Elevated levels of diesel range organic compounds in groundwater near Marcellus gas operations are derived from surface activities

Brian D. Drollettea, Kathrin Hoelzerb, Nathaniel R. Warnerc, Thomas H. Darrahd, Osman Karatumë, Megan P. O’Connore, Robert K. Nelsonf, Loretta A. Fernandezg, Christopher M. Reddyf Avner Vengoshh, Robert B. Jacksoni, Martin Elsnerb, Desiree L. Plataa

a Department of Chemical and Environmental Engineering, Yale University, New Haven, CT 06511
b Institute of Groundwater Ecology, Helmholtz Zentrum München, Germany
c Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802
d School of Earth Sciences, The Ohio State University, Columbus, OH 43210
e Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708
f Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543
g Departments of Civil and Environmental Engineering and Marine and Environmental Sciences, Northeastern University, Boston, MA 02115
h Division of Earth and Ocean Sciences, Duke University, Durham, NC 27708
i School of Earth, Energy, and Environmental Sciences, Woods Institute for the Environment, and Precourt Institute for Energy, Stanford University, Stanford, CA 94305

Corresponding Author: Desiree L. Plata; desiree.plata@yale.edu
Keywords: hydrophobic organic compounds, groundwater, high volume hydraulic fracturing, natural gas extraction, transport mechanisms
Abstract

Hundreds of organic chemicals are utilized during natural gas extraction via high volume hydraulic fracturing (HVHF). However, it is unclear if these chemicals, injected into deep shale horizons, reach shallow groundwater aquifers and impact local water quality, either from deep underground injection sites or from the surface or shallow subsurface. Here, we report detectable levels of organic compounds in shallow groundwater samples from private residential wells overlying the Marcellus Shale in northeastern Pennsylvania. Analyses of purgeable and extractable organic compounds from 64 groundwater samples revealed trace levels of volatile organic compounds, well below the Environmental Protection Agency’s maximum contaminant levels, and low levels of both gasoline range (GRO; 0-8 ppb) and diesel range organic compounds (DRO; 0-157 ppb). A compound-specific analysis revealed the presence of bis(2-ethylhexyl)phthalate, which is a disclosed HVHF additive, that was notably absent in a representative geogenic water sample and field blanks. Pairing these analyses with 1) inorganic chemical fingerprinting of deep saline groundwater, 2) characteristic noble gas isotopes, and 3) spatial relationships between active shale gas extraction wells and wells with disclosed environmental health and safety (EHS) violations, we differentiate between a chemical signature associated with naturally occurring saline groundwater and a one associated with alternative anthropogenic routes from the surface (e.g., accidental spills or leaks). The data support a transport mechanism of DRO to groundwater via accidental release of fracturing fluid chemicals derived from the surface rather than subsurface flow of these fluids from the underlying shale formation.
Significance Statement

Organic compounds found in drinking water aquifers above the Marcellus Shale and other shale plays could reflect natural geologic transport processes or contamination from anthropogenic activities, including enhanced natural gas production. Using analyses of organic compounds coupled with inorganic geochemical fingerprinting, estimates of groundwater residence time, and geospatial analyses of shale gas wells and disclosed safety violations, we determined that the dominant source of organic compounds to shallow aquifers was consistent with surface spills of disclosed chemical additives. There was no evidence of association with deeper brines or long-range migration of these compounds to the shallow aquifers. Encouragingly, drinking water sources impacted by disclosed surface spills could be targeted for treatment and monitoring to protect public health.

Introduction

Technological advances in high volume hydraulic fracturing (HVHF) have led to the expansion of unconventional fossil fuel extraction in the United States over the past decade [1-3]. In spite of the clear economic and national security benefits associated with domestic fuel production, the co-location of industrial practices with residential areas raises concerns for public and environmental health [4-6]. In particular, it is unclear if the organic chemicals that are used in relatively small proportions (but potentially large volumes) and injected into deep shale formations can contaminate shallow drinking-water aquifers. Several questions emerge: if organic chemicals are detected in groundwater, did they arrive via surface discharges, shallow subsurface pathways (e.g., leaking gas wells),
or deep transport routes? Furthermore, are organic compounds present in groundwater derived from naturally occurring, geogenic sources or associated with industrial activities, such as HVHF? Finally, what are the chemical fingerprints that enable one to make this distinction?

Although few studies have examined the occurrence and origin of organic contaminants in groundwater [7], the presence of light hydrocarbon gases (i.e., methane and ethane) and inorganic constituents has been investigated frequently. Osborn et al. (2011) and Jackson et al. (2013) demonstrated elevated methane levels within 1 km of unconventional gas wells over the Marcellus Shale [8, 9]. Further, Darrah et al. (2014) showed that stray gas contamination in a subset of groundwater wells likely resulted from poor well integrity (i.e., casing and cementing issues) [10]. In contrast, inorganic chemical constituents (e.g., Cl⁻ and Br⁻) in groundwater over the Marcellus Shale seem to reflect geogenic sources and provide evidence of hydraulic connectivity between shallow groundwaters and deeper formation brines on geological timescales in some areas in northeastern PA [11, 12]. This deep-origin, saline groundwater has a chemical and isotopic fingerprint similar to the Marcellus brines [11] but distinct from Marcellus flowback water [13]. However, it is unknown if this deep saline water carries a unique organic chemical fingerprint of either geogenic or anthropogenic origin.

The same mechanistic approaches taken to source apportion methane and inorganic compounds have not yet been applied to organic compounds in groundwater. While Gross et al. (2013) reported surface spills of hydraulic fracturing fluids and wastes that could impact groundwater with organic chemicals in Colorado [14], and the Pennsylvania Department of Environmental Protection (PA DEP) has assessed multiple
instances of local groundwater pollution by gas drilling operations [15], these studies relied on voluntary industry accident disclosure and did not probe for possible alternative exposure paths (i.e., through a broad geospatial sample set). In addition, they did not provide a detailed characterization of the water that could identify the dominant transport processes associated with the contamination (i.e., organic and inorganic markers, along with hydrocarbon composition, noble gas isotopes, and spatial distribution analysis).

Recently, Llewellyn et al. (2015) investigated a localized incident of stray gas groundwater pollution in Pennsylvania that exposed the potential for groundwater contamination from natural gas extraction practices, where the authors attributed the contamination to flow from HVHF wells through shallow subsurface pathways [7]. While critical and detailed, the targeted nature of this case study (i.e., sampling wells with documented contamination or close-to-contamination sites) precludes the identification of geologic transport mechanisms that may be occurring in the region. Thus, the mechanisms of organic chemical transport in groundwater associated with HVHF regions remain unclear.

To address this research gap, we sampled 64 private residential groundwater wells, ranging from 9 – 213m deep, over a three-year period (2012-2014) in northeastern Pennsylvania \((n = 62)\) and in southern New York \((n = 2)\) for analyses of gas chromatography (GC)-amenable organic compounds (Fig. 1). Fifty-nine samples were analyzed for volatile organic compounds (VOCs) and gasoline range organic compounds (GRO; defined as the hydrocarbons eluting between 2-methylpentane and 1,2,4-trimethylbenzene; approximately between \(nC_6\) and \(nC_{10}\)), and 41 were also analyzed for diesel range organic compounds (DRO; defined as the hydrocarbons eluting between...
Trace levels of GRO and DRO compounds were detected in 9 of 59 (0 – 8.8 ppb total GRO) and 23 of 41 (0 - 157.6 ppb total DRO) groundwater samples, respectively. While the highest concentrations of GRO and DRO were always detected within 1 km of active shale gas operations, this difference in concentration within 1 km ($n = 21$) and beyond 1 km ($n = 20$) from shale gas wells was only significantly higher in the case of DRO (Figure 2; $p = 0.01$, Mann-Whitney U test, SI Appendix Fig. S1). A sensitivity analysis indicated that the statistically significant difference in DRO levels persisted at 0.75 – 3.0 km (at 0.25 km intervals) from a gas well, as well as at 0.76 km away (the PA...
DEP’s suggested distance of evaluation; p < 0.01) [17]. Notably, although the highest GRO and highest DRO containing samples all occurred within 1 km of a shale gas well, the elevated GRO and DRO were not co-occurring (Figure 2, inset; discussed below). Finally, trace levels (<1 ppb) of VOCs, including BTEX compounds (benzene, toluene, ethylbenzene, and xylenes), were detected in six samples (10%) and well below the EPA’s drinking water maximum contaminant level (MCL) recommendations [18].

In this region, there are multiple potential sources of elevated DRO in groundwater, including: (a) upward migration of naturally occurring, formation derived organic compounds over geologic time, (b) lateral transport of drilling muds, flowback, or produced fluids from faulty wells, (c) leaking oil and gas waste containment ponds, (d) input of organic contaminants from surface spills of either raw chemicals or residual fracturing fluids, and (e) leaking underground storage containers or local traffic. To evaluate these sources systematically, we employed geochemical fingerprinting using inorganic constituents (i.e., Br/Cl ratios) [11], groundwater residence times (i.e., $^4$He concentration) [10], and dissolved methane concentrations [8, 9], coupled with our GRO and DRO analysis and geospatial analysis.

Upward Migration from Deep Formation. Warner et al. (2012) demonstrated that shallow groundwater in some areas in northeastern PA is saline with molar Br/Cl ratios similar to deeper Marcellus Formation water (designated “Type D” water), suggesting natural upward migration of deep saline water over geologic timescales [11]. Using this inorganic fingerprinting approach, we found no statistical difference in GRO or DRO contents based on water type (SI Appendix, Fig. S2; GRO, p > 0.05; DRO, p > 0.05; Kruskal-Wallis test), suggesting that the increased GRO and DRO signals were not a
result of upward migration of deeper naturally occurring formation fluids. Furthermore, samples with elevated GRO (>5 ppb), but lower DRO (<50 ppb), which might be considered geogenic and shale-derived considering the distinct transport rates of each (i.e., retarded transport of higher molecular weight compounds due to slower diffusion through and higher sorptivity to porous media) were not found in Type D waters uniformly (2 of 3 were not Type D).

Water migration from the Marcellus formation to shallow groundwater would also lead to significant enrichments in $^4$He and fractionation of air-saturated water noble gases (i.e., $^{20}$Ne/$^{36}$Ar) \[10\]. In contrast, we find that the highest concentration DRO and GRO samples occur in tritium-active groundwater (i.e., relatively young) and have the lowest $^4$He abundance (an integrated proxy for residence time, SI Appendix, Fig. S3) and, therefore, the lowest apparent crustal residence times \[10\]. This suggests that contamination occurs in the younger groundwater systems. Consequently, our results indicate that connectivity with deep subsurface brines is not a dominant source of organic compounds in the shallow groundwater.

**Lateral Transport from Faulty Wells.** An alternate source of organic compounds to shallow aquifers could be faulty gas well casings, as poor well integrity has been documented in gas wells targeting the Marcellus Shale \[6, 9, 10, 19, 20\]. Llewellyn et al. (2015) recently reported groundwater contamination of 2-$n$-butoxyethanol \[7\], while others documented stray gas contamination of light hydrocarbons (e.g., $nC_1$-$nC_3$) from poor well integrity \[8-10\]. Previous studies suggested that hydraulic fracturing fluids and denser formation brines could also migrate laterally, but on timescales longer than typical of methane transport \[5, 21, 22\]. Such migration through porous media from a well
casing would result in elevated GRO with lower levels of DRO (due to the higher
diffusivities and generally lower sorptivities of the lower molecular weight compounds),
along with higher levels of methane and a salinity signature similar to that of flowback or
produced waters. Methane abundance from paired samples or previous sampling
campaigns showed no correlation with GRO or DRO (SI Appendix, Fig. S4), and the
noble gas analysis provided no evidence for fugitive gas contamination in the elevated
GRO and DRO samples (e.g., low air-saturated water abundances ([36Ar], [N2]), or
4He/CH4 [10]). Furthermore, samples with elevated GRO (>5 ppb) had relatively low
methane and Br− (<1 ppm for both). Thus, leaky well casings are an unlikely source of
GRO compounds.

It is possible that a leaky well casing during slickwater injection could be a source
of elevated GRO or DRO without commensurate brine or methane inputs in a relatively
young well. If so, then one might expect some relationship between GRO or DRO
occurrence in groundwater and the age of the nearest HVHF well if and only if chemical
or bulk fluid transport times were fast relative to or on the same order as the well ages.
To entertain the possibility of a well age effect, we calculated time from the “spud date”
(the drilling date of the nearest HVHF well) to sample collection. Well ages ranged
broadly from 10 days to over 5 years with a fairly even temporal distribution, and levels
of GRO and DRO were not correlated with the age of the well (DRO, p > 0.05; GRO, p >
0.05; SI Appendix, Fig. S5).

Since the preceding well age argument relies on rapid fluid transport relative to
the well ages, we note that typical bulk groundwater velocities are highly variable in the
sampled aquifers and on the order of 0.1-8.2 km yr⁻¹ (spanning velocities in alluvium to
fractured bedrock aquifers) [23], and sorption-retarded transport velocities of the chemicals we detected (described below) would be on the order of 0.02-7.53 km yr\(^{-1}\) (approximately 2 mo - 50 yr to migrate a 1 km distance; see SI Appendix). However, depending on the topography, hydraulic connectivity, and large pressure gradients experienced during injection [5], transport times could be faster than predicted by simple porous media transport models. For example, Llwellyn et al. [7] argue that fracturing fluids could be driven 1-3 km in a 2-3 yr timeframe, which is reasonable for the fractured bedrock case. While it is not possible to put an exact timeframe on the fluid transport under the hypothetical condition of leaky casing during slickwater injection, two conditions emerge: either (1) chemical transport is slow and could not give rise to the elevated DRO compounds observed here (i.e., within 1 km and less than 2 years), or (2) transport is faster and a relationship between DRO and well age could have been observed, which it was not. In either case, our data suggest that leaky well is not a source of DRO to nearby groundwater wells.

Leaking Oil and Gas Waste Containment Ponds. Following hydraulic fracturing, the flowback and produced waters are often stored in polymer-lined, open waste containment pits, which are demonstrated sources of contamination to surface water and groundwater in cases where the liner integrity was compromised (e.g., torn, ripped, folded, or other failure due to a physical breach that allowed fluid to pass unrestricted) [24]. While many of these pits have been phased out voluntarily, many were still in use at the time of our study. Unfortunately, Pennsylvania does not maintain a publically available database of the location of the polymer-lined containment pits, and no spatial analysis between elevated DRO or GRO levels and containment pits is possible.
Nonetheless, these pits were designed to allow volatile compounds to outgas and particles to settle [25], and the residual wastes are often highly saline with a high organic content [24, 26]. Thus, leakage into groundwater from such containment basins would result in low GRO levels (due to volatile outgassing) and elevated DRO, such as observed in our samples (see SI Appendix for discussion of potential GRO/DRO fingerprints in groundwater and flowback water, SI Appendix, Fig. S8). Leaking from pits with compromised liners would give rise to elevated chloride and bromide in the high DRO samples, which was not observed (SI Appendix, Fig. S6). Therefore, diffusive transport of DRO through uncompromised liners could give rise to the observed chemical composition of the groundwater. However, the types of compounds revealed in our compound-specific analysis (detailed below) have very long transport times through model polymers characteristic of such liners. Considering the fastest-possible transport, the compound would not migrate through a 4-mm liner to the soil interface after 4 yr (only 2x10^{-27}% of the water-side content would migrate to 1 mm depth after 4 yr; model details provided in SI Appendix, Fig. S7). Over these long timescales, transport through intact pit liners could not have given rise the DRO observed in our samples. This implies that organic chemical transport through the liners was not the primary source of material in our samples.

Surface Spills of Hydraulic Fracturing Chemicals. State databases are maintained for disclosed releases of hydraulic fracturing components at the surface, as these present a direct route for surface water and groundwater contamination. Surface releases could result in low GRO due to volatilization within weeks [27, 28], with elevated levels of DRO leaching to groundwater. Such elevated DRO with low GRO was observed in our
samples, with higher DRO close to active shale gas wells (< 1km). In order to further evaluate the possible link between elevated DRO and proximity to the nearest well, we analyzed PA DEP violation reports. According to the PA DEP Oil and Gas Compliance Report, 5,791 violations were reported associated with 1,729 unconventional gas wells throughout the state between January 1, 2007 and June 1, 2014 [29] and classified as either “Administrative” or “Environmental Health & Safety” (EHS) violations. DRO concentrations were elevated significantly in groundwater samples in close proximity to EHS violations (p = 0.03, Spearman correlation; Figure 3), but GRO concentrations were not (p = 0.36). Furthermore, DRO occurrence in samples within 2 km of an unconventional well with an EHS violation (n = 20) was statistically higher than in samples further away than 2 km (n = 21; p = 0.03, Mann-Whitney U test), whereas GRO did not show the same relationship (p = 0.92). Neither DRO nor GRO levels were significant at the 1 or 0.76 km cutoff distances, perhaps due to the fact that the distribution of shale gas wells with an EHS violation is spatially diffuse compared to individual shale gas well locations.

Groundwater well depth could also provide information on the nature of the flow path of the compounds detected in our samples. For example, since vertical transport times are long, a deep or shallow source might give rise depth-dependent concentration gradient. There was no statistically significant difference between DRO or GRO concentrations in the shallowest (<100m) or deepest (>100m) sampled wells (DRO, p = 0.57; GRO, p = 0.89; Mann-Whitney U test), or at any other depth cutoff (50m, 75m, 125m), and neither GRO or DRO were correlated with depth (SI Appendix, Fig. S9). This could be an artifact of the scale and spatial resolution of the sampling effort. Designed to...
cover a large area (approximately 7,400 km$^2$) and constrained by well access, the groundwater samples were separated by widely varying lateral distances ($7 \pm 15$ and $11 \pm 18$ km for groundwater wells containing detectable DRO and GRO, respectively). As a result, any point source or spatially constrained “plume” of organic material could conceivably affect only a small population of groundwater wells, obfuscating any effect of well depth on the GRO or DRO concentration. Nevertheless, the samples with the highest GRO and DRO are found in groundwater wells less than 100m deep.

_Locking Underground Storage Tanks or Local Traffic._ We also explored the hypothesis that leaking underground storage tanks that typically contain gasoline, diesel, or fuel oil for both domestic and industrial use could provide a significant source of GRO and/or DRO. Leaking tank incident data obtained from PA DEP [30] showed no spatial correlations with DRO ($p = 0.95$, Spearman correlation) or GRO ($p = 0.81$) in the groundwater samples (SI Appendix, Fig. S10). In addition, the chemicals identified in the compound-specific analysis are not commonly stored in underground storage containers and are distinct from the chromatographic fingerprints of gasoline, diesel, or hydraulic fluids (SI Appendix, Fig S11). Indeed, were these materials present, they would be readily obvious via our analytical methods (detection limits near $100$ pg L$^{-1}$ or parts per quadrillion). Their absence implies that leaking underground storage tanks were not a source of material to the groundwater. Similarly, if local truck traffic were a source, then one might expect a distinct chemical fingerprint and correlation with distance to the nearest road. No such fingerprints (SI Appendix, Fig. S11) or correlations existed (DRO, $p = 0.78$; GRO, $p = 0.63$), suggesting that traffic was not responsible for the DRO observed in the studied groundwater.
Organic Chemical Fingerprinting via GC×GC-TOF-MS. In order to further evaluate the source of elevated DRO compounds, we conducted a compound-specific investigation using GC×GC with time-of-flight mass spectrometry (GC×GC-TOFMS). In particular, a subset ($n = 12$) of groundwater liquid-liquid extracts was analyzed, including those with the highest DRO levels, along with a field blank and one natural salt spring in Susquehanna County, PA, which is typically used to indicate the composition of natural gas and brine-rich fluids derived from shale rock sources in the northern Appalachian Basin [7-12]. The samples with the highest DRO ($n = 2$) contained bis(2-ethylhexyl) phthalate (Figure 4), whereas the salt spring, lower-DRO samples ($n = 8$), and field blank did not (SI Appendix, Fig. S12). Fatty acid phthalate esters (including bis(2-ethylhexyl) phthalate) are used in drilling and in hydraulic fracturing fluids [31], and bis(2-ethylhexyl) phthalate has been reported in Marcellus Shale, Barnett Shale, and Denver-Julesburg basin flowback waters [26, 31-33], in runoff and surface waters following an incident at a gas well site [34], and in multiple residential groundwater wells in Dimock, PA where the EPA has identified contamination directly from hydraulic fracturing operations (discerned by the timing onset of the contamination) [35].

Bis(2-ethylhexyl) phthalate is a ubiquitous chemical that is used in many industrial practices and materials, and it is difficult to attribute its presence solely to hydraulic fracturing activities. However, we present several lines of evidence that this particular phthalate is likely to be derived from HVHF activities. First, only our highest DRO samples contained bis(2-ethylhexyl) phthalate, suggesting that the compound was not derived from any step of our own analytical procedure. Second, if polyvinylchloride (PVC) pipes (known to contain phthalates and to be pervasive in water distribution
systems) were a source of the bis(2-ethylhexyl)phthalate, then one would expect a widespread presence in the analyzed samples \((n = 12)\). In contrast, it was only detected in the highest DRO samples \((n = 2)\). Third, compound-specific analysis of the natural salt spring did not contain the phthalate. Thus, the presence of bis(2-ethylhexyl) phthalate likely reflects its presence in the contamination source and is not an artifact of our sampling or preparation protocol.

Curiously, this particular phthalate has relatively low aqueous solubility. In chemical disclosure databases, bis(2-ethylhexyl) phthalate is reportedly used in “perfball” form (i.e., it is transported and injected as a solid). A solubilized form of a phthalate could be derived from surface spills of flowback or produced waters, or transport through containment pit liners. However, the former would carry a brine signature, which was not observed in the high DRO groundwater samples (SI Appendix, Fig. S6), and the latter has prohibitive transport timescales and could not give rise to the phthalate observed here (SI Appendix, Fig. S7). Consequently, our data suggests that some solubilized form of the phthalate (e.g., perfballs placed in a liquid carrier) is responsible for their appearance in shallow aquifers sampled in this study. Indeed, in all cases where the phthalate was detected, toluene was present as a co-contaminant. Further, since bis(2-ethylhexyl) phthalate is a disclosed additive in fracturing fluids, it is both (a) plausible that its presence in these samples is due to accidental surface releases of the parent fluids in the Marcellus region and (b) reasonable given our statistical spatial analysis using the disclosed spill database, as well as the complementary inorganic, methane, and helium abundance measurements. Nevertheless, one cannot rule out the possibility that the
bis(2-ethylhexyl) phthalate is derived from some non-HVHF source and just coincidentally correlated with proximity to disclosed HVHF EHS violations. Bis(2-ethylhexyl) phthalate is “reasonably anticipated to be a human carcinogen” [36]. Due to the analytical challenges of obtaining clean blanks and ubiquitous industrial use, the environmental fate of phthalates has been understudied since their presumable first appearances after the advent of plastics in the 1970s [37-40]. Detection of phthalates in environmentally derived samples, as well as their source apportionment, may have been overlooked out of fear of cross-contamination from other sources. However, if HVHF practices are employing phthalates (which are disclosed, but not with great frequency), the environmental geochemistry community is challenged to develop robust methods to track and source apportion these materials. Careful efforts to avoid contamination (i.e., the use of pre-combusted, all-glass or metal materials) and accountability for all other potential local industrial sources will be critical.

**Implications.** This is the first study of its kind to evaluate, on a regional scale, different possible mechanistic sources of organic compounds detected in drinking water wells in the Marcellus region using complementary inorganic chemical analyses and residence time approximations. Based on the evaluation of different possible mechanisms, our data are consistent with a surface-derived source of organic compounds in the study area, possibly from releases of hydraulic fracturing materials near drill sites. The emergent question arises: is the spill rate associated with unconventional shale gas development worse than any other industrial chemical or energy extraction activity? Unfortunately, a quantitative comparison can not be made due to the construct of the PA DEP disclosed violation reports, for which details are limited [41]. Often, ambiguous
language is used to describe the nature of the violation and volume estimates of reported releases are not provided. If volume data were available, an appropriate comparison of the environmental impacts of these releases could be calculated with the ratio of volume-of-spill to total-transported volume, as is done with crude oil (in 2014, 0.00007% of all oil transported by sea was spilled [42]). Incorporating volumetric data on releases from natural gas operations would allow direct comparisons to other energy industries. Clearly, such a report relies on accurate self-reporting or more robust monitoring [43]. Further, if the PA DEP required volume and chemical identity estimates, a more accurate assessment of the relative risks due to domestic energy extraction could be constructed.

Irrespective of the reporting nuances, it is clear that surface releases of fracturing fluids are usually accidental. Therefore, it is not necessarily the hydraulic fracturing process (i.e., the fluid injection) that can lead to groundwater contamination, but rather, the existence of the operation itself (i.e., the inherent risk associated with mechanical failure and human error in industrial practice). Domestic natural gas production necessitates co-location of residential areas with extraction facilities, and, like any industrial activity, the economic benefits come with some level of environmental and public health risk.

In summary, we show that some private residential groundwater wells contained trace concentrations of organic compounds (<200 ppb DRO) in close proximity to active shale gas wells and disclosed EHS violations. Surface sources are consistent with the presence of DRO compounds in groundwater with the lowest apparent groundwater residence times. We found no evidence for direct communication of deeper formation water or injected fracturing fluids with shallow drinking water wells due to upward
migration from shale horizons. This result is encouraging, as it implies there is some degree of temporal and spatial separation between injected fluids and drinking-water supply. However, shallow groundwater should be monitored over longer timescales [44] in areas of enhanced fracturing activities (e.g., where preferential faults could enhance deep-to-surface communication [45]). Future research should also focus on investigating chemical fingerprints of shale-derived organic matter via a careful comparison of raw fracturing fluids, flowback water, and geologic formation waters.

**Materials and Methods**

Shallow groundwater samples were collected in pre-combusted glass vials over three sampling campaigns from private residential groundwater wells. Wells were purged of stagnant water until stable readings of conductivity, pH, and temperature were recorded, upstream of any treatment system. The samples were fixed with acid, then stored on ice until analysis within 14-28 days. For the organic compound analysis, light hydrocarbons were analyzed using standard purging and preconcentration techniques (see SI Appendix for details) whereas heavier hydrocarbons were concentrated via liquid-liquid extraction into organic solvents. Compounds and compound classes were quantified via gas chromatography with flame ionization detection (GC-FID) and qualitatively identified with confirmed standards using GC-mass spectrometry (MS). A subset of liquid-liquid extracts was interrogated using comprehensive two-dimensional gas chromatography with time of flight mass spectrometry (GC×GC-TOF-MS). Inorganic constituents were analyzed by methods detailed in Warner et al. (2012) [11]. Methane was analyzed by methods detailed in Jackson et al. (2013) [9]. Noble gases were analyzed by methods detailed in Darrah et al. (2014) [10]. Maps and spatial data analysis were
prepared with ArcMap™ and all statistics were analyzed with the R statistical computing platform.

Acknowledgements. The authors thank Duke University’s Pratt School of Engineering and the National Science Foundation’s CBET Grant Number 1336702 and NSF EAGER (EAR-1249255) for financial support. Alissa White and Emma D’Ambro are thanked for their assistance and valuable discussion.


17. 25 Pa. Code § 38 Oil and Gas Wells.


Figure Legends

**Figure 1.** Shallow groundwater sample locations and the existing active shale gas wells at those times. Five samples were collected in December 2014 and included in the June 2014 data points. Shale gas well locations were obtained from the Pennsylvania Spatial Data Access.

**Figure 2.** Diesel range organic compounds (DRO, top) and gasoline range organic compounds (GRO, bottom) concentrations in shallow groundwater with respect to the distance from the nearest active shale gas well (black triangles) or gas well with an EHS violation (blue stars). DRO was significantly correlated ($p = 0.01$, Spearman correlation) with the distance to the nearest shale gas well and with the distance to the nearest EHS violation ($p = 0.03$). GRO was not correlated with distance to the nearest gas well ($p = 0.42$) or with the distance to the nearest EHS violation ($p = 0.36$). There was no correlation between GRO and DRO coming from the same sample (inset).

**Figure 3.** Locations of environmental health and safety (EHS) violations associated with unconventional gas well operations as reported by PA DEP Oil and Gas Reporting website [29].

**Figure 4.** GC×GC-TOF-MS extracted ion (m/z 41) chromatograms of two shallow groundwater samples (PAS311C and PAS310) that contained bis(2-ethylhexyl) phthalate (labeled as phthalate) and the natural salt spring that did not.
Groundwater Samples June 2014 (n = 25)
Groundwater Samples July 2013 (n = 16)
Groundwater Samples May 2012 (n = 23)
- Gas Wells July '13 - June '14
- Gas Wells May '12 - July '13
- Gas Wells pre-May '12
Groundwater Samples June 2014 (n = 25)
Groundwater Samples July 2013 (n = 16)
Groundwater Samples May 2012 (n = 23)
- Unconventional Gas Wells
- EHS Violations