



HEI Research Report 243

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

James E. Saiers et al.

Appendices A through D

Correspondence 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.

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.

© 2026 Health Effects Institute, One Beacon Street, Suite 21300, Boston, MA 02108

Contents

Appendix A. Testing Alternative Groundwater Model Configurations	2
Figure A-1. Recharge forcing from (A) ZS20 and (B) PFC2 datasets.....	2
Figure A-3. Model calibration results for domain (A) 0205, (B) 0501, (C) 0504, and (D) 0505.	4
Figure A-4. Distribution of path line lengths (meters) from particle-tracking simulations.	5
Appendix B. Evaluating Classification Accuracy of Ion-Ratio Pairs and Statistical Model Sensitivity Analysis via Random Permutations.....	6
Table B-1. Summary of Groundwater Quality Data for Evaluation of UOGD Produced Water- Like Imprints	6
Table B-2. Accuracy in Classifying Marcellus Produced Waters.....	7
Table B-3. Accuracy in Classifying Unimpacted Groundwaters.....	7
Table B-4. Sensitivity Analysis Results for the Permuted Dataset ($n = 16,090$).....	8
Appendix C. Companion Work: Assessing the Influence of Unsaturated Zone Flow	9
Figure C-1. Simulated flow path lengths from integrated surface-subsurface hydrologic modeling and particle tracking.....	10
Appendix D. Companion Work: Evaluation of Vulnerability to Abandoned Oil and Gas Wells	11
Figure D-1. Vulnerability to contamination from abandoned oil and gas wells.	12
References.....	13

Appendix A. Testing Alternative Groundwater Model Configurations

We explored additional sources of model uncertainty in terms of input forcing and parametrization. Specifically, we examined two versions of climatic forcing and two versions of geologic parametrization, resulting in four model setups derived from their combinations. These datasets came from two main sources: the Zell and Sanford¹ simulations, ZS20, and the spin-up simulation for ParFlow-CONUS2.0, PFC2.² The ZS20 forcing refers to long-term, average recharge estimates developed from United States Geological Survey streamflow observations³ (**Figure A-1**). The PFC2 forcing refers to long-term average precipitation minus evapotranspiration estimates derived from gridded meteorological datasets.⁴ We also adopted both the ZS20 and PFC2 approaches to geologic parametrization to represent the spatial distribution of hydraulic conductivity. In this case, ZS20 hydraulic conductivity zones are based on surficial geology maps,⁵ while the PFC2 zones are extracted from layer 4 of the ParFlow-CONUS2 hydrogeologic framework⁶ (**Figure A-2**). This layer corresponds to an aquifer unit underneath surficial soil layers.

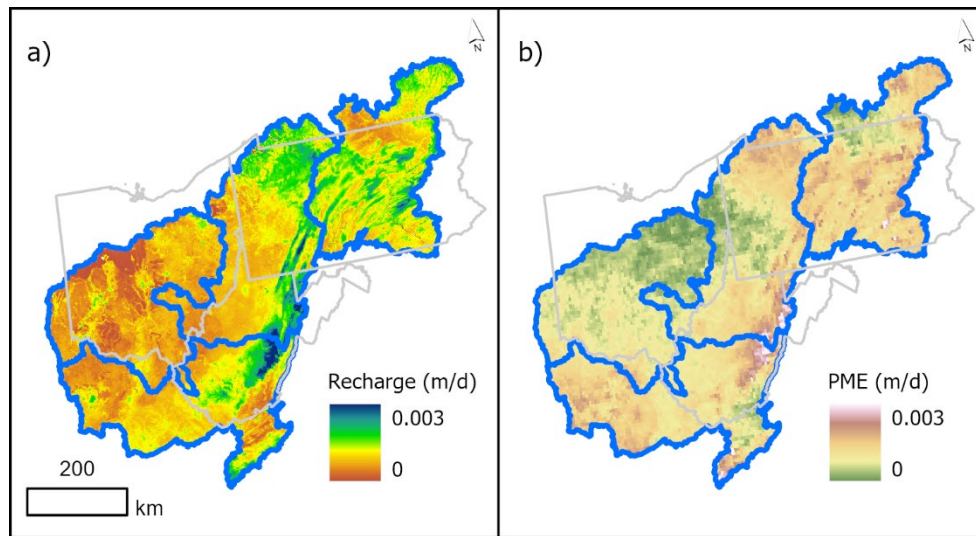


Figure A-1. Recharge forcing from (A) ZS20 and (B) PFC2 datasets.

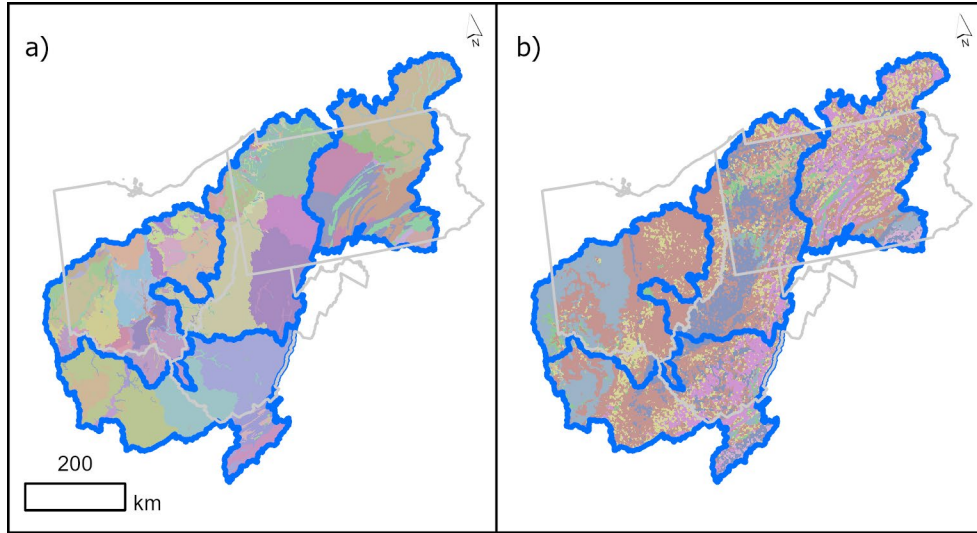


Figure A-2. Hydraulic conductivity parameter zones from (A) ZS20 and (B) PFC2 datasets.

Thus, the combinations of these datasets form four model setups that we tested in our analysis: c1 (ZS20 recharge + ZS20 geology), c2 (ZS20 recharge + PFC2 geology), c3 (PFC2 recharge + ZS20 geology), and c4 (PFC2 recharge + PFC2 geology). We ran PESTPP-IES for a maximum of three iterations to create calibrated ensembles for each setup, with each ensemble comprising 1,000 realizations of the hydraulic conductivity field.

All groundwater model setups converged to similar post-calibration performance after three PESTPP-IES optimization iterations, indicating satisfactory minimization of residuals between available long-term average water level observations and their simulated steady-state equivalents (**Figure A-3**).

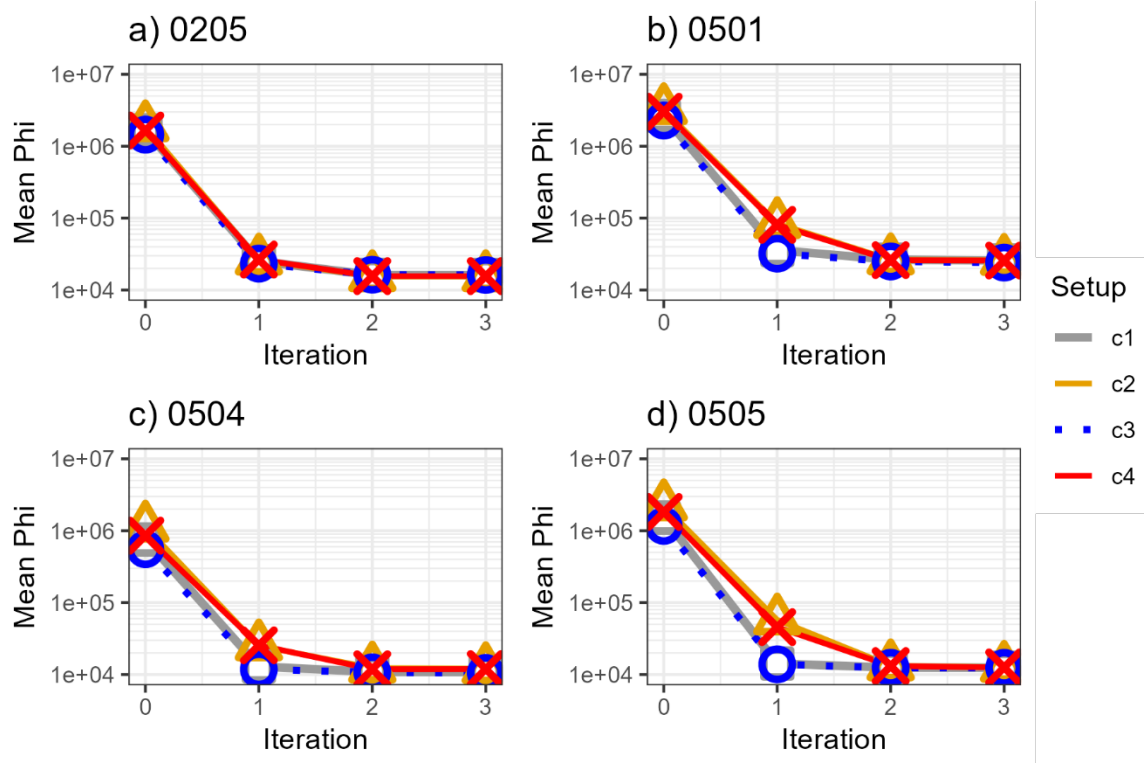


Figure A-3. Model calibration results for domain (A) 0205, (B) 0501, (C) 0504, and (D) 0505. The four setups tracked each other closely in domain 0205, while setups based on the PFC2 hydraulic conductivity parametrization (c2 and c4) exhibited higher residuals in the first iteration in domains 0501, 0504, and 0505. The difference in residuals across parameterizations decreased substantially in the second iteration, and by the end of the third iteration, the distribution of residuals was similar across the four setups. Phi is equivalent to the sum of squared residuals between groundwater level observations and their simulated equivalents. The number of observations used for calibration is 9,026 (0205), 10,995 (0501), 5,274 (0504), and 6,698 (0505). See the Groundwater Flow Model and Ensemble Calibration section in the main text for a description of the four simulation domains.

The four setups tracked each other closely in domain 0205, while setups based on the PFC2 hydraulic conductivity parametrization (c2 and c4) exhibited higher residuals in the first two iterations in domain 0501, 0504, and 0505. At the end of iteration 3, the distribution of residuals was similar across the four setups. These results indicate the robustness of the simulations despite the uncertainty incorporated through the multiple hydraulic conductivity parametrizations and recharge boundary conditions we specified. The particle-tracking results were similarly robust across the different model setups (**Figure A-4**). Thus, the main text only presents results based on setup c1, which is the original setup by Zell and Sanford.¹

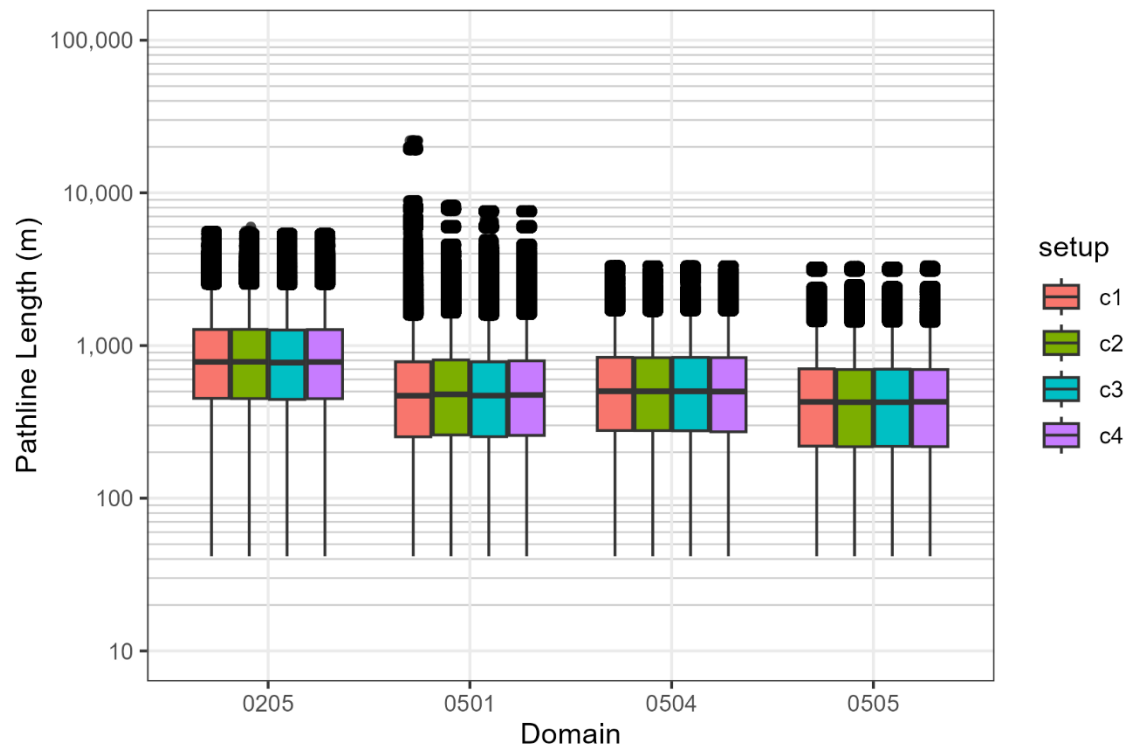


Figure A-4. Distribution of path line lengths (meters) from particle-tracking simulations. The central box represents the interquartile range (IQR), spanning from the first quartile (Q1) to the third quartile (Q3). The line inside the box indicates the median (Q2) of the data. “Whiskers” extend from the box to the smallest and largest values within 1.5 times the IQR from Q1 and Q3, respectively. Data points outside this range are plotted individually as outliers.

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

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 (**Table B-1**).

Table B-1. Summary of Groundwater Quality Data for Evaluation of UOGD Produced Water-Like Imprints

	Ba/Cl vs. Br/SO ₄	SO ₄ /Cl vs. Br	Mg/Li vs. SO ₄ /Cl	Mg/Li vs. Br/SO ₄	Ca/Sr vs. Ca/Mg	SO ₄ /Cl vs. Mg/Na
Number of samples	578	590	785	263	16,090	18,062
Number of PW- imprinted samples	49	40	65	69	865	136
Percentage of PW- imprinted samples	8.5%	6.8%	8.3%	26.2%	5.4%	0.8%

Note: Number of well water samples with measurements needed to compute a particular ion-ratio pair, number of samples with a produced water (PW) imprint, and percentage of samples with a PW imprint.

We evaluated the ability of the methodology described in Section 3.2.1 to accurately classify produced water samples from Marcellus oil and gas wells. The Marcellus well samples were obtained from the National Produced Water Database.⁸ Ba/Cl versus Br/SO₄ performed poorly, classifying only 56% of produced waters from Marcellus wells correctly (**Table B-2**). Classifications based on criteria SO₄/Cl versus Br, Mg/Li versus SO₄/Cl, and SO₄/Cl versus Mg/Na were better but had accuracies under 90%. Ca/Sr versus Ca/Mg exhibited the highest accuracy, correctly identifying 93.5% of the Marcellus-well samples as produced water.

Table B-2. Accuracy in Classifying Marcellus Produced Waters

	Ba/Cl vs. Br/SO₄	SO₄/Cl vs. Br	Mg/Li vs. SO₄/Cl	Mg/Li vs. Br/SO₄	Ca/Sr vs. Ca/Mg	SO₄/Cl vs. Mg/Na
Number of samples	158	158	154	153	353	158
Number correctly classified	89	141	136	142	330	131
Percentage correctly classified	56.3%	89.2%	88.3%	92.8%	93.5%	82.9%

Note: Number of produced water samples from Marcellus wells with measurements needed to calculate a particular ion-ratio pair, number of samples correctly classified as produced water, and percentage of correctly classified samples.

We also evaluated the ability of the ion-ratio pairs to correctly classify unimpacted, meteoric groundwaters from our study region. For this analysis, unimpacted groundwaters were assumed to be sampled from water wells in upland locations at least 2.5 km from the nearest documented UOGD spill. By excluding lower slope and valley locations, we reduced the probability of groundwater samples being influenced by natural upwelling of deep-basin brines capable of imparting a UOGD-like signature. Mg/Li versus Br/SO₄ performed poorly, classifying only 70% of the unimpacted samples correctly (**Table B-3**). SO₄/Cl versus Br, Ca/Sr versus Ca/Mg, and SO₄/Cl versus Mg/Na had the highest accuracy, each classifying at least 97.5% of unimpacted waters correctly, although the number of samples with ion measurements enabling evaluation of SO₄/Cl versus Br was small.

Table B-3. Accuracy in Classifying Unimpacted Groundwaters

	Ba/Cl vs. Br/SO₄	SO₄/Cl vs. Br	Mg/Li vs. SO₄/Cl	Mg/Li vs. Br/SO₄	Ca/Sr vs. Ca/Mg	SO₄/Cl vs. Mg/Na
Number of samples	75	76	63	10	3,366	4,831
Number correctly classified as unimpacted	69	76	61	7	3,283	4,823
Percentage correctly classified as unimpacted	92.0%	100.0%	96.8%	70.0%	97.5%	99.8%

Note: Number of unimpacted groundwater samples with measurements needed to calculate a particular ion-ratio pair, number of samples correctly classified as unimpacted, and percentage of correctly classified samples.

Of the six ion-ratio criteria considered, our analysis suggests that Ca/Sr versus Ca/Mg is most useful and reliable for identifying produced water imprints in well water samples. This criterion is based on combinations of ions that are reported and uncensored in a large number of samples within our database. Moreover, this criterion exhibits the greatest accuracy in classifying produced waters, and it is among the most accurate in classifying unimpacted waters.

Table B-4. Sensitivity Analysis Results for the Permuted Dataset ($n = 16,090$)

<i>Unadjusted</i>	Approach 1	Permuted 1	Permuted 2	Permuted 3
Vulnerability to UOGD spills	1.14 (<0.001)	0.98 (0.5065)	1 (0.8746)	0.99 (0.6537)
<i>Adjusted</i>				
Vulnerability to UOGD spills	1.13 (<0.001)	0.98 (0.470)	0.99 (0.842)	0.99 (0.643)
Upper slopes vs. peaks/summits	0.79 (0.430)	0.67 (0.079)	0.9 (0.673)	0.99 (0.966)
Lower slopes vs. peaks/summits	0.99 (0.967)	0.63 (<0.05)	1.03 (0.897)	1.06 (0.826)
Valleys vs. peaks/summits	1.77 (0.052)	0.72 (0.155)	0.96 (0.866)	1.01 (0.969)
Distance to geologic structures	0.82 (<0.01)	1.04 (0.214)	1.02 (0.602)	1.01 (0.733)
Presence of conventional oil and gas wells within 2 km ($n > 0$ vs. $n = 0$)	1.47 (<0.05)	1.04 (0.751)	0.85 (0.171)	1.08 (0.509)

Note: The findings from Approach 1 in Table 2 in the main report are shown for comparison. Statistically significant associations ($\alpha = 0.05$) are indicated in bold typeface.

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

We evaluated the importance of unsaturated zone processes in subsurface transport using ParFlow, an integrated hydrologic model,⁹ and EcoSLIM, a Lagrangian particle-tracking model specifically tailored to ParFlow output.^{2,10} ParFlow simultaneously solves the three-dimensional variably saturated subsurface flow using the three-dimensional Richards equation and two-dimensional overland flow using the shallow water equation.⁹ The ParFlow Continental United States 2.0 (ParFlow-CONUS2.0)¹¹ is a ParFlow simulation that covers the entire contiguous United States in a single computational domain, spanning approximately 7.85 million km². It features a horizontal resolution of 1 km and a subsurface depth of 392 m discretized into 10 layers with variable thicknesses of 200, 100, 50, 25, 10, 5, 1, 0.6, 0.3, and 0.1 m from bottom to top.^{6,12} The top four layers (2 m total) are soils, whereas the bottom six layers (390 m) represent various geologic units, including unconfined and confined aquifer systems. The model was driven by long-term average potential recharge (precipitation minus evapotranspiration) to achieve a quasi-steady state.¹²

Starting from the steady-state three-dimensional flow field generated by ParFlow-CONUS2.0, we extracted individual HUC4 subsets corresponding to the study area of Zell and Sanford using subset tools, a Python package to build and run smaller domain models.¹³ The seven major HUC4 units analyzed here are 0205 (Susquehanna), 0501 (Allegheny), 0502 (Monongahela), 0503 (Upper Ohio), 0504 (Muskingum), 0505 (Kanawha), and 0507 (Big Sandy Guyandotte). We then reran the spun-up ParFlow-CONUS2 model with a single time step to produce velocity fields for each HUC4 region. Forward particle tracking from unconventional wells to streams was performed using EcoSLIM. For computational efficiency, two particles were released at the land surface at each UOGD site to simulate spills. Detailed particle trajectories across the unsaturated and saturated zones were recorded.

Particle-tracking results from the HUC4 domains exhibited some variability in the distributions of both total and unsaturated flow path lengths, reflecting spatial heterogeneity in subsurface hydrologic conditions (**Figure C-1**). Median total flow path lengths across the seven HUC4s ranged from 780 m (0503) to 1,590 m (0502), while median unsaturated flow path lengths ranged from 16 m (0503) to 38 m (0507). Across all domains, the maximum total flow path length was 50,160 m. In all cases, mean values greatly exceeded medians, indicating highly right-skewed distributions with a small number of exceptionally long flow paths that contribute to pronounced long tails. The ratio of unsaturated to total flow path length is consistently less than 10% in all HUC4s. These patterns indicate that unsaturated flow paths represent only a small fraction of the overall transport pathway. These results suggest that unsaturated zone processes are less significant relative to those in the saturated zone in our study area, allowing us to focus our analysis on saturated zone dynamics.

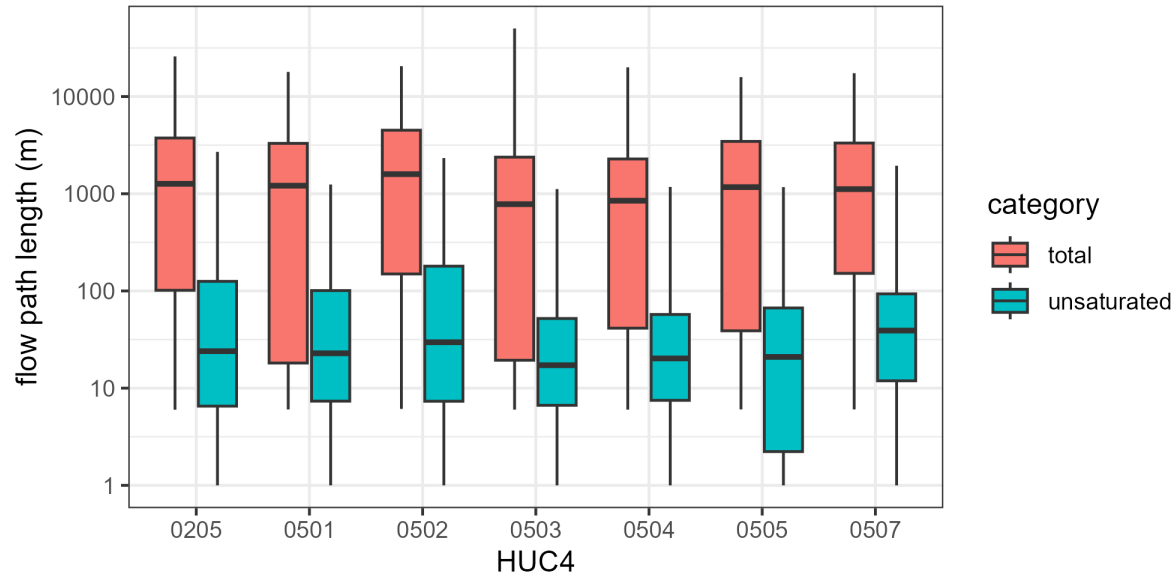


Figure C-1. Simulated flow path lengths from integrated surface-subsurface hydrologic modeling and particle tracking. Total flow path length includes flow across the full particle trajectory from release at the land surface to discharge at a stream cell. Unsaturated flow path length refers to the portion of the total flow path above the water table. Note that HUC4s in this figure correspond to official United States Geological Survey designations, not to our MODFLOW domain IDs (cf. Table 2 in main report).

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

We adopted the methodology developed in the main report to quantify vulnerability to contamination from abandoned oil and gas wells. Abandoned wells refer to wells that are out of operation and have not been adequately decommissioned through methods like plugging or sealing. These wells typically belong to the conventional oil and gas well category. While surface spills are unlikely in these inactive locations, abandoned wells may pose risks to groundwater if they leak toxic contaminants into aquifers through faulty casings. For this analysis, we used a dataset of verified abandoned wells in the United States.¹⁴ The researchers who assembled the dataset estimate that this corresponds to only 3% of abandoned wells in the country, such that actual risks may be more than what we can estimate from our analysis. The dataset contains a total of 29,584 abandoned wells located within our simulation domains.

Simulated path lengths were similar to those of unconventional wells, with 82% of path lengths less than 1 km. The median path length was 567 m. The area with vulnerability ≥ 0.001 spans 8,816 sq km, constituting 2.9% of the study region (**Figure D-1**). Within these areas, an estimated population of 402,948 individuals who depend on domestic groundwater wells may be impacted. These area and population estimates are over two times greater than those for UOGD as described in the main report and principally are due to the larger number and spatial spread of abandoned wells serving as potential sources.

Vulnerability to abandoned oil and gas wells is concentrated in different regions compared with that for UOGD. The vulnerability to abandoned wells is concentrated in the north and southwest portion of domain 0501 and the northeast of domain 0504, with very little vulnerability within domain 0205. These patterns align with the distribution of conventional oil and gas fields. Nevertheless, there are areas where vulnerability to both UOGD and abandoned oil and gas overlap, including some areas in southwestern Pennsylvania and along the Ohio–West Virginia boundary. Additional research is needed to elucidate the cumulative impacts of active UOGD and legacy oil and gas activity on groundwater quality and public health in such communities. Further research to quantify hazards, such as the likelihood and magnitude of aqueous contaminant releases from abandoned wells, is needed for a more holistic view of risks.

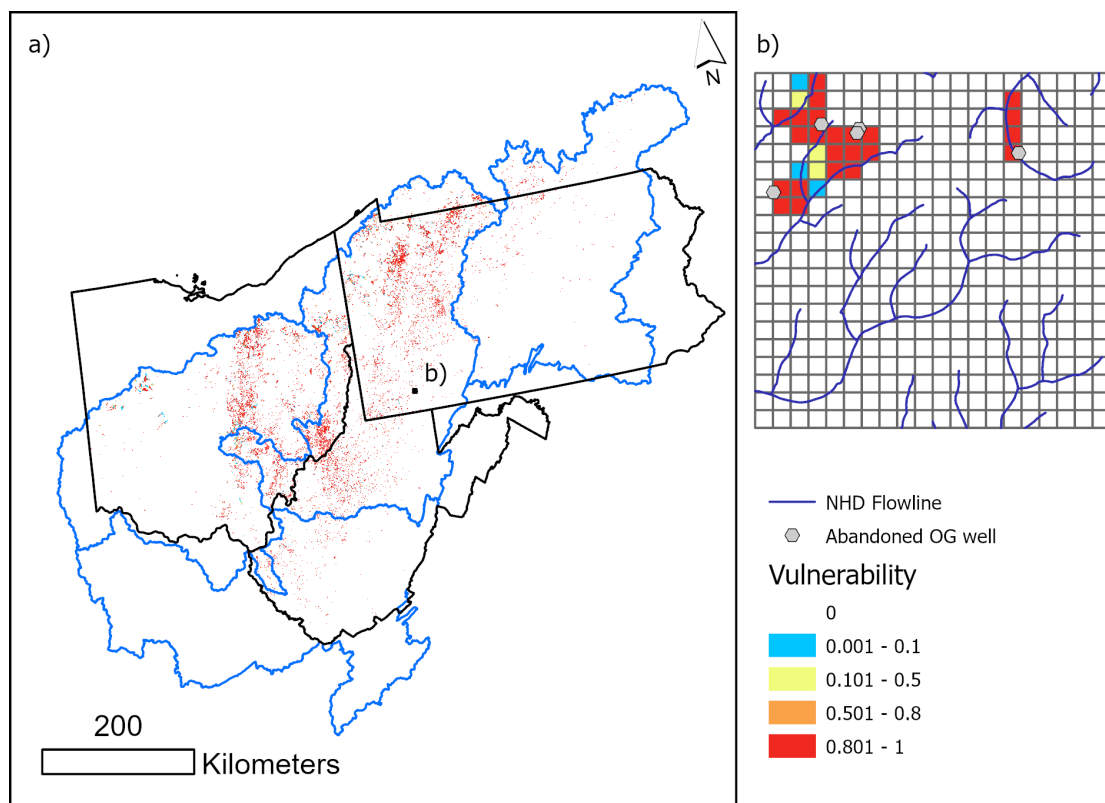


Figure D-1. Vulnerability to contamination from abandoned oil and gas wells. (A) Results for entire 305,000 sq km study region. (B) Detailed view of results for a 25 sq km window.

References

1. 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>.
2. Yang C, Condon LE, Maxwell RM. 2025. Unravelling groundwater–stream connections over the continental United States. *Nature Water* 3(1):70–79, <https://doi.org/10.1038/s44221-024-00366-8>.
3. Reitz M, Sanford WE, Senay GB, Cazenias J. 2017. Annual estimates of recharge, quick-flow runoff, and evapotranspiration for the contiguous US using empirical regression equations. *J Am Water Resour Assoc* 53(4):961–983, <https://doi.org/10.1111/1752-1688.12546>.
4. Livneh B, Bohn TJ, Pierce DW, Munoz-Arriola F, Nijssen B, Vose R, et al. 2015. A spatially comprehensive, hydrometeorological dataset for Mexico, the US, and Southern Canada 1950–2013. *Sci Data* 2(1):150042, <https://doi.org/10.1038/sdata.2015.42>.
5. Soller DR, Reheis MC, Garrity CP, Van Sistine DR. 2009. Map database for surficial materials in the conterminous United States. US Geological Survey Data Series 425, scale 1:5,000,000. Available: <https://pubs.usgs.gov/ds/425/>.
6. Swilley JS, Tijerina-Kreuzer D, Tran HV, Zhang J, Yang C, Condon LE, et al. 2024. Continental-scale hydrostratigraphy: comparing geologically informed data products to analytical solutions. *Groundwater* 62(1):75–92, <https://doi.org/10.1111/gwat.13354>.
7. 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>.
8. Blondes MS, Knierim KJ, Croke MR, Freeman P, Doolan C, Herzberg AS, et al. 2023. US Geological Survey National Produced Waters Geochemical Database (ver. 3.0, December 2023): US Geological Survey data release. Available: <https://doi.org/10.5066/P9DSRCZJ>
9. Kollet SJ, Maxwell RM. 2005. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv Water Resour* 29(7):945–958, <https://doi.org/10.1016/j.advwatres.2005.08.006>.
10. Maxwell RM, Condon LE, Danesh-Yazdi M, Bearup LA. 2019. Exploring source water mixing and transient residence time distributions of outflow and evapotranspiration with an integrated hydrologic model and Lagrangian particle tracking approach. *Ecohydrology* 12(1):e2042, <https://doi.org/10.1002/eco.2042>.
11. Yang C, Tijerina-Kreuzer DT, Tran HV, Condon LE, Maxwell RM. 2023. A high-resolution, 3D groundwater-surface water simulation of the contiguous US: Advances in the integrated ParFlow CONUS 2.0 modeling platform. *J Hydrol* 626:130294, <https://doi.org/10.1016/j.jhydrol.2023.130294>.
12. Tijerina-Kreuzer D, Swilley JS, Tran HV, Zhang J, West B, Yang C, et al. 2024. Continental-scale hydrostratigraphy: Basin-scale testing of alternative data-driven approaches. *Groundwater* 62(1):93–110, <https://doi.org/10.1111/gwat.13357>.

13. Triplett AK, Artavanis G, Hasling WM, Maxwell RM, Defnet A, Johnson AM, et al. 2024. SubsetTools: A Python package to subset data to build and run ParFlow hydrologic models. J Open Source Softw 9(99):6752, <https://doi.org/10.21105/joss.06752>.
14. Boutot J, Peltz AS, McVay R, Kang M. 2022. Documented orphaned oil and gas wells across the United States. Environ Sci Technol 56(20):14228–14236, <https://doi.org/10.1021/acs.est.2c03268>.