Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction

Robert B. Jacksona,b,1, Avner Vengosh1, Thomas H. Darrah1, Nathaniel R. Warner1, Adrian Downa,b, Robert J. Poreda1, Stephen G. Osborn1, Kaiguang Zhaoa,b, and Jonathan D. Karra,b

*Division of Earth and Ocean Sciences, Nicholas School of the Environment and 1Center on Global Change, Duke University, Durham, NC 27708; 1Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY 14627; and 4Geological Sciences Department, California State Polytechnic University, Pomona, CA 91768

Edited by Susan E. Trumbore, Max Planck Institute for Biogeochemistry, Jena, Germany, and approved June 3, 2013 (received for review December 17, 2012)

Horizontal drilling and hydraulic fracturing are transforming energy production, but their potential environmental effects remain controversial. We analyzed 141 drinking water wells across the Appalachian Plateaus physiographic province of northeastern Pennsylvania, examining natural gas concentrations and isotopic signatures with proximity to shale gas wells. Methane was detected in 82% of drinking water samples, with average concentrations six times higher for homes <1 km from natural gas wells (P = 0.0006). Ethane was 23 times higher in homes <1 km from gas wells (P = 0.0013); propane was detected in 10 water wells, all within approx. 1 km distance (P = 0.01). Of three factors previously proposed to influence gas concentrations in shallow groundwater (distances to gas wells, valley bottoms, and the Appalachian Structural Front, a proxy for tectonic deformation), distance to gas wells was highly significant for methane concentrations (P = 0.007; multiple regression), whereas distances to valley bottoms and the Appalachian Structural Front were not significant (P = 0.27 and P = 0.11, respectively). Distance to gas wells was also the most significant factor for Pearson and Spearman correlation analyses (P < 0.01). For ethane concentrations, distance to gas wells was the only statistically significant factor (P < 0.005). Isotopic signatures (δ13C-CH4, δ13C-C2H6, and δ13C-C3H8), hydrocarbon ratios (methane to ethane and propane), and the ratio of the noble gas 4He to CH4 in groundwater were characteristic of a thermally postmature Marcellus-like source in some cases. Overall, our data suggest that some homeowners living <1 km from gas wells have drinking water contaminated with stray gases.

Unconventional sources of gas and oil are transforming energy supplies in the United States (1, 2). Horizontal drilling and hydraulic fracturing are driving this transformation, with shale gas and other unconventional sources now yielding more than one-half of all US natural gas supply. In January of 2013, for instance, the daily production of methane (CH4) in the United States rose to ~2 × 1012 m3, up 30% from the beginning of 2005 (3).

Along with the benefits of rising shale gas extraction, public concerns about the environmental consequences of hydraulic fracturing and horizontal drilling are also growing (4, 5). These concerns include changes in air quality (6), human health effects for workers and people living near well pads (5), induced seismicity (7), and controversy over the greenhouse gas balance (8, 9). Perhaps the biggest health concern remains the potential for drinking water contamination from fracturing fluids, natural formation waters, and stray gases (4, 10–12).

Despite public concerns over possible water contamination, only a few studies have examined drinking water quality related to shale gas extraction (4, 11, 13). Working in the Marcellus region of Pennsylvania, we published peer-reviewed studies of the issue, finding no evidence for increased concentrations of salts, metals, or radioactivity in drinking water wells accompanying shale gas extraction (4, 11). We did find higher methane concentrations and less negative δ13C-CH4 signatures, consistent with a natural gas source, in water for homeowners living <1 km from shale gas wells (4). Here, we present a more extensive dataset for natural gas in shallow water wells in northeastern Pennsylvania, comparing the data with sources of thermogenic methane, biogenically derived methane, and methane found in natural seeps. We present comprehensive analyses for distance to gas wells and ethane and propane concentrations, two hydrocarbons that are not derived from biogenic activity and are associated only with thermogenic sources. Finally, we use extensive isotopic data [e.g., δ13C-CH4, δ4He/CH4, δ13C-C2H6, δ13C-C3H8, δ13C-dissolved inorganic carbon (δ13C-DIC), and δ13C-H2O] and helium analysis (4He/CH4) to distinguish among different sources for the gases observed (14–16).

Our study area (Figs. S1 and S2) is within the Appalachian Plateaus physiographic province (17, 18) and includes six counties in Pennsylvania (Bradford, Lackawanna, Sullivan, Susquehanna, Wayne, and Wyoming). We sampled 81 new drinking water wells from the three principle aquifers (Alluvium, Catskill, and Lock Haven) (Fig. S1) (11). We combined the data with results from 60 previously sampled wells in Pennsylvania (4) and included a few wells from the Genesee Formation in Otsego County of New York (4). The typical depth of drinking water wells in our study was 60–90 m (11). We also sampled a natural methane seep at Salt Springs State Park in Franklin Forks, Pennsylvania (N 41.91397, W 75.8663; Susquehanna County) to compare with drinking water from homes in our study, some located within a few kilometers of the spring.

Descriptions of the underlying geology, including the Marcellus Formation found 1,500–2,500 m underground, are presented in refs. 4 and 11 and Figs. S1 and S2. Previous researchers have characterized the region’s geology and aquifers (19–23). Briefly, the two major bedrock aquifers are the Upper Devonian Catskill Formation, comprised primarily of a deltaic clastic wedge gray-green to gray, red sandstone, siltstone, and shale, and the underlying Lock Haven Formation, consisting of interbedded fine-grained sandstone, siltsite, and shale, and the underlying Alluvium Formation, consisting of interbedded fine-grained sandstone, siltsite, and silt shale (19, 22, 24). The two formations can be as deep as ~1,000 m in the study area and have been exploited elsewhere for oil and gas historically. The sedimentary sequences are gently folded and dip shallowly (1–3°) to the east and south (Fig. S2), creating alternating exposures of synclines and anticlines at the surface (17, 23, 25). These formations are over lain by the Alluvium aquifer, comprised of unconsolidated glacial till, alluvium sediments, and postglacial deposits found primarily in valley bottoms (20, 22).


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. E-mail: jackson@duke.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221635110/-/DCSupplemental.
Concentrations were greater and $C_{3}/C_1 < 0.0006$, Kruskal-Wallis test $P < 0.005$, regardless of distance to the Appalachian Structural Front (ASF; an index for the trend in increasing thermal maturity and degree of tectonic deformation) has been suggested to influence dissolved gas concentrations. Of these factors, distance to gas wells was the dominant statistical factor in our analyses for both methane ($P = 0.0007$) (Table 1, multiple regression analysis) and ethane ($P < 0.005$) (Table 1). In contrast, neither distance to the ASF ($P = 0.11$) nor distance to valley bottom streams ($P = 0.27$) was significant for methane concentrations analysis using linear regression. For single correlation factors, distance to gas wells was again the dominant statistical term ($P = 0.0003$ and $P = 0.001$ for Pearson and Spearman coefficients, respectively). Distance to the ASF was slightly significant by Pearson and Spearman correlation analyses ($P = 0.04$ and $P = 0.02$, respectively), whereas distance to valley bottom streams was slightly significant only for the nonparametric Spearman analysis ($P = 0.22$ for Pearson and $P = 0.01$ for Spearman) (Table 1). For observed ethane concentrations, distance to gas wells was the only factor in our dataset that was statistically significant ($P < 0.005$, regardless of whether analyzed by multiple regression, Pearson correlation, or Spearman analyses) (Table 1).

Along with distance to gas wells (4), proximity to both valley bottom streams (i.e., discharge areas) (28) and the Appalachian Structural Front (ASF; an index for the trend in increasing thermal maturity and degree of tectonic deformation) has been suggested to influence dissolved gas concentrations. Of these factors, distance to gas wells was the dominant statistical factor in our analyses for both methane ($P = 0.0007$) (Table 1, multiple regression analysis) and ethane ($P < 0.005$) (Table 1). In contrast, neither distance to the ASF ($P = 0.11$) nor distance to valley bottom streams ($P = 0.27$) was significant for methane concentrations analysis using linear regression. For single correlation factors, distance to gas wells was again the dominant statistical term ($P = 0.0003$ and $P = 0.001$ for Pearson and Spearman coefficients, respectively). Distance to the ASF was slightly significant by Pearson and Spearman correlation analyses ($P = 0.04$ and $P = 0.02$, respectively), whereas distance to valley bottom streams was slightly significant only for the nonparametric Spearman analysis ($P = 0.22$ for Pearson and $P = 0.01$ for Spearman) (Table 1). For observed ethane concentrations, distance to gas wells was the only factor in our dataset that was statistically significant ($P < 0.005$, regardless of whether analyzed by multiple regression, Pearson correlation, or Spearman analyses) (Table 1).

**Results and Discussion**

Dissolved methane was detected in the drinking water of 82% of the houses sampled (115 of 141). Methane concentrations in drinking water wells of homes <1 km from natural gas wells (59 of 141) were six times higher on average than concentrations for homes farther away ($P = 0.0006$, Kruskal–Wallis test) (Fig. 1 and Fig. S3). Of 12 houses where CH$_4$ concentrations were greater than 28 mg/L (the threshold for immediate remediation set by the US Department of the Interior), 11 houses were within 1-km distance of an active shale gas well (Fig. 1). The only exception was a home with a value of 32 mg CH$_4$/L at 1.4-km distance.

Similar to the results for methane, concentrations of ethane (C$_2$H$_6$) and propane (C$_3$H$_8$) were also higher in drinking water of homes near natural gas wells (Fig. 1). Ethane was detected in 40 of 133 homes (30%; 8 fewer homes were sampled for ethane and propane than for methane). Propane was detected in water wells in 10 of 133 homes, all approximately <1 km from a shale gas well ($P = 0.01$) (Fig. 1, Lower Inset). Ethane concentrations were 23 times higher on average for homes <1 km from a gas well: 0.18 compared with 0.008 mg C$_2$H$_6$/L ($P = 0.001$, Kruskal–Wallis). Seven of eight C$_2$H$_6$ concentrations >0.5 mg/L were found <1 km from a gas well (Fig. 1), with the eighth point only 1.1 km away (Fig. 1). Moreover, the higher ethane concentrations all occurred in groundwater with methane concentrations >15 mg/L ($P = 0.003$ for the regression of $C_2$ and $C_1$) (Fig. S4), although not all higher methane concentration waters had elevated ethane.

**Fig. 1.** Concentrations of (Upper) methane, (Lower) ethane, and (Lower Inset) propane (milligrams liter$^{-1}$) in drinking water wells vs. distance to natural gas wells (kilometers). The locations of natural gas wells were obtained from the Pennsylvania DEP and Pennsylvania Spatial Data Access databases (54). The gray band in Upper is the range for considering hazard mitigation recommended by the US Department of the Interior (10–28 mg CH$_4$/L); the department recommends immediate remediation for any value >28 mg CH$_4$/L.

Ratios of ethane to methane ($C_2/C_1$) and propane to methane ($C_3/C_1$) were much higher for homes within ∼1 km of natural gas wells (Fig. 2). Our high $C_3/C_1$ samples were also an order of magnitude greater than in salt-rich waters from a natural methane seep at the nearby Salt Springs State Park (mean [C$_3$/C$_1$] = 0.000029 and [C$_3$] = 0.0022 mg/L for the salt spring samples). Because microbes effectively do not produce ethane or propane in the subsurface (26, 27), our observed values within ∼1 km of drilling seem to rule out a biogenic methane source, and they are consistent with both wetter (higher C$_2$ + C$_3$ content) gases found in the Marcellus Formation and our earlier observation of methane in drinking water wells in the region (4).

Along with distance to gas wells (4), proximity to both valley bottom streams (i.e., discharge areas) (28) and the Appalachian Structural Front (ASF; an index for the trend in increasing thermal maturity and degree of tectonic deformation) has been suggested to influence dissolved gas concentrations. Of these factors, distance to gas wells was the dominant statistical factor in our analyses for both methane ($P = 0.0007$) (Table 1, multiple regression analysis) and ethane ($P < 0.005$) (Table 1). In contrast, neither distance to the ASF ($P = 0.11$) nor distance to valley bottom streams ($P = 0.27$) was significant for methane concentrations analysis using linear regression. For single correlation factors, distance to gas wells was again the dominant statistical term ($P = 0.0003$ and $P = 0.001$ for Pearson and Spearman coefficients, respectively). Distance to the ASF was slightly significant by Pearson and Spearman correlation analyses ($P = 0.04$ and $P = 0.02$, respectively), whereas distance to valley bottom streams was slightly significant only for the nonparametric Spearman analysis ($P = 0.22$ for Pearson and $P = 0.01$ for Spearman) (Table 1). For observed ethane concentrations, distance to gas wells was the only factor in our dataset that was statistically significant ($P < 0.005$, regardless of whether analyzed by multiple regression, Pearson correlation, or Spearman analyses) (Table 1).

**Fig. 2.** The ratio of ethane to methane ($C_2/C_1$) and (Inset) propane to methane ($C_3/C_1$) concentrations in drinking water wells as a function of distance to natural gas wells (kilometers). The data are plotted for all cases where [CH$_4$], [C$_2$H$_6$], and [C$_3$H$_8$] were above detection limits or [CH$_4$] was >0.5 mg/L but [C$_2$H$_6$] or [C$_3$H$_8$] was below detection limits using the detection limits of 0.0005 and 0.0001 mg/L for [C$_2$H$_6$] and [C$_3$H$_8$], respectively.
Isotopic signatures and gas ratios provide additional insight into the sources of gases in groundwater. Signatures of δ¹³C-CH₄ > −40‰ (reference to Vienna Pee Dee Belemnite standard) generally suggest a thermogenic origin for methane, whereas δ¹³C-CH₄ values < −60‰ suggest a biogenically derived methane source (27, 29, 30). Across our dataset, the most thermogenic δ¹³C-CH₄ signatures (i.e., most enriched in ¹³C) in drinking water were generally found in houses with elevated [CH₄] < 1 km from natural gas wells (Fig. 3A). In fact, all drinking water wells with methane concentrations >10 mg/L, the US Department of Interior’s threshold for considering remediation, have δ¹³C-CH₄ signatures consistent with thermogenic natural gas. Our data also show a population of homes near natural gas wells with water that has δ¹³C-CH₄ signatures that seem to be microbial in origin, specifically those homes shown in Fig. 3A, lower left corner. The combination of our δ¹³C-CH₄ (Fig. 3A) and δ²H-CH₄ data (Fig. 3B) overall, however, suggests that a subset of homes near natural gas wells has methane with a higher thermal maturity than homes farther away.

Analyses of δ¹³C-CH₄ and δ¹³C-C₂H₆ can help constrain potential sources of thermally mature natural gases (14, 15, 30). Because organic matter cracks to form oil and then natural gas, the gases initially are enriched in higher aliphatic hydrocarbons C₂ and C₃ (e.g., C₁ > C₂ > C₃; i.e., a relatively wet gas). With increasing thermal maturity, the heavier hydrocarbons are progressively broken down, increasing the C₂/C₃ ratio and leading to isotopic compositions that become increasingly heavier or enriched (31). In most natural gases, the isotopic composition (δ¹³C) of C₁ > C₂ > C₃ (i.e., δ¹³C of ethane is heavier than methane). In thermally mature black shales, however, this maturity trend reverses, creating diagnostic isotopic reversals in which the δ¹³C-CH₄ becomes heavier than δ¹³C-C₂H₆ (Δ¹³C = δ¹³C-CH₄ − δ¹³C-C₂H₆ > 1) (14, 15, 28, 30, 32).

For 11 drinking water samples in our dataset with sufficient ethane to analyze isotopic signatures, 11 samples were located <1.1 km from drilling, and 6 samples exhibited clear isotopic reversals similar to Marcellus production gases (Fig. 4). Conversely, five drinking water samples and spring water from Salt Springs State Park showed the more common trend consistent with Upper Devonian production gases (Fig. 4). In the study area, these isotopic values suggest multiple sources for hydrocarbon gases. The Upper Devonian gases are likely introduced into the shallow crust either by natural processes over geologic time or through leakage around the casing in the annular space of the production well. In contrast, natural gas with heavy δ¹³C-CH₄ and Δ¹³C > 0 likely stems from Marcellus production gases or a mixture of Marcellus gases and other annulus gases that migrated to the surface during drilling, well completion, or production.

Similar to our data, independent CH₄ measurements taken by the US Environmental Protection Agency (EPA) in Dimock, Pennsylvania (Residential Data Reports found at http://www.epaosc.org/site/doc_list.aspx?site_id=7555) in January of 2012 also show three δ¹³C-CH₄ values in drinking water wells between −24.98‰ and −29.36‰ δ¹³C-CH₄ and five samples with δ¹³C-CH₄ values in the range of Marcellus gas defined in ref. 28. The heaviest methane isotopic signatures in the EPA samples
Fig. 4. Stable isotope signatures (% VPDB) of methane ($\delta^{13}$C-$\text{CH}_4$) vs. $\delta^{13}$C for methane minus ethane ($\Delta^{13}$C = $\delta^{13}$C-$\text{CH}_4$ – $\delta^{13}$C-$\text{CH}_2$); 6 of 11 drinking water samples exhibited isotopic reversals and $\delta^{13}$C-$\text{CH}_4$ values consistent with Marcellus production gas (14, 28, 55). In contrast, five drinking water samples and the salt spring at Salt Springs State Park (filled square) had $\delta^{13}$C-$\text{CH}_4$ and $\delta^{13}$C $<$ 0 consistent with Upper Devonian production gases (14, 55). Eleven drinking water samples had sufficient ethane concentrations for isotopic determinations. Ten of the samples were $<$1 km distance from shale gas wells, and one sample is at 1.1 km distance (the point in the lower left corner of the plot).

($-24.98\%_o \delta^{13}$C-$\text{CH}_4$) exceeded the values observed for ethane ($-31.2\%_o \delta^{13}$C-$\text{C}_2\text{H}_6$), an isotopic reversal ($\Delta^{13}$C = 6.22%), characteristic of Marcellus or other deeper gas compared with gases from Upper Devonian sequences (14, 28).

Helium is an inert noble gas with a radiogenic isotope, $^{4}$He, that is a major component of thermogenic natural gas. Similar to hydrocarbon components, the abundance and isotopic composition of helium can help distinguish between potential sources and/or residence times of fluids in the crust, including natural gases (15, 16, 33). Across our dataset, the ratio of $^{3}$He/$\text{CH}_4$ in most drinking water wells showed a typical range between $\sim 2 \times 10^{-3}$ and $1 \times 10^{-2}$, independent of distance to natural gas wells (Fig. 5). In contrast, a subset of points with elevated [CH$_4$] has a $^{3}$He/$\text{CH}_4$ ratio significantly below the range established for shallow drinking water in the region and consistent with a mixture between shallow groundwater and Marcellus production gases there ($\sim 2 - 5 \times 10^{-4}$) (Fig. 5) (15).

The relative proportions of methane to higher-chain hydrocarbons, such as ethane and propane, can also be used to help differentiate biogenically and thermogenically derived methane as well as different thermogenic sources of natural gas (34). As described above, low ratios of methane to higher-chain hydrocarbons ($< 100$) in water typically suggest a hydrocarbon gas derived from a thermogenic source, whereas ratios of methane to higher-chain hydrocarbons $> 1,000$ suggest a microbial origin for the gas (27). Across our hydrocarbon dataset, $\sim 15$ samples seem to fall within the range corresponding to thermogenic gas, whereas the composition of 5 or 6 samples seems to be microbial in origin (Fig. 3C). These other points fall on two intermediate trajectories. One trajectory is simple mixing between thermogenically and biogenically derived gas (lower curve in Fig. 3C). The other trajectory reflects either diffusive migration or a more complex, three-component mixture between Middle and Upper Devonian gases and shallow biogenic sources (30, 35) (upper trajectory in Fig. 3C).

The relative distribution of ethane and propane provides additional insight into the source and mixture of gases. The ratio of propane to methane concentrations plotted against [C$_2$H$_6$] (Fig. 5S) shows that at least 6 of 10 water samples with detectable [C$_2$H$_6$] had an order of magnitude greater [C$_2$]/[C$_4$] ratio and [C$_2$] content than spring water from the natural methane seep at the Salt Springs State Park. The salt spring is the only location for which we found detectable [C$_3$] outside of our 11 samples (mean [C$_3$]/[C$_1$] = 0.000029 and [C$_3$] = 0.0022 mg/L for the Salt Springs samples) (Fig. 5S).

The abundance and relative proportions of aliphatic hydrocarbons (i.e., propane and ethane) and methane in groundwater are also useful for comparing with production gases (14, 36) and samples from the Salt Springs State Park. Ratios of propane to ethane (C$_3$/C$_2$) in our dataset were generally higher than ratios for the Salt Springs State Park, and ratios of methane to ethane (C$_1$/C$_2$) were generally lower (Fig. 5S), approaching ratios for Marcellus gases in some cases (Fig. 5S). We also observed that the highest methane concentrations coincided with increased abundances of ethane and propane and a higher proportion of propane relative to ethane (Fig. 5S). The observed gas composition in groundwater samples also had a substantially higher proportion of propane relative to ethane than water from the Salt Springs State Park, which is known to have historic methane-rich discharges (11, 37) (Fig. 5S). Based on limited available production data, the Marcellus production gases have a wetness (C$_3$/C$_2$) of at least 1–2% and C$_3$/C$_2$ of $> 0.03\%$, whereas Upper Devonian gases, specifically those gases observed in Upper Devonian aquifers before shale gas development (30), tend to be relatively depleted in wetter gases; samples from the Salt Springs State Park had intermediate wetness, which is discussed above (14, 30). As a result, increasing proportions of C$_3$/C$_2$ tend to be more representative of gases from Marcellus-producing wells (Fig. 5S) than Upper Devonian Formations or Salt Springs State Park.

An enrichment of $^{13}$C in DIC (e.g., $\delta^{13}$C-DIC > +10‰) and positive correlations between $\delta^{13}$C-DIC and $\delta^{13}$C-$\text{CH}_4$ and between $^{3}$H-$\text{H}_2$O and $^{3}$H-$\text{CH}_4$ have all been used as indicators of microbial methane sourced from relatively shallow depths ($< 500$ m) (38, 39). Most of our $\delta^{13}$C-DIC values were $20–25\%$ lighter (more negative) than typical for DIC influenced by microbially derived methane in shallow groundwater, and the $\delta^{13}$C-$\text{CH}_4$ values of the samples showed no evidence of a positive relationship with $\delta^{13}$C-DIC (and even a slight negative relationship; $P = 0.003$) (Fig. 5S, Upper). We also found no statistical relationship between the $^{3}$H values of methane and $^{3}$H of water (Fig. 5S, Lower). Based on these data and similar to the observations in the work by Osborn et al. (4), most of the methane in our samples does not.

Fig. 5. The ratio of $^{3}$He/$\text{CH}_4$ concentrations in drinking water wells vs. distance to gas wells (kilometers). The values are compared with water samples (mean ± SE) from the salt spring at Salt Springs State Park ($n = 3$) and Marcellus ($n = 4$) and Upper Devonian ($n = 5$) production gases (15).
appear to be derived locally in the shallow aquifers, and the gas composition is not consistent with extensive microbial production from methanogenesis or sulfate reduction. Methanotrophy also does not seem to be occurring broadly across our dataset; it would decrease $[\text{CH}_4]$ and $C_1-C_2$ ratios and increase $\delta^{13}\text{CH}_4$ values, reducing the differences that we observed for distance to gas wells. Overall, the combined results suggest that natural gas, derived at least in part from thermogenic sources consistent with a Middle Devonian origin, is present in some of the shallow water wells <1 km away from natural gas wells.

The two simplest explanations for the higher dissolved gas concentrations that we observed in drinking water are (i) faulty or inadequate steel casings, which are designed to keep the gas and any water inside the well from leaking into the environment, and (ii) imperfections in the cement sealing of the annulus or gaps between casings and rock that keep fluids from moving up the outside of the well (4, 40–42). In 2010, the Pennsylvania Department of Environmental Protection (DEP) issued 90 violations for faulty casing and cementing on 64 Marcellus shale gas wells; 119 similar violations were issued in 2011.

Distinguishing between the two mechanisms is important because of the different contamination to be expected through time. Casing leaks can arise from poor thread connections, corrosion, thermal stress cracking, and other causes (43). If the protective casing breaks or leaks, then stray gases could be the first sign of contamination, with less mobile salts and metals from formation waters or chemicals from fracturing fluids potentially coming later. In contrast, faulty cement can allow methane and other gases from intermediate layers to flow into, up, and out of the annulus into shallow drinking water layers. In such a scenario, the geochemical and isotopic compositions of stray gas contamination would not necessarily match the target shale gas, and no fracturing chemicals or deep formation waters would be expected, because a direct connection to the deepest layers does not exist; also, such waters are unlikely to migrate upward. Comprehensive analyses of well integrity have shown that sustained casing pressure from annular gas flow is common. A comprehensive analysis of ~15,500 oil and gas wells (43) showed that 12% of all wells drilled in the outer continental shelf area of the Gulf of Mexico had sustained casing pressure within 1 y of drilling, and 50–60% of the wells had it from 15 y onward. For our dataset, there is a weak trend to higher methane concentrations with increasing age of the gas wells and 0.067 for $[\text{CH}_4]$ vs. time since initial drilling). This result could mean that the number of drinking water problems may grow with time or that drilling practices are improving with time; more research is needed before firm conclusions can be drawn.

In addition to well integrity associated with casings or cementing, two other potential mechanisms for contamination by hydraulic fracturing/horizontal drilling include enhancing deep-to-shallow hydraulic connections and intersecting abandoned oil and gas wells. Horizontal drilling and hydraulic fracturing can stimulate fractures or mineralized veins, increasing secondary hydraulic connectivity. The upward transport of gases is theoretically possible, including pressure-driven flow through open, dry fractures and pressure-driven buoyancy of gas bubbles in aquifers and water-filled fractures (44, 45). Reduced pressures after the fracturing activities could also lead to methane exsolving rapidly from solution (46). If methane were to reach an open fracture pathway, however, the gas should redissolve into capillary-bound water and/or formation water, especially at the lithostatic and hydrostatic pressures present at Marcellus depths. Legacy or abandoned oil and gas wells (and even abandoned water wells) are another potential path for rapid fluid transport. In 2000, the Pennsylvania DEP estimated that it had records for only 141,000 of 325,000 oil and gas wells drilled historically in the state, leaving the status and location of ~184,000 abandoned wells unknown (47). However, historical drilling activity is minimal in our study area of north-eastern Pennsylvania, making this mechanism unlikely there.

This study examined natural gas composition of drinking water using concentration and isotope data for methane, ethane, propane, and $^3\text{He}$. Based on the spatial distribution of the hydrocarbons (Figs. 1 and 2), isotopic signatures for the gases (Figs. 3 and 4), wetness of the gases (Fig. 2 and Figs. S5, S6, and S7), and observed differences in $^3\text{He}:\text{CH}_4$ ratios (Fig. 5), we propose that a subset of homeowners has drinking water contaminated by drilling operations, likely through poor well construction. Future research and greater data disclosure could improve understanding of these issues in several ways. More research is needed across the Marcellus and other shale gas plays where the geological characteristics differ. For instance, a new study by Duke University and the US Geological Survey showed no evidence of drinking water contamination in a part of the Fayetteville Shale with a less fractured or tectonically deformed geology than the Marcellus and good confining layers above and below the drinking water layers (48). More extensive predrilling data would also be helpful. Additional isotopic tools and geochemical tracers are needed to determine the source and mechanisms of stray gas migration that we observed. For instance, a public database disclosing yearly gas compositions (molecular and isotopic $\delta^{13}\text{C}$ and $\delta^{18}\text{H}$ for methane and ethane) from each producing gas well would help identify and eliminate sources of stray gas (49). In cases where carbon and hydrogen isotopes may not distinguish deep Marcellus-derived methane from shallower, younger Devonian methane, the geochemical fingerprint of $^3\text{He}$ and other noble gases provides a promising approach (15, 50).

Another research need is a set of detailed case studies of water-quality measurements taken before, during, and after drilling and hydraulic fracturing. Such studies are underway, including partnerships of EPA- and Department of Energy-based scientists and industry in Pennsylvania, Texas, and North Dakota. In addition to predrilling data, disclosure of data from mud-log gases and wells to regulatory agencies and ideally, publicly would build knowledge and public confidence. Ultimately, we need to understand why, in some cases, shale gas extraction contaminates groundwater and how to keep it from happening elsewhere.

**Methods**

A total of 81 samples from drinking water wells were collected in six counties in Pennsylvania (Bradford, Lackawanna, Sullivan, Susquehanna, Wayne, and Wyoming). Field and lab results were validated with 60 previously reported data compiled by the work by Osborn et al. (4). The samples were obtained from homeowner associations and contacts with the goal of sampling Alluvium, Catskill, and Lock Haven groundwater wells across the region. For analyses of $^3\text{He}$ (Fig. 5), samples from 30 drinking water wells were used to estimate concentration ratios of $^3\text{He}:\text{CH}_4$. Wells were purged to remove stagnant water and then measured for $pH$, electrical conductivity, and $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values were recorded. Samples were collected upstream of any treatment systems and as close to the well water as possible, preserved in accordance with procedures detailed in SI Text, and returned immediately to Duke University for analyses. The chemical and isotopic ($\delta^{13}\text{C}-\text{DIC}$, $\delta^{18}\text{O}-\text{H}_2\text{O}$, and $\delta^{13}\text{O}-\text{D}_2\text{O}$) compositions of the collected waters were measured at Duke University’s Environmental Stable Isotope Laboratory. Values of $\delta^{13}\text{O}-\text{H}_2\text{O}$ and $\delta^{2}D-\text{H}_2\text{O}$ were measured using temperature conversion elemental analysis/continuous flow isotope ratio MS using a ThermoFinnigan temperature conversion elemental analyzer and Delta-XP mass spectrometer and normalized to Vienna Standard Mean Ocean Water (analytical precision of ± 0.1% for $\delta^{13}\text{O}$ and ± 1.5% for $\delta^{2}D$, respectively). Samples were collected in refrigeration-grade copper tubes flushed with water before sealing with stainless steel clamps and analyzed using a VG 5400 MS at the University of Rochester (15, 51).

Dissolved gas samples were collected in the field using procedures detailed by Isotex Laboratories (52), stored on ice until delivery to their facilities, and analyzed for concentrations and isotopic compositions of methane, ethane, and propane. Procedures for gas analyses are summarized in ref. 4. Isotex Laboratories uses chromatographic separation followed by combustion and dual-inlet isotope ratio MS to measure dissolved gas concentrations of $\delta^{13}\text{C}-\text{CH}_4$ and $\delta^{13}\text{C}-\text{C}_2\text{H}_6$, and detection limits for $C_2$, $C_3$, and $C_4$ are 0.001, 0.0005, and 0.0001 mol %, respectively. Dissolved $\text{CH}_4$ and $\delta^{13}\text{C}-\text{CH}_4$ were also determined by cavity-ring-down spectroscopy in the Duke Environmental Stable Isotope Laboratory on eight samples using a Picarro G2112i.
Dissolved [CH\textsubscript{4}] was equilibrated using a head-space equilibration method (S3) and diluted when necessary using zero air. A set of 33 groundwater samples with a range of [CH\textsubscript{4}] and \textdelta\textsuperscript{13}C-CH\textsubscript{4} were collected in duplicate and analyzed at both Duke University and Isotech Laboratories (Fig. S9). Hydrocarbon concentrations in groundwater were converted to milligrams [C\textsubscript{1}-L\textsuperscript{-1}] from a correlation with mol % (R\textsuperscript{2} = 0.95). As in refs. 4 and 11, the derived distances to gas wells represent planimetric lengths from sampling locations to nearest gas wells and do not account for the direction or extent of horizontal drilling underground. Distances to streams were determined as the shortest lengths from sampled locations to valley centerlines using the national stream network as the base map; distance to geologic strata was measured using GIS software. Statistical analyses were performed using MATLAB and R software.

ACKNOWLEDGMENTS. W. Chameides, the Jackson laboratory, and anonymous reviewers provided helpful suggestions on the work. We acknowledge financial support from the Nicholas School of the Environment and Center on Global Change and Fred and Alice Stanback to the Nicholas School. We thank William Chameides, Dean of the Nicholas School, for supporting this research.