

LETTER • OPEN ACCESS

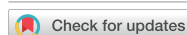
Large-sample evidence for surface and groundwater quality associations with marginal, orphaned, and unconventional oil and gas wells across the United States

To cite this article: Ruta Basijokaite *et al* 2026 *Environ. Res. Lett.* **21** 124042

View the [article online](#) for updates and enhancements.

You may also like

- [Autler–Townes splitting in Rydberg atoms: transition dipole matrix element extraction and field efficiency analysis](#)
Brian C Holloway, Gavin M Chase, Lee E Harrell *et al.*
- [ICRH modelling of DTT in full power and reduced-field plasma scenarios using full wave codes](#)
A Cardinali, C Castaldo, F Napoli *et al.*
- [Seasonal contrasts in the long-term evolution of rapidly intensifying tropical cyclones over the Western North Pacific](#)
Haeun Jo, Dong-Hyun Cha, Minkyu Lee *et al.*

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED

13 March 2026

REVISED

5 June 2026

ACCEPTED FOR PUBLICATION

11 June 2026

PUBLISHED

25 June 2026

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Large-sample evidence for surface and groundwater quality associations with marginal, orphaned, and unconventional oil and gas wells across the United States

Ruta Basijokaite^{1,*} , Mary Kang² and Tao Wen^{1,*} ¹ Department of Earth and Environmental Sciences, Syracuse University, Syracuse, NY, 13244, United States of America² Department of Civil Engineering, McGill University, Montreal, Canada

* Authors to whom any correspondence should be addressed.

E-mail: ruta.basijokaite@gmail.com and twen08@syr.edu**Keywords:** water quality, surface water, groundwater, active and orphaned wells, oil and gas developmentSupplementary material for this article is available [online](#)

Abstract

Oil and gas wells pose long-term risks to water resources, yet their cumulative impact remains poorly constrained, particularly for marginal (low productivity) and orphaned sites. Using approximately three million water-quality records across New York, Pennsylvania, Colorado, and Texas, we quantify how well proximity and density are associated with concentrations of oil-and-gas-relevant analytes, including barium, chloride, sodium, sulfate, strontium, and specific conductivity, in surface and groundwater. Unlike prior large-sample studies focused primarily on unconventional development, this analysis jointly evaluates marginal, orphaned, and unconventional wells using a uniform statistical framework. Although these analytes can also arise from natural and non-oil-and-gas sources, statistically significant relationships were widespread, with negative correlations most directly consistent with elevated concentrations near wells. Signals were stronger and more consistent in surface water than in groundwater, and subregional analyses in Texas show that statewide assessments can mask localized patterns controlled by hydrogeology and aquifer connectivity. Together, these results indicate that oil and gas infrastructure leaves a measurable geochemical imprint on water systems and highlight the need for hydrogeology-informed monitoring to detect, predict, and mitigate contamination from legacy and active oil and gas wells.

1. Introduction

Detecting and quantifying the environmental impacts from oil and gas production have long been at the center of scientific and public concern. Although this sector remains essential for meeting domestic and global energy demand [1] we still do not fully understand the environmental and societal risks associated with aging infrastructure and emerging production needs [2, 3]. As hydrocarbon extraction intensifies [4] and regulatory policies evolve [5], the potential of these risks continues to increase, underscoring the urgent need for improved detection and mitigation strategies.

Poorly managed oil and gas wells—both active and inactive—can degrade quality of groundwater, surface water, and air through multiple pathways, including produced-water spills and leaks, historic

pits and impoundments, wellhead leakage, faulty well casing, or poor abandonment practices, posing significant human health risks [6–8]. These wells allow hydrocarbons, salts, trace metals, and other toxic compounds to migrate into aquifers and surface water bodies [9–12]. Structural failures, well age, production rates, and maintenance history all influence leakage [13–29]. Abandoned or ‘orphaned’ wells are particularly concerning because they often remain unplugged and unmonitored due to funding shortfalls [19, 30].

Previous studies have documented elevated concentrations of barium (Ba), bromide (Br), chloride (Cl), sodium (Na), strontium (Sr), specific conductance (SC), and other constituents downstream from unconventional wells and treatment facilities in oil and gas producing states (e.g. Texas, Pennsylvania) across the U.S [7, 8, 31–41]. However, surface

water evidence remains limited due to its dynamic hydrology [42], while groundwater contamination signals often vary with geology and well integrity [34, 43–45]. Some investigations report no measurable region-wide degradation in water quality [34, 35, 46], likely reflecting local hydrogeologic heterogeneity, limited sampling, or impacts that are localized rather than regional in scale.

Most relevant previous large-scale studies focus on unconventional wells [3, 47–52], but far fewer studies consider marginal wells—low-production wells that yield no more than 15 barrels of oil equivalent per day and/or 90 000 cubic feet of natural gas per day over a 12 month period [53, 54]. Marginal wells account for a small percentage of total production yet a disproportionately large share of methane emissions [27, 55, 56], suggesting that aging infrastructure and limited maintenance may also be relevant for water-quality risks. Together with orphaned wells, these sites represent a widespread but undercharacterized legacy-infrastructure concern. As production declines and wells age, additional wells may transition into marginal production or orphaned status, increasing the need to understand whether these legacy and low-production wells leave detectable water-quality signals. However, no study has evaluated whether marginal, orphaned, and unconventional wells are associated with measurable water-quality impacts across multiple U.S. basins using uniform statistical methods.

Here, we conduct a comprehensive assessment of how marginal, orphaned, and unconventional wells influence surface and groundwater quality across Colorado, New York, Pennsylvania, and Texas (figure 1). Oil and gas production practices, rates, and regulations drastically vary across these states. These four states collectively capture the diversity of U.S. oil and gas development: Pennsylvania, Colorado, and Texas are among the nation's leading producers of unconventional hydrocarbons and have long histories of extraction, resulting in dense populations of both marginal and orphaned wells. In contrast, New York—where hydraulic fracturing has been banned since 2010—serves as an adjacent, low-production counterpart that enables comparison of hydrogeologic controls across shared aquifers and geologic provinces [57, 58]. Using approximately three million records of Ba, Cl, Na, Sr, SC, and sulfate (SO₄) that are often associated with oil and gas production [43, 59–61] collected over 70 years (1948–2024), we quantify relationships between water chemistry and well proximity and density. By integrating hydrogeologic context into our spatial analyses, we reveal how local geology and aquifer connectivity modulate detectable contamination signals. To our knowledge, this study represents the most extensive data-driven evaluation of oil and gas well impacts on water quality conducted to date.

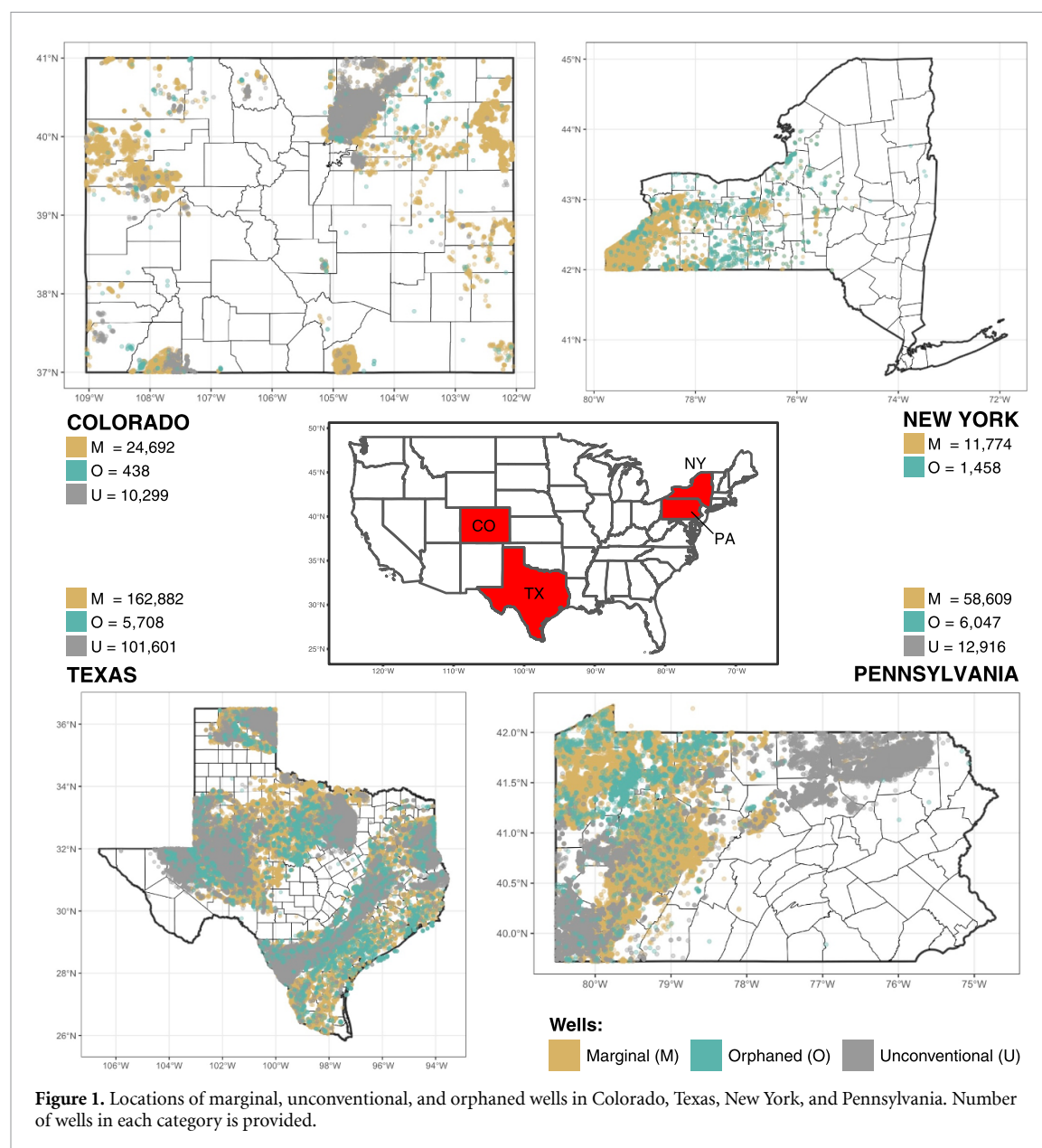
2. Materials and methods

2.1. Experimental design

Our analysis was divided into several categories. First, we chose four states with vast differences in economic contributions from oil and gas industry—Colorado, New York, Pennsylvania, and Texas—to explore how it affects water quality. Second, our analysis was focused on identifying distinct impacts from active (unconventional and marginal [54]) as well as orphaned oil and gas wells—wells that are unplugged and nonproducing with no solvent owner of record to plug and mitigate them [62]. Conventional active wells that were neither classified as marginal nor orphaned were not analyzed as a separate category. We focused on marginal, orphaned, and unconventional wells because these classes represent major policy priorities and because the water-quality impacts of marginal and orphaned wells remain comparatively undercharacterized at large spatial scales. Third, we explored surface as well as ground water data by analyzing six water quality analytes (Ba, Cl, Na, SC, SO₄, Sr) associated with oil-gas activity.

We retrieved oil and gas well information for each category from several databases. Information about marginal wells in Colorado, Pennsylvania and Texas was downloaded from Enverus database, while New York's data came from New York State Department of Environmental Conservation. Unconventional well data was downloaded from various national- and state-level databases: in Texas data was obtained from Enverus, in Pennsylvania from Pennsylvania Department of Environmental Protection [63], and in Colorado from Energy and Carbon Management Commission (ECMC) [64]. New York was excluded from this category since unconventional drilling was banned in 2010. All orphaned well data was obtained from Boutot *et al* [62].

As operating procedures differ from well to well, we chose six water quality analytes often associated with oil and gas production [43, 59–61] to analyze how oil-gas activities are impacting surface and ground water quality. Measurements of Barium (Ba), Chloride (Cl), Sodium (Na), Sulfate (SO₄), Strontium (Sr) as well as Specific Conductivity (Sp. Cond. or SC) concentrations were downloaded from a mix of state-wide and national databases. Water Quality Portal [65], which integrates water quality data from United States Geological Survey and the Environmental Protection Agency, was used to download all available surface and ground water samples collected over 70 years (1948–2024). In addition, we used state-level water quality databases to maximize the number of samples available for our statistical analysis: ECMC [64] in Colorado, Shale Network [66] in Pennsylvania, and Texas Water Development Board in Texas [67]. Number of available water quality samples (table S1) as well as their concentrations varied



greatly from state to state (figure S1) likely influenced by local geology and other environmental factors.

Water quality measurements underwent rigorous data quality control and samples not meeting our screening criteria were removed to avoid influencing statistical analysis results. Both national and state-level databases had samples collected at a variety of mediums. However, for our analysis we selected only water measurements collected from natural surface (river, stream) and groundwater (seep, spring, well) sources to avoid samples affected by anthropogenic sources or measurements that could be affected by stagnation. Samples that were flagged as rejected or contaminated were removed. In addition, concentrations below detection limit were also removed. Each dataset was manually inspected for outliers, and all abnormal concentration values were deleted. Basic statistics like mean, median, standard deviation, kurtosis and skewness for each analyte

aggregated by state are reported in table S2. To support reproducibility, all data-processing scripts and intermediate outputs used to document the number of records retained or removed at each screening step are provided in the project GitHub repository. These materials allow readers to trace the full screening workflow for each analyte, water type, and state.

2.2. Statistical analysis

We conducted geospatial analysis to examine the relationship between changes in water quality and proximity to oil and gas wells. Four proximity and density-based proxies were used to analyze whether analyte concentrations are higher near oil-gas wells which would signal negative impact. Spud dates were available for all wells included in the analysis except 88 orphaned wells in Colorado. As the Colorado orphaned-well dataset was small ($N = 438$), these

wells were assigned a placeholder spud date of 1900-01-01 to avoid excluding potentially relevant water-quality records. First, we calculated distances between unique water quality sampling locations and wells using *get.knnx* function from *FNN* package in R. Analyte measurements that were taken before well spud date were removed from further analysis as it would not be suitable for this impact analysis. These distance calculations were used to derive four spatial proxies: (1) distance to the closest well (in meters); (2) summed distance of 10 closest wells (in meters); (3) number of oil and gas wells within 1 kilometer radius from water sampling location; (4) number of oil and gas wells within 3 km radius from water sampling location. Proxies 1,3 and 4 were utilized in similar research studies [46, 68] and based on hydrogeologic literature assessing contaminant transport distances [49, 69, 70], whereas proxy 2 is newly proposed by this study. As concentration and proxy data were not normally distributed (table S2), we log-transformed our data before performing statistical analysis. Then, to analyze the relationship between log-transformed analyte concentrations and spatial proxies, we calculated Pearson correlation coefficient (R) and its significance level. We considered correlation results to be statistically significant if p -value < 0.005 . This significance threshold was chosen to control the family-wise error rate given the large number of simultaneous tests performed across analytes, states, and well classes (Bonferroni-motivated adjustment). We report these correlations results for each analyte and proxy analyzed separately among states, well types and water sources (figures 2–4 and S3). To evaluate whether observed correlations could arise from spatially structured but non-causal associations, we performed spatial permutation tests in which nearest-well-distance values were randomly shuffled within each analyte–state–well-type combination while holding water-quality measurements fixed at their original location. Correlations were recalculated for each permuted configuration and compared with the observed correlation coefficients. This procedure provided an empirical check on whether the observed relationships exceeded those expected under randomized spatial association. Details of the permutation procedure and summary results are provided in supplementary text S1 and table S4. In addition, we explored how Na/Cl molar ratios varied with distance from oil and gas wells as an additional diagnostic check on whether road salting could explain the observed Na, Cl, and specific conductivity patterns (figure S5). This analysis was used to evaluate, but not fully eliminate, the potential influence of non-oil-and-gas sources on the observed spatial relationships. We focused this diagnostic analysis on marginal wells because marginal wells occur in all four study states and represent a central well class in this study.

2.3. Visualization and hierarchical clustering

Hierarchical clustering was applied to determine whether states, well types, or analytes with similar spatial correlation structures grouped together in an unsupervised, data-driven manner providing a complement to the pairwise correlation results that does not depend on pre-specified hypotheses. This analysis was performed using *heatmap* function in *stats* package in R. This function allows users to represent values contained in a matrix using colors and has built-in statistical tools like clustering algorithm to visualize the results with dendrograms. It uses *hclust* function to perform hierarchical cluster analysis using Euclidean distance as the distance measure. A set of dissimilarities for the n objects in a matrix being clustered is used to perform hierarchical clustering. Initially, each value in a matrix is assigned to its own cluster and then iteratively two most similar clusters are joined until only one cluster remains. Dendrogram is displayed to represent clustering results. The algorithm used in *hclust* orders subtree showcasing the tighter cluster on the left. Here, we performed hierarchical clustering on correlation results between analyte concentrations and spatial proxies discussed above. The rows of the correlation matrix (figure 2) were re-ordered based on hierarchical clustering results, putting similar values close to each other (figure S3).

2.4. Impact of data selection in Texas

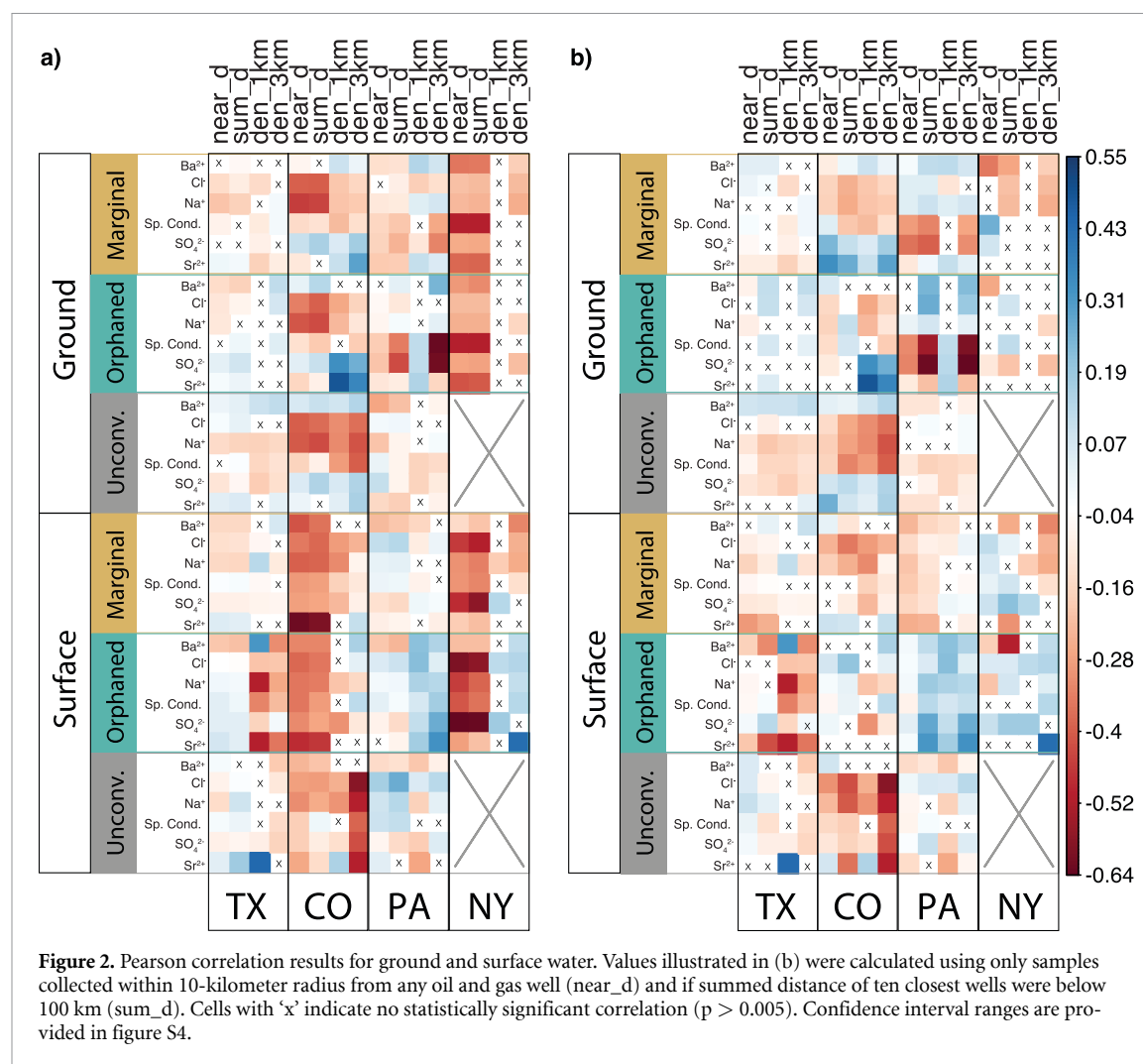
The water quality dataset was split based on major aquifers defined by Texas Water Development Board [71]. We chose to focus on aquifers that expand large subsurface areas to maximize the number of data points available per each aquifer: Ogallala, Pecos Valley, Edwards-Trinity (Plateau), Carrizo-Wilcox, Trinity, Edwards (Balcones Fault Zone), Gulf Coast. Water quality measurements were divided based on counties associated with each aquifer. This type of analysis allowed us to examine if unique aquifer properties need to be considered doing this type of large scale geospatial analysis as background analyte concentrations can vary considerably [72].

Our geospatial analysis framework allowed us to compare how marginal, unconventional, and orphaned oil and gas wells impact surface and ground water quality in Texas, Colorado, Pennsylvania, and New York. In addition, comparison of regional as well as sub-regional results allowed us to examine how data selection impacts our ability to detect oil-gas impact on water resources using statistical analysis alone.

3. Results

3.1. Sampling limits detection of contamination

Our dataset includes approximately three million records of six analytes (Ba, Cl, Na, Sr, SC, and



SO₄)—collected from 1948 to 2024 across New York, Pennsylvania, Colorado, and Texas (figure 1). Sample availability varied widely among states and between surface and groundwater systems (table S1, figures S1 and S2). Overall, surface water data were more abundant (836,019 samples in Texas; 583 561 in Colorado; 457 156 in Pennsylvania; 348 035 in New York) compared to groundwater data (284,045; 130 313; 158 846; 129 501 samples, respectively; tables S1 and S2).

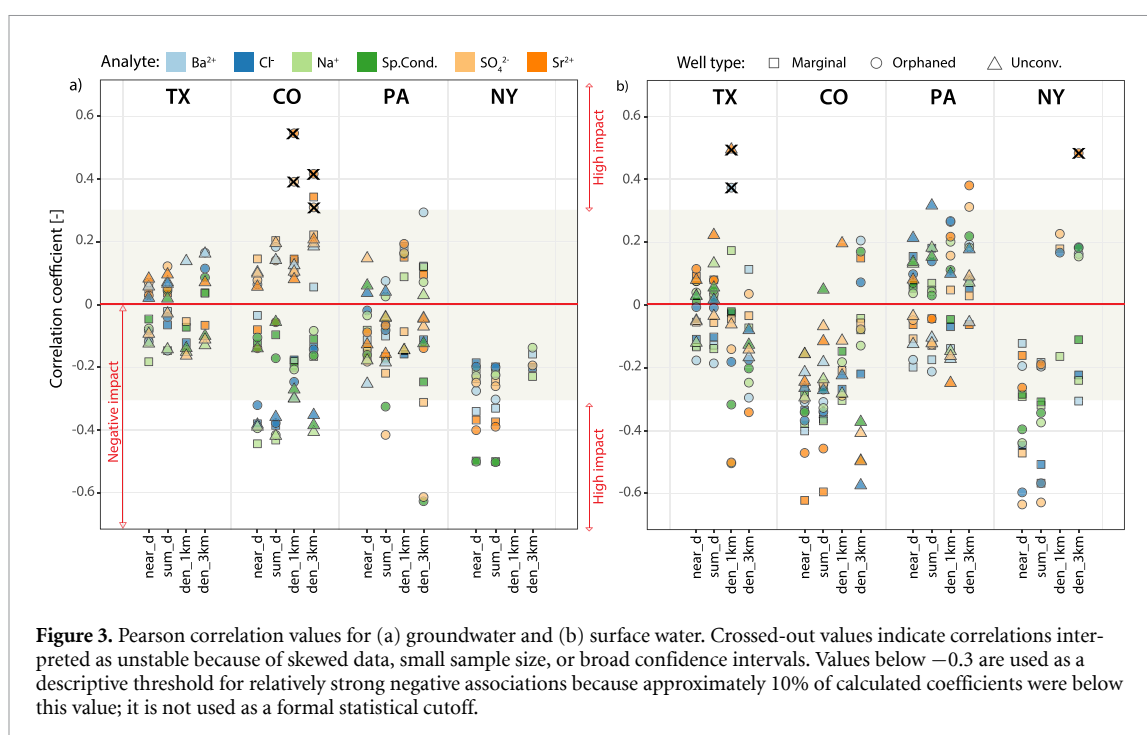
However, only a small fraction of these samples was collected near oil and gas wells. Water sampling coverage within one kilometer radius ranged from 1%–28% of available data depending on well type (figure S2). Ba and Sr measurements were extremely scarce near wells—often below 5% of total available samples—limiting statistical confidence for these species. This spatial bias emphasizes how localized data availability constrains our ability to detect contamination, particularly in heterogeneous hydrogeologic settings.

3.2. Surface waters show strong impacts

We quantified correlation between analyte concentrations and multiple spatial proxies of oil and gas

activity, including distance- and density-based metrics (figure 2). Coefficient values from density-based proxies were multiplied by -1 so that negative correlations consistently indicate degraded water quality near wells.

Across all states, oil and gas well proximity and density were frequently significantly associated with water quality variation (figure 2). Statistically significant correlations ($p < 0.005$), including both positive and negative relationships, were detected for 83.3% of surface water and 76.9% of groundwater analyses (figure 2(a)). Of all tests, 54.5% of surface-water analyses and 53.0% of groundwater analyses were both statistically significant and negative, and therefore most directly consistent with elevated concentrations near wells or in areas of higher well density. Because negative coefficients indicate higher analyte concentrations closer to wells or in areas of higher well density, negative correlations are most directly consistent with water-quality degradation near oil and gas infrastructure. The number of significant relationships declined when only proximity-limited samples were analyzed: from 76.9% to 67.8% for ground and from 83.3% to 75.8% for surface water analysis (figure 2(b)). This reduction—most pronounced



in New York (from 60.4% to 35.4% for ground and from 77.1% to 68.8% for surface water)—illustrates how limiting the spatial window can remove regionally coherent signals. Spatial permutation tests further supported the robustness of the main correlation patterns: observed nearest-well-distance correlations for key analytes exceeded those obtained from permuted configurations, indicating that the strongest relationships were unlikely to arise solely from chance spatial overlap. Because the proximity-limited subset reduced sample sizes and produced broader confidence intervals for several analytes (figure S4), subsequent analyses focus on the full dataset shown in figure 2(a). We interpret these results cautiously, emphasizing patterns that are consistent across analytes, spatial proxies, and regions rather than relying on statistical significance alone. We also focus on negative relationships that were observed in the proximity-restricted analysis because these results are less likely to reflect broad regional gradients unrelated to oil and gas development.

Statistical significance also varied with state and water type. Percentage of detected significant relationships varied more for groundwater (66.7% for TX, 91.7% for CO, 83.3% for PA, 73.6% for NY) as compared to surface water (83.3%, 84.7%, 86.1%, 84.7%) among states. Roughly 42 (16%) surface water coefficients (4 for TX, 23 for CO, 0 for PA, 15 for NY) and 31 (12%) groundwater coefficients (0, 15, 5, 11, respectively) were lower than -0.3 , indicating relatively strong negative associations (figure 3). We use -0.3 as a descriptive threshold for relatively strong negative associations because approximately 10% of calculated coefficients were below this value; it is not used as a formal statistical cutoff.

These numbers suggest that surface water is more impacted in Texas, Colorado, and New York, but not in Pennsylvania as none of the coefficient values exceeded -0.3 threshold. In addition, overall percentage of relationships indicating negative effects from oil and gas wells in PA were far greater for groundwater (58.3%) as compared to surface water (33.3%). Distance-based proxies produced more consistent patterns than density-based ones (85.2% vs. 56.8% achieving $p < 0.005$). Among statistically significant coefficients, 67.4% for groundwater and 68.2% for surface water were negative, consistent with elevated concentrations near wells or in areas of higher well density. Unconventional wells showed slightly weaker correlations than marginal or orphaned wells. In particular, marginal and orphaned wells had the highest negative effect in NY [$\text{mean}_{\text{marginal}} = -0.32$ (surface), $\text{mean}_{\text{orphaned}} = -0.4$ (surface), $\text{mean}_{\text{marginal}} = -0.31$ (ground), $\text{mean}_{\text{orphaned}} = -0.31$ (ground)] and CO [$\text{mean}_{\text{marginal}} = -0.37$ (surface), $\text{mean}_{\text{orphaned}} = -0.34$ (surface)]. Analyte-specific trends also emerged. Sr and SO₄ displayed the widest correlation ranges (-0.62 – -0.54 and -0.63 – -0.39 , respectively), while Na showed relatively uniform but weaker variations (-0.5 – -0.2). Surface water correlations were particularly strong in Colorado (mean = -0.3) and New York (mean = -0.36), whereas groundwater effects were more spatially variable but the strongest in NY (mean = -0.31). Overall, results were highly variable among states and water types likely indicating region-specific influence.

Significant positive correlations were also observed for some analytes and well classes. These relationships should not be interpreted as evidence of water-quality improvement near wells. Instead, they

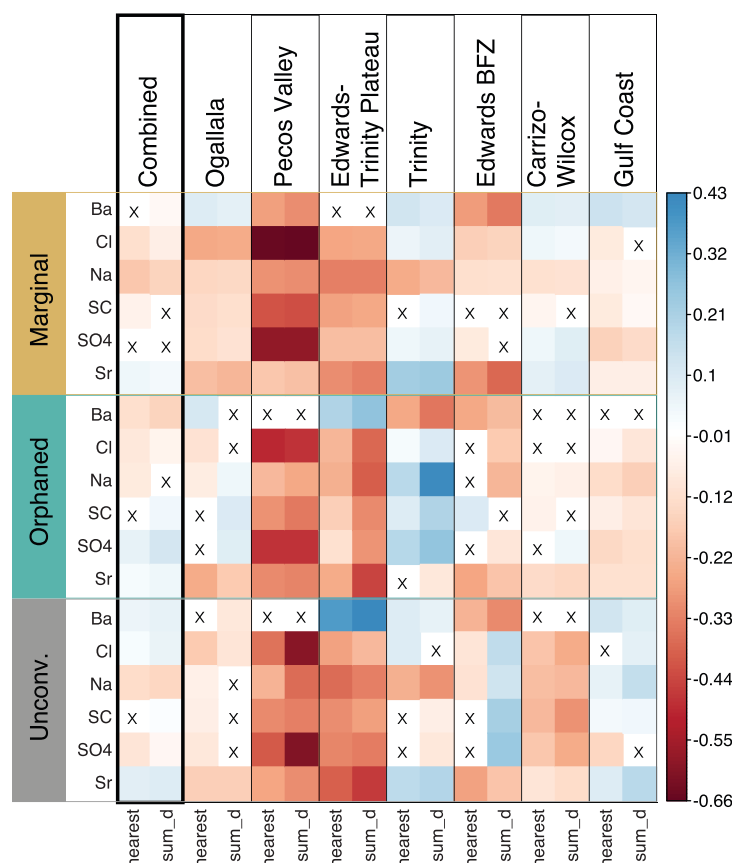


Figure 4. Subregional correlation results in Texas reveal varying contamination patterns across seven major aquifers. While state level (combined) trends were weak ($|R| < 0.2$), Pecos Valley and Edwards-Trinity Plateau aquifers results suggest major water quality degradation near marginal, orphaned, and unconventional wells.

may reflect low sample density, broad confidence intervals, background geochemical gradients, dilution or mixing effects, or geochemical reactions that limit the mobility of specific analytes. For example, Ba may have lower dissolved mobility in sulfate-rich waters because of barite precipitation, which could weaken or reverse simple spatial correlations between Ba and oil and gas proximity. We therefore emphasize consistent negative correlations across multiple analytes and proxies when interpreting potential oil-and-gas-related signals.

3.3. Hydrogeology and production practice drive distinct hydrogeochemical patterns

Hierarchical clustering of correlation coefficients revealed contrasting results between surface and groundwater systems (figure S3). As New York State does not have unconventional wells, it was excluded from this analysis. For groundwater, correlation results grouped primarily by analyte type, forming three clusters: (1) Na, Cl, and SC, (2) SO₄, and (3) Ba and Sr, reflecting cumulative longer-term effects from oil and gas wells as well as baseline geochemistry. In contrast, surface water results clustered by well type, suggesting that production characteristics and operational differences exert stronger controls on surface water systems. In particular, marginal and

orphaned wells formed a cohesive cluster, indicating similar geochemical impacts despite different production practices. Several points (i.e. Ba, SO₄, Sr) did not follow this pattern, likely due to sparse sampling (table S3). Clustering results were consistent when states and different proxy types were analyzed separately. Overall, clustering patterns imply that surface water quality reflects short-term and production-driven effects, while groundwater signals integrate longer-term geochemical and hydrologic processes.

3.4. Subregional trends reveal hidden signals

We conducted a subregional analysis focusing on Texas, which hosts the largest and most diverse population of oil and gas wells in the U.S. Unlike the other three states, Texas encompasses a wide range of hydrogeologic settings—from arid basins to carbonate aquifers [72]—and contains both long-established conventional fields and extensive unconventional developments. The state also provides a spatially comprehensive water-quality record (table S1), making it ideal for testing whether large-scale averaging obscures localized contamination patterns.

Although Texas hosts the largest number of oil and gas wells in the U.S., statewide correlations between oil-gas production activity and groundwater chemistry were surprisingly weak ($|R| < 0.2$). To

test whether statewide averaging obscured localized impacts, we reanalyzed the data by aquifer (figure 4). This finer-scale approach revealed pronounced contamination in two major aquifers: Pecos Valley and Edwards-Trinity Plateau. In particular, Pecos Valley correlation ranges were the highest among the seven aquifers studied (-0.66 – -0.18). Concentration of two analytes—Cl and SO₄—were particularly strongly correlated with proximity to oil and gas wells in all three well categories averaging -0.54 and -0.52 , respectively. Meanwhile, in Edwards-Trinity Plateau correlations were comparable among five analytes (Cl, Na, SC, SO₄, Sr) as well as all three well types. These relationships were undetectable in statewide analyses because weak or absent signals in other aquifers diluted the overall trend. In contrast, aquifers such as Carrizo-Wilcox showed weaker but spatially coherent negative correlations (-0.3 – -0.1), particularly around clusters of unconventional wells ($\text{mean}_{\text{unconv}} = -0.2$).

These results demonstrate that regional hydrogeology governs the strength and detectability of oil and gas impacts on groundwater quality. Importantly, expanding data coverage alone does not guarantee improved inference—rather, analytical resolution must align with hydrologic boundaries to reveal meaningful relationships.

4. Discussion

4.1. Consistent correlations indicate measurable impacts

Our analysis reveals statistically robust and spatially coherent relationship between marginal, unconventional, as well as orphaned wells and elevated concentrations of key analytes—Ba, Cl, Na, SC, SO₄, and Sr—in both surface and groundwater. While correlations vary in strength across well types, water types, analytes, as well as states, the consistent negative trends near wells suggest measurable contamination or salinization processes linked to oil and gas infrastructure. Even though the relationships are correlative rather than strictly causal, the high fraction of statistically significant results indicates a pervasive influence that cannot be attributed solely to random spatial coincidence. Our findings are broadly consistent with prior site-specific studies that documented increased salt and metal concentrations downstream or down-gradient of oil and gas activities in Pennsylvania, Texas, New York, and other producing regions [31, 34–41, 73]. Numerous studies found evidence of oil and gas impacts on surface and groundwater resources in Pennsylvania [34, 35, 38–41, 73, 74]. Warner *et al* [75] found increased Cl, Ba and Sr concentration downstream of a brine treatment facility. Similarly, Olmstead *et al* [31] found increased Cl concentration using large scale geospatial analysis. Contamination evidence was also detected in

the Barnett Shale region in Texas relating increased strontium concentrations to unconventional oil and gas activity [36]. Together, these results indicate that oil and gas wells, including marginal and orphaned sites often overlooked in monitoring programs, are associated with widespread and detectable water-quality degradation signals. Because spatial correlation does not prove causal attribution, we cannot fully exclude alternative ion sources such as road salts, agriculture, and mine drainage. The analytes considered here also differ in how directly they reflect produced-water inputs. Chloride, sodium, strontium, and specific conductivity are commonly elevated in produced waters and are relatively conservative during transport, making them more likely to produce coherent spatial signals. In contrast, barium and sulfate can be strongly affected by local redox conditions, mixing, mineral saturation, and barite precipitation, which may explain why Ba and SO₄ show less consistent behavior across regions. Thus, discrepancies among analytes do not necessarily indicate an absence of oil-and-gas influence but instead may reflect differences in produced-water chemistry and post-release geochemical processing. These geochemical differences may also explain why Ba and SO₄ do not always cluster with the more conservative indicators in the correlation analysis (figure S3).

Across the full analysis, the most interpretable signals were those that were consistently negative across multiple analytes, well classes, or spatial proxies, rather than isolated coefficients. We therefore place greatest emphasis on relationships that met three criteria: (I) negative correlations, indicating higher concentrations closer to wells or in areas of higher well density; (ii) consistency across multiple proxies or chemically related analytes; and (iii) persistence, where possible, in the proximity-restricted analysis. Under these criteria, the clearest signals include all six analytes in Colorado surface water, especially near marginal and orphaned wells, as well as Cl and Na in Colorado groundwater; SC and SO₄ in groundwater near marginal and orphaned wells in Pennsylvania; Ba, SC, Sr in groundwater, as well as Cl, Na, SC and SO₄ in surface water, in New York; Cl, Na, SC, SO₄ and Sr in groundwater near all three well types analyzed in Pecos Valley and Edwards-Trinity Plateau in Texas with a notable Ba signal only near marginal wells in Pecos Valley; Ba and Sr in Edwards (Balcones Fault Zone) aquifer; Cl, Na, SC, SO₄ in Carrizo-Wilcox aquifer uniquely near unconventional wells (figures 2(a) and 4). In addition, despite limitations associated with proximity restricted analysis (figure 2(b)) several water-quality signals identified in figure 2(a) were also observed in figure 2(b): Cl, Na, SC, SO₄, Sr in surface water and Cl, Na in groundwater near unconventional wells in Colorado; SC and SO₄ in groundwater near marginal and orphaned wells in Pennsylvania; and Ba in groundwater near marginal wells as well

as in surface water near orphaned wells in New York. In contrast, isolated positive correlations or single-analyte responses are interpreted cautiously because they may reflect background geochemical gradients, dilution, sparse sampling, or analyte-specific mobility rather than oil-and-gas-related degradation.

4.2. Hydrogeology governs spatial variability of contamination

Spatial variability in contamination patterns reflects the combined influence of hydrologic connectivity, geologic features, and sampling scale. For groundwater, differences among aquifers largely explain the inconsistency of correlation strengths: shallow, rapidly recharging systems are more likely to dilute contamination signals, whereas confined or low-permeability formations retain dissolved ions over longer timescales. Such controls have been demonstrated in previous regional studies of the Appalachian Basin and other sedimentary systems [8, 34, 37, 39, 45, 76–78].

The contrasting clustering behavior between surface and groundwater further highlights these mechanisms. Surface waters respond to short-term events—spills, impoundment leaks, and runoff from well pads—while groundwater integrates long-term chemical exchange between brines, rocks, and aquifers. Olmstead *et al* [31] found evidence of statistically significant correlations between Cl concentrations in streams and 4 day cumulative precipitation diluting observed concentration. These findings combined with our results highlight the importance of collecting samples near oil-gas development to avoid impacts of dilution and natural attenuation that occurs due to surface- and subsurface-processes.

Consequently, the detectability of oil and gas impacts depends not only on well proximity but also on the hydrogeologic context that governs fluid flow and solute transport. Texas provides a clear example: state-level analyses masked strong sub-regional contamination signatures within specific aquifers because they differ markedly in lithology and hydraulic connectivity.

4.3. Data and methodological limitations constrain signal detection

Although our analysis is the most comprehensive geospatial assessment of its kind, several limitations should be acknowledged. First, uneven and/or scarce sampling—both spatially and temporally—reduces statistical power [34, 43, 61, 79, 80], especially for trace metals like Ba and Sr that are infrequently measured. Second, water-quality datasets compiled across multiple agencies often differ in analytical precision, detection limits, and metadata completeness, introducing additional noise into correlation analyses (refer to Materials and Methods). Third, correlation-based methods cannot always distinguish among contamination sources. Mine drainage, road

salts, discharge from wastewater treatment plants, agricultural runoff, as well as industrial discharge can all contribute to elevated ion concentrations [81, 82]. Road salting and other non-oil-and-gas sources could confound relationships between water chemistry and oil and gas wells, particularly for Na, Cl, and specific conductivity. The Na/Cl molar-ratio analysis helps evaluate this possibility and may explain why correlations in PA were relatively weak despite substantial oil and gas activity (figures 2 and S5). Ratios below 1 are more consistent with chloride-rich brines or produced-water influence than with halite-dominated road salt alone, whereas elevated Na/Cl ratios farther from wells may reflect rock weathering or other non-oil-and-gas sources. In addition, if an inflow from contamination source is low and/or contains analyte concentrations close to background levels found in the shallow aquifer, the impact will be untraceable [83]. Lastly, orphaned-well inventories are incomplete and continue to evolve. Our analysis uses documented orphaned wells from the compiled geospatial dataset available for this study, but undocumented or newly identified orphaned wells are not captured and could lead to underestimation or spatial misclassification of orphaned-well associations. We did not conduct comparable subregional hydrogeologic analyses for all four states because (i) the uneven spatial distribution of water quality samples left too few observations within individual hydrogeologic units to produce reliable correlation estimates, precluding detection of subregional trends; and (ii) additional regional-specific interpretation in the other three states is beyond the scope of this Letter.

Because the analytes considered here are primarily salts and metals rather than dissolved gases, the correlations we observe are more plausibly associated with produced-water-related pathways rather than methane migration alone. Potential mechanisms include historical unlined pits and impoundments, produced-water storage and handling, spills or leaks from tanks and pipelines, blowouts, surface runoff from well pads, and, in some settings, subsurface leakage along compromised wellbores. Our spatial proxies cannot distinguish among these pathways, nor can they isolate the role of casing or cementing failures without direct well-integrity data. We therefore interpret the observed correlations as evidence of spatially coherent water-quality associations with oil and gas infrastructure, rather than as attribution to a single leakage mechanism. Previous works have estimated a wide range for the percentage of wells that experience integrity issues, which varies by region, some being as high as 27% [23, 84]. This suggests that detectable impacts likely represent only a fraction of total compromised sites.

Finally, density-based proxies can produce right-skewed distributions, especially in low-well-density regions, which may lead to false associations if not corrected, like several coefficients that were identified

in figure 2 and later removed. Hence, it is crucial to use multiple proxies while analyzing this type of data as comparing trends from different methods is most likely to lead to meaningful and realistic conclusions. These constraints also highlight the importance of coupling large datasets with localized process studies to validate inferred contamination pathways. A further limitation is that surface-water samples were analyzed using the same radial proximity and density metrics as groundwater samples. This choice allowed direct comparison between surface-water and groundwater results using a uniform spatial framework, but it does not explicitly account for upstream drainage area or flow direction. Future surface-water analyses could improve hydrologic attribution by incorporating watershed- or HUC-based oil and gas metrics. Another limitation is that proximity and density metrics were calculated within state boundaries. Samples near state borders may be influenced by wells in adjacent states, particularly in regions such as Pennsylvania where nearby oil and gas development also occurs in Ohio and West Virginia. Future region-specific analyses that are not constrained by state boundaries could better evaluate cross-border oil-and-gas influences.

4.4. Integrating geologic context can improve predictive models and monitoring

Our findings underscore that more data alone do not ensure better detection of contamination. Rather, data resolution must align with the relevant hydrogeologic and spatial scales. Incorporating aquifer lithology, hydrologic properties such as permeability, and flow direction into statistical and machine learning modeling can substantially improve predictive power and help distinguish anthropogenic signals from natural variability. Our subregional analysis (figure 4) highlighted the Pecos and Edwards-Trinity Plateau as the Texas aquifers with the strongest negative correlation. These patterns are consistent with the arid to semi-arid climate of the Pecos River region, limited dilution, high evaporation, and interaction with salt-rich stratigraphic units that can elevate dissolved salts and sulfate [72]. Longer groundwater residence times or more limited hydrologic connectivity in these aquifers may allow produced-water-derived solutes to persist, whereas more hydraulically connected systems may dilute or disperse such signals more rapidly. Elevated sulfate may also reduce dissolved Ba mobility through barite precipitation, helping explain why Ba behaves differently from Cl, Na, SC, and Sr in some aquifers.

Future contamination assessment efforts should integrate data-driven geospatial analysis with physics-based models and machine learning approaches that explicitly account for subsurface heterogeneity and connectivity [34, 68, 69, 85, 86]. In particular, surface and subsurface flow paths can

act as conduits from well sites to distant aquatic ecosystems [29, 50, 87, 88]. In addition, preferential flow paths could amplify the extent of a groundwater contamination [8, 89] which would complicate its detection, while unconsolidated aquifers would have a more uniform contaminant dispersion [70, 90]. Process-based system analysis would complement statistical studies like this one and inform specific hydrogeologic conditions that should not be ignored in data-driven analysis while improving predictive models aimed at minimizing emissions. Such models can be trained on regional datasets like ours to identify high-risk zones for leakage or contamination and guide cost-effective monitoring of unconventional, marginal, and orphaned wells. Ultimately, combining statistical and mechanistic perspectives will enhance our ability to forecast contamination, minimize associated risks, prioritize remediation, and inform regulatory strategies for sustainable energy production.

5. Conclusion

This study demonstrates that oil and gas wells—including marginal, unconventional, and orphaned—are associated with a measurable geochemical imprint on both surface and groundwater systems. Using approximately three million water-quality records from New York, Pennsylvania, Colorado, and Texas, the analysis reveals widespread statistically significant associations between well proximity/density and elevated concentrations of analytes such as chloride, sodium, sulfate, strontium, barium, and specific conductivity. We find statistically significant and spatially coherent negative associations across all four states, with 54.5% of surface-water analyses and 53.0% of groundwater analyses showing significant negative correlations most directly consistent with elevated concentrations near wells. Subregional analysis in Texas shows that statewide assessments can mask strong aquifer-specific contamination patterns governed by geology and hydrologic connectivity.

Together, these findings highlight that legacy as well as active wells all contribute to detectable degradation of water quality, underscoring the need for monitoring frameworks that explicitly account for local geology and flow pathways. The results also emphasize that more data alone do not necessarily improve contamination detection; rather, sampling resolution must align with hydrogeologic boundaries to reveal meaningful trends. By combining geologic context with large-scale statistical analysis, future monitoring and predictive models can more effectively identify high-risk zones, guide cost-efficient mitigation, and support regulatory efforts to reduce long-term contamination from active and abandoned oil and gas wells.

Acknowledgment

We thank the reviewers whose comments helped to improve this manuscript.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.18379747> [91]. The project GitHub repository includes data-processing scripts, intermediate screening outputs, and analysis code needed to reproduce the data-cleaning workflow and statistical results. Supplementary Materials include a PDF file for supplementary figures, tables and text.

Supporting Information available at <https://doi.org/10.1088/1748-9326/ae7c17/data1>.

Funding

This work was funded by the National Science Foundation grant OAC-2209864 (T.W.).

Author contributions

Ruta Basijokaite  0000-0003-2645-4505
Conceptualization (lead), Data curation (lead),
Formal analysis (lead), Investigation (lead),
Methodology (lead), Validation (lead),
Visualization (lead), Writing – original draft (lead)

Mary Kang  0000-0001-9142-384X
Data curation (supporting), Writing – review &
editing (equal)

Tao Wen  0000-0002-6113-7532
Conceptualization (supporting), Data
curation (supporting), Funding acquisition (lead),
Methodology (supporting), Project
administration (lead), Resources (lead),
Supervision (lead), Writing – original
draft (supporting)

References

- [1] U.S. Energy Information Administration 2025 *Annual Energy Outlook*
- [2] Colborn T, Kwiatkowski C, Schultz K and Bachran M 2011 Natural gas operations from a public health perspective *Hum. Ecol. Risk Assess.* **17** 1039–56
- [3] Vengosh A, Jackson R B, Warner N, Darrah T H and 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** 8334–48
- [4] U.S. Energy Information Administration 2025 *US Oil and Gas Wells by Production Rate*
- [5] US EPA O 2025 EPA launches biggest deregulatory action in U.S. history
- [6] Nordbotten J M, Kavetski D, Celia M A and Bachu S 2009 Model for CO₂ leakage including multiple geological layers and multiple leaky wells *Environ. Sci. Technol.* **43** 743–9
- [7] Hammond P A, Wen T, Brantley S L and Engelder T 2020 Gas well integrity and methane migration: evaluation of published evidence during shale-gas development in the USA *Hydrogeol. J.* **28** 1481–502
- [8] Hammond P A, Wen T, Woda J, Oakley D and Qingchao L 2024 Pathways and environmental impacts of methane migration: case studies in the Marcellus Shale, USA *Geofluids* **2024** e9290873
- [9] Brownlow J W, Yelderman J C Jr and James S C 2017 Spatial risk analysis of hydraulic fracturing near abandoned and converted oil and gas wells *Groundwater* **55** 268–80
- [10] Dusseault M and Jackson R 2014 Seepage pathway assessment for natural gas to shallow groundwater during well stimulation, in production, and after abandonment *Environ. Geosci.* **21** 107–26
- [11] Flewelling S A, Tymchak M P and Warpinski N 2013 Hydraulic fracture height limits and fault interactions in tight oil and gas formations *Geophys. Res. Lett.* **40** 3602–6
- [12] Wen T, Castro M C, Nicot J-P, Hall C M, Larson T, Mickler P and Darvari R 2016 Methane sources and migration mechanisms in shallow groundwaters in parker and Hood Counties, Texas—a heavy noble gas analysis *Environ. Sci. Technol.* **50** 12012–21
- [13] Barclay I, Pellenbarg J, Tettero F, Pfeiffer J, Slater H, Staal T, Stiles D, Tilling G and Whitney C 2001 The beginning of the end: a review of abandonment and decommissioning practices *Oilfield Rev.* **13** 28–41
- [14] Allen D T et al 2013 Measurements of methane emissions at natural gas production sites in the United States *Proc. Natl Acad. Sci. USA* **110** 17768–73
- [15] Jackson R B 2014 The integrity of oil and gas wells *Proc. Natl Acad. Sci. USA* **111** 10902–3
- [16] Ingraffea A R, Wells M T, Santoro R L and Shonkoff S B C 2014 Assessment and risk analysis of casing and cement impairment in oil and gas wells in *Proc. Natl Acad. Sci.* **111** 10955–60
- [17] Lackey G, Rajaram H, Sherwood O A, Burke T L and Ryan J N 2017 Surface casing pressure as an indicator of well integrity loss and stray gas migration in the Wattenberg Field, Colorado *Environ. Sci. Technol.* **51** 3567–74
- [18] Robertson A M, Edie R, Snare D, Soltis J, Field R A, Burkhart M D, Bell C S, Zimmerle D and Murphy S M 2017 Variation in methane emission rates from well pads in four oil and gas basins with contrasting production volumes and compositions *Environ. Sci. Technol.* **51** 8832–40
- [19] Kang M, Mauzerall D L, Ma D Z and Celia M A 2019 Reducing methane emissions from abandoned oil and gas wells: strategies and costs *Energy Policy* **132** 594–601
- [20] Riddick S N, Mauzerall D L, Celia M A, Kang M, Bressler K, Chu C and Gum C D 2019 Measuring methane emissions from abandoned and active oil and gas wells in West Virginia *Sci. Total Environ.* **651** 1849–56
- [21] Jackson R E, Dusseault M B, Frape S, Phan T and Steelman C 2020 Investigating the origin of elevated H₂S in groundwater discharge from abandoned gas wells, Norfolk County, Ontario *Geoconvention 2020*
- [22] Robertson A M, Edie R, Field R A, Lyon D, McVay R, Omara M, Zavala-Araiza D and Murphy S M 2020 New Mexico permian basin measured well pad methane emissions are a factor of 5–9 times higher than U.S. EPA estimates *Environ. Sci. Technol.* **54** 13926–34
- [23] Wisen J, Chesnaux R, Werring J, Wendling G, Baudron P and Barbecot F 2020 A portrait of wellbore leakage in northeastern British Columbia, Canada *Proc. Natl Acad. Sci. USA* **117** 913–22
- [24] Sandl E, Cahill A G, Welch L and Beckie R 2021 Characterizing oil and gas wells with fugitive gas migration

- through Bayesian multilevel logistic regression *Sci. Total Environ.* **769** 144678
- [25] Williams J P, Regehr A and Kang M 2021 Methane emissions from abandoned oil and gas wells in Canada and the United States *Environ. Sci. Technol.* **55** 563–70
- [26] El Hachem K and Kang M 2022 Methane and hydrogen sulfide emissions from abandoned, active, and marginally producing oil and gas wells in Ontario, Canada *Sci. Total Environ.* **823** 153491
- [27] Omara M, Zavala-Araiza D, Lyon D R, Hmiel B, Roberts K A and Hamburg S P 2022 Methane emissions from US low production oil and natural gas well sites *Nat. Commun.* **13** 2085
- [28] El Hachem K and Kang M 2023 Reducing oil and gas well leakage: a review of leakage drivers, methane detection and repair options *Environ. Res.* **3** 012002
- [29] Lackey G, Dyer A, Pfander I, Shih C Y and Dilmore R M 2025 Understanding drivers of oil and gas well integrity issues in the greater Wattenberg area of Colorado *Int. J. Greenhouse Gas Control* **141** 104296
- [30] Williams J P, Omara M, Himmelberger A, Zavala-Araiza D, MacKay K, Benmergui J, Sargent M, Wofsy S C, Hamburg S P and Gautam R 2025 Small emission sources in aggregate disproportionately account for a large majority of total methane emissions from the US oil and gas sector *Atmos. Chem. Phys.* **25** 1513–32
- [31] Olmstead S M, Muehlenbachs L A, Shih J-S, Chu Z and Krupnick A J 2013 Shale gas development impacts on surface water quality in Pennsylvania *Proc. Natl Acad. Sci.* **110** 4962–7
- [32] Akob D M, Mumford A C, Orem W, Engle M A, Kluges J G, Kent D B and Cozzarelli I M 2016 Wastewater disposal from unconventional oil and gas development degrades stream quality at a West Virginia injection facility *Environ. Sci. Technol.* **50** 5517–25
- [33] Bonetti P, Leuz C and Michelon G 2021 Large-sample evidence on the impact of unconventional oil and gas development on surface waters *Science* **373** 896–902
- [34] Wen T, Niu X, Gonzales M, Zheng G, Li Z and Brantley S L 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** 7149–59
- [35] Li Z, You C, Gonzales M, Wendt A K, Wu F and Brantley S L 2016 Searching for anomalous methane in shallow groundwater near shale gas wells *J. Contaminant Hydrol.* **195** 23–30
- [36] Fontenot B E et al 2013 An evaluation of water quality in private drinking water wells near natural gas extraction sites in the Barnett shale formation *Environ. Sci. Technol.* **47** 10032–40
- [37] Kreuzer R L et al 2018 Structural and hydrogeological controls on hydrocarbon and brine migration into drinking water aquifers in Southern New York *Groundwater* **56** 225–44
- [38] Osborn S G, Vengosh A, Warner N R and Jackson R B 2011 Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing *Proc. Natl Acad. Sci. USA* **108** 8172–6
- [39] Darrah T H, Vengosh A, Jackson R B, Warner N R and Poreda R J 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** 14076–81
- [40] Llewellyn G T, Dorman F, Westland J L, Yoxtheimer D, Grieve P, Sowers T, Humston-fulmer E and Brantley S L 2015 Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development *Proc. Natl Acad. Sci.* **112** 6325–30
- [41] Jackson R B, Vengosh A, Darrah T H, Warner N R, Down A, Poreda R J, Osborn S G, Zhao K and Karr J D 2013 Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction *Proc. Natl Acad. Sci.* **110** 11250–5
- [42] Basijokaite R and Kelleher C 2021 Time-varying sensitivity analysis reveals relationships between watershed climate and variations in annual parameter importance in regions with strong interannual variability *Water Resour. Res.* **57** e2020WR028544
- [43] Vidic R D, Brantley S L, Vandenbossche J M, Yoxtheimer D and Abad J D 2013 Impact of shale gas development on regional water quality *Science* **340** 1235009
- [44] McMahon P B, Barlow J R B, Engle M A, Belitz K, Ging P B, Hunt A G, Jurgens B C, Kharaka Y K, Tollett R W and Kresse T M 2017 Methane and benzene in drinking-water wells overlying the eagle ford, fayetteville, and haynesville shale hydrocarbon production areas *Environ. Sci. Technol.* **51** 6727–34
- [45] McMahon P B, Landon M K, Davis T A, Wright M T, Rosecrans C Z, Anders R, Land M T, Kulongoski J T and Hunt A G 2021 Relative risk of groundwater-quality degradation near California (USA) oil fields estimated from 3H, 14C, and 4He *Appl. Geochem.* **131** 105024
- [46] Siegel D I, Azzolina N A, Smith B J, Perry A E and Bothun R L 2015 Methane concentrations in water wells unrelated to proximity to existing oil and gas wells in Northeastern Pennsylvania *Environ. Sci. Technol.* **49** 4106–12
- [47] Brandt A R, Heath G A and Cooley D 2016 Methane leaks from natural gas systems follow extreme distributions *Environ. Sci. Technol.* **50** 12512–20
- [48] Duren R M et al 2019 California's methane super-emitters *Nature* **575** 180–4
- [49] Soriano M A, Siegel H G, Gutches K M, Clark C J, Li Y, Xiong B, Plata D L, Deziel N C and Saiers J E 2020 Evaluating domestic well vulnerability to contamination from unconventional oil and gas development sites *Water Resour. Res.* **56** e2020WR028005
- [50] Soriano M A, Siegel H G, Johnson N P, Gutches K M, Xiong B, Li Y, Clark C J, Plata D L, Deziel N C and Saiers J E 2021 Assessment of groundwater well vulnerability to contamination through physics-informed machine learning *Environ. Res. Lett.* **16** 084013
- [51] Cusworth D H et al 2022 Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States *Proc. Natl Acad. Sci. USA* **119** e2202338119
- [52] Sherwin E D et al 2024 US oil and gas system emissions from nearly one million aerial site measurements *Nature* **627** 328–34
- [53] National Energy Technology Laboratory 2024 Methane measurement guidelines for marginal conventional wells
- [54] Interstate Oil and Gas Compact Commission 2016 Marginal wells: fuel for economic growth 2016 report
- [55] Environmental Defense Fund 2021 By the numbers: marginal oil and gas wells
- [56] Deighton J A, Townsend-small A, Sturmer S J, Hoschouer J and Heldman L 2020 Measurements show that marginal wells are a disproportionate source of methane relative to production *J. Air Waste Manage. Assoc.* **70** 1030–42
- [57] ShaleXP All oil & gas producing states in the United States (available at: <https://www.shalexp.com/states>) (Accessed 15 November 2025)
- [58] U.S. Energy Information Administration Crude oil production (available at: www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbbl_a.htm) (Accessed 15 November 2025)
- [59] Barbot E, Vidic N S, Gregory K B and Vidic R D 2013 Spatial and temporal correlation of water quality parameters of produced waters from Devonian-age shale following hydraulic fracturing *Environ. Sci. Technol.* **47** 2562–9
- [60] Rosenblum J, Nelson A W, Ruyle B, Schultz M K, Ryan J N and Linden K G 2017 Temporal characterization of flowback and produced water quality from a hydraulically fractured oil and gas well *Sci. Total Environ.* **596–597** 369–77
- [61] Agarwal A, Wen T, Chen A, Zhang A Y, Niu X, Zhan X, Xue L and Brantley S L 2020 Assessing contamination of stream

- networks near shale gas development using a new geospatial tool *Environ. Sci. Technol.* **54** 8632–9
- [62] Boutot J, Peltz A S, McVay R and Kang M 2022 Documented orphaned oil and gas wells across the United States *Environ. Sci. Technol.* **56** 14228–36
- [63] Anon Oil and gas inventory—report extracts (available at: <https://greenport.pa.gov/ReportExtracts/OG/OilGasWellInventoryReport>) (Accessed 14 November 20)
- [64] Colorado Energy & Carbon Management Commission COGIS—sample site inquiry (available at: <https://ecmc.colorado.gov/data-maps-reports/cogis-database/cogis-sample-site-inquiry>) (Accessed 12 November 2025)
- [65] Read E K, Carr L, De Cicco L, Dugan H A, Hanson P C, Hart J A, Kreft J, Read J S and Winslow L A 2017 Water quality data for national-scale aquatic research: the Water Quality Portal *Water Resour. Res.* **53** 1735–45
- [66] Brantley S 2015 Shale network data (available at: https://hiscentral.cuahsi.org/pub_network.aspx?n=228) (Accessed 1 November 2015)
- [67] Texas Water Development Board GWDB reports and downloads (available at: www.twdb.texas.gov/groundwater/data/gwdbbrpt.asp) (Accessed 12 November 2025)
- [68] Clark C J et al 2022 Assessing unconventional oil and gas exposure in the appalachian basin: comparison of exposure surrogates and residential drinking water measurements *Environ. Sci. Technol.* **56** 1091–103
- [69] Rogers J D, Burke T L, Osborn S G and Ryan J N 2015 A framework for identifying organic compounds of concern in hydraulic fracturing fluids based on their mobility and persistence in groundwater *Environ. Sci. Technol. Lett.* **2** 158–64
- [70] Taherdangkoo R, Tatomir A and Sauter M 2020 Modeling of methane migration from gas wellbores into shallow groundwater at basin scale *Environ. Earth Sci.* **79** 432
- [71] Texas Water Development Board Major aquifers (available at: www.twdb.texas.gov/groundwater/aquifer/major.asp) (Accessed 14 November 2025)
- [72] Balley N Texas aquifers study (available at: www.twdb.texas.gov/groundwater/docs/studies/TexasAquifersStudy_2016.pdf) (Accessed 1 November 2015)
- [73] Li Z, You C, Gonzales M, Wendt A K, Wu F and Brantley S L 2017 Corrigendum to “Searching for anomalous methane in shallow groundwater near shale gas wells” (*J. Contam. Hydrol.* (2016) 195 (23–30) (S0169772216300985) (<https://doi.org/10.1016/j.jconhyd.2016.10.005>) *J. Contaminant Hydrol.* **207** 50–51
- [74] Shaheen S W, Wen T, Herman A and Brantley S L 2022 Geochemical evidence of potential groundwater contamination with human health risks where hydraulic fracturing overlaps with extensive legacy hydrocarbon extraction *Environ. Sci. Technol.* **56** 10010–9
- [75] Warner N R, Darrah T H, Jackson R B, Millot R, Kloppmann W and Vengosh A 2014 New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations *Environ. Sci. Technol.* **48** 12552–60
- [76] Molofsky L J, Connor J A, McHugh T E, Richardson S D, Woroszylo C and Alvarez P J 2016 Environmental factors associated with natural methane occurrence in the appalachian basin *Groundwater* **54** 656–68
- [77] McMahon P B, Galloway J M, Hunt A G, Belitz K, Jurgens B C and Johnson T D 2021 Geochemistry and age of groundwater in the Williston Basin, USA: assessing potential effects of shale-oil production on groundwater quality *Appl. Geochem.* **125** 104833
- [78] Woda J, Wen T, Oakley D, Yoxtheimer D, Engelder T, Castro M C and Brantley S L 2018 Detecting and explaining why aquifers occasionally become degraded near hydraulically fractured shale gas wells *Proc. Natl Acad. Sci.* **115** 12349–58
- [79] Brantley S L, Vidic R D, Brasier K, Yoxtheimer D, Pollak J, Wilderman C and Wen T 2018 Engaging over data on fracking and water quality *Science* **359** 395–7
- [80] Brantley S L, Yoxtheimer D, Arjmand S, Grieve P, Vidic R, Pollak J, Llewellyn G T, Abad J and Simon C 2014 Water resource impacts during unconventional shale gas development: the Pennsylvania experience *Int. J. Coal Geol.* **126** 140–56
- [81] Chapman E C, Capo R C, Stewart B W, Kirby C S, Hammack R W, Schroeder K T and Edenborn H M 2012 Geochemical and strontium isotope characterization of produced waters from Marcellus shale natural gas extraction *Environ. Sci. Technol.* **46** 3545–53
- [82] Lautz L K, Hoke G D, Lu Z, Siegel D I, Christian K, Kessler J D and Teale N G 2014 Using discriminant analysis to determine sources of salinity in shallow groundwater prior to hydraulic fracturing *Environ. Sci. Technol.* **48** 9061–9
- [83] Dresel P E and Rose A W 2010 Chemistry and origin of oil and gas well brines in western Pennsylvania
- [84] Lackey G, Rajaram H, Bolander J, Sherwood O A, Ryan J N, Shih C Y, Bromhal G S and Dilmore R M 2021 Public data from three US states provide new insights into well integrity *Proc. Natl Acad. Sci. USA* **118** e2013894118
- [85] Soriano M A, Deziel N C and Saiers J E 2022 Regional scale assessment of shallow groundwater vulnerability to contamination from unconventional hydrocarbon extraction *Environ. Sci. Technol.* **56** 12126–36
- [86] Mallants D, Doble R and 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
- [87] Rice A K, Lackey G, Proctor J and Singha K 2018 Groundwater-quality hazards of methane leakage from hydrocarbon wells: a review of observational and numerical studies and four testable hypotheses *Wiley Interdiscip. Rev.* **5** e1283
- [88] Rice A K, McCray J E and Singha K 2021 Numerical investigation of wellbore methane leakage from a dual-porosity reservoir and subsequent transport in groundwater *Water Resour. Res.* **57** e2019WR026991
- [89] Hartmann A et al 2021 Risk of groundwater contamination widely underestimated because of fast flow into aquifers *Proc. Natl Acad. Sci.* **118** e2024492118
- [90] Gelhar L W, Welty C and Rehfeldt K R 1992 A critical review of data on field-scale dispersion in aquifers *Water Resour. Res.* **28** 1955–74
- [91] Basijokaite R 2026 Data and code for ERL (Version v2.0) *Water Quality Near Wells* (<https://doi.org/10.5281/zenodo.1837974>)