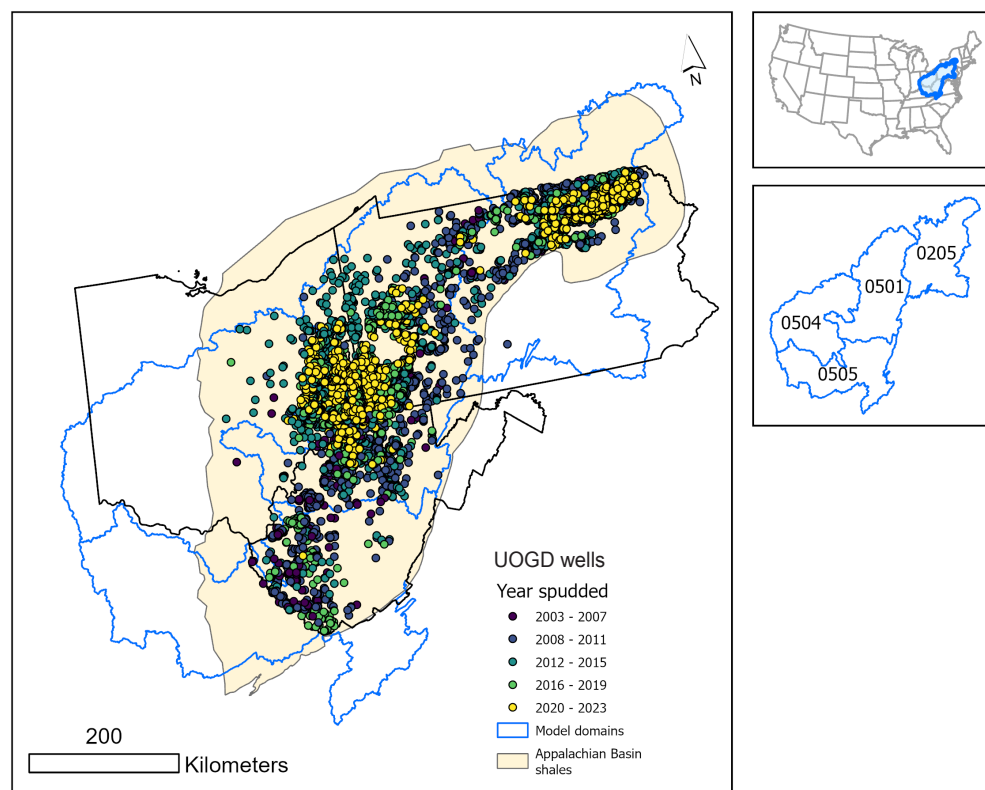
Groundwater vulnerability was quantified through hydrologic modeling coupled with particle tracking across calibration-constrained parameter ensembles. Our groundwater flow models were based on MODFLOW 6 models of two-dimensional flow through the unconfined, surficial aquifer systems of the conterminous US (CONUS) developed by Zell and Sanford.³⁹ Their CONUS-wide simulation consists of 75 separate model domains defined by individual **HUC4 subregions** (four-digit hydrologic unit codes that represent large watershed subregions) or groups of contiguous HUC4 polygons. We selected the four simulation domains encompassing our study region: domain 0205, 0501, 0504, and 0505 (Figure 2).

Each simulation domain was discretized into a Cartesian grid of 250 m × 250 m cells. The top boundary corresponds to the land surface, with elevations upscaled from the 30 m National Elevation Dataset. Bottom elevations were varied spatially, resulting in variable domain thicknesses with a median of 74 m in our study region. While we retained the geometry of the Zell and Sanford models, we also explored potential sources of uncertainty in forcing and parameterization. Specifically, we examined two versions of climatic forcing and two versions of geological parametrization, resulting in four model setups derived from their combinations (Appendix A).

Ensembles comprising 1,000 realizations of hydraulic conductivity parameter fields were developed for each setup through model calibration using PESTPP-IES,* which implements the Gauss-Marquardt-Levenberg algorithm with an iterative ensemble smoother.⁴⁰

We used the same set of target observations collated by Zell and Sanford³⁹ from the National Water Information System, National Hydrography Dataset, and National Wetlands Inventory. The National Water Information System observations

*Calibration refers to the systematic adjustment of model parameters to minimize residuals between available field (target) observations and their corresponding model outputs.



Model Domain ID	HUC4 Units	Area (km ²)	<i>n</i> UOGD Wells
0205	Susquehanna (0205)	71,000	7,523
0501	Allegheny (0501), Monongahela (0502), Upper Ohio (0503)	84,000	15,386
0504	Muskingum (0504), Scioto (0506), Great Miami (0508), Middle Ohio (0509)	75,000	1,488
0505	Kanawha (0505), Big Sandy-Guyandotte (0507), Kentucky-Licking (0510)	75,000	1,220

Figure 2. Study region, model domains, HUC4 units, and UOGD wells. Footprint of Appalachian Basin shales includes Marcellus, Utica-Point Pleasant, and Upper Devonian shales. Smaller panels depict location of study region within the United States and the identifiers for computational domains of groundwater models.

were average groundwater levels from wells with a time series of at least 10 measurements. We considered, but ultimately did not use, static water levels reported in water well drill logs because these measurements are often collected before newly installed wells have reached hydraulic equilibrium and therefore may not accurately represent ambient groundwater levels. Perennial, first-order streams from the National Hydrography Dataset and freshwater wetlands larger than 20 hectares from the National Wetlands Inventory were interpreted as land surface expressions of the steady-state water table, such that a water table depth of zero was imposed at their centroids for calibration. The number of target observations used for calibration was 9,026 (0205), 10,995 (0501), 5,274 (0504), and 6,698 (0505).

Particle Tracking and Vulnerability Assessment

We performed particle tracking for each realization of the hydraulic conductivity field in our calibrated ensembles. Forward particle tracking was implemented through MODPATH 7, with particles initialized in two ways: (1) at grid cells containing UOGD wells and (2) at grid cells containing the subset of UOGD wells from which spills were documented and could be positively geolocated with UOGD well pads. In each scenario, the released particles were tracked forward until they were discharged from the groundwater flow system. Each grid cell containing UOGD was assigned nine particles arranged in a 3×3 configuration to account for all possible directions of the hydraulic gradient. The resulting

ensemble of path lines documents the steady-state trajectories of conservative solutes released from UOGD sites transported through the groundwater system by advection (**Figure 3**). A path line terminates when the particle is discharged, typically at a stream grid cell in which the simulated hydraulic head is higher than the land surface elevation. We used the default setting of the Pollock method, wherein tracking terminates upon encountering any face of a discharging cell, and particles that originate in discharge cells terminate upon release.⁴¹

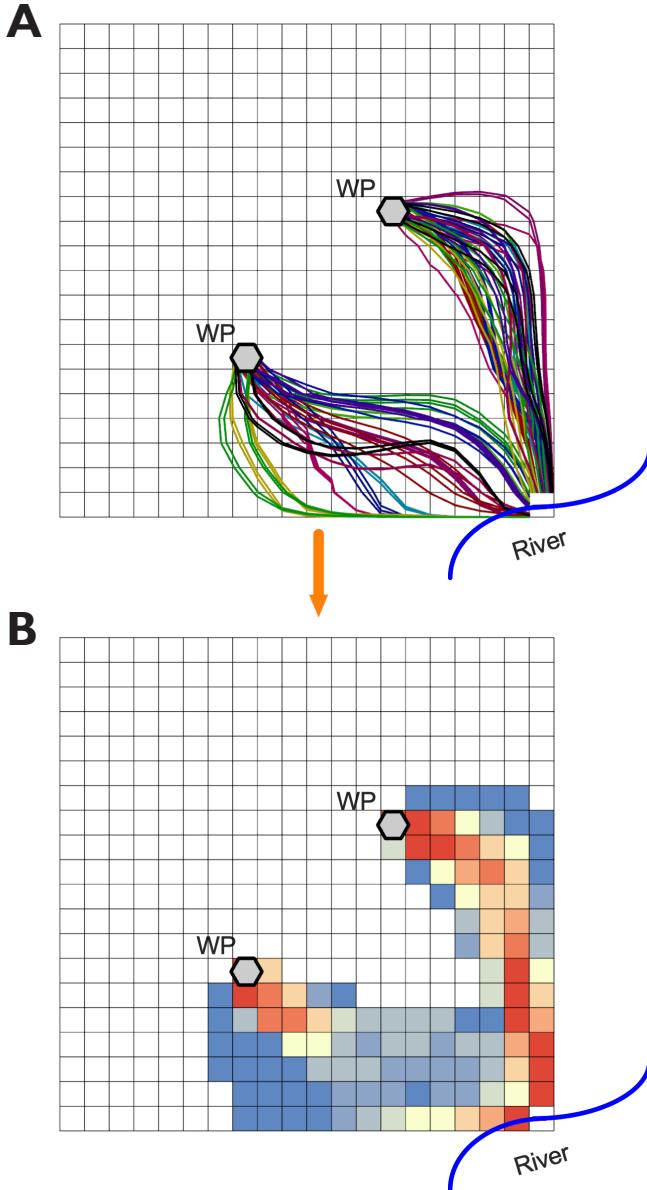


Figure 3. (A) Path lines from multiple realizations showing contaminant migration from two well pads (WP), and (B) vulnerability (V) computed with equation (2) using realizations shown in (A), where white, blue, and red grid cells denote areas of no, low, and high V, respectively.

For each grid cell i , we operationalized vulnerability to contamination from a UOGD spill as

$$V_i = n_{p,i}/n_T \quad (1)$$

where the numerator of (1), $n_{p,i}$, is the number of realizations in which a path line intersected grid cell i , while the denominator, n_T , is the total number of retained realizations in the ensemble. V_i ranges from 0 to 1, with higher values indicating greater agreement in the ensemble that a grid cell lies along an advective-transport path from UOGD sites (Figure 3). Areas with higher vulnerability are at greater risk of being affected by contaminant releases from UOGD sources, should they occur.

We calculated the total area and populations dependent on domestic groundwater wells within vulnerable grid cells across our study region using 2010 estimates by Johnson and colleagues (2019) using their block group method.⁴² In addition, we aggregated the original gridded results to census tracts and standard 30 km hexagonal polygons to examine broader spatial patterns.

UOGD Spill Data

We collated data on UOGD spills tabulated in databases and maintained by Ohio, Pennsylvania, and West Virginia regulatory agencies.^{43–45} Whenever possible, we extracted data on timing, location, volume, type, and pathway of fluid releases, although data on spill characteristics for individual events were commonly incomplete. Only spills assigned to coordinate locations and colocated with UOGD well pad locations were reported in public databases (<https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=1088>, <https://gis-odnr.opendata.arcgis.com/documents/3dce2c55a9e1486b9483d-b25710ae588/about>, and <https://tagis.dep.wv.gov/home/gis%20data>) and included in the final dataset.

Ohio We collected spill data through the OneCall and Emergency Program managed by Ohio's Department of Natural Resources. The Department's Division of Oil & Gas Resources Management provided these records in response to Public Records Request 8509 and through prior correspondence with the division's enforcement coordinator. The dataset includes details such as type of material released (e.g., brine, crude oil, natural gas), incident classification, estimated quantities released, and affected media (e.g., soil, water). In the initial phase of processing, we identified and corrected format inconsistencies, particularly for location descriptions, while also standardizing date formats and material classifications. We geocoded the dataset using the provided addresses and descriptive location data. For the remaining records that could not be geocoded, primarily because of incorrect or missing addresses, we attempted to link them to their corresponding well pad API numbers, when available. The processed dataset encompassed spills from 2016 to 2023.

Pennsylvania We used two government databases to gather information on unconventional oil and gas drilling-related spills associated with well inspections from 2015 to 2023. The first

database, PA.gov Compliance (<https://greenport.pa.gov/Reports/Extracts/OG/OilComplianceReport>), contained summaries of spill-related information, including inspection notes in tabular format. However, the database lacked geographic locations. We obtained coordinates by matching the inspection identification number to the second database (http://cedatareporting.pa.gov/Reportserver/Pages/ReportViewer.aspx?/Public/DEP/OG/SSRS/OG_Inspection_Docs). To identify inspections related to spills, we filtered the inspection descriptions for specific keywords indicating a potential spill (“dumped,” “spill,” “deposited,” “leak,” “dispose,” “leaking,” “discharge,” “failure to plug,” “release,” “overflow”). We observed that if multiple violations were reported, multiple rows would be linked to the same inspection IDs. We removed these duplicates, such that no spill would be represented twice. In the absence of an automated filtering method, we manually examined these cases to determine the spill details, including the amount, substance, and location. After a thorough analysis of the inspection descriptions, we excluded cases that did not indicate any spills. The processed dataset encompassed spills from 2015 to 2023. We supplemented these spill data with UOGD spills collated by Patterson and colleagues (see Supporting Information)²² for the period 2007–2014.

West Virginia We obtained spill data for West Virginia from two databases provided by the West Virginia Department of Oil and Gas upon request. The Spill System, managed by West Virginia Homeland Security and Emergency Management since 2009, collates reports through a public hotline and includes numerous kinds of incidents, some of which are unrelated to oil and gas development. The Risk-Based Data Management System, introduced in 2012 by the West Virginia Office of Oil and Gas, records spills from oil and gas operations and partially overlaps with the Spill System. We cleaned and organized the data by removing duplicate entries and irrelevant records, while standardizing formats and extracting relevant information, when available, on spill coordinates and characteristics. The processed dataset encompassed spills from 2016 to 2023.

ASSOCIATIONS BETWEEN VULNERABILITY AND WELL WATER CHEMISTRY

Water Quality Data: Chemical Fingerprints of UOGD Impacts

We did not collect new water quality data for this project but relied instead on published observations. Data on the composition of domestic well waters for Ohio, Pennsylvania, and West Virginia were acquired from the Shale Network (www.shalenetwork.org), as well as from peer-reviewed studies reporting water quality measurements that could be assigned corresponding coordinate locations.^{26,30,46} The water quality data were analyzed for potential fingerprints of UOGD-related chemical impairments stemming from inputs of oil- and gas-produced waters, which comprise a substantial fraction of fluids spilled from UOGD well pads. We used published prescriptions of boundary lines within a bivariate space of

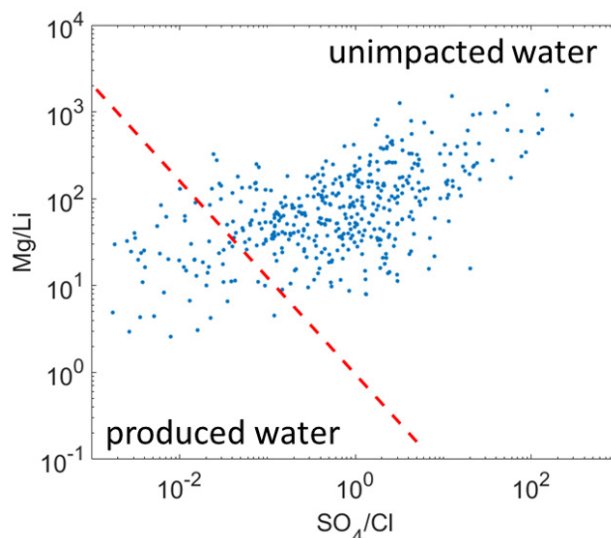


Figure 4. Bivariate plot of ion ratios for distinguishing unimpacted groundwaters from groundwaters influenced by produced-water inputs.

ion-ratio pairs (e.g., Mg/Li vs. SO_4/Cl in **Figure 4**) to delineate fields of unimpacted groundwater and groundwater affected by mixing with oil- and gas-produced waters.^{8,26,31,32} We then classified water samples as unimpacted or UOGD-imprinted depending on which field, according to their ion ratios, they occupied. We conducted this classification procedure with six different combinations of ion-ratio pairs (i.e., Ba/Cl vs. Br/ SO_4 , SO_4/Cl vs. Br, Ca/Sr vs. Ca/Mg, SO_4/Cl vs. Mg/Na, Mg/Li vs. SO_4/Cl , and Mg/Li vs. Br/ SO_4). Interpretation of these elemental ratios could overestimate the number of spill-imprinted water samples because deep-basin brines, a component of oil- and gas-produced waters, may occur naturally, particularly in groundwaters beneath valleys or near fault zones where natural upwelling of deeper groundwaters may occur. We attempted to control for these effects in our regression analysis as explained in the Statistical Modeling section.

Statistical Modeling

We used logistic regression to quantify the association between UOGD fingerprints at well locations across the study region and our corresponding estimates of vulnerability to documented spills. Specifically, the regression model is given as

$$Y_{it} | p_{it} \sim \text{Bernoulli}(p_{it}), i = 1, \dots, n; t = 1, \dots, m_i \text{ with } \text{logit}(p_{it}) = \beta_0 + \beta_1 V_{it} + \mathbf{x}_i^T \boldsymbol{\eta} + \mathbf{z}_i^T \boldsymbol{\gamma} \quad (2)$$

where Y_{it} is the binary indicator of a UOGD fingerprint in measurement t taken from well i (n total wells with m_i measures per well); p_{it} is the probability of observing a UOGD fingerprint; V_{it} is the location- and time-specific vulnerability accumulated from all spills reported before the year that the sample was collected; and $\mathbf{x}_i$ is a vector of covariates, including

an intercept, a categorical variable representing topographic position (peaks/summits, upper slopes, lower slopes, and valleys), Euclidean distance to geological structures (faults and folds), and a categorical variable indicating the presence or absence of conventional oil and gas wells within 2 km of the sample. Reference categories were peaks/summits for topographic position and absence of conventional oil and gas wells within 2 km for the well indicator. (The choice of reference category does not affect statistical inference, as any pairwise comparison between categories remains estimable.) The 2-km distance was selected as it encompassed over 90% of our simulated path lines. Spills were converted to year-specific indicator rasters aligned with the vulnerability maps, and sample vulnerability was determined based on whether spills were documented prior to the sample collection year. Topographic position was assigned using the multiscale topographic position index.⁴⁷ Geological structures^{45,48–50} and conventional oil and gas wells^{43–45} were compiled from publicly available shapefiles, and Euclidean distance rasters from such features were produced using the same grid as the vulnerability estimates.

Models were fit with and without $\mathbf{x}_i$ (i.e., adjusted and unadjusted) to better understand the association with UOGD spill risk before and after accounting for other known factors that may influence the probability of the presence of a UOGD fingerprint. All unadjusted and adjusted models also accounted for geographical location through inclusion of latitude, longitude, and their interaction as additional covariates; each included in the $\mathbf{z}_i$ vector. These variables were standardized prior to model fitting to stabilize estimation. Sensitivity analyses were performed to evaluate the robustness of the findings by randomly permuting the dependent variable (Y_{it}) so that outcomes and covariates were no longer recorded at the same spatial location. Under this spatial misalignment, we would expect to see generally null associations between vulnerability and UOGD fingerprints.

RESULTS

DRINKING WATER VULNERABILITY TO UOGD SPILLS

Groundwater Flow Simulation

Fields of simulated hydraulic head showed a pronounced influence of topography, with the water table configuration following the land surface elevations (**Figure 5**). Thus, groundwater in the unconfined shallow aquifer flowed from upland areas, where the water table was comparatively high but beneath the surface, toward valleys where the water table intersected the surface and groundwater discharged to the stream network. This topography-controlled flow system is consistent with previous large-scale analyses^{51,52} and conceptual models of local groundwater hydrology for this region.⁵³ The post-calibration mean absolute error, calculated by comparing simulated heads with observations at 31,993 locations, ranged from 0.1 to 1.0 m, while the root mean squared error ranged from 0.9 to 1.6 m across the simulation domains and

all four model configurations (variable hydraulic conductivity parametrization and input forcing), indicating robust performance (Appendix A).

Particle Tracking

The hydraulic head fields from the ensemble of calibrated models ($n = 1,000$ realizations) were processed in a particle-tracking algorithm to delineate the steady-state trajectory of contaminants released via surface spills at UOGD well pads. The resulting distribution of path lines exhibited a positive skew, with median and mean values of 532 m and 671 m, respectively (**Figure 6**). The higher frequency of values in the first histogram bin reflects particles that originated within discharge cells and therefore had a path line length of zero. The distribution also reflects a downward bias of less than 177 m (diagonal length to a grid cell's center) in the path line lengths, as tracking terminates at the faces of discharge cells rather than the cell center. Although path lines longer than 5,000 m were simulated, most path line lengths did not exceed 3,000 m, with the 95th percentile at 1,790 m. The range of values reflects the heterogeneity of the flow field throughout the large study region. Results were consistent across alternative model setups, with differences apparent only in the upper ends of the distributions (Appendix A).

Vulnerability Assessment

We calculated groundwater vulnerability (V) from the particle-tracking results as the proportion of realizations for which a path line intersected a given location relative to the total number of realizations (see **Figure 3** and the Particle Tracking and Vulnerability Assessment section). Thus, $0 \leq V \leq 1$ is a measure of the agreement between different realizations that a given location is along the trajectory of contaminants released at unconventional oil and gas well pads. Considering only well pads with documented spills as sources, the area with $V \geq 0.1$ (i.e., intersected in at least 100 of 1,000 realizations) equaled 590 km². Meanwhile, considering all UOGD well pads as sources, the area with $V \geq 0.1$ expanded sixfold to 3,538 km², corresponding to 1.2% of the modeled area. With a lower threshold of $V \geq 0.001$, the vulnerable area increased to 616 km² and 3,758 km² considering documented spill locations and all UOGD wells, respectively. Most intersected grid cells (i.e., $V > 0$) had high vulnerability ($V \geq 0.5$), indicating robust agreement across realizations on the locations at risk, which could be attributed to the variability in the path lines mostly occurring at the subgrid scale.

Combining vulnerability with estimates of populations served by domestic groundwater wells,⁴² we calculated the aggregate number of people who rely on high-vulnerability ($V \geq 0.5$) groundwater. Our estimates ranged from 22,000 to 123,000, considering vulnerability to documented spills and to all UOGD wells, respectively. Considering all UOGD wells, we also found spatial clustering of populations with highly vulnerable groundwater in Northeastern Pennsylvania and in the tri-state area, with additional hotspots in Southern West Virginia (**Figure 7**). When expressed as a proportion of

Figure 5. Simulated hydraulic head (setup c1, realization 1 in ensemble). (A) Results for entire 305,000 sq km study region. Note the pronounced imprint of topography. (B) Detailed view of results for a 25 sq km window. Flowlines from the National Hydrography Dataset are also shown to illustrate the model's general accuracy in capturing the stream network.

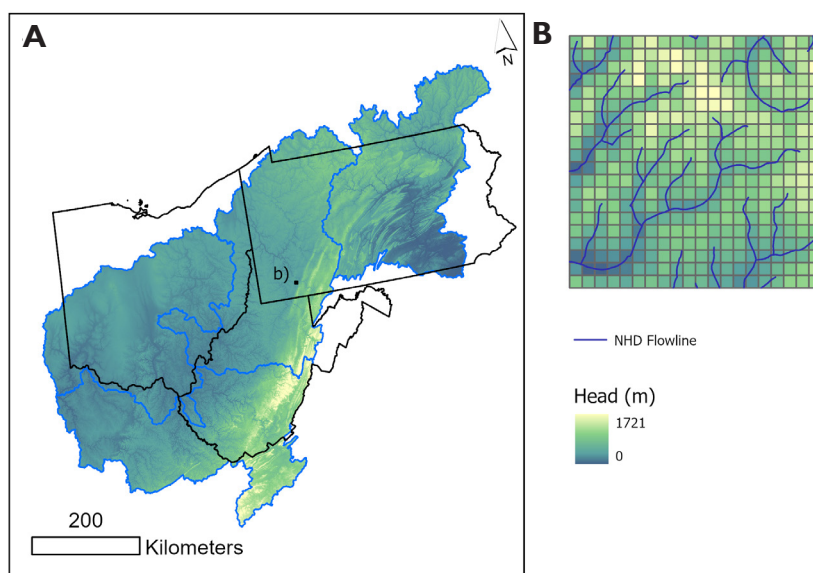


Figure 6. Histogram of path line length. Path lines represent the entire trajectory of a contaminant from initialization at a UOGD site to discharge from the groundwater system. The solid red line, dashed black line, and dotted gray line indicate the median, mean, and 95th percentile of the distribution, respectively. The final bin shown also includes path lines >5,000 m.

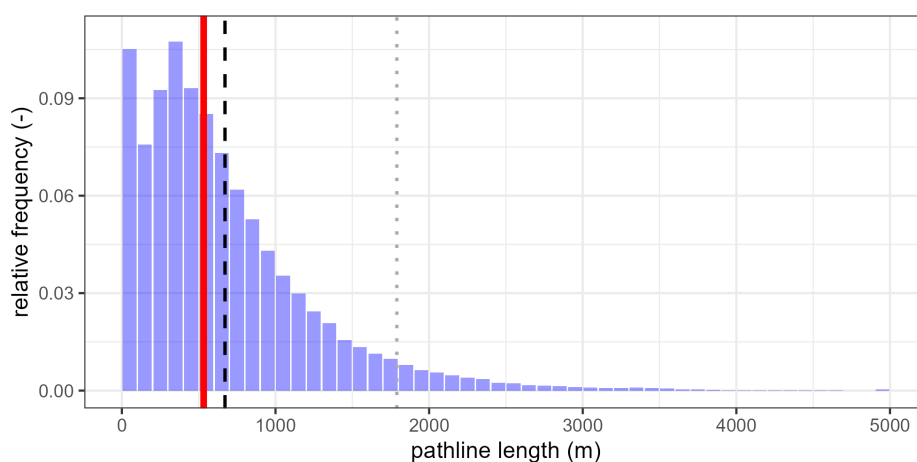
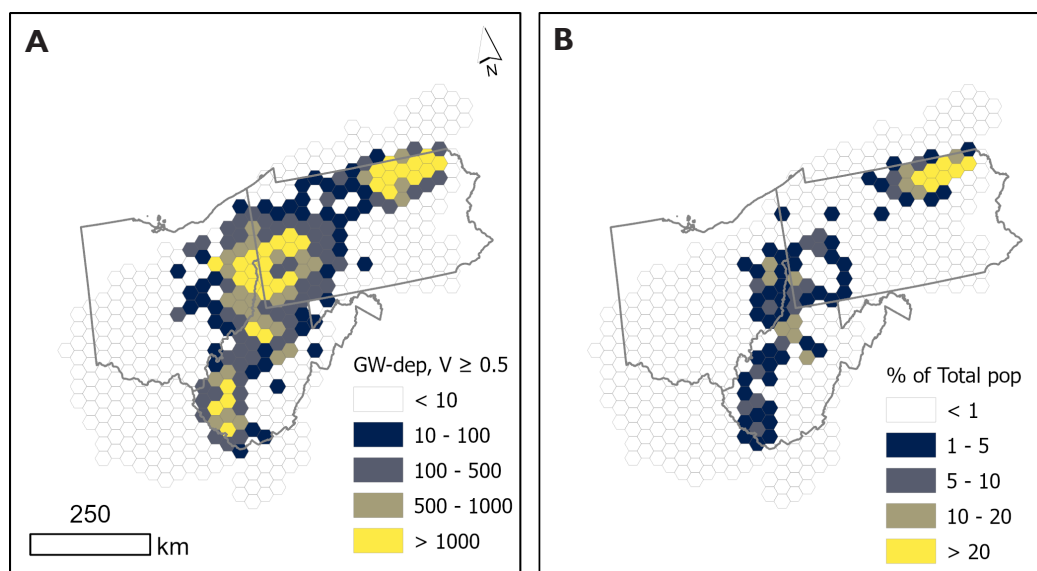


Figure 7. Groundwater-dependent populations served by domestic wells in locations highly vulnerable to contamination from UOGD. (A) Absolute numbers, and (B) percentage of total population within each 30-km hexagon.



total population (not just those served by domestic wells), Northeastern Pennsylvania displayed clustering of areas where more than 20% (up to a maximum of 39%) of local populations were colocated with groundwaters predicted to be highly vulnerable to contamination from UOGD. Several locales in the tri-state area also exhibited clusters where such proportions exceeded 10% (up to a maximum of 17%). The emergence of these hotspots may arise, in part, from regional-scale heterogeneity in local groundwater flow patterns and more substantially from the accumulation of vulnerability from multiple well pads within subregions where UOGD has been historically intense.

ASSOCIATION BETWEEN VULNERABILITY AND FINGERPRINTS OF UOGD SPILLS

We compared predicted groundwater vulnerability to documented spills with water quality data on ion-ratio pairs, which have been used by others to detect chemical imprints of oil and gas wastewaters in well water samples.^{8,26,31–33} The full dataset comprised water quality measurements on 27,318 samples collected between 2008 and 2020; however, missing or censored measurements reduced the number of samples for which ion ratios could be computed (Appendix B, Table B-1). Fewer than 800 water quality analyses were suitable for evaluating Ba/Cl versus Br/SO₄, SO₄/Cl versus Br, Mg/Li versus SO₄/Cl, and Mg/Li versus Br/SO₄. According to these four criteria, between 40 and 69 samples exhibited UOGD chemical imprints. Ca/Sr versus Ca/Mg and SO₄/Cl versus Mg/Na had fewer censored or missing measurements and, hence, could be used to assess the potential for produced water imprints in a large sample. Five percent of 16,090 samples exhibited produced water imprints according to Ca/Sr versus Ca/Mg, while 0.8% of 18,062 samples exhibited a produced water signature according to SO₄/Cl versus Mg/Na. Both criteria exhibited high accuracy in correctly classifying unaffected groundwaters (97.5% for Ca/Sr vs. Ca/Mg; 99.8% for SO₄/Cl vs. Mg/Na), while Ca/Sr versus Ca/Mg exhibited considerably higher accuracy (93.5% vs. 82.9%) in correctly identifying produced water samples from the Marcellus Shale (Appendix B, Tables B-2 and B-3). As Ca/Sr versus Ca/Mg was more accurate in detecting produced water, our subsequent statistical analyses thus focused on evaluating this criterion against predicted groundwater vulnerability to documented spills.

Logistic regression models showed significant positive associations between groundwater vulnerability to spills and the likelihood of detecting UOGD produced water–like imprints in a water sample. In these models, we treated *V* as either (1) a continuous variable ($0 \leq V \leq 1$) or (2) a binary variable indicating whether a sample was highly vulnerable ($V \geq 0.5$ vs. $V < 0.5$). Under Approach 1, unadjusted models showed that a 0.10 increase in vulnerability was associated with a 14% increase in the odds of exhibiting

a UOGD fingerprint ($P < 0.001$) (Table 2). Adjusted models accounting for topographic position, distance to geological structures (faults and folds), and presence of nearby conventional oil and gas wells showed slightly reduced (13%) but still statistically significant associations between vulnerability and UOGD imprint ($P < 0.001$). Under Approach 2, models showed that highly vulnerable samples had over three times higher odds of exhibiting UOGD imprints ($P < 0.001$).

The distance to geological structures was also a significant predictor in adjusted models, with associations indicating a lower likelihood of detecting UOGD-like imprints as distance from geological structures increases. Meanwhile, the binary variable indicating the presence of conventional oil and gas wells within 2 km of a water sample was also a statistically significant predictor of increased likelihood of detecting produced water imprint. These covariates suggest possible complex sources of produced water–like chemistry in our groundwater samples. Specifically, the significance of conventional oil and gas suggests that such sources cannot be ruled out, especially because of their ubiquitous spatial distribution and historical presence in Appalachia. However, we note that if we compare binary definitions (i.e., Approach 2), the increase in odds of detecting produced water–like imprints associated with high vulnerability to UOGD spills is more than double that of the presence of nearby conventional oil and gas (3.24 vs. 1.47). Sensitivity analyses via random permutations of the data confirmed the robustness of associations between a higher likelihood of detecting produced water imprints and

Table 2. Associations Between Simulated Vulnerability to Spills and Detection of UOGD Produced Water–Like Imprints in Groundwater Samples^a

	Odds Ratio (<i>P</i> Value)	
	Approach 1	Approach 2
Unadjusted		
Vulnerability to UOGD spills	1.14 (<0.001)	3.71 (<0.001)
Adjusted		
Vulnerability to UOGD spills	1.13 (<0.001)	3.24 (<0.001)
Upper slopes vs. peaks/summits	0.79 (0.430)	0.79 (0.430)
Lower slopes vs. peaks/summits	0.99 (0.967)	0.99 (0.969)
Valleys vs. peaks/summits	1.77 (0.052)	1.77 (0.052)
Distance to geological structures	0.82 (<0.01)	0.82 (<0.01)
Presence of conventional oil and gas wells within 2 km ($n > 0$ vs. $n = 0$)	1.47 (<0.05)	1.47 (<0.05)

UOGD = unconventional oil and gas development.

^aStatistically significant associations ($\alpha = 0.05$) indicated in bold typeface. Under Approach 1, results for *V* are based on a 0.10 increase as a continuous variable. Under Approach 2, results are based on a binary classification of highly vulnerable versus not (i.e., $V \geq 0.5$ vs. $V < 0.5$). Both unadjusted and adjusted models accounted for geographic location through inclusion of latitude, longitude, and their interaction.

our estimates of vulnerability to contamination from UOGD spills (Appendix B, Table B-4). The results suggest that our primary findings shown in Table 2 are unlikely to have occurred by chance, as all the sensitivity analyses show no association between vulnerability and UOGD imprint.

DISCUSSION

CONTAMINANT TRANSPORT TRAJECTORIES IN CONTEXT OF SETBACK DISTANCES

Our analysis enables inferences about where contaminants spilled from UOGD wells are likely to travel by groundwater flow. The particle-tracking simulations indicate transport distances of contaminants originating from UOGD well pads generally exceed regulatory setback distances between UOGD and drinking water wells. Current setback distances are 152 m for Pennsylvania, 15 m for Ohio, and 76 m for West Virginia.^{54–56} In 2020, Pennsylvania’s 43rd Statewide Investigating Grand Jury made a recommendation to increase setback distances between UOGD and water wells to 2,500 ft (762 m).⁵⁷ The jury’s recommended distance is exceeded by 34% of the transport distances in our particle-tracking analysis. A rulemaking petition filed with Pennsylvania’s Environmental Quality Board to increase the setback distance to 3,281 ft (1,000 m) was approved in 2025, triggering further evaluation of the petition by the Pennsylvania Department of Environmental Protection.⁵⁸ This distance is exceeded by 21% of our simulated transport distances. Taken together, these outcomes suggest that our analysis may provide a physically based and systematic framework to inform advocacy and policy deliberations for drinking water protection.

VULNERABLE POPULATIONS AND VULNERABILITY HOT SPOTS

The risk UOGD poses to domestic well contamination is partly shaped by hydrologic vulnerability, which quantifies the probability of a contaminant reaching a drinking water receptor (domestic well) upon release from a source (UOGD well pad). Our analysis indicates that more than 20,000 people dependent on domestic wells reside in areas characterized by high vulnerability ($V \geq 0.5$) to spills included in our analysis. This analysis accounted for 1,761 spills — those that could be positively geolocated with UOGD well pads and had information on spill timing — but our initial search also yielded an additional 2,092 likely spills in state regulatory databases that, owing to incomplete or vague descriptions, could not be included for further analysis. While an indeterminate portion of these excluded spills were likely associated with COGD activities, our database of included spills may underrepresent the number of spills occurring at UOGD well pads. Thus, the actual at-risk population from spills at UOGD well pads may be somewhat greater than our estimate suggests.

According to our model-based estimates, over 120,000 people who rely on domestic wells reside in areas that have high hydrologic vulnerability ($V \geq 0.5$) when considering all

UOGD well pads as potential contaminant sources. This vulnerable population, an upper-bound estimate for our hydrologic-based analysis, constitutes 3% of the 3.9 million people who use domestic wells within the Marcellus region, yet it is less than previous estimates that were based solely on household proximity to UOGD wells.^{59,60} Using commonly adopted radial distances of 1 km and 2 km around UOGD wells, the population served by domestic wells potentially at risk is more than 160,000 and 389,000, respectively. Our estimate of vulnerable populations is lower because it excludes domestic wells that are unlikely to be hydrologically connected to a UOGD well pad.

It is important to identify where and what communities are vulnerable, recognizing that hydrologic vulnerability to UOGD and its impacts are distributed unevenly.³⁸ Although the vulnerable proportion is small when taken across the entire study area, it exceeds 20% of local populations in Northeastern Pennsylvania, with additional vulnerability hot spots in the tri-state area and Southern West Virginia (Figure 7). Hotspots of V arise from heterogeneities in groundwater flow patterns as well as the distribution of UOGD sources. When we consider the 30 km spatial units used in Figure 7, we find a correlation coefficient of 0.91 between aggregate mean V and the number of UOGD wells. Thus, while vulnerability is influenced by hydrologic heterogeneities at local scales, it is predominantly controlled by the accumulation of potential contaminant sources at larger scales.

VULNERABILITY IS ASSOCIATED WITH CHEMICAL FINGERPRINTS OF UOGD SPILLS

We applied logistic regression models to evaluate associations between model-computed vulnerability and the presence/absence of UOGD chemical imprints in well water samples. UOGD chemical imprints were identified based on the ratios of inorganic ion concentrations. Our results show that vulnerability to spills is strongly associated with a higher likelihood of detecting chemical imprints of oil and gas wastewaters, even when covariates accounting for other factors capable of imparting such signatures to groundwater are introduced. These results support the appropriateness of the physically based vulnerability assessment framework and are consistent with the hypothesis that spills from UOGD well pads have had a systematic and detectable effect on domestic groundwater chemistry in our study region.

Although statistically significant associations between modeled risk and UOGD fingerprints are encouraging from the standpoint of model performance, it is important to recognize that UOGD fingerprints, as defined here, are indicators of the mixing of unimpaired groundwaters with fluids having chemical characteristics resembling produced waters. These fingerprints are not, however, equivalent to exceedances of either health-based standards (e.g., maximum contaminant levels) or aesthetics-based standards (e.g., secondary contaminant levels). It would be informative to evaluate our framework against other classes of chemicals that are known to occur in oil and gas wastewaters and are of human health

concern, such as semivolatile or volatile organic compounds, should groundwater datasets that include georeferenced measurements of these analytes become more widely available.

LIMITATIONS

The analysis could likely be improved through a more sophisticated treatment of hazards that considers spill characteristics other than location and timing, such as volume and composition.^{61,62} A more comprehensive risk assessment would also need to consider additional pathways of contamination, including spills at wastewater injection wells, leaks from abandoned oil and gas wells, and unintentional releases during wastewater transport by pipelines and tanker trucks, as well as during “frac-outs” where active drilling within the target shale formation unintentionally releases contaminants to groundwater via existing wells.⁶³ While these additional source characteristics and pathways are important in shaping contaminant fate and concentrations detected in drinking water receptors, they are rarely reported in regulatory databases.

Our analysis accounted for spills that could be positively geolocated with known UOGD well pads and had information about spill timing, but there were more spills in state regulatory databases that could not be included in our analysis. Processing and cleaning spill data reported in state regulatory databases was time-consuming, in some cases requiring harmonizing information stored on different databases. We encountered numerous spill records with ambiguous location information. American Petroleum Institute (API) numbers, which would enable geolocation of oil and gas wells through additional lookups, as well as information on spilled material and volume, were commonly missing. Agency personnel were cooperative and conscientious in responding to our data inquiries, but this could not compensate for deficiencies in record keeping.

The vulnerability assessment should be considered in light of simplifying assumptions that were made to address data sparsity and other data limitations, as well as to enable execution of regional-scale flow-and-transport simulation. The MODFLOW 6 simulations reproduced the key features of topographically controlled groundwater flow, but the two-dimensional setup does not simulate underflow beneath small receiving streams that could promote contaminant dispersal. That deeper flow paths (associated with underflow) are less prone to entrain surface-derived contaminants and are more likely to lie beneath domestic wells may mitigate the effects of neglected underflow on modeled vulnerability. The MODFLOW 6 simulations also assume steady-state flow, implying that groundwater flow patterns remain constant over time. The steady-state approximation provides a reasonable representation of average groundwater flow conditions, yet variation in recharge, associated with storm events, for example, may lead to perturbations in hydraulic head gradients. As these perturbations are ephemeral, typically decaying over hours to weeks,⁶⁴ their effect on our vulnerability estimates is likely to be minor.

MODPATH 7, which uses flow fields from MODFLOW 6 to simulate the transport of spilled solutes, also incorporates several simplifications. In particular, our application of MODPATH 7 assumes that spilled solutes are transported conservatively. While this assumption provides a reasonable approximation for the transport of the inorganic ions considered in the fingerprinting analysis, it will cause overestimation of the vulnerability for many organic compounds that sorb more strongly or are susceptible to degradation. Unfortunately, information on the chemical composition of spills needed to estimate solute-specific parameters for reactive transport is rarely available. Thus, vulnerability inferences based on the conservative-transport assumption are likely to be most defensible, as they align with the precautionary principle, which prioritizes protection against harm under uncertainty.

MODPATH 7 further simplifies solute transport by accounting only for advection and neglecting hydrodynamic dispersion. Hydrodynamic dispersion causes spreading and dilution of a contaminant plume and hence, relative to advection alone, this process would increase the area down-gradient of a UOGD spill in which solute concentrations exceed background levels. Although not represented mechanistically, the effects of dispersion on the vulnerability fields are approximated by our ensemble particle-tracking approach that simulates a distribution in advective-transport pathways as a means of accounting for uncertainty in predictions of the groundwater flow field.

In regional-scale applications, such as the one presented in this study, it is practically infeasible to relax the simplifying assumptions (described above) owing to computational constraints and data limitations. At smaller scales, within locales where UOGD-related concerns may be particularly acute, perhaps in response to a large spill or an elevated frequency of water quality impairments, application of vulnerability frameworks of greater complexity may be warranted. In these cases, modeling efforts would require allocation of human and financial resources for detailed hydrogeologic characterization and the transparent exchange of information on spill composition, timing, duration, and volume.

CONCLUSIONS

According to our assessment, 0.5% to 3% of the population reliant on domestic wells reside in vulnerable areas, depending on whether only 1,761 known spills are considered or all UOGD well pads are considered as potential sources of well water contamination. This analysis is consistent with observational studies that have found groundwater impairments stemming from UOGD activities occur rarely within the Appalachian Basin.^{23,30,42,45,65–67} The modeling framework advanced for this study identifies where these impairments, although seemingly infrequent at least by historical standards, are most likely to occur. Our analysis identifies hot spots in which vulnerable population proportions are as much as tenfold greater than the all-basin average. The framework is an improvement over proximity-based metrics of risk that fail to account for

groundwater flow patterns that regulate contaminant dispersal, and it is applicable over large scales and at reasonably high spatial resolution.⁶⁸ Our approach could be extended to explore other water quality issues associated with historic or contemporary oil and gas development, such as vulnerability to contamination from abandoned hydrocarbon wells and produced water disposal via underground injection. It is also amenable to application outside of the Appalachian Basin, although some of the data limitations that we encountered in this study may be more severe in other oil- and gas-producing regions. Given renewed policy attention on fossil fuel production in the United States, improvements in collection and management of records on industry violations associated with accidental contaminant releases and regulatory monitoring of vulnerable drinking water wells would benefit from future studies of potential linkages between UOGD, groundwater quality, and human health.

DATA AVAILABILITY STATEMENT

Groundwater quality data are available via the Shale Network Database (www.shalenetwork.org). Scripts for groundwater modeling and calibration were modified from previously published code by Zell and Sanford (2020)⁶⁹ and Hugman and colleagues (2022).⁷⁰ Original R code used for our statistical analysis and aggregated results for vulnerability, populations, and spills are available in a Zenodo repository (<https://zenodo.org/records/21381071>).

ACKNOWLEDGMENTS

The authors thank Dr. Susan Brantley, Dr. Tao Wen, and Dr. Samuel Shaheen for assistance with the Shale Network Database. The simulations presented in this article were performed on computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University. Mario Soriano Jr. also acknowledges support from a faculty start-up grant from the National University of Singapore.

REFERENCES

1. US Energy Information Administration (EIA). US Oil and Natural Gas Wells by Production Rate. Appendix C: Drilling Production Report. Available: <https://www.eia.gov/petroleum/wells/>.
2. Jackson RB, Vengosh A, Darrah TH, Warner NR, Down A, Poreda RJ, et al. 2013. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proc Natl Acad Sci USA* 110:11250–11255, <https://doi.org/10.1073/pnas.1221635110>.
3. Osborn SG, Vengosh A, Warner NR, Jackson RB. 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc Natl Acad Sci USA* 108:8172–8176, <https://doi.org/10.1073/pnas.1100682108>.
4. Heilweil VM, Grieve PL, Hynek SA, Brantley SL, Solomon DK, Risser DW. 2015. Stream measurements locate thermogenic methane fluxes in groundwater discharge in an area of shale-gas development. *Environ Sci Technol* 49:4057–4065, <https://doi.org/10.1021/es503882b>.
5. Woda J, Wen T, Oakley D, Yoxtheimer D, Engelder T, Castro MC, et al. 2018. Detecting and explaining why aquifers occasionally become degraded near hydraulically fractured shale gas wells. *Proc Natl Acad Sci USA* 115:12349–12358, <https://doi.org/10.1073/pnas.1809013115>.
6. Darrah TH, Vengosh A, Jackson RB, Warner NR, Poreda RJ. 2014. Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proc Natl Acad Sci USA* 111:14076–14081, <https://doi.org/10.1073/pnas.1322107111>.
7. Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environ Sci Technol* 48:8334–8348, <https://doi.org/10.1021/es405118y>.
8. Brantley SL, Yoxtheimer D, Arjmand S, Grieve P, Vidic R, Pollak J, et al. 2014. Water resource impacts during unconventional shale gas development: The Pennsylvania experience. *Int J Coal Geol* 126:140–156, <https://doi.org/10.1016/j.coal.2013.12.017>.
9. Vidic RD, Brantley SL, Vandenbossche JM, Yoxtheimer D, Abad JD. 2013. Impact of shale gas development on regional water quality. *Science* 340:1235009, <https://doi.org/10.1126/science.1235009>.
10. Barth-Naftilan E, Aloysius N, Saiers JE. 2015. Spatial and temporal trends in freshwater appropriation for natural gas development in Pennsylvania's Marcellus Shale Play. *Geophys Res Lett* 42:6348–6356, <https://doi.org/10.1002/2015GL065240>.
11. Gallegos TJ, Varela BA, Haines SS, Engle MA. 2015. Hydraulic fracturing water use variability in the United States and potential environmental implications. *Water Resour Res* 51:5839–5845, <https://doi.org/10.1002/2015WR017278>.
12. Gregory KB, Vidic RD, Dzombak DA. 2011. Water management challenges associated with the production of shale gas by hydraulic fracturing. *Elements* 7:181–186, <https://doi.org/10.2113/gselements.7.3.181>.
13. Nicot J-P, Scanlon BR, Reedy RC, Costley RA. 2014. Source and fate of hydraulic fracturing water in the Barnett Shale: A historical perspective. *Environ Sci Technol* 48:2464–2471, <https://doi.org/10.1021/es404050r>.
14. Cozzarelli IM, Kent DB, Briggs M, Engle MA, Benthem A, Skalak KJ, et al. 2021. Geochemical and geophysical indicators of oil and gas wastewater can trace potential exposure pathways following releases to surface waters.

- Sci Total Environ 755:142909, <https://doi.org/10.1016/j.scitotenv.2020.142909>.
15. Kondash AJ, Lauer NE, Vengosh A. 2018. The intensification of the water footprint of hydraulic fracturing. *Sci Adv* 4:eaar5982, <https://doi.org/10.1126/sciadv.aar5982>.
 16. Warner NR, Christie CA, Jackson RB, Vengosh A. 2013. Impacts of shale gas wastewater disposal on water quality in Western Pennsylvania. *Environ Sci Technol* 47:11849–11857, <https://doi.org/10.1021/es402165b>.
 17. Shih J-S, Saiers JE, Anisfeld SC, Chu Z, Muehlenbachs LA, Olmstead SM. 2015. Characterization and analysis of liquid waste from Marcellus Shale gas development. *Environ Sci Technol* 49:9557–9565, <https://doi.org/10.1021/acs.est.5b01780>.
 18. Deziel NC, Brokovich E, Grotto I, Clark CJ, Barnett-Itzhaki Z, Broday D, et al. 2020. Unconventional oil and gas development and health outcomes: A scoping review of the epidemiological research. *Environ Res* 182:109124, <https://doi.org/10.1016/j.envres.2020.109124>.
 19. Lauer NE, Harkness JS, Vengosh A. 2016. Brine spills associated with unconventional oil development in North Dakota. *Environ Sci Technol* 50:5389–5397, <https://doi.org/10.1021/acs.est.5b06349>.
 20. 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>.
 21. 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>.
 22. Patterson LA, Konschnik KE, Wiseman H, Fargione J, Maloney KO, Kiesecker J, et al. 2017. Unconventional oil and gas spills: Risks, mitigation priorities, and state reporting requirements. *Environ Sci Technol* 51:2563–2573, <https://doi.org/10.1021/acs.est.6b05749>.
 23. Xiong B, Soriano Jr. MA, Gutchess KM, Hoffman N, Clark CJ, Siegel HG, et al. 2022. Groundwaters in Northeastern Pennsylvania near intense hydraulic fracturing activities exhibit few organic chemical impacts. *Environ Sci Process Impacts* 24:252–264, <https://doi.org/doi: 10.1039/D1EM00124H>.
 24. Drollette BD, Hoelzer K, Warner NR, Darrah TH, Karatum O, O'Connor MP, et al. 2015. Elevated levels of diesel range organic compounds in groundwater near Marcellus gas operations are derived from surface activities. *Proc Natl Acad Sci USA* 112:13184–13189, <https://doi.org/10.1073/pnas.1511474112>.
 25. Llewellyn GT, Dorman F, Westland JL, Yoxheimer 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>.
 26. Soriano Jr. MA, Siegel HG, Johnson NP, Gutchess KM, Xiong B, Li Y, et al. 2021. Assessment of groundwater well vulnerability to contamination through physics-informed machine learning. *Environ Res Lett* 16:084013, <https://doi.org/10.1088/1748-9326/ac10e0>.
 27. 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>.
 28. Warner NR, Jackson RB, Darrah TH, Osborn SG, Down A, Zhao K, et al. 2012. Geochemical evidence for possible natural migration of Marcellus formation brine to shallow aquifers in Pennsylvania. *Proc Natl Acad Sci USA* 109:11961–11966, <https://doi.org/10.1073/pnas.1121181109>.
 29. Siegel DI, Smith B, Perry E, Bothun R, Hollingsworth M. 2015. Pre-drilling water-quality data of groundwater prior to shale gas drilling in the Appalachian Basin: Analysis of the Chesapeake Energy Corporation dataset. *Appl Geochem* 63:37–57, <https://doi.org/10.1016/j.apgeochem.2015.06.013>.
 30. Siegel HG, Soriano Jr. MA, Clark CJ, Johnson NP, Wulsin HG, Deziel NC, et al. 2022. Natural and anthropogenic processes affecting domestic groundwater quality within the Northwestern Appalachian Basin. *Environ Sci Technol* 56:13761–13773, <https://doi.org/10.1021/acs.est.2c04011>.
 31. Cantlay T, Bain DJ, Stolz JF. 2020. Determining conventional and unconventional oil and gas well brines in natural samples III: Mass ratio analyses using both anions and cations. *J Environ Sci Health A Tox Hazard Subst Environ Eng* 55:24–32, <https://doi.org/10.1080/10934529.2019.1666562>.
 32. Tisherman R, Bain DJ. 2019. Alkali earth ratios differentiate conventional and unconventional hydrocarbon brine contamination. *Sci Total Environ* 695:133944, <https://doi.org/10.1016/j.scitotenv.2019.133944>.
 33. Wilson JM, Wang Y, VanBriesen JM. 2014. Sources of high total dissolved solids to drinking water supply in Southwestern Pennsylvania. *J Environ Eng (New York)* 140:B4014003, [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000733](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000733).
 34. Darrah TH, Jackson RB, Vengosh A, Warner NR, Poreda RJ. 2015. Noble gases: A new technique for fugitive gas investigation in groundwater. *Ground Water* 53:23–28. PMID: 25713829.
 35. Warner NR, Darrah TH, Jackson RB, Millot R, Kloppmann W, Vengosh A. 2014. New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environ Sci Technol* 48:12552–12560, <https://doi.org/10.1021/es5032135>.
 36. Soriano Jr. MA, Siegel HG, Gutchess KM, Clark CJ, Li Y, Xiong B, et al. 2020. Evaluating domestic well vulnerability to contamination from unconventional oil and gas development sites. *Water Resour Res* 56:e2020WR028005, <https://doi.org/10.1029/2020WR028005>.

37. Soriano Jr. MA, Deziel NC, Saiers JE. 2022. Regional scale assessment of shallow groundwater vulnerability to contamination from unconventional hydrocarbon extraction. *Environ Sci Technol* 56:12126–12136, <https://doi.org/10.1021/acs.est.2c00470>.
38. Soriano Jr. MA, Warren JL, Clark CJ, Johnson NP, Siegel HG, Deziel NC, et al. 2023. Social vulnerability and groundwater vulnerability to contamination from unconventional hydrocarbon extraction in the Appalachian Basin. *GeoHealth* 7:e2022GH000758, <https://doi.org/10.1029/2022GH000758>.
39. Zell WO, Sanford WE. 2020. Calibrated simulation of the long-term average surficial groundwater system and derived spatial distributions of its characteristics for the contiguous United States. *Water Resour Res* 56:e2019WR026724, <https://doi.org/10.1029/2019WR026724>.
40. White JT, Hunt RJ, Fienen MN, Doherty JE. 2020. Approaches to Highly Parameterized Inversion: PEST++ Version 5, a Software Suite for Parameter Estimation, Uncertainty Analysis, Management Optimization and Sensitivity Analysis. Water Availability and Use Science Program. Available: <https://pubs.usgs.gov/tm/07/c26/tm7c26.pdf>.
41. Pollock DW. 2016. User Guide for MODPATH Version 7 — A Particle-Tracking Model for MODFLOW: US Geological Survey Open-File Report. 1086:35. Available: <http://dx.doi.org/10.3133/ofr20161086>.
42. Johnson TD, Belitz K, Lombard M. 2019. Estimating domestic well locations and populations served in the contiguous US for years 2000 and 2010. *Sci Total Environ* 687:1261–1273, <https://doi.org/10.1016/j.scitotenv.2019.06.036>.
43. Pennsylvania Spatial Data Access (PASDA). 2026. Data Summary: Oil Gas Locations — Conventional Unconventional. Available: <https://www.pasda.psu.edu/uci/Data-Summary.aspx?dataset=1088>.
44. Ohio Department of Natural Resources. 2024. Oil and Gas Wells Locations of Ohio (Statewide). Available: <https://gis-odnr.opendata.arcgis.com/documents/3dce-2c55a9e1486b9483db25710ae588/about>.
45. West Virginia Department of Environmental Protection (WVDEP) Geographic Information Server. GIS Data. Available: <https://tagis.dep.wv.gov/home/gis%20data>.
46. Bugher NA, Xiong B, Gentles RI, Glist LD, Siegel HG, Johnson NP, et al. 2024. Domestic groundwater wells in Appalachia show evidence of low-dose, complex mixtures of legacy pollutants. *Environ Sci Process Impacts* 26:2250–2263, <https://doi.org/10.1039/D4EM00364K>.
47. Theobald DM, Harrison-Atlas D, Monahan WB, Albano CM. 2015. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PLoS One* 10:e0143619, <https://doi.org/10.1371/journal.pone.0143619>.
48. Horton JD. 2017. The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States (ver. 1.1, August 2017): US Geological Survey Data Release. Available: <https://doi.org/10.5066/F7WH2N65>.
49. PA Department of Conservation and Natural Resources. Updated February 24, 2026. PA Geology Folds. Available: <https://newdata-dcnr.opendata.arcgis.com/datasets/DCNR::pennsylvania-folds/about>.
50. New York State Museum. Geology — GIS/Maps. Available: <https://nysm.nysed.gov/research-collections/geology/gis>.
51. Yang C, Condon LE, Maxwell RM. 2025. Unravelling groundwater-stream connections over the continental United States. *Nature Water* 3:70–79, <https://doi.org/10.1038/s44221-024-00366-8>.
52. Gleeson T, Smith L, Moosdorf N, Hartmann J, Dürr HH, Manning AH, et al. 2011. Mapping permeability over the surface of the Earth. *Geophysical Res Lett* 38, <https://doi.org/10.1029/2010GL045565>.
53. Wen T, Niu X, Gonzales M, Zheng G, Li Z, Brantley SL. 2018. Big groundwater data sets reveal possible rare contamination amid otherwise improved water quality for some analytes in a region of Marcellus Shale development. *Environ Sci Technol* 52:7149–7159, <https://doi.org/10.1021/acs.est.8b01123>.
54. 58 PA Cons Stat § 3215 Well location restrictions. In 2024. Available: <https://www.legis.state.pa.us/WU01/LI/LI/CT/HTM/58/00.032.015.000..HTM>.
55. OH Rev Code § 1509.021 Surface Locations of New Wells. In 2024. Available: <https://codes.ohio.gov/ohio-revised-code/section-1509.021>.
56. WV Code § 22-6A-12 Well Location Restrictions. In 2024. Available: <https://code.wvlegislature.gov/22-6A-12/>.
57. PA Office of the Attorney General. 2020. Report 1 of the Forty-Third Statewide Investigating Grand Jury. Available: https://www.legis.state.pa.us/WU01/LI/CSM/2021/0/35558_27878.pdf.
58. Phillips, A. and K. Edwards. PA Board Accepts Petition to Require No-Drill Zones for Further Study. Environmental Integrity Project. Available: https://environmentalintegrity.org/news/pa-board-accepts-petition-to-require-no-drill-zones-for-further-study/?utm_source=chatgpt.com.
59. Jasechko S, Perrone D. 2017. Hydraulic fracturing near domestic groundwater wells. *Proc Natl Acad Sci USA* 114:13138–13143, <https://doi.org/10.1073/pnas.1701682114>.
60. Czolowski ED, Santoro RL, Srebotnjak T, Shonkoff SBC. 2017. Toward consistent methodology to quantify populations in proximity to oil and gas development: A national spatial analysis and review. *Environ Health Perspect* 125:086004, <https://doi.org/10.1289/ehp1535>.
61. Mallants D, Doble R, Beiraghdar Y. 2022. Fate and transport modelling framework for assessing risks to soil and groundwater from chemicals accidentally released during surface operations: An Australian example application from shale gas developments. *J Hydrol* 604:127271, <https://doi.org/10.1016/j.jhydrol.2021.127271>.
62. Mallants D, Kirby J, Golding L, Apte S, Williams M. 2022. Modelling the attenuation of flowback chemicals for a soil-groundwater pathway from a hypothetical spill accident. *Sci Total Environ* 806:150686, <https://doi.org/10.1016/j.scitotenv.2021.150686>.

63. Cantlay T, Bain DJ, Dombrowski C, Rykaceski M, Bayto M, Miller K, et al. 2025. Contamination of private water supplies after a well communication event (frac-out) in Southwest Pennsylvania. *Sci Rep* 15:32352, <https://doi.org/10.1038/s41598-025-16976-5>.
64. Townley LR. 1995. The response of aquifers to periodic forcing. *Adv Water Resour* 18:125–146, [https://doi.org/10.1016/0309-1708\(95\)00008-7](https://doi.org/10.1016/0309-1708(95)00008-7).
65. McMahon PB, Lindsey BD, Conlon MD, Hunt AG, Belitz K, Jurgens BC, et al. 2019. Hydrocarbons in upland groundwater, Marcellus Shale Region, Northeastern Pennsylvania and Southern New York, USA. *Environ Sci Technol* 53:8027–8035, <https://doi.org/10.1021/acs.est.9b01440>.
66. Harkness JS, Darrah TH, Warner NR, Whyte CJ, Moore MT, Millot R, et al. 2017. The geochemistry of naturally occurring methane and saline groundwater in an area of unconventional shale gas development. *Geochim Cosmochim Acta* 208:302–334, <https://doi.org/10.1016/j.gca.2017.03.039>.
67. 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 Sci Technol* 58:17862–17873, <https://doi.org/10.1021/acs.est.4c03371>.
68. Deziel NC, Clark CJ, Casey JA, Bell ML, Plata DL, Saiers JE. 2022. Assessing exposure to unconventional oil and gas development: Strengths, challenges, and implications for epidemiologic research. *Curr Environ Health Rep* 9:436–450, <https://doi.org/10.1007/s40572-022-00358-4>.
69. Zell WO, Sanford WE. 2020. MODFLOW 6 Models Used to Simulate the Long-Term Average Surficial Groundwater System for the Contiguous United States. In: US Geological Survey Data Release, 2020. Available: <https://www.sciencebase.gov/catalog/item/631405c5d34e36012efa3190>.
70. Hugman R, White JT, Fienen MN, Hunt RJ. 2022. Parameter Estimation, Uncertainty Analysis, and Optimization with the PEST++ Family of Codes: Tutorial Jupyter Notebooks: US Geological Survey Data Release. Available: <https://doi.org/10.5066/P901F9G0>.

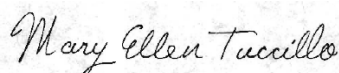
HEI QUALITY ASSURANCE STATEMENT

Eastern Research Group, Inc. (ERG) provided an independent quality assurance (QA) audit for this study. ERG staff members Dr. Mary Ellen Tuccillo and Dr. William Morgan conducted a remote audit of the final report and associated appendices. An on-site audit was not conducted for this study. The ERG auditors are experienced in QA oversight and the subject matter relevant to this study, including geochemistry, hydraulic fracturing, contaminant transport, statistical methods, and groundwater flow modeling. Dr. Tuccillo, an environmental scientist, served as the lead auditor and interacted with principal investigators, reviewed the text for clarity of methods and results, and reviewed all aspects of the final report. Dr. Morgan reviewed all groundwater flow modeling, vulnerability assessments, and statistical modeling.

The remote audit of the final report began in early 2026, with most work conducted in March to May 2026 as files from the research team became available. The objectives of this audit were to ensure that study methods were well documented, the final report was understandable, reported results were accurate, and key study findings and limitations were highlighted. To meet these objectives, the ERG auditors reviewed the final report text, tables, and figures to verify their accuracy and clarity. Raw data; Python, MATLAB, and R code; input and output files for groundwater flow modeling; input and output files for statistical modeling; and files for generation of figures were reviewed. The auditors confirmed that reported results and figures could be accurately reproduced with the files provided and that the results and methods were accurately presented in the final report. The audit also evaluated how the final report compared with protocols and methods described in the QA/QC Plan.

Auditors identified minor discrepancies and areas where the report text or figures could be improved for clarity and accuracy. The research team has addressed these issues. The audit findings did not affect the overall results of this research study.

Audit findings were documented in an ERG audit report, and the resulting responses by the research team have been documented. The overall approach in the final report has been well described, key results were clearly presented in tables and figures, and discussions and conclusions are clear and consistent with the results of the analyses. The ERG audit team is satisfied that the final study report is representative of the research conducted and that the study followed a valid set of procedures.



Mary Ellen Tuccillo, PhD, Environmental Scientist, Quality Assurance Auditor



William Morgan, PhD, Environmental Modeler, Quality Assurance Auditor

SUPPLEMENTARY APPENDICES

Supplementary Appendices A through D contain additional information not included in the main report. They are available on the [HEI website](#).

Appendix A. Testing Alternative Groundwater Model Configurations

Appendix B. Evaluating Classification Accuracy of Ion-Ratio Pairs and Statistical Model Sensitivity Analysis via Random Permutations

Appendix C. Companion Work: Assessing the Influence of Unsaturated Zone Flow

Appendix D. Companion Work: Evaluation of Vulnerability to Abandoned Oil and Gas Wells

ABOUT THE AUTHORS

James E. Saiers, lead principal investigator, is the Clifton R. Musser Professor of Hydrology at the Yale School of the Environment. His recent research addresses drinking water impacts of fossil fuel development, natural carbon capture and carbon cycling within watersheds, and migration of contaminants through groundwater aquifers. This research relies on experimentation and mathematical modeling and is conducted in collaboration with ecologists, geochemists, engineers, and epidemiologists. His overarching goal is to generate new observations and to develop predictive approaches that can be used to inform water resource management approaches intended to safeguard freshwater, ecosystems, and public health. He earned a BS in geosciences from Indiana University of Pennsylvania and an MS and PhD in environmental sciences from the University of Virginia.

Mario Soriano Jr., co-principal investigator, is an assistant professor in the Department of Geography at the National University of Singapore. Before joining the National University of Singapore, he was a postdoctoral research associate and environmental fellow at the High Meadows Environmental Institute at Princeton University, where he was also an affiliated researcher at the Integrated GroundWater Modeling Center and Blue Lab, an environmental multimedia storytelling collective. He received his PhD in hydrology and water resources from Yale University. His research draws on theoretical frameworks and methodologies from hydrogeology, sustainability science, geography, and applied data science to examine the movement of water and contaminants through the robust integration of observations and high-fidelity simulation models. His interests explore the intersections of human activities and values that shape narratives of water as a natural resource and natural hazard, a driver of environmental processes, and a pillar of social well-being. He has experience in collaborative projects on the water-energy nexus, sustainable food systems, and environmental communication.

Joshua Warren, co-principal investigator, is an associate professor in the Department of Biostatistics at the Yale School of Public Health. Previously, he received his PhD in statistics from North Carolina State University in 2011 and was a postdoctoral researcher in the Department of Biostatistics at the University of North Carolina at Chapel Hill before joining the faculty at Yale in 2014. His research focuses on the development of hierarchical Bayesian methods for spatially/spatiotemporally referenced data, understanding the impact of environmental exposures on human health, and characterizing the spread of infectious diseases. He also has an interest in developing and applying spatiotemporal models in

several collaborative settings in medicine and public health. His theoretical and methodological interests include multiple topics in spatial/spatiotemporal modeling and Bayesian nonparametrics.

Reed Maxwell, co-principal investigator, is the William and Edna Macaleer Professor of Engineering and Applied Science, and a professor in civil and environmental engineering at the High Meadows Environmental Institute at Princeton University. He also directs the Integrated GroundWater Modeling Center. His research interests are focused on understanding connections within the hydrologic cycle and how they relate to water quantity and quality under anthropogenic stresses. Before joining Princeton, he was on the faculty at the Colorado School of Mines and worked as a postdoctoral researcher and later as a staff scientist at the Hydrologic Sciences Group at Lawrence Livermore National Laboratory. He holds a PhD in environmental water resources from the Civil and Environmental Engineering Department at the University of California, Berkeley.

Shuqi Lin is a master's student in biostatistics at Yale University, specializing in Bayesian and spatial statistical methods under the guidance of Joshua Warren. Her recent research includes developing spatiotemporal approaches for optimizing infectious disease testing and creating spatial meta-regression models to examine environmental health relationships. She currently collaborates on interdisciplinary projects studying how climate patterns and geographical factors affect human health outcomes. In her future work, Shuqi aims to develop innovative statistical approaches with applications across diverse fields, especially those addressing critical environmental and health challenges.

Yingcan Cathy Wang is a PhD candidate in the Department of Civil and Environmental Engineering and at Princeton University, working with Reed Maxwell in the Integrated GroundWater Modeling Center. Her research focuses on reactive transport and understanding hydrogeochemical processes in the variably saturated subsurface. She is currently investigating the fate of surplus nitrogen from agricultural lands to streams, with an emphasis on evaluating the long-term impacts of nitrogen transport and denitrification at the continental scale. She earned a BS in environmental science and policy from the College of William and Mary and an MS in civil and environmental engineering from Stanford University.

David van Velden is a civil and environmental engineering undergraduate student at Princeton University interested in sustainable water management. Raised near Amsterdam, a city known for its advanced water infrastructure, he developed a strong interest in environmental issues from an early age. He integrates this background with his professional rowing experience, giving him a unique perspective and connection to water. At Princeton, he is a research assistant at the High Meadows Environmental Institute, working with Reed Max-

well in the Integrated Groundwater Modeling Institute. For his senior thesis, he researched the impact of abandoned oil and gas wells on groundwater vulnerability. In the summer of 2025, he interned as a water engineer for Arcadis, a company he plans to rejoin after graduation.

Katherine Baker is a master's student in environmental management at Yale School of the Environment, specializing in water resource science and management. She works as a part-time research assistant in James Saiers's hydrology lab. Beyond this project, her recent research contributions include water quality monitoring, urban stream restoration analysis, and using environmental DNA to monitor coastal cyanobacteria fluctuations. In future work, she aims to continue utilizing her laboratory analysis and visual communication experience to enhance sustainable water access for vulnerable communities. She earned a BS in environmental studies from the University of Richmond.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Soriano Jr. MA, Warren JL, Wang YC, van Velden D, Lin S, Maxwell RM, Saiers JE. 2026. Higher odds of produced-water-like chemistry in groundwater vulnerable to shale gas spills. *Nat Sustain*, <https://doi.org/10.1038/s41893-026-01933-5>.

Research Report 243, *A Groundwater Modeling Framework for Understanding Potential Population Exposures to Groundwater Contamination Associated with Oil and Gas Development Across the Marcellus Region*, by J.E. Saiers et al.

INTRODUCTION

The advent of high-volume and multistage hydraulic fracturing combined with horizontal drilling, known as unconventional oil and gas development (UOGD),* in the early 2000s has resulted in a marked increase in onshore oil and natural gas development in the United States. This increase in the rate and scale of development has created additional instances for potential human exposure to chemical and nonchemical stressors and has generated concerns about their effect on human health.

UOGD releases to groundwater or surface water may occur under accidental conditions (e.g., spills), posing a risk of groundwater contamination. Thus, they have been the subject of much public attention and concern, especially in regions such as the Marcellus of Appalachia where there is substantial UOGD alongside populations that rely heavily on groundwater for drinking water. Although a review by the US EPA¹ did not find extensive contamination of water associated with UOGD, it did cite instances where contamination of groundwater had occurred. It concluded that understanding remained limited regarding how UOGD releases to groundwater and surface water could lead to population exposures through contaminated water. Data on the likelihood, magnitude, frequency, or duration of exposure to water potentially contaminated by UOGD are limited, highlighting the need for studies quantifying potential population exposures to water contamination associated with UOGD processes.²

In 2023, HEI Energy issued a Request for Qualifications [RFQ E23-2](#): Regional Groundwater Modeling to Understand Community Exposures Associated with Oil and Gas Development. The goal of the RFQ was to develop a novel groundwater quality modeling approach to better understand the potential for documented UOGD spills in the Marcellus region to adversely affect water quality and lead to population exposures.

James E. Saiers's 1-year study, "Potential Groundwater Vulnerability to Spills Associated with Unconventional Oil and Gas Development Across the Marcellus Region," began in May 2024. Total expenditures were \$250,000. The draft Investigators' Report from Saiers and colleagues was received for review in March 2025. A revised report, received in October 2025, was accepted by the HEI Energy Review Committee in November 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.

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

In November 2023, Dr. James E. Saiers submitted the application, "A Groundwater Modeling Framework for Elucidating Community Exposures Across the Marcellus Region to Contamination Associated with Oil and Gas Development." The Energy Research Committee thought the proposed study was very responsive to the RFQ and had several notable strengths, one being the clearly articulated modeling approach. This study is the only study funded under [RFQ E23-2](#).

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

SCIENTIFIC AND REGULATORY BACKGROUND

UOGD OVERVIEW

UOGD refers to the development and production of oil and natural gas through multistage hydraulic fracturing in horizontal wells.³ UOGD processes occur on and off the well pad and include the following:

- **Field development:** exploration; pad preparation; vertical and horizontal drilling; well completion (casing and cementing, perforating, acidizing, hydraulic fracturing, millout, flowback, and well testing) in preparation for production; and waste management.
- **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.

Although some UOGD operations are regulated at the federal level under the Clean Air Act, the Clean Water Act, and the Safe Drinking Water Act, state-level regulations play a larger role in governing UOGD more generally. Such regulations vary among states, with some states defining minimum setback distances between UOGD and specific land uses, such as residential homes and schools, to protect local populations. For example, in Ohio, Pennsylvania, and West Virginia, which form part of the Marcellus region examined in this study, setback distances separating well pads from buildings or occupied dwellings are 150 feet, 500 feet, and 625 feet, respectively.⁴⁻⁶

UOGD AND WATER QUALITY

UOGD releases to groundwater and surface water may occur through authorized or accidental releases. Authorized releases include permitted discharges of treated UOGD wastewater to surface water or the use of produced water outside oil and gas operations for applications such as road treatment.^{2,7} Accidental releases, or spills, include those that may occur because of wellbore failure, malfunctioning equipment, or other circumstances.⁸ The potential effect of a spill on groundwater or surface water quality depends on the chemical and physical properties of the spilled constituents and the characteristics of the receiving surface water body or subsurface aquifer.

Prior research has documented thousands of spills related to UOGD across several states, and other studies have demonstrated associations between documented UOGD spills and contaminants in nearby groundwater.^{8–11} Quantifying any UOGD-related exposures is challenging because it can be difficult to distinguish potential exposures associated with UOGD from other sources, such as conventional oil and gas development, other anthropogenic sources, or naturally occurring conditions.² Some researchers have used geochemical signatures to detect effects on water quality from surface spills attributable to UOGD.¹² However, detecting changes in water quality from UOGD-related spills to the subsurface is more difficult given the complexity and uncertainty in understanding hydrogeologic conditions, which affect potential migration or transformation of any contaminants. Novel groundwater modeling studies have begun to make progress on such issues.¹³

Nonetheless, efficient frameworks and methods are still needed that could directly address the challenges in isolating and evaluating UOGD effects on water quality and the potential for population exposures associated with water quality impacts, a need this study aimed to address.

STUDY OBJECTIVES

The goal of Saiers's study was to develop a modeling framework for evaluating the vulnerability of groundwater to potential contamination from spills associated with UOGD in the Marcellus region states of Ohio, Pennsylvania, and West Virginia. The study had two specific aims:

1. **Aim 1.** Develop a physics-based model that is applicable across the region to quantify the vulnerability of groundwater that supplies domestic wells (hereafter "well water vulnerability") to potential contamination from UOGD spills.
2. **Aim 2.** Evaluate associations between UOGD chemical fingerprints measured in well waters and the estimates of UOGD-related well water vulnerability from Aim 1.

For Aim 1, Saiers and colleagues coupled a physics-based groundwater flow model with a particle-tracking algorithm to simulate the transport trajectories of contaminants through

the groundwater flow system in the Marcellus region following their release at a UOGD well pad. They performed multiple model simulations and integrated the results to quantify the probability of a potential contaminant associated with a UOGD spill reaching a drinking water receptor (i.e., estimates of well water vulnerability across the Marcellus region states of Ohio, Pennsylvania, and West Virginia). Well water vulnerability estimates were computed for two scenarios: (1) for all UOGD wells in the region ($n = 25,617$), and (2) for a subset of UOGD wells in the region ($n = 1,761$) with documented spills using information collected from state data repositories in Ohio, Pennsylvania, and West Virginia. The investigators also calculated the total potentially vulnerable land area and population for both scenarios.

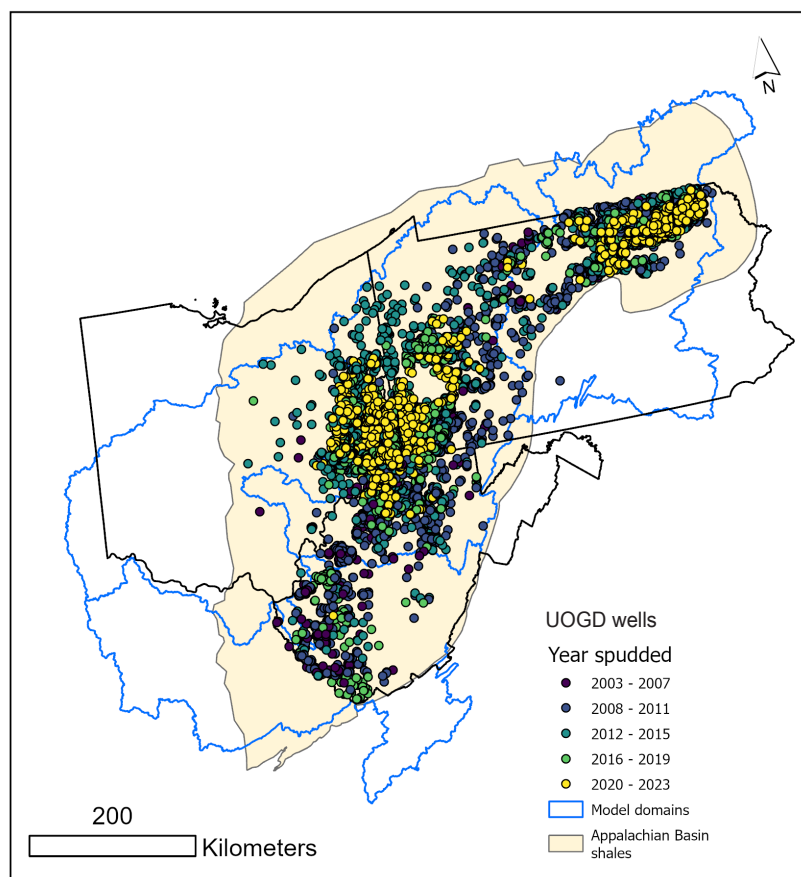
For Aim 2, Saiers and colleagues collated previously published data on well water (which is supplied by groundwater) in these states that included information on chemical signatures. They then used logistic regression models to evaluate associations between the presence or absence of UOGD chemical fingerprints in well water and their computed estimates of well water vulnerability from Aim 1 in areas with documented cases of UOGD spills.

SUMMARY OF METHODS AND STUDY DESIGN

AIM 1. ESTIMATING WELL WATER VULNERABILITY TO POTENTIAL CONTAMINATION FROM UOGD IN THE MARCELLUS REGION

Saiers and colleagues combined two-dimensional physics-based hydrologic models of the study region to simulate advective transport (i.e., the movement of a substance by bulk flow of a liquid) of potential contaminants through the groundwater after their release at UOGD well pads. Specifically, they used MODFLOW 6, which was based on prior models of aquifer systems across the contiguous United States¹⁴ with a particle-tracking algorithm, MODPATH 7.¹⁵ The portion of the Marcellus that constituted the study region spanned four hydrologic model domains and included over 26,000 UOGD wells (**Commentary Figure 1**).

The investigators calibrated the groundwater flow models using multiple hydraulic conductivity fields (i.e., different representations of how easily water moves through subsurface materials such as soil or rock) and performed multiple model simulations of particle-tracking for two scenarios: (1) a particle was released at all grid cells containing a UOGD well in the region, and (2) at grid cells for a subset of UOGD wells in the region with documented spills. Of a total 25,617 wells in the region, approximately 3,853 documented UOGD spills were compiled, broadly covering the 2007–2023 period. Data were obtained from regulatory agencies in Ohio, Pennsylvania, and West Virginia, and supplemented with peer-reviewed publications. Documented spills were not included if they could not be geolocated or collocated with known UOGD well pads, which resulted in a final dataset of 1,761 documented spills across the region that were analyzed. The



Commentary Figure 1. Map of the study region, including the hydrologic model domains and number of UOGD wells (colors indicate the year drilling began [“year spudded”]). Source: Adapted from Investigators’ Report Figure 3.

particle-tracking results across simulations were integrated to calculate estimates of well water vulnerability in each model grid cell for both scenarios. The results were reported as the probability (0–1) of the path line of a particle released from a UOGD well pad intersecting a grid cell across model simulations.

In addition, the investigators calculated the total land area and population dependent on groundwater wells and potentially at risk.¹⁶ At-risk populations are located within “vulnerable” grid cells, as defined by multiple vulnerability thresholds. Specifically, vulnerability thresholds of 0.001, 0.1, or 0.5 were used, with higher values indicating greater agreement among the model simulations that a grid cell lies along the trajectory of potential contaminants released from a UOGD well pad, thereby indicating greater vulnerability. An estimate of 0.5 or more indicated “high vulnerability.” They also aggregated these results to 30-km polygons to explore potential spatial clustering of well water vulnerability in the region.

AIM 2. EVALUATING ASSOCIATIONS BETWEEN WELL WATER VULNERABILITY AND UOGD CHEMICAL FINGERPRINTS IN WELL WATER MEASUREMENTS

Saiers and colleagues next evaluated potential associations between chemical fingerprints indicating the presence of UOGD-related chemicals in well water measurements in the study region and the estimates of well water vulnerability. They used logistic regression models to determine such associations for the subset of UOGD wells with documented spills from Aim 1. Water quality data collected across Ohio, Pennsylvania, and West Virginia between the years 2008 and 2020 were obtained from a publicly available database¹⁷ and from peer-reviewed publications.^{18–20} These data consisted of approximately 27,000 measurements of well water supplied by groundwater.

Well water measurements were classified as either having UOGD chemical fingerprints or not, depending on their chemical compositions. Groundwater studies¹⁸ have used analysis of inorganic ion ratios to determine whether groundwater had been affected by mixing with produced waters (which make up a large fraction of fluids in spills from UOGD well pads). Here, the investigators analyzed the concentration ratios of six different combinations of inorganic ion-ratio pairs (**Commentary Table 1**) to explore their potential for determining chemical fingerprints associated with UOGD. They focused on chemicals often found in high concentrations within produced waters.

Next, the investigators implemented both unadjusted and adjusted logistic regression models to estimate odds ratios (ORs) and *P* values. Adjusted models included covariates such as geographic location (e.g., latitude and longitude), distance to geological structures (e.g., folds, faults), topographic position (e.g., summits, slopes, valleys), and the presence of

Commentary Table 1. Combinations of Inorganic Ion-Ratio Pairs Analyzed in This Study to Classify UOGD Chemical Signatures in Well Water Measurements.

Ion-Ratio Pair Combinations
Barium/chloride vs. bromide/sulfate
Sulfate/chloride vs. bromide
Magnesium/lithium vs. sulfate/chloride
Magnesium/lithium vs. bromide/sulfate
Calcium/strontium vs. calcium/magnesium
Sulfate/chloride vs. magnesium/sodium

nearby (less than 2 km) conventional oil and gas wells. These covariates were introduced to control for other factors known to impart produced water–like fingerprints, such as natural upwelling of deep-basin brines beneath valleys or near fault zones. They ran two sets of models: (1) vulnerability was included as a continuous variable (i.e., $0 \leq V \leq 1$, as computed in Aim 1); and (2) vulnerability was included as a binary variable indicating well water vulnerability was “high” or not (defined as $V \geq 0.5$ or $V < 0.5$).

SUMMARY OF KEY RESULTS

AIM 1. WELL WATER VULNERABILITY TO POTENTIAL CONTAMINATION FROM UOGD IN THE MARCELLUS REGION

Groundwater simulations from the hydrologic models demonstrated robust performance in capturing local groundwater hydrology for the Marcellus region. Estimated mean and median lengths of particle path lines for the trajectory of potential contaminants through the groundwater flow system were 671 m and 532 m, respectively, with a 95th percentile length of 1,790 m, consistent with previous work.¹³ Thus, potential chemical contamination of groundwater associated with UOGD well pads was not expected to occur at distances greater than 2 km. However, the length of these predicted path lines generally exceeds regulatory setback distances between UOGD and wells supplying drinking water in Pennsylvania (152 m), Ohio (15 m), and West Virginia (76 m).⁴⁻⁶

Saiers and colleagues estimated that well water vulnerability to potential contamination from UOGD spills for scenario 1 (i.e., considering all documented UOGD wells in the region as potential sources of spills) encompassed a land area that ranged from 3,538 km² using a vulnerability threshold of 0.1 (i.e., the probability that a particle path line from a UOGD well pad intersected a grid cell in at least 100 of 1,000 model simulations) to 3,758 km² using a lower vulnerability threshold of 0.001. Estimates of well water vulnerability for scenario 2 (i.e., considering only UOGD wells in the region with documented spills) encompassed a total land area ranging from 590 km² to 616 km², using vulnerability thresholds of 0.1 and 0.001, respectively.

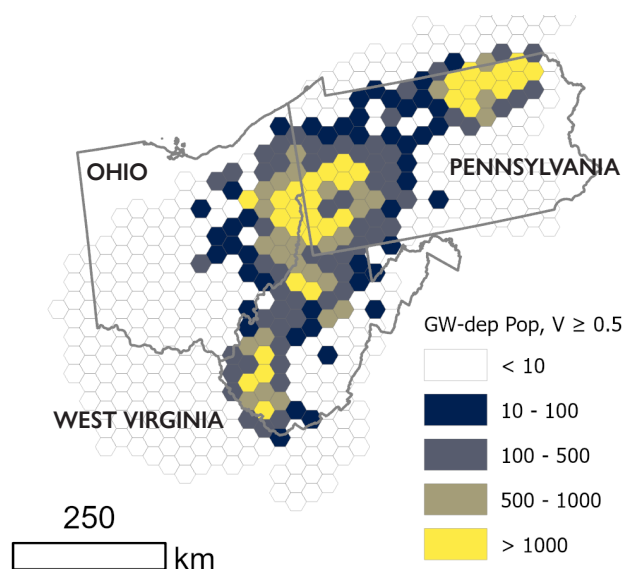
Using existing estimates of populations dependent on groundwater wells in the region,¹⁶ the investigators calculated that for scenario 1, about 3%, or 122,000 people, live in areas that would be considered vulnerable to potential well water contamination from a spill (using a high vulnerability threshold: 0.5). For scenario 2, this estimate would amount to 0.5% of that population, or about 22,000 people.

To estimate potential clustering of exposed populations, they aggregated the data to 30 km polygons. Under scenario 1, they observed that estimates of populations living in locations that have high well water vulnerability ($V > 0.5$) tended to be spatially clustered throughout Northeastern

Pennsylvania, Southern West Virginia, and throughout the tri-state area where the three states meet (**Commentary Figure 2**). The investigators note that such clustering may suggest both regional-scale heterogeneity in local groundwater flow patterns and the accumulation of estimated vulnerability in locations where there are multiple UOGD well pads in proximity to one another.

AIM 2. ASSOCIATIONS BETWEEN UOGD CHEMICAL FINGERPRINTS IN WELL WATER MEASUREMENTS AND WELL WATER VULNERABILITY

Saiers and colleagues analyzed the concentration ratios of six combinations of inorganic ion-ratio pairs in the water quality measurements data to assess their potential for determining chemical fingerprints associated with UOGD (Commentary Table 1). Because of the number of missing or censored measurements, the combinations of ion-ratio pairs that could be evaluated as criteria for identifying UOGD chemical fingerprints in well water measurements were limited to calcium/strontium versus calcium/magnesium and sulfate/chloride versus magnesium/sodium. Of these, calcium/strontium versus calcium/magnesium demonstrated the highest accuracy in correctly identifying samples with UOGD chemical signatures when evaluated against water samples from oil and gas wells in the Marcellus region. Thus, the investigators selected calcium/strontium versus calcium/magnesium as their criterion for classifying whether water quality measurements exhibited the presence of potential UOGD chemical fingerprints. A total of 16,090 water quality measurements were included in their subsequent statistical analysis, of which 865 (5.4%) were classified as having poten-



Commentary Figure 2. Estimated population dependent on groundwater (GW-dep), aggregated to 30 km polygons, with high well water vulnerability ($V \geq 0.5$) to potential contamination from UOGD spills in the Marcellus region. Source: Adapted from Investigators' Report Figure 7.

tial UOGD chemical fingerprints.

The investigators reported positive associations (i.e., statistically significant ORs > 1) between the odds of detecting a chemical fingerprint indicating the presence of UOGD in well water measurements and well water vulnerability (Commentary Table 1). In unadjusted models, when vulnerability was estimated as a continuous variable between 0 and 1, the odds of well water measurements exhibiting UOGD chemical fingerprints increased 14% for every 0.1 increase in well water vulnerability. When vulnerability was estimated as a binary variable indicating high well water vulnerability (or not), the odds of the well water measurement exhibiting UOGD chemical fingerprints increased 271% for high well water vulnerability.

The investigators also ran adjusted models that accounted for covariates such as geographic location, topography, and the presence of nearby conventional oil and gas wells. Using adjusted models, these associations became attenuated but remained elevated and statistically significant (**Commentary Table 2**). For example, in adjusted models, every 0.1 increase in well water vulnerability was associated with a 13% increase in the odds that a well water measurement also exhibited UOGD chemical fingerprints. In both sets of models, the presence of conventional oil and gas wells within 2 km was a statistically significant predictor of increased odds of detecting UOGD chemical signatures in a well water measurement. Given the difficulty in distinguishing between chemicals released from unconventional versus conventional sources without intensive water quality analyses, this result highlights the complexity of how different sources of water influence groundwater chemistry.

Commentary Table 2. Associations Between the Presence of Potential UOGD Chemical Signatures in Marcellus Region Well Water Measurements and Estimates of Well Water Vulnerability^a

Variable	Odds Ratio (P Value)	
	0 ≤ V ≤ 1	V ≥ 0.5
Unadjusted Models		
Well water vulnerability	1.14 (<0.001)	3.71 (<0.001)
Adjusted Models		
Well water vulnerability	1.13 (<0.001)	3.24 (<0.001)
Distance to geological structures	0.82 (<0.01)	0.82 (<0.01)
Presence of conventional oil and gas wells within 2 km	1.47 (<0.05)	1.47 (<0.05)
Upper slopes vs. peaks/summits	0.79 (0.430)	0.79 (0.430)
Lower slopes vs. peaks/summits	0.99 (0.967)	0.99 (0.969)
Valleys vs. peaks/summits	1.77 (0.052)	1.77 (0.052)

^aVulnerability estimates are modeled as either a continuous variable (0–1), or as a binary variable (0 or 1 depicting no versus high well water vulnerability, defined as V > 0.5), with select covariates. Associations are presented as ORs and P values.

HEI ENERGY REVIEW COMMITTEE'S EVALUATION

This study estimated domestic well water vulnerability to potential contamination from UOGD-related spills in three states — Ohio, Pennsylvania, and West Virginia — in the Marcellus region, using combined physics-based hydrologic models with a particle-tracking algorithm. It also examined associations between estimated well water vulnerability and UOGD chemical fingerprints in well water sampled in locations with documented UOGD spills in the same region. Saiers and colleagues found that 0.5% to 3% of populations who rely on groundwater-supplied well water in the region live in locations that may be considered highly vulnerable to potential contamination from UOGD spills. They also found that the odds of detecting chemical fingerprints of UOGD are greater in hydrologically vulnerable locations, even when controlling for other factors known to impart such fingerprints.

In its independent evaluation, the HEI Energy Review Committee found that this study provided an effective modeling framework for examining groundwater vulnerability to potential contamination from UOGD spills that may be useful to inform decision-making. They also appreciated that the framework could be extended to other regions in the United States and applied to other research questions related to examining water quality and UOGD. The Committee broadly agreed with the results but noted that data quality issues for publicly available data on documented UOGD spills were the main limitation of this work. Further details on the strengths and limitations of the study are discussed below.

STUDY DESIGN, DATASETS, AND ANALYTICAL APPROACHES

The Committee identified the overall modeling approach as a key strength of the study. It appreciated the curation of a dataset of water quality measurements from publicly available resources for the three study region states that could be used in future studies. It also appreciated the multiple analytical approaches used to explore well water vulnerability.

Several limitations were highlighted by the Committee. In particular, the Committee thought that the high percentage of missing or incomplete information for documented spills from state regulatory databases (approximately 54% of spills) that were therefore excluded from the analyses was the main limitation of this work. This missing information likely biased the investigators' findings toward an underestimation of well water vulnerability in the region. However, the Committee appreciated that the investigators analyzed all UOGD well pad locations in the region to partially address this limitation. The analysis provides an upper threshold of well water vulnerability to potential contamination from spills. This approach was an appropriate and clever way to address the limitations of the publicly available data on spills.

Second, the Committee remarked that the use of a two-dimensional instead of a three-dimensional hydrologic model to simulate groundwater flow in the region might limit the generated estimates of well water vulnerability. Importantly, a three-dimensional model would have captured considerations of flow across vertical variations. In addition, to simulate transport trajectories of potential contaminants through the groundwater flow system, the investigators used a particle-tracking algorithm that was based on advective transport rather than advective-dispersive transport. Advective transport is based on bulk flow, whereas advective-dispersive transport would account for mixing and spread of contaminants. Using advective transport might also limit the generated estimates of well water vulnerability. However, the Committee noted that the investigators appropriately acknowledged these limitations, noting that simplifying assumptions in the modeling approaches were used to address data and computational limitations.

Third, the Committee commented on the statistical analyses conducted to evaluate associations between estimates of well water vulnerability and chemical fingerprints of UOGD in water quality measurements. It observed that the inclusion of geographic location (latitude and longitude) as a covariate to account for potential spatial autocorrelation likely only captures larger-scale spatial correlation. This approach thus might not capture smaller, local-scale spatial correlation given that such geographic coordinates do not represent the nuance of local-scale distances and relationships.

FINDINGS AND INTERPRETATION

In their estimates of well water vulnerability, Saiers and colleagues reported a range of results that depended on different thresholds of vulnerability and scenarios of UOGD spills. For example, they considered all UOGD well pads as potential spill sources and considered only UOGD well pads with documented spills as sources. This approach affected the estimates for the vulnerable land area (up to 3,758 km²) and groundwater well-dependent populations (up to 122,000 people). The Committee thought that the findings were presented both prudently and thoughtfully given some of the limitations of the study's datasets and analytical approaches. In addition, the investigators highlighted the utility of the findings in informing future decision-making concerning UOGD and water quality. For example, they discussed results in the context of regulatory setback distances.

In statistical analyses evaluating associations between well water vulnerability and UOGD chemical fingerprints in well water measurements, Saiers and colleagues reported a large (up to 271%) increase in the odds of detecting UOGD chemical fingerprints in well water samples. This result came from analyses when considering estimated vulnerability as a binary indicator (i.e., highly vulnerable well water, or not). However, the investigators recognized that the chemical fingerprints defined as indicators of UOGD in this analysis did not constitute the range of chemicals known to occur in spills that are known to be of human health concern, such as some semivolatile or volatile organic compounds, which the Committee appreciated.

CONCLUSIONS

In summary, Saiers and colleagues found that up to 3% of people who depend on groundwater-supplied well water in the region live in areas that may be "highly vulnerable" to well water contamination. This number resulted from an analysis that considered all UOGD well pads as potential contaminant sources, regardless of whether actual spills had been recorded. The affected area corresponded to clusters of locations among Northeastern Pennsylvania, Southern West Virginia, and the area where those two states meet with Ohio.

Saiers and colleagues also reported a higher likelihood of detecting chemical fingerprints of UOGD in water quality measurements and when well water vulnerability was high for locations with documented UOGD spills in the region. The study is limited by the quality of publicly available data on documented spills in the region and by some simplifying assumptions related to the groundwater flow and contaminant transport simulations. However, overall, this study provides a strong modeling framework and valuable datasets that can be used to inform decision-making, for example by helping to predict the location of drinking water wells that are most vulnerable to UOGD spills.

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 Cloelle Danforth for oversight of the study, to Yasmin Romitti 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 for her role in preparing this Research Report for publication.

REFERENCES

1. 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, 2016. Washington, DC: US EPA.
2. HEI Energy Research Committee. 2020. Human Exposure to Unconventional Oil and Gas Development: A Literature Survey for Research Planning. Communication 1. Boston, MA: Health Effects Institute.
3. Legislative Service Commission. Ohio Laws and Administrative Rules. OH Rev Code § 1509.021 Surface Locations of New Wells. Available: <https://codes.ohio.gov/ohio-revised-code/section-1509.021>.
4. PA General Assembly. 58 PA Cons Stat § 3215 Well Location Restrictions. Available: <https://www.legis.state.pa.us/WU01/LI/LI/CT/HTM/58/00.032.015.000..HTM>.
5. West Virginia Code. WV Code § 22-6A-12 Well Location Restrictions. Available: <https://code.wvlegislature.gov/22-6A-12/>.
6. HEI Special Scientific Committee on Unconventional Oil and Gas Development in the Appalachian Basin. 2015. Strategic Research Agenda on the Potential Impacts of 21st Century Oil and Natural Gas Development in the Appalachian Region and Beyond. Boston, MA: Health Effects Institute.
7. 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(4):e2024GH001263, <https://doi.org/10.1029/2024GH001263>.
8. 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* 581:369–377, <https://doi.org/10.1016/j.scitotenv.2016.12.142>.
9. Drollette BD, Hoelzer K, Warner NR, Darrah TH, Karatum O, O'Connor MP, et al. 2015. Elevated levels of diesel range organic compounds in groundwater near Marcellus gas operations are derived from surface activities. *Proc Natl Acad Sci USA* 112:13184–189, <https://doi.org/10.1073/pnas.1511474112>.
10. 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>.
11. Patterson LA, Konschnik KE, Wiseman H, Fargione J, Maloney KO, Kiesecker J, et al. 2017. Unconventional oil and gas spills: Risks, mitigation priorities, and state reporting requirements. *Environ Sci Technol* 51:2563–2573, <https://doi.org/10.1021/acs.est.6b05749>.
12. 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>.
13. Soriano Jr. MA, Deziel NC, Saiers JE. 2022. Regional scale assessment of shallow groundwater vulnerability to contamination from unconventional hydrocarbon extraction. *Environ Sci Technol* 56:12126–12136, <https://doi.org/10.1021/acs.est.2c00470>.
14. Zell WO, Sanford WE. 2020. Calibrated simulation of the long-term average surficial groundwater system and derived spatial distributions of its characteristics for the contiguous United States. *Water Resources Research* 6(8):e2019WR026724, <https://doi.org/10.1029/2019WR026724>.
15. Pollock DW. 2016. User guide for MODPATH Version 7 — A particle-tracking model for MODFLOW. Open-File Report 2016–1086. Reston, Virginia: US Geological Survey. Available: <https://dx.doi.org/10.3133/ofr20161086>.
16. Johnson TD, Belitz K, Lombard MA. 2019. Estimating domestic well locations and populations served in the contiguous US for years 2000 and 2010. *Sci Total Environ* 687:1261–1273, <https://doi.org/10.1016/j.scitotenv.2019.06.036>.
17. Shale Network. Status of the Shale Network Database. Available: <https://shalenetwork.org/>.
18. Soriano Jr. MA, Siegel HG, Johnson NP, Gutchess KM, Xiong B, Li Y, et al. 2021. Assessment of groundwater well vulnerability to contamination through physics-informed machine learning. *Environ Res Lett* 16:084013, <https://doi.org/10.1088/1748-9326/ac10e0>.
19. Siegel HG, Soriano Jr. MA, Clark CJ, Johnson NP, Wulsin HG, Deziel NC, et al. 2022. Natural and anthropogenic processes affecting domestic groundwater quality within the Northwestern Appalachian Basin. *Environ Sci Technol* 56:13761–13773, <https://doi.org/10.1021/acs.est.2c04011>.
20. Bugher NA, Xiong B, Gentles RI, Glist LD, Siegel HG, Johnson NP, et al. 2024. Domestic groundwater wells in Appalachia show evidence of low-dose, complex mixtures of legacy pollutants. *Environ Sci: Process Impacts* 26:2250–2263, <https://doi.org/10.1039/D4EM00364K>.

ABBREVIATIONS AND OTHER TERMS

DISSOLVED IONS

Ba	barium
Br	bromide
Ca	calcium
Cl	chloride
Li	lithium
Mg	magnesium
Na	sodium
SO ₄	sulfate
Sr	strontium
P_{it}	probability of observing a UOGD fingerprint
R	risk
V	vulnerability
Y_{it}	binary indicator of a UOGD fingerprint in measurement taken from well

OTHER ABBREVIATIONS

COGD	conventional oil and gas development
HUC	hydrologic unit code
UOGD	unconventional oil and gas development
c1, c2, c3, c4	four calibration setups of the groundwater flow model

HEI BOARD, ENERGY COMMITTEES, and STAFF

BOARD OF DIRECTORS

Richard A. Meserve, Chair Senior of Counsel, Covington and Burling LLP; President Emeritus, Carnegie Institution for Science; former Chair, US Nuclear Regulatory Commission

Stephen Corman President, Corman Enterprises

Martha J. Crawford Operating Partner, Macquarie Asset Management

Ana V. Diez Roux Dana and David Dornsife Dean and Distinguished University Professor of Epidemiology, Dornsife School of Public Health, Drexel University; Director, Drexel Urban Health Collaborative

Alan I. Leshner CEO Emeritus, American Association for the Advancement of Science

Catherine L. Ross Regents' Professor Emerita, City and Regional Planning and Civil and Environmental Engineering, Georgia Institute of Technology; Chairman of the Board of Directors of the Auto Club Group, American Automobile Association

Martha E. Rudolph Environmental Attorney, Former Director of Environmental Programs, Colorado Department of Public Health and Environment

Karen C. Seto Frederick C. Hixon Professor of Geography and Urbanization Science, Yale School of the Environment, Yale University

Jared L. Cohon In Memoriam (1947–2024)

ENERGY COMMITTEE OF THE HEI BOARD OF DIRECTORS

Martha E. Rudolph, Chair Environmental Attorney, Former Director of Environmental Programs, Colorado Department of Public Health and Environment

Richard A. Meserve Senior of Counsel, Covington and Burling LLP; President Emeritus, Carnegie Institution for Science; former Chair, US Nuclear Regulatory Commission

Catherine L. Ross Regents' Professor Emerita, City and Regional Planning and Civil and Environmental Engineering, Georgia Institute of Technology; Chairman of the Board of Directors of the Auto Club Group, American Automobile Association

Jared L. Cohon In Memoriam (1947–2024)

HEI ENERGY RESEARCH COMMITTEE

George Hornberger, Chair Distinguished Professor Emeritus, Vanderbilt University

Shari Dunn-Norman* Chancellor's Professor Emeritus, Petroleum Engineering, Missouri University of Science and Technology

Stefanie Ebelt* Professor, Department of Environmental Health, Rollins School of Public Health, Emory University

Alfred Eustes Associate Professor Emeritus, Petroleum Engineering Department, Colorado School of Mines

Julia Haggerty Professor of Geography, Department of Earth Sciences, Montana State University

Howard Hu* Professor, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles

Kirsten Koehler Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University

Judy S. LaKind* Adjunct Associate Professor, Department of Epidemiology and Public Health, University of Maryland School of Medicine

Christopher Paciorek Adjunct Professor and Research Computing Consultant, Department of Statistics, University of California, Berkeley

*Shari Dunn-Norman, Stefanie Ebelt, Howard Hu, and Judy S. LaKind rotated off the Research Committee before this report's publication.

HEI BOARD, ENERGY COMMITTEES, and STAFF

Armistead (Ted) G. Russell *Howard T. Tellepsen Chair and Regents Professor of Civil and Environmental Engineering, Georgia Institute of Technology*

Peter S. Thorne *University of Iowa Distinguished Chair and Professor, Department of Occupational and Environmental Health, University of Iowa College of Public Health*

Yifang Zhu *Professor of Environmental Health Sciences, UCLA Fielding School of Public Health, University of California, Los Angeles*

HEI ENERGY REVIEW COMMITTEE

Isabelle Cozzarelli, Chair *Senior Research Scientist Emerita, USGS Geology, Energy, and Minerals Science Center*

Jim Crompton *Affiliate Professor, Petroleum Engineering Department, Colorado School of Mines*

Elizabeth Mannshardt* *Adjunct Associate Professor, Department of Statistics, North Carolina State University*

Albert Presto *Director, Center for Atmospheric Particle Studies; Research Professor, Department of Mechanical Engineering, Carnegie Mellon University*

Christine Wiedinmyer *Director of UCAR Community Programs, University Corporation for Atmospheric Research (UCAR)*

Jun Wu *Professor, Department of Environmental and Occupational Health, University of California, Irvine*

STAFF AND CONSULTING SCIENTISTS

Elena Craft *President and CEO*

Ellen K. Mantus *Director of Science*

Jacqueline C. Rutledge *Director of Finance and Administration*

Pallavi Pant *Director of Global Initiatives*

Anna Rosofsky *Interim Manager of HEI Energy*

Emily Alden *Corporate Secretary*

Daniel S. Greenbaum *President Emeritus, In Memoriam (1952–2024)*

Robert M. O'Keefe *Vice President Emeritus*

Annemoon M. van Erp *Deputy Director of Science Emerita*

Amy Andreini *Senior Science Communications Specialist*

Hanna Boogaard *Consulting Principal Scientist*

Jacki Collins *Senior Accountant*

Dan Crouse *Senior Scientist*

Cloelle Danforth *Senior Scientist*

Philip J. DeMarco *Controller*

Alexa Diehl *Administrative Coordinator*

Kristin C. Eckles *Senior Editorial Manager*

Hlina Kiros *Research Assistant*

Lissa McBurney *Senior Science Administrator*

*Elizabeth Mannshardt rotated off the Review Committee before this report's publication.

HEI BOARD, ENERGY COMMITTEES, and STAFF

(Staff and Consulting Scientists, continued)

Samantha Miller *Research Associate*

Victor Nthusi *Program Manager (Africa)*

Allison P. Patton *Senior Scientist*

Yasmin Romitti *Staff Scientist*

Abinaya Sekar *Program Manager (South Asia)*

Robert Shavers *Operations Manager*

Eva Tanner *Staff Scientist*

Alexis Vaskas *Digital Communications Manager*

Donna J. Vorhees *Senior Advisor for HEI Energy*

Derek Walker *Senior Advisor*

RESEARCH REPORT

NUMBER 243
OCTOBER 2026



HEI
energy

Health Effects Institute

One Beacon Street Suite 21300
Boston, Massachusetts 02108
USA
+1-617-488-2300

www.healtheffects.org

www.HEIenergy.org