Supporting Information for

Geochemical evidence of potential groundwater contamination with human health risks where hydraulic fracturing overlaps with extensive legacy hydrocarbon extraction

Samuel W. Shaheen[a], Tao Wen[b], Alison Herman[c], Susan L. Brantley[a,c]

[a] Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA

[b] Department of Earth and Environmental Sciences, Syracuse University, Syracuse, NY, 13244, USA

[c] Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA

Table of Contents (38 pages, 12 Supporting Texts, 15 Figures, 4 Tables)

Supporting Text

Supporting Materials and Methods:

Text S1. Extended study area background

As of 10/2020, 3,610 unconventional oil and gas (UOG) wells (wells emplaced with both hydraulic fracturing and horizontal drilling of shale), including the first such well in the state, were drilled in Washington, Beaver, and Greene counties, Pennsylvania (Figure S1A).[1,2] This tri-county area considered in the study is henceforth referred to as SWPA. Over 8,250 conventional oil/gas (COG) wells have also been drilled in the area (Figure S1A), alongside possibly tens of thousands of undocumented abandoned COG wells.[3] Underlain by regional coal beds at varying depths, the region of SWPA and Washington County specifically are now approximately 35.1% and >53% mined, respectively (Figure S1B).[4,5] Reflecting this overlap, 89% of UOG wells in SWPA are located in "coal areas", defined as areas of workable coal seam(s). These wells are thus likely to be drilled through coal beds. At least 950 UOG wells in SWPA are drilled directly through pillars within coal mines.[2]

In the SWPA region, like much of the Appalachian Basin, Ca-HCO_3 or Ca-Mg-HCO_3 are typically the dominant ions at shallow depths and in upland regions.[6] Groundwaters evolve to Na-HCO_3 or Na-Cl type water with increasing depth or residence time due to ion exchange and/or mixing with Appalachian Basin brine.[7] Groundwater flow paths in SWPA are typically intermediate to local flow systems with saline groundwater present at depths of tens of meters below land surface.[6] Furthermore, while very few faults outcrop at the surface in the study area, anticlinal and synclinal folding is found throughout Washington and Greene counties (Figure S1C).

Text S2. Additional databases used

Additional spatial data used in our study were obtained by downloading (1) unconventional oil and gas well, conventional oil and gas well, and coal mining area locations from the PA DEP Open Data Portal, (2) bedrock lithology, fault, syncline, and anticline locational data from the United States Geological Survey PA Geologic Map database and (3) primary and secondary highway locations from the U.S. Census Bureau TIGER/Line Shapefiles database.[2,8,9]

Text S3. Additional spatial analysis details

The great-circle (Haversine) distances between groundwater samples and the nearest coal mining areas, highways, faults, anticlines, and synclines were calculated in ArcGIS Pro 2.6.2 using the "Near" feature. The distances between groundwater samples and oil and gas wells and the density of oil and gas wells within 1km were calculated using R 3.6.1. For distance calculations, we calculated the distance to the closest oil and gas well, and the density of UOG wells was determined by calculating the number of oil and gas wells drilled within 1km of a given water sample. Our distance and density analyses considered only oil and gas wells that were drilled prior to groundwater sample collection, and excluded wells with a status listed as "Operator Reported Not Drilled" or "Proposed But Never Materialized". Some wells in the PA DEP database do not report a spud date, and we excluded such wells from our analyses because we could not ensure they were drilled prior to groundwater sample collection. To ensure this was justified, we investigated and found the majority of wells reported as "Active" but lacking spud dates were wells for which permits had been issued but were not yet drilled.

The 5x5km window size and 200m steps used in our sliding window analysis were selected to be consistent and to allow comparisons with past work that analyzed correlations between CH_4 concentrations and proximity to anticlines, faults, and oil and gas wells using these exact window and step sizes[10,11]. The 5x5km window and 200m step size were initially selected based on a consideration of computational complexity and heatmap smoothness. Our sliding window algorithm also necessitates at least 10 samples are located within a window before the Kendall rank correlation is calculated, and a 5x5km window size generally ensures adequate sample numbers to identify statistically significant correlations, if present.

Text S4. Geochemical protocol to identify recently migrated methane

To delineate longstanding CH_4 from newly migrated ("anomalous") CH_4 that may be attributable to UOGD, we implemented a geochemical protocol adapted from Wen et al. (2019).[5] This protocol identifies potentially anomalous CH_4 based on salinity and redox-related parameters. First, samples were screened to include only those with $[CH_4] \geq 1$ mg/L. Next, samples were screened based on ion ratios to exclude samples that contain CH_4 that may be present due to naturally occurring Appalachian brine migration, retaining only samples with a calcium to sodium mass ratio ($[Ca] / [Na]) \geq 0.52$ and $[Cl] \leq 30$ mg/L. Past studies in the Appalachian Basin have documented $[CH_4] \geq 1$ mg/L rarely occurs naturally in non-Na dominated groundwater because Appalachian basin brines are NaCl-rich.[12,13] Thus when $[CH_4] \geq 1$ mg/L is measured in waters with Ca-dominated chemistry, the CH_4 is more likely to be anomalous. Finally, a redox-related filter was applied to further detect samples that contain recently migrated CH_4. Past work has argued the anaerobic oxidation of CH_4 in groundwater is coupled to the reduction of sulfate (SO_4) and iron (Fe); thus, when CH_4 is present in groundwater for long time periods, both $[SO_4]$ and [Fe] are low.[14] In contrast, transiently high $[SO_4]$ and [Fe]

has been observed in groundwaters containing recently migrated CH_4 from leaking UOGD wells.[5,15] Thus, samples that contained $[SO_4] \geq 6$ mg/L or $[SO_4] \geq 6$ mg/L + $[Fe] \geq 0.3$ mg/L were thus considered potentially impacted by recent CH_4 migration, following the criteria of Wen et al.[5]

The results of this protocol are summarized as follows: 382 / 6,991 groundwater samples in our dataset contain $[CH_4]$ over 1 mg/L and were thus investigated for redox and salinity species concentrations that might indicate recent CH_4 migration from a leaking UOG well (shown in red in the workflow in Figure 1). Out of those 382 samples, only 32 contain a [Ca] / [Na] mass ratio ≥ 0.52 and $[Cl] \leq 30$ mg/L, indicating for the majority of the samples, elevated $[CH_4]$ is likely caused by natural brine migration in SWPA. Of the 32 samples that are unlikely to be impacted by natural brine migration, 22 show $[SO_4] \geq 6$ mg/L and 11 show both $[SO_4] \geq 6$ mg/L plus $[Fe] \geq 0.3$ mg/L. Because sulfate and iron concentrations are typically transiently high following a recent influx of CH_4 from leaking oil and gas wells, the 22 samples that were CH_4- and Ca-rich and met either of these redox filters were considered potentially impacted by recently migrated CH_4 (Figure S3).

Considering our results in tandem with sliding window, 3 of the 22 samples are located within hotspots where $[CH_4]$ significantly increases nearby UOG wells (Figure 3A). As such, multiple lines of evidence suggest CH_4 may have recently migrated from UOG wells. The lack of hotspots overlapping with the other 19 samples does not necessarily indicate that these samples were not impacted by UOGD. This geochemical protocol may be better suited for identifying isolated incidents (e.g., only one water supply well contaminated), as a significant correlation between $[CH_4]$ and proximity to UOGD may not be identified by sliding window if only one sample is impacted. Alternatively, the timescales over which CH_4 migration may induce redox

chemistry changes is not well-understood, so sliding window may better detect contaminated localities where a transient spike in iron has already occurred, or where electron acceptors (e.g., iron oxides or sulfate) were already depleted prior to CH_4 migration. It is also possible that our geochemical protocol flags some false positives. For example, long well screens can enable mixing of groundwaters and induce similar changes to redox chemistry. In theory, water wells drilled nearby fractures that enable fluid mixing (e.g., deep anoxic and shallower oxic groundwater) could also produce a similar effect.

Text S5. Additional details on Non-negative Matrix Factorization

Trace ions (e.g., iodide (I), bromide (Br)) and isotopes capable of distinguishing sources of salts in groundwater[15,16] are rarely analyzed in the PADEP dataset and, even when analyzed, are frequently not detected to be above reporting limit. As such, traditional geochemical ratio analysis is not well suited for distinguishing UOGD-related brine contamination from background Cl sources (such as road salt or organic wastes) across all samples in our dataset. To delineate Cl sources using major ion data, we applied a machine learning method, non-negative matrix factorization (NMF).

Our NMF methodology adapted a previously published approach[17] for analyzing sources of SO_4 in streams. NMF derives the mixing proportions and compositions of endmember water types for each analyte by decomposing the matrix multiplication equation V = W x H, where V is the groundwater sample matrix, W is a matrix of the mixing proportions of endmember sources at each location, and H is a matrix comprising the chemical compositions of endmember water types. To prepare the data for NMF, we calculated the molar ratios of Ba, Ca, Mg, SO_4, and Na to Cl and normalized each ratio to its highest respective value.

While NMF does not require prior knowledge of endmember compositions, it does require a known number of endmembers. We thus defined the number of endmembers as the minimum number of components needed to explain 90% of the variance in the data in a Principal Component Analysis. For Cl, this required 3 endmembers. After defining the number of endmembers, we ran NMF using 10,000 model iterations with random initiation. For the 10,000 model deconvolutions run, only model outputs where mixing proportions summed to 1.00 ± 0.05 were retained. These model outputs were then filtered to retain only the top 5% best fitting models, based on the calculated sum square error values (see Shaughnessy et al. (2021)[17] for the SSE equation used). The chemistry and mixing proportions of endmember sources within each sample were subsequently calculated based on the mean and standard deviation of the filtered model outputs.

Supporting Results and Discussion:

Text S6: Correlations between conventional wells and groundwater species

To investigate the impacts of conventional drilling relative to UOGD, we first considered region-wide correlations between [CH_4] or [Cl] vs. conventional well proximity or density. Analyzing the Kendall rank correlation, we found significant correlations (p-values < 0.05) between both [CH_4] and [Cl] and the distance to the nearest COG well across SWPA, but insignificant correlations between these species and COG well density (Table S3). We thus identify a significant correlation between [CH_4] and proximity to COG wells, but not distance to UOG wells (Table 1), in SWPA. Li et al. identified the same correlation for COG wells in their dataset from NEPA, where the number of COG wells is much smaller.[10] These results, across

regions with very different histories of hydrocarbon extraction, suggest CH_4 leakage from COG wells may be widespread and more frequent than leakage from UOG wells.

Despite the regionally significant correlations, we only identify 3 hotspots in the sliding window analysis of [Cl] vs. proximity to COG wells and 2 hotspots in our analysis of [CH_4] vs. proximity to conventional wells that meet the threshold for significance (Figure S2). Additionally, only 2 hotspots where [CH_4] increases with COG well density and 4 hotspots where [Cl] increases with density were identified (Figure S3).

It may seem a puzzle that we see significant region-wide correlations between the distance to conventional wells and both [CH_4] and [Cl], but fewer hotspots potentially associated with CH_4 or Cl contamination from COGD (Figures S2, S3) than UOGD (Figures 2, 3). We hypothesize this seemingly contradictory finding may be a function of both well leakage rates and the volume of gas produced by COG vs. UOG wells. In terms of atmospheric emissions, a greater percentage of COG wells in the Appalachian Basin leak CH_4 to the atmosphere relative to UOG wells, but the magnitude of emissions from COG wells is typically much smaller.[18] Similarly, it's possible that CH_4 leakage to groundwater is more common from COG wells than UOG wells, but smaller volumes of CH_4 reach aquifers from leaking COG wells. Correspondingly, CH_4 leakage from COG wells may be more widespread (reflected in the region-wide correlation we identify), but impact smaller areas and fewer water supply wells due to less gas volume leaked (resulting in fewer hotspots).

Despite evidence for CH_4 and Cl leakage to groundwater from COG wells, we are reasonably confidence that misattribution of COG-related contamination to UOGD is minimal. Overlap between density or distance heatmaps for COG vs. UOG wells is limited to partial overlap of one hotspot for [Cl] vs. well distance and one hotspot for [Cl] vs. well density.

Because we believe the regional correlation between [Cl] and proximity to UOGD is driven by hotspots, the relative lack of overlap with COG well hotspots implies that this regional correlation is not an artifact of COGD contamination.

While we suggest the regional correlations we identify between proximity to COG wells and [CH_4] / [Cl] may indicate more diffuse contamination not detected in sliding window analysis, we suspect such impacts may in fact be less common in the vicinity of UOG wells. Specifically, unconventional operators are required by PADEP regulations (78a.52a and 78a.73) to identify and monitor any orphan or abandoned wells within 1000 feet of the vertical or horizontal sections of a UOG well to avoid communication between wells. Abandoned wells are perhaps the COG wells most likely to contaminate water supplies, as such wells are typically unplugged and in disrepair, and thus the monitoring and potential abatement of such issues during UOGD may minimize the potential for COGD contamination misattributed to UOGD.

Text S7. Assessing the relationship between topography, unconventional gas well locations, and [Cl]

To assess the relationship between topography, [Cl], and UOG well locations, topographic position index (TPI) was calculated for SWPA and Bradford County (in NEPA) following Weiss (2001).[19] Using TPI, the topographic positions of water samples and UOG wells were classified as valley, lower slope, flat slope, middle slope, upper slope, or ridge (listed in order of low to high topographic positions). In SWPA, UOG wells are disproportionately drilled at higher topographic positions, with 63% of wells drilled at ridgetops and 92% of wells drilled in upslope (middle slope, upper slope, or ridge) positions (Figure S8A). In Bradford County, UOG wells are more evenly distributed across topographic positions, although a majority (54%) are still drilled in upslope classes (Figure S8B).

To assess the role topography plays in the distribution of saline groundwater, we compared differences in [Cl] across TPI classes in our SWPA dataset (Figure S8C) and an 11,244-sample groundwater dataset from Bradford County (Figure S8D). The Bradford County dataset, available on the Shale Network database[20], is essentially identical to the dataset used to study SWPA. We assessed whether differences in [Cl] across the 6 topographic classes were statistically significant based on a Tukey's Honest Significant Differences test. No significant differences in [Cl] were observed between TPI classes in SWPA ($p > 0.05$), documenting saline groundwater is relatively evenly distributed regardless of topographic position. This is consistent with past observations of i) a relatively uniform depth to brine across valleys and ridges in SWPA and ii) less topographic relief in SWPA vs. NEPA and thus a lower driving force for brine upwelling in valley bottoms.[6] In contrast, [Cl] in Bradford County was significantly higher in valley bottoms compared to other TPI classes ($p < 0.05$). This finding is also consistent with i) shallower depths to saline groundwater in valley bottoms than ridgetops in NEPA[6] or ii) brine upwelling into groundwater at low topographic positions in NEPA.[21] Either or both could contribute to the prevalence of higher [Cl] in valley bottom groundwaters in the northeast region.

Text S8. Sliding window analysis of [Cl] vs. UOG well distance and density in Bradford County, PA

Unlike in SWPA, we identified a statistically significant decrease in [Cl] nearby UOG wells (Kendall's $\tau = 0.017$, p-value = 0.007) and statistically insignificant increase in [Cl] with increasing UOG well density (Kendall's $\tau = 0.007$, p-value = 0.088) in Bradford County. Hotspots where [Cl] increases with proximity to UOG wells or with UOG well density were mapped across Bradford County using sliding window analysis of the 11,244-sample Bradford County dataset.[20] Three hotspots with significant [Cl] increases were identified in Bradford

County for both distance- (Figure S11) and density-based (Figure S12) analysis. To assess the per-well increase in [Cl] in density hotspots in Figure S12 we calculated the Akritas-Theil-Sen (ATS) slope, Kendall's Tau and p-value of the correlation between [Cl] and UOG well density in hotspot samples (n = 736). These samples display a significant positive correlation between [Cl] and density (ATS slope = 19.2, Kendall's $\tau = 0.171$, p-value < 0.001). The calculated ATS slope thus predicts a 19.2mg/L increase in [Cl] per UOG well drilled. Multiplying this predicted increase (19.2mg/L) by the highest density of UOG wells in these hotspots (n = 5), the maximum predicted increase in [Cl] due to UOGD is 96.0mg/L.

However, the largest hotspots on both the distance and density maps plot nearby the Bridge Street fault and Towanda anticline. This is the area where Wen et al. (2018) observed correlations between [CH_4] and the distance to UOG wells, anticlines, and faults.[11] Although these hotspots do not entirely overlap, we cannot eliminate the possibility these correlations between UOGD and [Cl] are related to natural brine migration along anticlines or faults rather than UOGD because of the many observations of such migration in similar localities.

Text S9. Testing the sliding window geospatial tool (SWGT) on Susquehanna County, PA groundwater methane data

While SWGT has previously been applied to identify potentially CH_4 contaminated sites in the Marcellus Shale region,[10,11] we sought to further validate and exemplify the analysis using a test dataset of groundwater samples from Susquehanna County in NEPA. Dimock Township in Susquehanna County was the site of a contentious methane migration incident following UOGD.[22] Our dataset contains 1,946 samples publicly available in the Shale Network database,[20] as well as 20 samples from Dimock Township discussed in previous literature.[22] Sliding window

was run with and without the Dimock Township data to test the ability of SGWT to detect locations of known CH_4 migration.

When sliding window is run without Dimock Township data, one significant hotspot is identified in Auburn Township (Figure S13A). This hotspot is <1km from a well previously cited by the Pennsylvania Department of Environmental Protection for casing/cementing violations showing signs of gas migration.[23] With Dimock Township data[22] included, an additional hotspot is identified that is centered around the portion of Dimock Township (nearby Burdick and Meshoppen Creeks) associated with gas migration into homeowner wells (Figure S13B). As such, we conclude sliding window can successfully identify confirmed UOGD-related gas migration incidents within a large dataset when the appropriate water analyses are available.

Text S10. Testing our Non-negative Matrix Factorization model using the Lautz et al. (2014) synthetic dataset

To test the viability of using NMF to analyze sources of Cl in a large data set of groundwater chemistry spanning a large geographic area, we applied our methodology to the synthetic "high Cl" data set from Lautz et al. (2014).[15] This data set (n = 8,107) was created by the authors by conservatively mixing a random sample from a bootstraped low-Cl (<20mg/L) groundwater dataset (n = 3,000) with a randomly selected sample from synthetic data sets of road salt, basin brine, septic effluent, or animal waste endmember sources of Cl, with the exact mixing proportions varying between samples and with different endmembers in order to create a realistic data set of high-Cl groundwater.

For our analysis, we considered the same analytes as Lautz et al. (2014): Ba, Br, Ca, I, K, Mg, Na, and Sr.[15] Our model identified three endmembers that we attributed as brine, road salt, and waste-derived sources of chloride. The brine endmember was attributed based on high I/Cl

and Br/Cl molar ratios, road salt was identified based on high Na/Cl, and the waste-derived Cl endmember source was attributed based on high K/Cl, Mg/Cl, and Ca/Cl (Table S4). These endmembers align closely with the sources of Cl used to construct the data set (road salts, Appalachian Basin Brine, septic effluent, and animal waste). However, our NMF model grouped the septic effluent and animal waste Cl sources as one endmember (perhaps due to fewer inorganic species strongly associated with these sources of Cl compared to brine- or road salt-derived Cl).

We assessed our model's performance relative to that of the Lautz et al. linear discriminant analysis (LDA) model[15] as a rough measure of its accuracy, though we caution these approaches are not directly comparable. For the comparison, we classified the predominant source of Cl in every sample based on which endmember displayed the largest mixing proportion (e.g., $\alpha_{brine} > \alpha_{road\ salt}, \alpha_{waste\text{-}derived}$ was classified as a sample with brine-sourced Cl). Our NMF model classified the correct mixing endmember Cl source for 85% of brine-derived samples, 86% of road salt derived samples, and 64% of waste-derived samples. These results show similar accuracy to that described when the Lautz et al. (2014) LDA model was applied to other regions of the U.S. with known sources of Cl contamination.[24]

Text S11. Groundwater flow in SWPA vs. NEPA

Hydrogeologic regimes vary between SWPA and NEPA, potentially influencing UOGD contaminant transport and domestic well vulnerability to contamination. Greater topographic relief in NEPA produces greater hydraulic head, and more regional groundwater flow compared to SWPA.[6] Additionally, recharge is generally slower in domestic wells in SWPA compared to NEPA.[25] Slower flow in SWPA could act to limit the transport of contaminants from a UOG

wellpad to a domestic well, or alternatively reduce the dilution of contaminants in contaminated aquifers. However, the exact vulnerability of a domestic well in SWPA or NEPA is ultimately shaped by multiple factors including the depth of the well, aquifer(s) from which it draws water, topographic position, and the presence of nearby preferential flowpaths (e.g., fractures or coal seams).

When broadly assessed on a regional scale, domestic groundwater supplies in both SWPA and NEPA are potentially more vulnerable to surface and subsurface sources of UOGD contamination than other shale gas basins in the United States. Groundwater age studies in both NEPA and a portion of the Appalachian Basin bordering SWPA (northern West Virginia) both report typical groundwater in domestic wells is a mix of fresh water recharged post-1950 and waters recharged pre-1950 with higher salinities and total dissolved solids.[26,27] In contrast, domestic wells in the Eagle Ford, Bakken, and Haynesville Shale plays (U.S.A.) predominantly contain entirely pre-1950 waters, making UOGD-related contamination from surface sources (e.g., leaking impoundment pits or spills) unlikely in the short term.[28,29] In the Williston Basin, surbsurface hydrogeologic transport of contamination from a UOGD wellpad into water supplies >0.5km is unlikely given the short amount of time since the onset of UOGD and long timescales necessary for contaminant transport.[28] However, domestic wells in both NEPA and SWPA, where groundwaters typically contain modern recharge, may be vulnerable to contamination from both land surface activities (e.g., leaking impoundments) and subsurface leaks.

As such, we conclude that relative to other shale gas basins, SWPA and NEPA may be more likely to experience contamination but hydrogeologic differences between the two settings are unlikely to dramatically alter the frequency of contamination. Future work could pair our

analysis methods with hydrogeologic models to predict domestic well vulnerability to UOGD.[30,31]

Text S12: Gas geochemistry in a small subset of sample sites

Although our datasets are large, they and other studies using PA DEP pre-drill measurements[10,11,12] generally do not provide all the data that can be collected in location-specific case studies to delineate sources of CH_4. Additionally, while ethane concentrations were measured in 6,957 / 6,991 samples in the data set, ethane detection limits are >2.5 mg/L in 2,157 of the samples in our data set and >1 mg/L in 3,899 of the samples. Resultantly, ethane concentrations in only 77 samples are reported above the limit, limiting the utility of calculated C_1/C_2+ ratios across the region. Other measurements such as CH_4 isotopic compositions and water table depths are generally not reported, and water sample or well depths are only sometimes reported. Therefore, no analyses of these important observations can be made from our dataset. This points to the importance of case studies in smaller regions to test hypotheses or to develop multiple lines of evidence to explain incidents of contamination when they occur.

However, more detailed data on dissolved gas concentrations and gas isotopic compositions are available for a small number of sites (n = 58) in our data set. Generally, the anions and cations discussed in the main body of this paper were measured on separate samples, and isotopic measurements were completed on samples collected from the same water sources on later dates. We have no knowledge of why additional samples were taken, but one possibility is that the regulator was testing the waters because of higher [CH_4]. Within these samples, $\delta^{13}C$-CH_4 ranges from -49.75‰ to -76.57‰, $\delta^{2}H$-CH_4 ranges from -152.9‰ to -317.5‰, $\delta^{13}C$-C_2H_6 ranges from -18.84‰ to -36.20‰, and ratios of CH_4 (C1) to ethane and propane ($C_2 + C_3$) range

from 29 to 36,320 (Figure S14). [CH_4] in the corresponding samples in our geochemical data set range from 6.54 to 73.40mg/L.

Of these 58 samples, 7 were collected from the hotspot circled in Figure 3A where we suspected a shale gas well may be leaking CH_4 into groundwater. For these samples, measured values of $\delta^{13}C$-CH_4 (-62.35‰ to -66.38‰) and $\delta^{2}H$-CH_4 (-189.7‰ to -200.5‰) are consistent with a biogenic origin of CH_4. However, $C_1 / (C_2 + C_3)$ ratios (116 - 315) fall within the range typically associated with thermogenic CH_4 or a mix of thermogenic and biogenic CH_4.[32] Additionally, values of [CH_4] (8.05 – 22.50 mg/L) are much higher than the concentrations typically associated with biogenic CH_4 in Pennsylvania.[33,34] We focus on three possible explanations to explain the $\delta^{13}C$-CH_4 and $C_1 / (C_2 + C_3)$ observed in these samples: CH_4 oxidation, the mixing of thermogenic and biogenic gases, or the presence of coal bed gas.

We calculated how oxidation of a theoretical biogenic CH_4 alters $\delta^{13}C$-CH_4, $\delta^{2}H$-CH_4, and $C_1 / (C_2 + C_3)$ based on equations 1-3:

$$\delta^{13}C_t = \delta^{13}C_i - \varepsilon_C * \log(f) \quad (1)$$

$$\delta^{2}H_t = \delta^{2}H_i - \varepsilon_D * \log(f) \quad (2)$$

$$C_1 / (C_2+C_3)_t = C_1 / (C_2+C_3)_i * f \quad (3)$$

where ε_C and ε_D are the kinetic fractionation factors for carbon & hydrogen, respectively, and f represents the fraction of initial CH_4 remaining. $\delta^{13}C_i$, $\delta^{2}H_i$, and $C_1 / (C_2 + C_3)_i$ represent the initial gas composition, while $\delta^{13}C_t$, $\delta^{2}H_t$, and $C_1 / (C_2 + C_3)_t$ represent the gas composition at a given f. We constructed our model using values for ε_C (3) and ε_D (54) reported by Schout et al.,[35] and biogenic CH_4 with an initial $\delta^{13}C$-CH_4 of -70‰, $\delta^{2}H$-CH_4 of -290‰, and $C_1 / (C_2 + C_3)$ of 1,000. Our calculation indicates the oxidation of such CH_4 would produce $\delta^{13}C$-CH_4, $\delta^{2}H$-CH_4, and $C_1 / (C_2 + C_3)$ within the range observed in our samples (Figure S15) at an f of 0.16-0.19

(e.g., 81-84% of original CH_4 oxidized). While oxidation could thus plausibly explain the seemingly biogenic $\delta^{13}C\text{-}CH_4$ and thermogenic $C_1 / (C_2 + C_3)$ ratios in these samples, a trend in $\delta^{13}C\text{-}CH_4$ vs. $\delta^{2}H\text{-}CH_4$ or $\delta^{13}C\text{-}CH_4$ vs. $C_1 / (C_2 + C_3)$ indicative of progressive oxidation is not especially apparent within the hotspot samples or the full isotopic dataset (Figure S14).

Next, we evaluated whether the mixing of thermogenic gas with biogenic gas could produce a gas geochemistry consistent with what we observe in the hotspot samples, particularly focusing on a scenario where thermogenic CH_4 leaked from a UOG well mixes with biogenic CH_4 present in an aquifer prior to drilling. Mixing Marcellus shale gas ($\delta^{13}C\text{-}CH_4$ = -37.7‰, $\delta^{2}H\text{-}CH_4$ = -165.6‰, $C_1 / (C_2 + C_3)$ = 37.55, based on averages reported in Sharma et al.[36]) with biogenic gas ($\delta^{13}C\text{-}CH_4$ = -75‰, $\delta^{2}H\text{-}CH_4$ = -205‰, $C_1 / (C_2 + C_3)$ = 1,000) also produces $\delta^{13}C\text{-}CH_4$, $\delta^{2}H\text{-}CH_4$, and $C_1 / (C_2 + C_3)$ that fit within the range in our samples at a mix of 24-30% Marcellus gas (Figure S15).

Alternatively, the observed $\delta^{13}C\text{-}CH_4$ and $C_1 / (C_2 + C_3)$ may simply reflect the presence of coal bed CH_4. While coal bed gases are often predominantly CH_4, $C_1 / (C_2 + C_3) < 100$ has been reported for the Pittsburgh coal seam in SWPA.[37] However, $\delta^{13}C\text{-}CH_4$ values in gas samples from the Freeport and Pittsburgh coals in Pennsylvania are typically isotopically heavier than our samples, although coal bed methane $\delta^{13}C\text{-}CH_4$ can vary considerably.[38] Methanogenesis within coals could also produce secondary biogenic gas that mixes with more thermogenic coal bed CH_4 to produce a mixed isotopic signature,[38] perhaps explaining the more depleted $\delta^{13}C\text{-}CH_4$.

Despite the ambiguous provenance of CH_4 in this locality, some of the scenarios explored are consistent with a potential impact of UOGD. A mix of thermogenic and biogenic gas could conceivably result from leakage of thermogenic gas from a UOG well. Additionally, it's possible

that UOG wells could enable the upward migration of coalbed gas into groundwater if they lack coal casings. As we documented UOG wells intersecting coal mining areas that lacked coal casings in this hotspot, this mechanism is especially plausible in these samples.

Supporting Information References:

(1) Carter, K. M.; Harper, J. A.; Schmid, K. W.; Kostelnik, J. Unconventional Natural Gas Resources in Pennsylvania: The Backstory of the Modern Marcellus Shale Play. *Environ. Geosci.* **2011**, *18* (4), 217–257. https://doi.org/10.1306/eg.09281111008.

(2) PA Department of Environmental Protection Open Data Portal https://newdata-padep-1.opendata.arcgis.com/ (accessed October 1, 2021).

(3) Dilmore, R. M.; Sams, J. I.; Glosser, D.; Carter, K. M.; Bain, D. J. Spatial and Temporal Characteristics of Historical Oil and Gas Wells in Pennsylvania: Implications for New Shale Gas Resources. *Environ. Sci. Technol.* **2015**, *49* (20), 12015–12023. https://doi.org/10.1021/acs.est.5b00820.

(4) Wilkin, R. T.; Lee, T. R.; Ruybal, C. J.; Rectenwald, D. J. Retrospective Case Study in Southwestern Pennsylvania: Study of the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources. U.S. Environmental Protection Agency 2015.

(5) Wen, T.; Woda, J.; Marcon, V.; Niu, X.; Li, Z.; Brantley, S. L. Exploring How to Use Groundwater Chemistry to Identify Migration of Methane near Shale Gas Wells in the Appalachian Basin. *Environ. Sci. Technol.* **2019**, *53* (15), 9317–9327. https://doi.org/10.1021/acs.est.9b02290.

(6) Siegel, D. I.; Smith, B.; Perry, E.; Bothun, R.; Hollingsworth, M. 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.* **2015**, *63*, 37–57. https://doi.org/10.1016/j.apgeochem.2015.06.013.

(7) Piper, A. M. *Ground Water in Southwestern Pennsylvania*; 4th series, Bulletin W-1; Pennsylvania Geological Survey, 1933.

(8) Pennsylvania geologic map data https://mrdata.usgs.gov/geology/state/state.php?state=PA (accessed October 1, 2021).

(9) Bureau, U. C. TIGER/Line Shapefiles https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html (accessed October 1, 2021).

(10) Li, Z.; You, C.; Gonzales, M.; Wendt, A. K.; Wu, F.; Brantley, S. L. Searching for Anomalous Methane in Shallow Groundwater near Shale Gas Wells. *J. Contam. Hydrol.* **2016**, *195*, 23–30. https://doi.org/10.1016/j.jconhyd.2016.10.005.

(11) Wen, T.; Niu, X.; Gonzales, M.; Zheng, G.; Li, Z.; Brantley, S. L. 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.* **2018**, *52* (12), 7149–7159. https://doi.org/10.1021/acs.est.8b01123.

(12) Molofsky, L. J.; Connor, J. A.; Wylie, A. S.; Wagner, T.; Farhat, S. K. Evaluation of Methane Sources in Groundwater in Northeastern Pennsylvania. *Groundwater* **2013**, *51* (3), 333–349. https://doi.org/10.1111/gwat.12056.

(13) McPhillips, L. E.; Creamer, A. E.; Rahm, B. G.; Walter, M. T. Assessing Dissolved Methane Patterns in Central New York Groundwater. *J. Hydrol. Reg. Stud.* **2014**, *1*, 57–73. https://doi.org/10.1016/j.ejrh.2014.06.002.

(14) Woda, J.; Wen, T.; Oakley, D.; Yoxtheimer, D.; Engelder, T.; Castro, M. C.; Brantley, S. Detecting and Explaining Why Aquifers Occasionally Become Degraded near Hydraulically Fractured Shale Gas Wells. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 201809013. https://doi.org/10.1073/pnas.1809013115.

(15) Lautz, L. K.; Hoke, G. D.; Lu, Z.; Siegel, D. I.; Christian, K.; Kessler, J. D.; Teale, N. G. Using Discriminant Analysis to Determine Sources of Salinity in Shallow Groundwater Prior to Hydraulic Fracturing. *Environ. Sci. Technol.* **2014**, *48* (16), 9061–9069. https://doi.org/10.1021/es502244v.
(16) Llewellyn, G. T. Evidence and Mechanisms for Appalachian Basin Brine Migration into Shallow Aquifers in NE Pennsylvania, USA. *Hydrogeol. J.* **2014**, *22* (5), 1055–1066. https://doi.org/10.1007/s10040-014-1125-1.
(17) Shaughnessy, A. R.; Gu, X.; Wen, T.; Brantley, S. L. Machine Learning Deciphers CO2 Sequestration and Subsurface Flowpaths from Stream Chemistry. *Hydrol. Earth Syst. Sci.* **2021**, *25* (6), 3397–3409. https://doi.org/10.5194/hess-25-3397-2021.
(18) Omara, M.; Sullivan, M. R.; Li, X.; Subramanian, R.; Robinson, A. L.; Presto, A. A. Methane Emissions from Conventional and Unconventional Natural Gas Production Sites in the Marcellus Shale Basin. *Environ. Sci. Technol.* **2016**, *50* (4), 2099–2107. https://doi.org/10.1021/acs.est.5b05503.
(19) Weiss, A. Topographic Position and Landform Analysis *Poster presentation, ESRI user conference, San Diego, CA,* 2001.
(20) Brantley, S. L. Shale Network Database. *Consortium for Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI)* **2011**. https://doi.org/10.4211/his-data-shalenetwork.
(21) Warner, N.; Jackson, R.; Darrah, T.; Osborn, S.; Down, A.; Zhao, K.; White, A.; Vengosh, A. Geochemical Evidence for Possible Natural Migration of Marcellus Formation Brine to Shallow Aquifers in Pennsylania. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 11961–11966. https://doi.org/10.1073/pnas.1121181109.
(22) Hammond, P. A. The Relationship between Methane Migration and Shale-Gas Well Operations near Dimock, Pennsylvania, USA. *Hydrogeol J* **2016**, *24* (2), 503–519. https://doi.org/10.1007/s10040-015-1332-4.
(23) Oil and Gas Compliance - Report Extracts https://www.depgreenport.state.pa.us/ReportExtracts/OG/OilComplianceReport (accessed October 13, 2021).
(24) Chien, N. P.; Lautz, L. K. Discriminant Analysis as a Decision-Making Tool for Geochemically Fingerprinting Sources of Groundwater Salinity. *Sci. Total Environ.* **2018**, *618*, 379–387. https://doi.org/10.1016/j.scitotenv.2017.11.019.
(25) Trapp Jr., H.; Horn, M. A. *Ground Water Atlas of the United States: Segment 11, Delaware, Maryland, New Jersey, North Carolina, Pennsylvania, Virginia, West Virginia*; Hydrologic Atlas; Report 730L; 1997; pp L1–L24. https://doi.org/10.3133/ha730L.
(26) McMahon, P. B.; Lindsey, B. D.; Conlon, M. D.; Hunt, A. G.; Belitz, K.; Jurgens, B. C.; Varela, B. A. Hydrocarbons in Upland Groundwater, Marcellus Shale Region, Northeastern Pennsylvania and Southern New York, USA. *Environ. Sci. Technol.* **2019**, *53* (14), 8027–8035. https://doi.org/10.1021/acs.est.9b01440.
(27) Harkness, J. S.; Darrah, T. H.; Warner, N. R.; Whyte, C. J.; Moore, M. T.; Millot, R.; Kloppmann, W.; Jackson, R. B.; Vengosh, A. The Geochemistry of Naturally Occurring Methane and Saline Groundwater in an Area of Unconventional Shale Gas Development. *Geochim. Cosmochim. Acta* **2017**, *208*, 302–334. https://doi.org/10.1016/j.gca.2017.03.039.
(28) McMahon, P. B.; Caldwell, R. R.; Galloway, J. M.; Valder, J. F.; Hunt, A. G. Quality and Age of Shallow Groundwater in the Bakken Formation Production Area, Williston Basin,

Montana and North Dakota. *Groundwater* **2015**, *53* (S1), 81–94. https://doi.org/10.1111/gwat.12296.
(29) 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.; Kresse, T. M. Methane and Benzene in Drinking-Water Wells Overlying the Eagle Ford, Fayetteville, and Haynesville Shale Hydrocarbon Production Areas. *Environ. Sci. Technol.* **2017**, *51* (12), 6727–6734. https://doi.org/10.1021/acs.est.7b00746.
(30) Soriano, Jr/, M. A.; Siegel, H. G.; Gutchess, K. M.; Clark, C. J.; Li, Y.; Xiong, B.; Plata, D. L.; Deziel, N. C.; Saiers, J. E. Evaluating Domestic Well Vulnerability to Contamination From Unconventional Oil and Gas Development Sites. *Water Resour. Res.* **2020**, *56* (10), e2020WR028005. https://doi.org/10.1029/2020WR028005.
(31) Soriano, Jr., M. A.; Siegel, H. G.; Johnson, N. P.; Gutchess, K. M.; Xiong, B.; Li, Y.; Clark, C. J.; Plata, D. L.; Deziel, N. C.; Saiers, J. E. Assessment of Groundwater Well Vulnerability to Contamination through Physics-Informed Machine Learning. *Environ. Res. Lett.* **2021**, *16* (8), 084013. https://doi.org/10.1088/1748-9326/ac10e0.
(32) McIntosh, J. C.; Hendry, M. J.; Ballentine, C.; Haszeldine, R. S.; Mayer, B.; Etiope, G.; Elsner, M.; Darrah, T. H.; Prinzhofer, A.; Osborn, S.; Stalker, L.; Kuloyo, O.; Lu, Z.-T.; Martini, A.; Lollar, B. S. A Critical Review of State-of-the-Art and Emerging Approaches to Identify Fracking-Derived Gases and Associated Contaminants in Aquifers. *Environ. Sci. Technol.* **2019**, *53* (3), 1063–1077. https://doi.org/10.1021/acs.est.8b05807.
(33) Osborn, S. G.; Vengosh, A.; Warner, N. R.; Jackson, R. B. Methane Contamination of Drinking Water Accompanying Gas-Well Drilling and Hydraulic Fracturing. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (20), 8172–8176. https://doi.org/10.1073/pnas.1100682108.
(34) Jackson, R. B.; Vengosh, A.; Darrah, T. H.; Warner, N. R.; Down, A.; Poreda, R. J.; Osborn, S. G.; Zhao, K.; Karr, J. D. Increased Stray Gas Abundance in a Subset of Drinking Water Wells near Marcellus Shale Gas Extraction. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (28), 11250–11255. https://doi.org/10.1073/pnas.1221635110.
(35) Schout, G.; Hartog, N.; Hassanizadeh, S. M.; Griffioen, J. Impact of an Historic Underground Gas Well Blowout on the Current Methane Chemistry in a Shallow Groundwater System. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (2), 296–301. https://doi.org/10.1073/pnas.1711472115.
(36) Sharma, S.; Mulder, M. L.; Sack, A.; Schroeder, K.; Hammack, R. Isotope Approach to Assess Hydrologic Connections During Marcellus Shale Drilling. *Groundwater* **2014**, *52* (3), 424–433. https://doi.org/10.1111/gwat.12083.
(37) Kim, A. G. *The Composition of Coalbed Gas*; U.S. Department of the Interior, Bureau of Mines: Washington, 1973.
(38) Laughrey, C. D.; Baldassare, F. J. Geochemistry and Origin of Some Natural Gases in the Plateau Province, Central Appalachian Basin, Pennsylvania and Ohio1. *Am. Assoc. Pet. Geol. Bull.* **1998**, *82* (2), 317–335. https://doi.org/10.1306/1D9BC403-172D-11D7-8645000102C1865D.
(39) Siegel, D. I.; Azzolina, N. A.; Smith, B. J.; Perry, A. E.; Bothun, R. L. Methane Concentrations in Water Wells Unrelated to Proximity to Existing Oil and Gas Wells in Northeastern Pennsylvania. *Environ. Sci. Technol.* **2015**, *49* (7), 4106–4112. https://doi.org/10.1021/es505775c.

Supporting Figures:

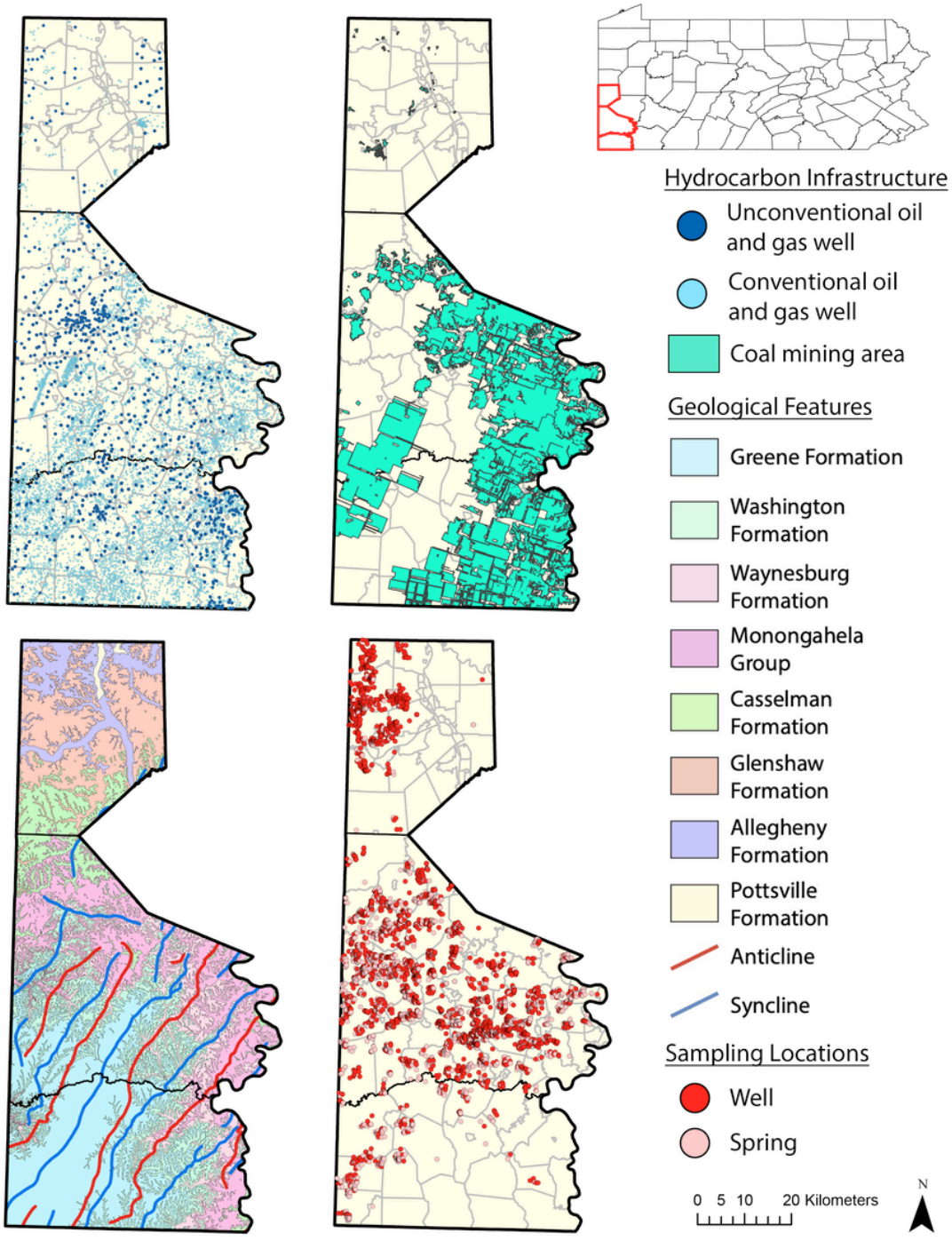

Figure S1. Maps of the locations of unconventional and conventional oil and gas wells, coal mining areas identified by the PADEP, regional geology with regards to bedrock lithology and folding, and the location of groundwater samples (collected from wells or springs) included in our dataset within Beaver (top), Washington (middle) and Greene (bottom) counties in SWPA.

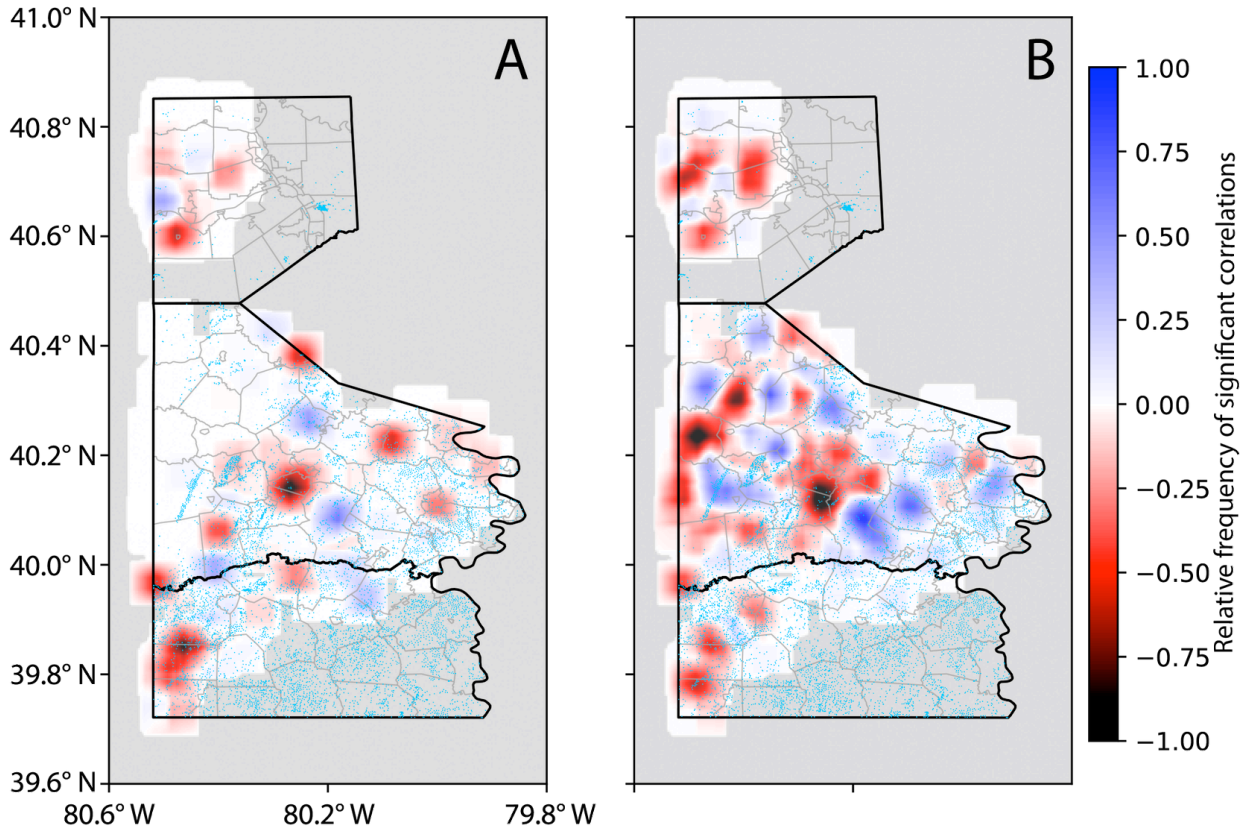

Figure S2. Sliding window heatmaps displaying correlations between the distance to conventional oil and gas (COG) wells and (A) [CH_4] or (B) [Cl] within Beaver (top), Washington (middle) and Greene (bottom) counties in SWPA. The locations of COG wells are shown in light blue. Areas shaded red indicate negative correlations ([analyte] increases as the distance to a COG well decreases) while areas shaded blue indicate positive correlations ([analyte] decreases as the distance to a COG well decreases). Darker colors imply stronger correlation between water chemistries at a given point on the map and proximity to the feature of interest. Areas shaded white are locations where sliding window was run and those shaded grey are localities without sufficient data density for sliding window analysis.

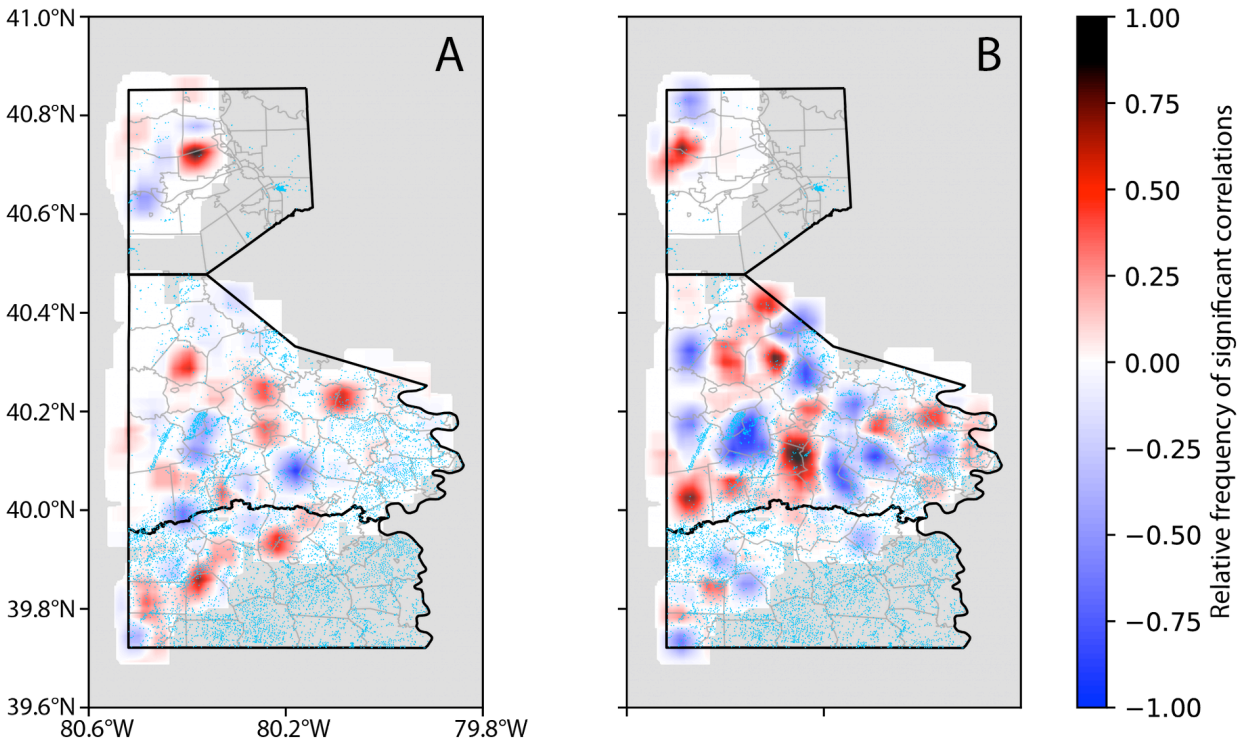

Figure S3. Sliding window heatmaps displaying correlations between the density of conventional oil and gas (COG) wells within 1km of a respective water sample and (A) [CH_4] or (B) [Cl]. The locations of COG wells are shown in light blue. Areas shaded blue indicate negative correlations ([analyte] decreases as COG well density increases) while areas shaded red indicate positive correlations ([analyte] increases as COG well density increases), with color intensity corresponding to the relative frequency of significant positive (red) or negative (blue) correlations. We consider areas with a relative frequency of significant correlation > 0.7 as "hotspots" (maroon to black coloration). See Figure S2 for a description of the significance of areas shaded white and grey.

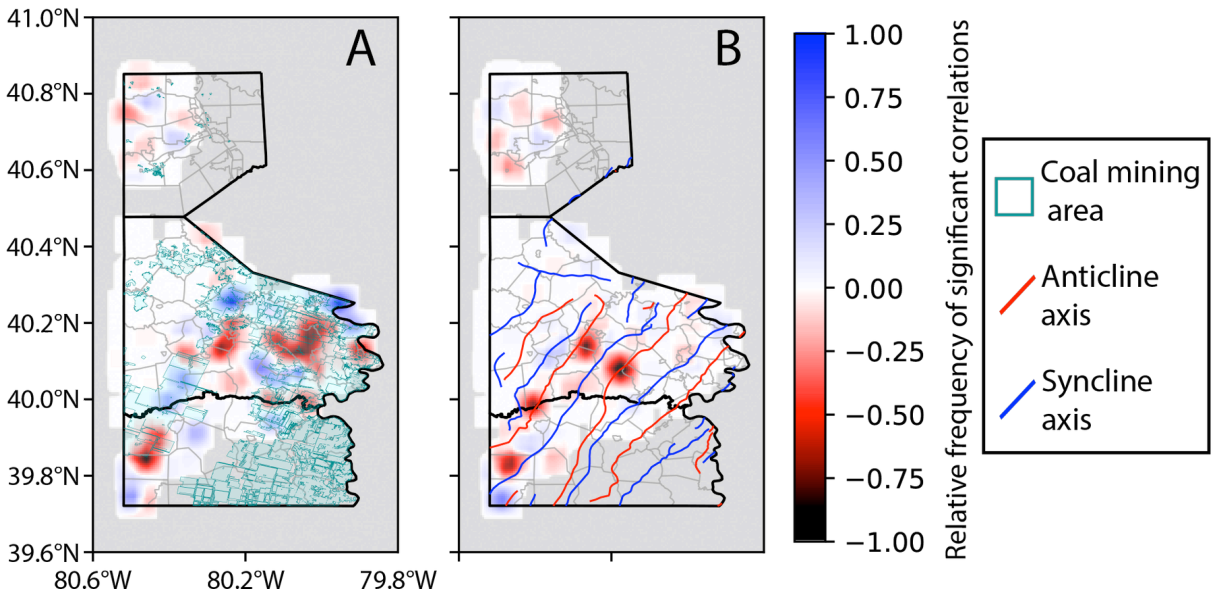

Figure S4. Sliding window heatmaps displaying correlations between [CH_4] and the distance to (A) coal mining areas (outlined in blue-green) and (B) anticlinal fold axes within Beaver (top), Washington (middle) and Greene (bottom) counties in SWPA. See Figure S2 for descriptions of color shading and the significance of areas shaded white and grey.

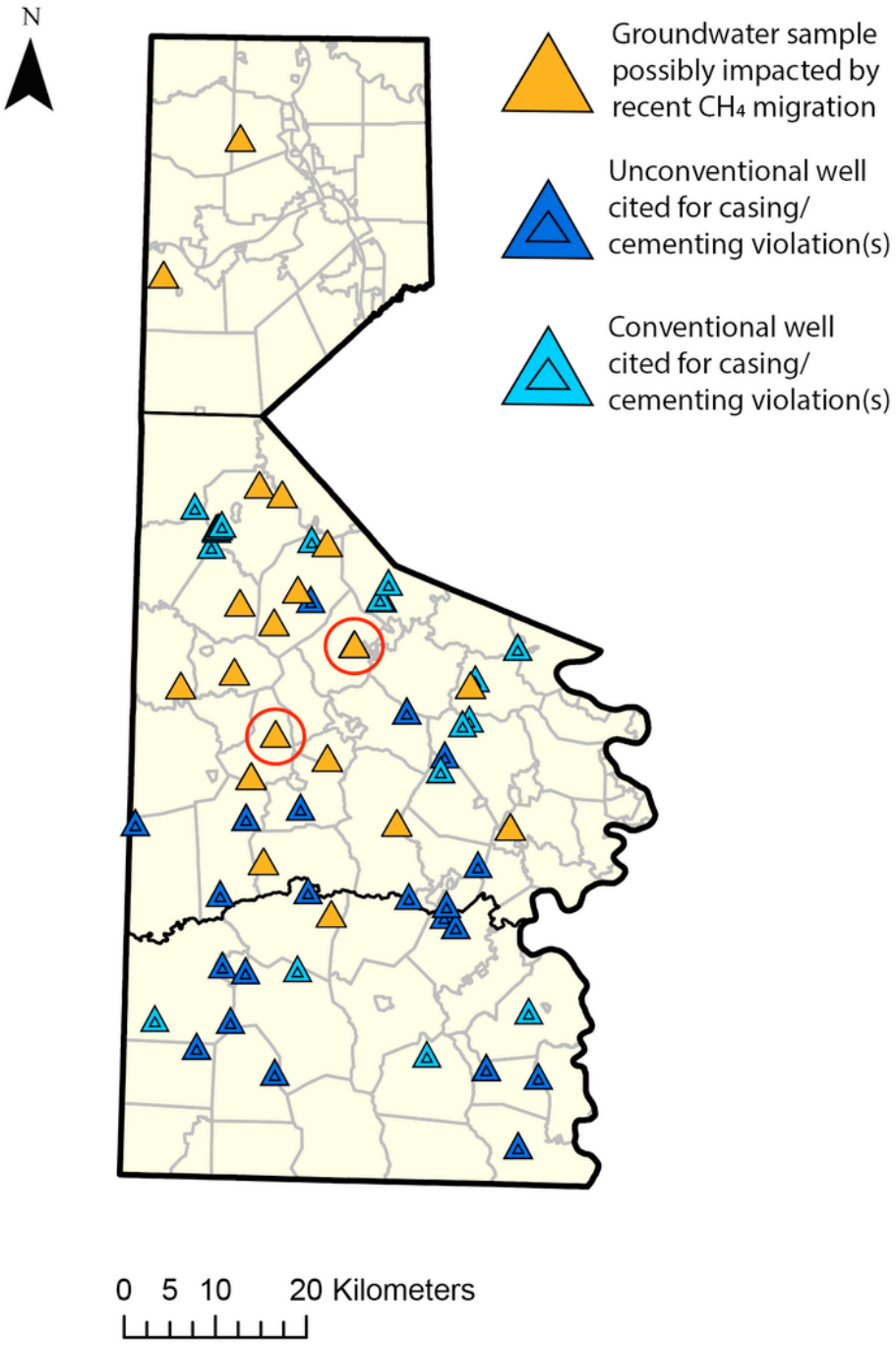

Figure S5. The locations of the 22 samples flagged by the geochemical protocol as being potentially impacted by recent CH_4 migration (yellow triangle) along with locations of unconventional (dark blue) and conventional (light blue) oil and gas wells cited for casing or cementing violations in the PA DEP compliance report database [22]. Beaver (top), Washington (middle) and Greene (bottom) counties are outlined. Samples circled in red are located within hotspots where [CH_4] increases with proximity to UOG wells (Figure 3A).

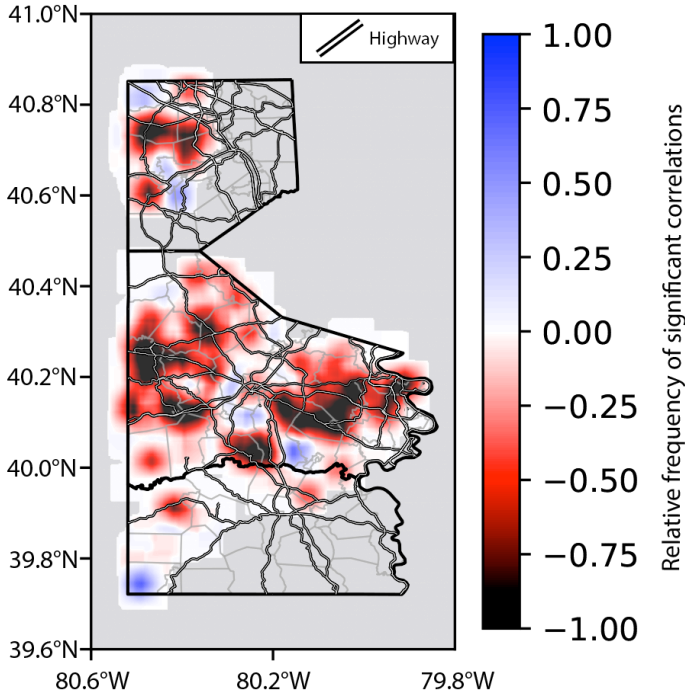

Figure S6. Sliding window heatmap displaying correlations between [Cl] and the distance to U.S. Census Bureau TIGER/Line shapefile roads in SWPA. See Figure S2 for descriptions of color shading and the significance of areas shaded white and grey.

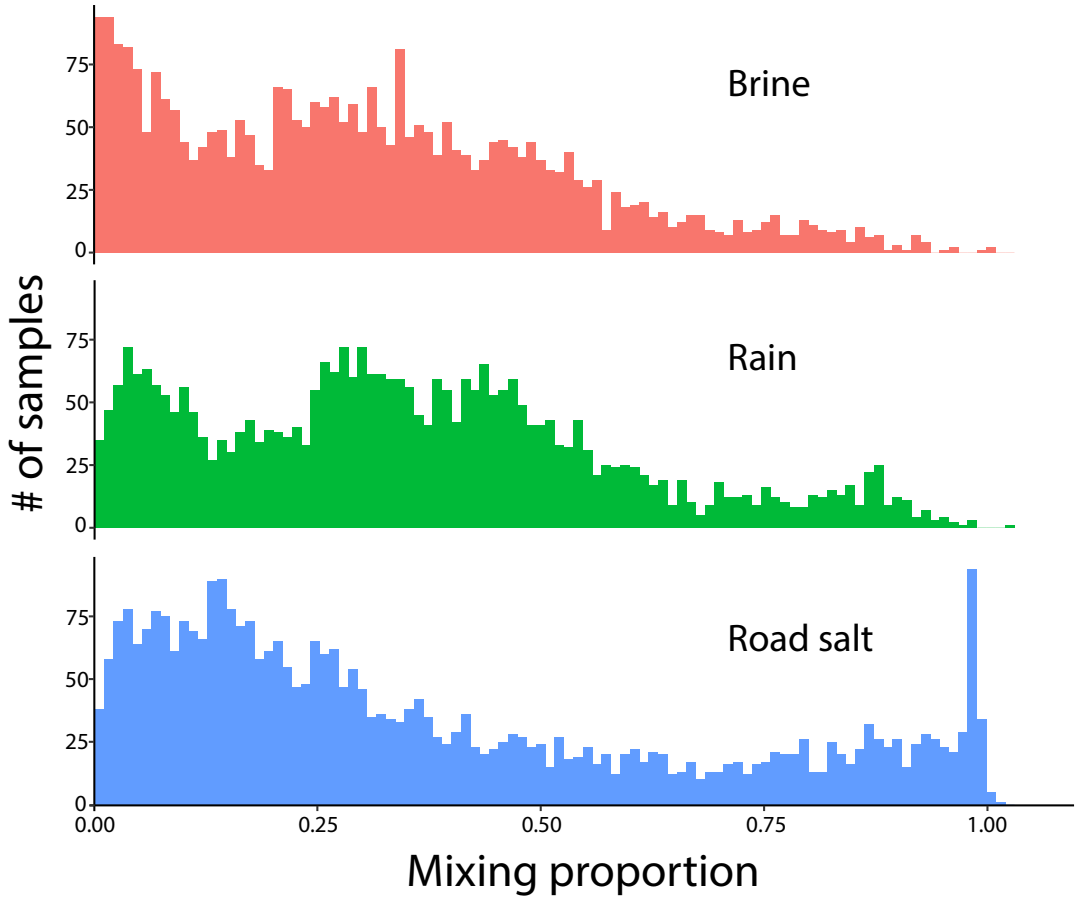

Figure S7. Histograms of NMF-derived mixing proportions for endmember sources of chloride, depicting the mixing proportions for all samples included in the final model. The contribution of road salt to samples is highly skewed, with a small percentage of samples appearing highly impacted (i.e., 16% of samples with $\alpha_{\text{road salt}} > 0.75$ vs. 44% of samples with $\alpha_{\text{road salt}} < 0.25$) by road salting. Only 2% and 4% of samples contain $\alpha_{\text{brine}} > 0.75$ or $\alpha_{\text{MR}} > 0.75$, respectively.

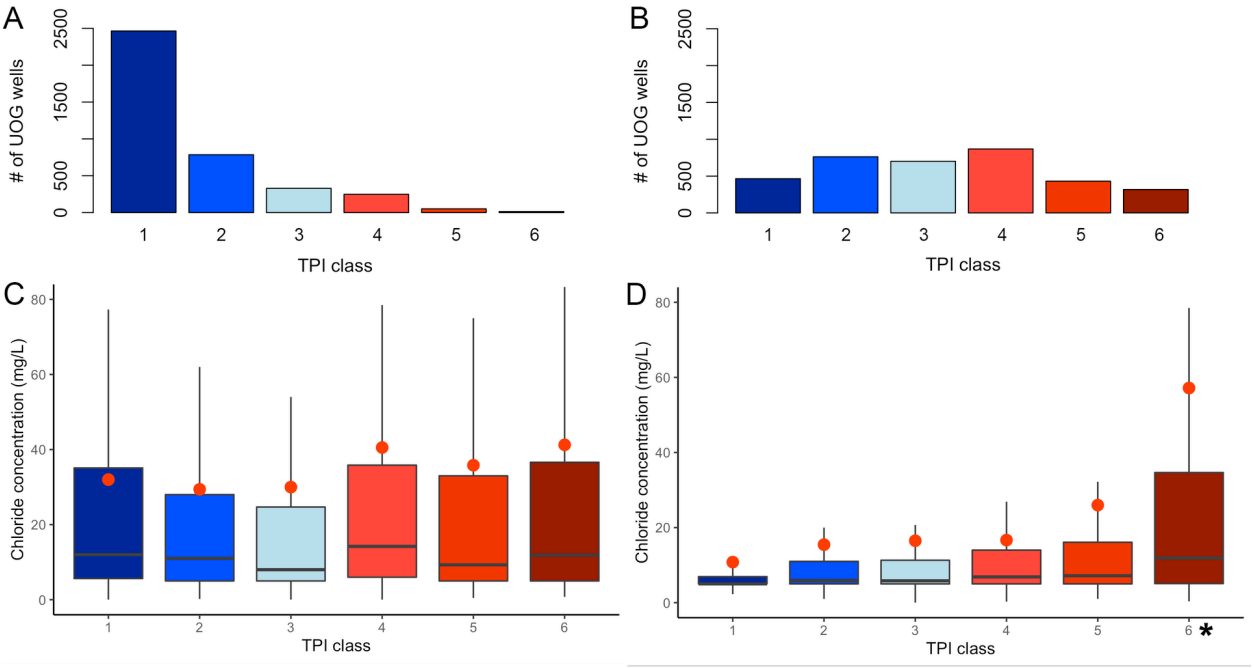

Figure S8. UOG well locations in SWPA (A) and Bradford County (B) grouped by their respective topographic position. Topographic positions were assigned based on Weiss (2001), where classes 1-6 correspond to a topographic position of ridge, upper slope, middle slope, flat slope, lower slope, and valley, respectively [16]. Plots of chloride concentrations across topographic positions in SWPA (C) and Bradford County (D). The lower and upper bounds of each box correspond to the first (Q1) and third (Q3) quantile [Cl], respectively, while the solid line through the box represents the median [Cl]. The upper whisker corresponds to Q3 + 1.5*(Q3-Q1), while the lower whisker corresponds to Q1– 1.5*(Q3-Q1). Outlier values are not shown in the plot. The mean [Cl] in each topographic position is denoted using a red circle. A * corresponds to a statistically significant difference in mean [Cl] ($p < 0.05$), assessed for every topographic position using a Tukey's Honest Significant Differences test. The only significant difference identified was for TPI Class 6 (valleys) in Bradford County, where [Cl] was significantly higher than any other TPI class.

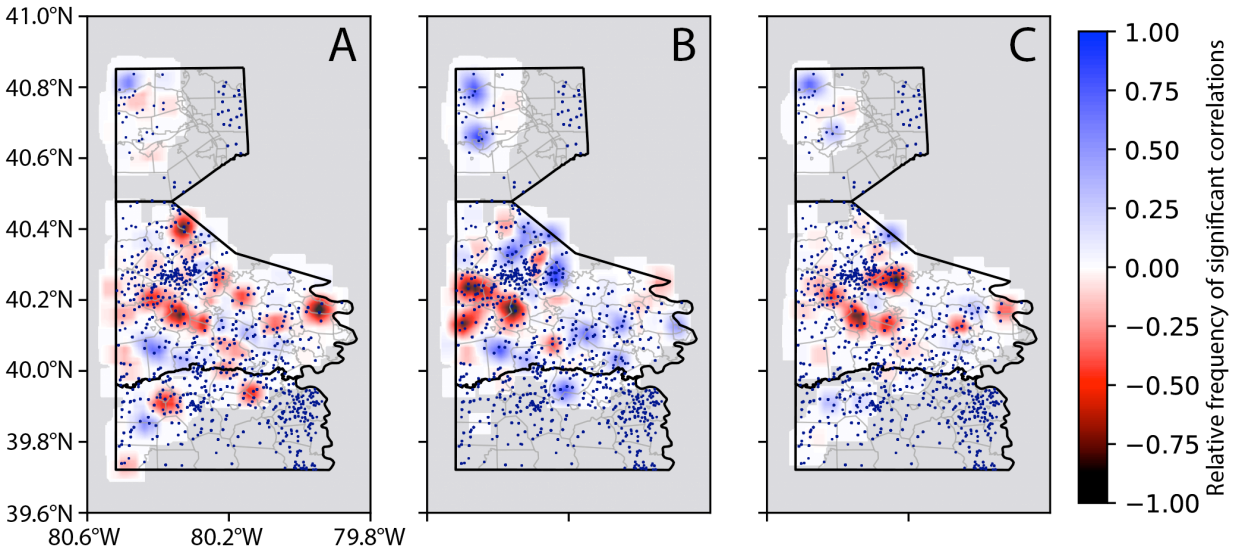

Figure S9. Sliding window heatmaps for correlations between the proximity to unconventional wells and concentrations of barium (A), strontium (B), and NMF-derived concentrations of chloride attributable to a brine source (C). The locations of unconventional wells are shown in dark blue dots on the heatmap. Hotspots on the barium and strontium heatmaps generally align with hotspots identified for Cl (Figure 3B), supporting a brine source of Cl in these localities.

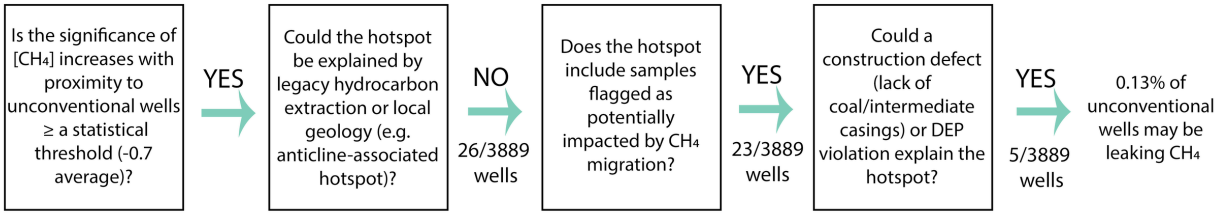

Figure S10. Schematic diagram of the workflow used to derive a conservative estimate of the number of potentially leaking wells in the study area based on sliding window outputs.

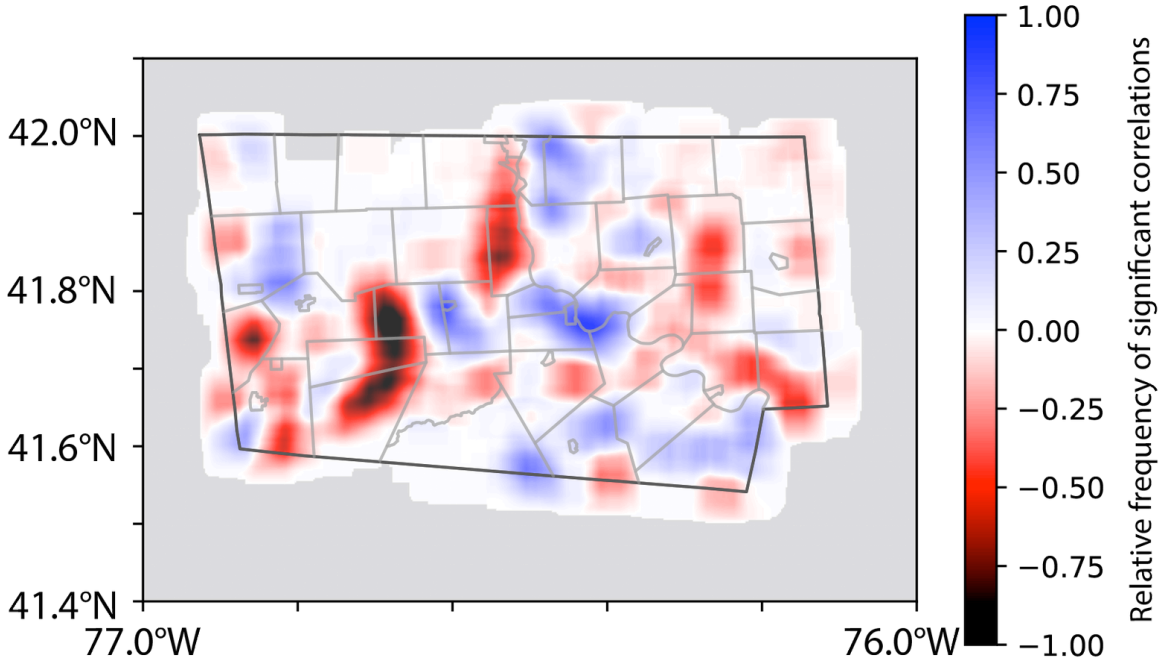

Figure S11. Sliding window heatmap displaying correlations between chloride concentrations and the proximity to unconventional wells in Bradford County.

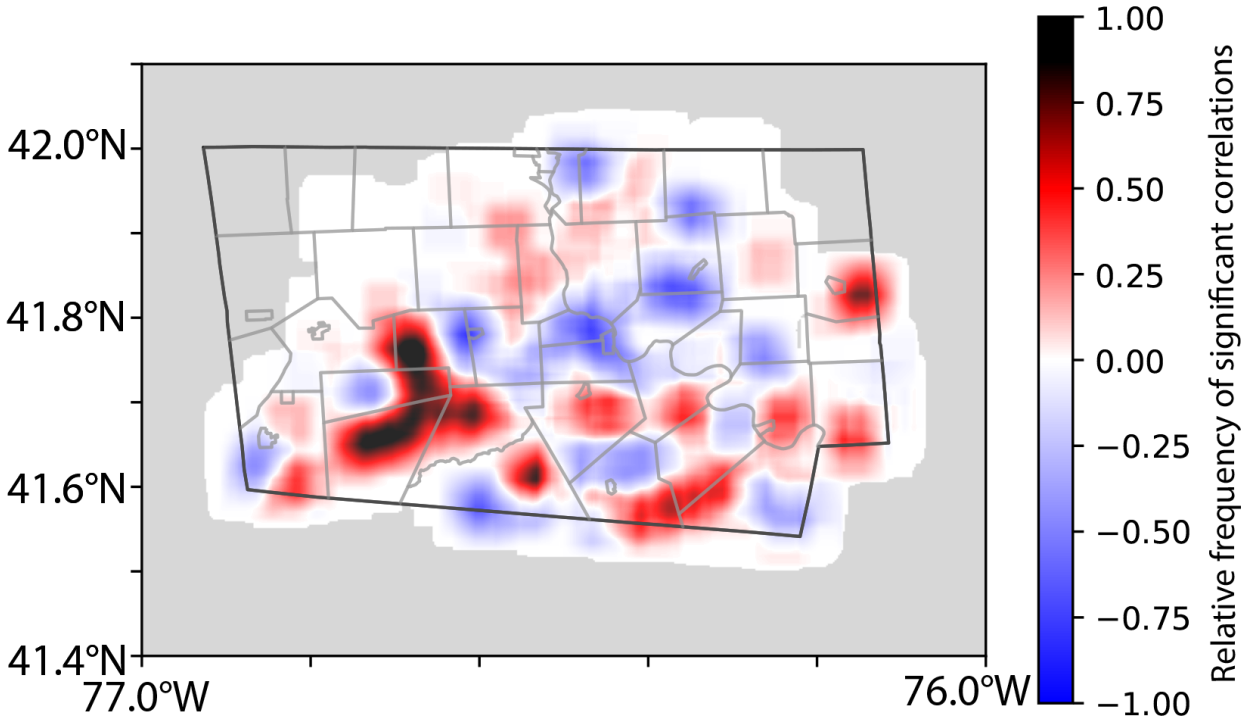

Figure S12. Sliding window heatmap displaying correlations between chloride concentrations and the density of unconventional wells in Bradford County. See Figure S3 for an explanation of the color scheme.

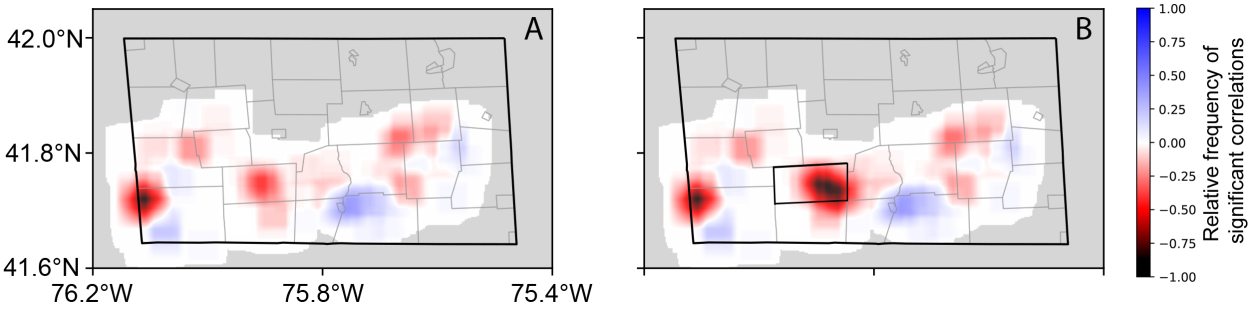

Figure S13. Sliding window heatmaps for correlations between methane concentrations and the distance to unconventional wells in Susquehanna County. Analysis without Dimock Township samples from Hammond (2016) is shown in (A), while analysis including Dimock samples are displayed in (B). The location of Dimock Township is outlined in black in (B).

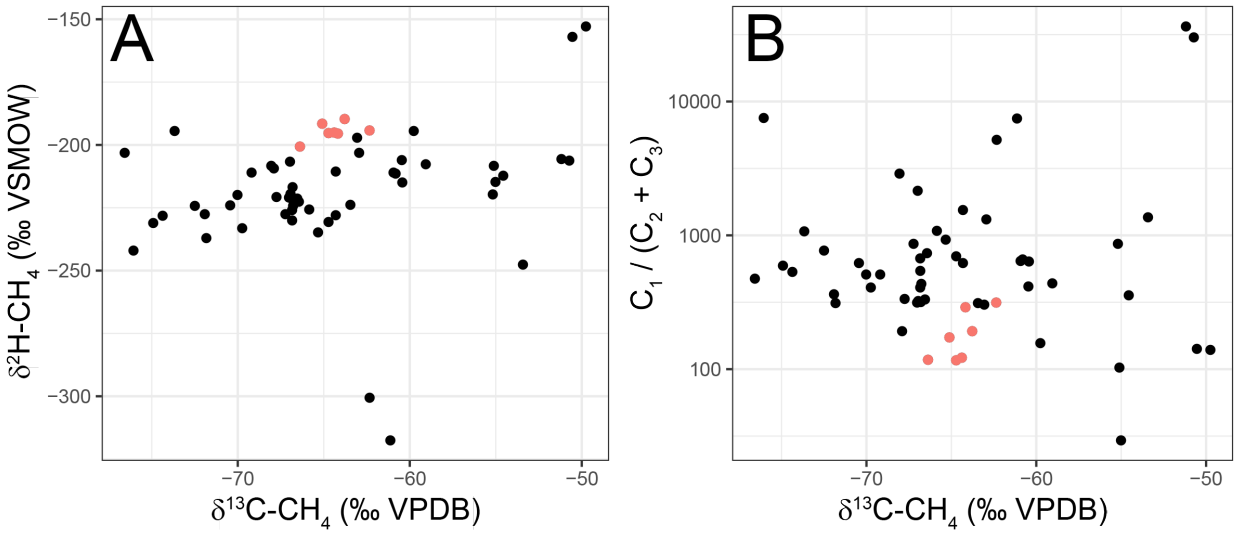

Figure S14. Methane carbon isotopic composition (δ^{13}C-CH_4) plotted against (A) ratios of methane (C_1) to ethane (C_2) + propane (C_3) or (B) methane hydrogen isotopic composition (δ^{2}H-CH_4). Samples located in the hotspot circled in Figure 3A are denoted in red, while samples outside the hotspot are shown in black.

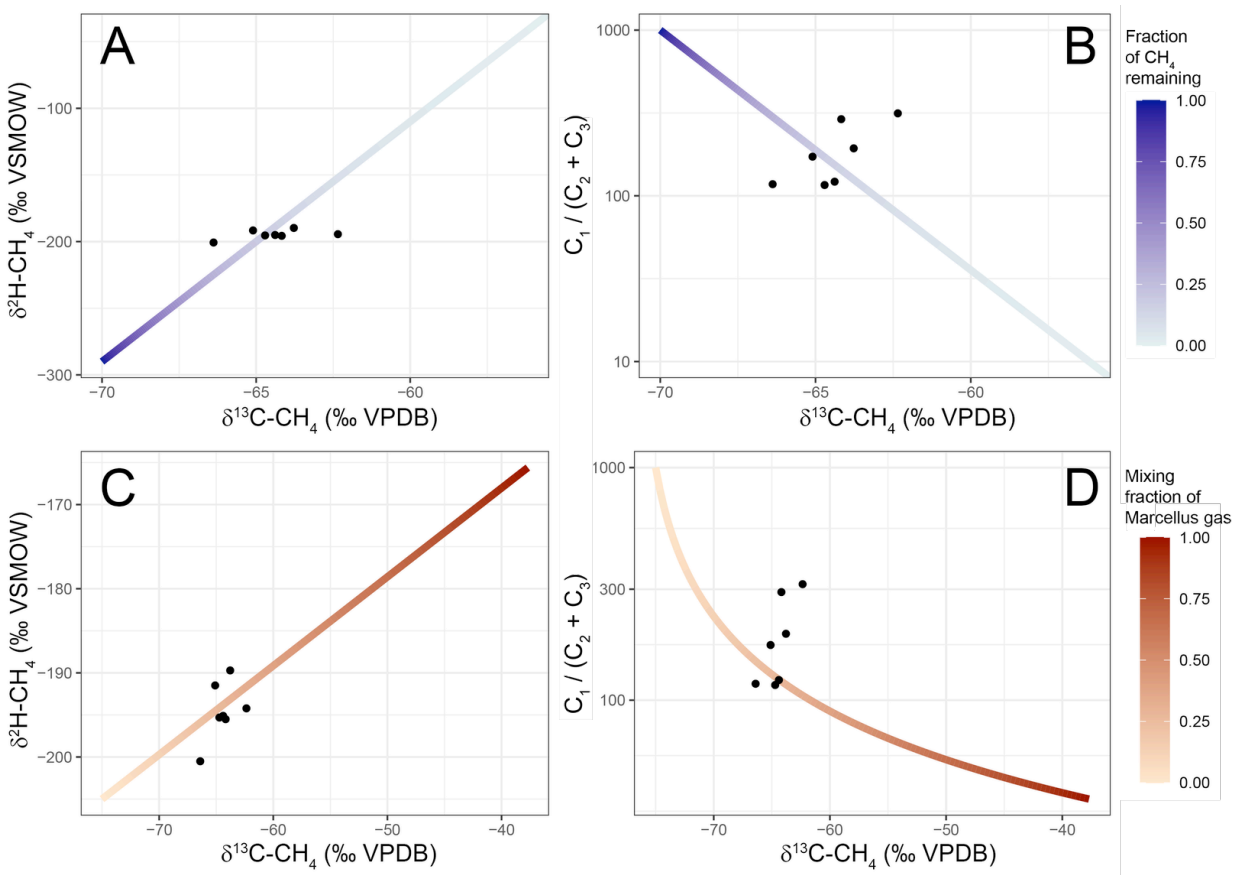

Figure S15. Gas chemistry (δ^{13}C-CH_4, δ^{2}H-CH_4, C_1 / (C_2 + C_3)) data from the hotspot circled in Figure 3A fit to models for the oxidation of biogenic methane (A & B) or the mixing of biogenic and Marcellus Shale gas (C & D). See Text S12 for an explanation of the compositions of endmember gases and model designs.

Supporting Tables

Table S1. Endmember sources of chloride and their respective molar ratios

Endmember	Ba / Cl	Ca / Cl	Mg / Cl	SO_4 / Cl	Na / Cl	Interpretation
1	0.0065 ± 0.0027	4.9 ± 2.0	0.6 ± 0.3	0	0.3 ± 0.3	Brine
2	0.0005 ± 0.0003	0.1 ± 0.9	0.3 ± 0.3	0.2 ± 0.3	3.8 ± 1.8	Road salt
3	0	9.1 ± 3.9	3.3 ± 1.4	3.0 ± 1.3	0.5 ± 0.4	Rain

Table S2. Summary of additional brine species (X) in SWPA produced water, the predicted increases in concentration (in mg/L) per UOGD well drilled, and the highest predicted concentration in hotspots relative to EPA limits.

Species	[X]/[Cl] ratio	[X] increase per UOG well	Highest [X] predicted in hotspots	EPA MCL/SMCL		% EPA limit at highest predicted [X]
TDS	2.22E+00	7.98E+00	9.58E+01	500	SMCL	19.16%
Ag	8.82E-05	3.17E-04	3.81E-03	0.1	SMCL	3.81%
Al	3.89E-04	1.40E-03	1.68E-02	0.05	SMCL	33.64%
As	1.76E-04	6.34E-04	7.61E-03	0.01	MCL	76.06%
B	3.84E-04	1.38E-03	1.66E-02			
Ba	1.76E-02	6.35E-02	7.62E-01	2	MCL	38.12%
Be	7.06E-05	2.54E-04	3.05E-03	0.004	MCL	76.26%
Br	1.01E-02	3.62E-02	4.34E-01			
Ca	1.48E-01	5.33E-01	6.39E+00			
Cd	8.81E-05	3.17E-04	3.81E-03	0.005	MCL	76.12%
Cl	1.00E+00	3.60E+00	4.32E+01	250	SMCL	17.28%
Co	5.38E-04	1.94E-03	2.32E-02			
Cr	8.77E-05	3.16E-04	3.79E-03	0.1	MCL	3.79%
Cs	3.32E-06	1.20E-05	1.44E-04			
Cu	2.18E-04	7.86E-04	9.43E-03	1	SMCL	0.94%
F	2.97E-04	1.07E-03	1.28E-02	2	SMCL	0.64%
Fe (total)	1.54E-03	5.54E-03	6.65E-02	0.3	SMCL	22.15%
Hg	1.61E-06	5.80E-06	6.96E-05	0.002	MCL	3.48%
I	3.13E-04	1.13E-03	1.35E-02			
K	2.51E-02	9.05E-02	1.09E+00			
Li	9.26E-04	3.33E-03	4.00E-02			
Mg	1.62E-02	5.84E-02	7.01E-01			
Mn	1.40E-04	5.05E-04	6.06E-03	0.05	SMCL	12.11%
Mo	7.02E-04	2.53E-03	3.03E-02			
NO_2	3.58E-04	1.29E-03	1.55E-02	1	MCL	1.55%
NO_3	8.47E-04	3.05E-03	3.66E-02	10	MCL	0.37%
NH_4	2.54E-02	9.13E-02	1.10E+00			
Na	4.11E-01	1.48E+00	1.78E+01			
Ni	4.49E-04	1.62E-03	1.94E-02	0.1	MCL	19.38%
PO_4	7.81E-04	2.81E-03	3.37E-02			
Pb	4.40E-05	1.58E-04	1.90E-03	0.015	Action Level	12.66%
Rb	1.10E-05	3.96E-05	4.75E-04			
S	1.02E-02	3.66E-02	4.40E-01			
SO_3	1.12E-01	4.04E-01	4.85E+00			
SO_4	1.83E-01	6.60E-01	7.92E+00	250	SMCL	3.17%
Sb	1.76E-04	6.35E-04	7.62E-03			
Se	8.83E-05	3.18E-04	3.81E-03	0.05	MCL	7.63%
Sr	2.30E-02	8.30E-02	9.96E-01			
Ti	8.78E-04	3.16E-03	3.80E-02			
Tl	1.78E-04	6.40E-04	7.68E-03	0.002	MCL	383.85%
Zn	4.53E-05	1.63E-04	1.96E-03	5	SMCL	0.04%
Alkalinity (as HCO_3)	1.44E-01	5.18E-01	6.22E+00			
DIC	1.16E-03	4.17E-03	5.00E-02			
DOC	3.84E-01	1.38E+00	1.66E+01			

Table S3. Correlations between [CH_4] or [Cl] and distance to/density of conventional oil & gas (COG) wells

Calculation	Distance to COG wells vs.		Density of COG wells vs.	
Species	[CH_4]*	[Cl]*	[CH_4]	[Cl]
Kendall's τ	-0.018	-0.019	0.010	0.010
p-value	0.019	0.015	0.201	0.214

* denotes statistically-significant correlations

Table S4. NMF-derived endmember sources of chloride using the Lautz et al. (2014) synthetic high-Cl dataset and their respective molar ratios used to interpret endmember identities.

End member	I / Cl	Na / Cl	K / Cl	Mg / Cl	Ca / Cl	Ba / Cl	Sr / Cl	Interpretation
1	9.1E-5 ± 1.9E-5	0.66 ± 0.14	0	0.11 ± 0.02	0.40 ± 0.08	0.0015 ± 0.0003	0.0073 ± 0.0016	Brine
2	7.2E-6 ± 3.0E-6	2.6 ± 1.1	0.029 ± 0.012	0	0.014 ± 0.005	0.0011 ± 0.0005	0.0031 ± 0.0012	Road salt
3	1.8E-5 ± 4.7E-6	0.0085 ± 0.0192	0.13 ± 0.03	0.99 ± 0.26	2.43 ± 0.65	0.00062 ± 0.00016	0.0034 ± 0.0009	Organic waste