Electronic Supplementary Information for:

Organic chemical contamination of groundwaters near hydraulic fracturing activities in northeastern Pennsylvania

[1]Boya Xiong, [2]Mario A. Soriano Jr., [2]Kristina M. Gutchess, [1]Nicholas Hoffman, [3]Cassandra J. Clark, [2]Helen G. Siegel, [1]Glen Andrew D De Vera, [1]Yunpo Li, [1]Rebecca J. Brenneis, [1]Austin J Cox, [3,4]Emma C. Ryan, [1]Andrew J. Sumner, [3]Nicole C. Deziel, [2]James E. Saiers, [1]Desiree L. Plata

[1] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

[2] School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut, United States

[3] Department of Environmental Health Sciences, School of Public Health, Yale University, New Haven, Connecticut, United States

[4] Tufts University, Department of Public Health and Community Medicine
136 Harrison Avenue, Boston MA 02111 USA

Participant Recruitment. The approach to recruiting participants, permissions and Institutional Review Board documentation are described in Clark *et al*.[1] and Li *et al*. [2]. Briefly, mailers were sent to all residents in pre-selected zip codes and participants voluntarily elected to participate in the study by calling to schedule a visit. All participants provided informed consent prior to water sample collection. The study protocol was approved by the Institutional Review Board of Yale University (HIC #2000021809) and reviewed and approved by the US Environmental Protection Agency (HSR-001162). The study area of Bradford County, PA was selected based on several criteria including the relatively high number of unconventional wells, low number of conventional wells, high proportion of domestic water wells, prior oil and gas-related contamination events The number of samples was a consequence of participation rates, ability to sample the region in one field season given constraints on cost and personnel, and a goal of approximately 100 for sufficient distribution of sampling sites.

Table S1. **Detection limits for Diesel range organic (DRO), gasoline range organic (GRO) and targeted volatile organic compounds.**

Compound	LOQ (ppb)	LOD (ppb)
DRO		0.94 ± 0.28
GRO*		0.12 ± 0.06
Chloromethane	0.05	
Vinyl chloride	0.05	
Bromomethane	0.08	
Chloroethane	0.09	
Trichlorofluoromethane	0.05	
Sum of 1,1 dichloroethene and trans 1,2 dichloroethene	0.09	
Methylene chloride (dichloromethane)	0.08	
1,1-Dichloroethane	0.12	
Sum of 2,2-Dichloropropane and cis 1,2-dichloroethene	0.05	
Bromochloromethane	0.17	
Chloroform	0.18	
1,1,1-Trichloroethane	0.13	
Sum of 1,1-Dichloropropene and carbon tetrachloride	0.12	
Sum of 1, 2-dichloroethane and benzene	0.06	

Trichloroethene	0.05
1,2-Dichloropropane	0.14
Bromodichloromethane	0.77
Dibromomethane	0.35
cis-1,3-Dichloropropene	0.19
Toluene	0.05
Trans-1,3-Dichloropropene	0.82
1,1,2-Trichloroethane	0.10
Tetrachloroethene	0.13
1,3-Dichloropropane	0.18
Dibromochloromethane	0.78
1,2-Dibromoethane (EDB)	0.77
Chlorobenzene	0.11
1,1,1,2-Tetrachloroethane	0.20
Ethylbenzene	0.05
Sum of m-Xylene and p-xylene	0.06
Sum of o-Xylene, styrene, and bromoform	0.05
1,1,2,2-Tetrachloroethane	0.48
Isopropylbenzene (cumene) and n-Propylbenzene	0.06
Bromobenzene	0.15
1,2,3-Trichloropropane	0.17
1,3,5-Trimethylbenzene	0.10
2-Chlorotoluene	0.05
4-Chlorotoluene	0.11
Tert-Butylbenzene	0.05
1,2,4-Trimethylbenzene	0.10
sec-Butylbenzene	0.05
1,3-Dichlorobenzene	0.12
p-Isopropyltoluene (p-Cymene)	0.11
1,4-Dichlorobenzene	0.12
n-Butylbenzene and 1,2-Dichlorobenzene	0.12
1,2-Dibromo-3-chloropropane	1.63
1,2,4-Trichlorobenzene	0.12

Hexachlorobutadiene	0.22
Naphthalene	0.11
1,2,3-Trichlorobenzene	0.13

*GRO is calibrated by the sum of 1,2-dichloroethane, benzene, toluene, ethylbenzene, m-xylene, p-xylene, isopropylbenzene, n-propylbenzene, 1,3,5-Trimethylbenzene, tert-butylbenzene, and 1,2,4-Trimethylbenzene. GRO includes all compounds that elute between 2-methylpentane and 1,2,4 trimethylbenzene.

Table S2. Provisional violations category assignment used in this study (adopted from Rahm *et al.* [3]).

Violation	Enforcement Code
Spills/Potential Spills	301, 402, 91.34A, 401CSL, 401CLS, 307CSL, 78.54, 691.WPD, 92.3, CSL201, 105.11, 205B, 78.56(a), 78.56(1), 78.54, 91.33A, 91.33B, 91.34(A), 91.34A,
Erosion/Potential Erosion	102.4, 102.11, 102.22, 102.4HQBMP, 78.53
Cementing, Casing	78.73A, 78.85, 79.12CW, 207B 78.73B, 78.81D1, 78.81D2, 78.83A, 78.83COALCSG, 78.83GRNDWTR, 78.84, 78.81(a)2
Impoundments	78.56PITCNST, 78.56LINER, 78.56FRBRD, 78.56(3), 78.57
Site Restoration	206C, 210IMPRPLUG, 210UNPLUG, 78.65, 206, 208A, 78.51(A), OGA3216(A), OGA3216(C), 206REST
Solid Waste	SWMA301, 601.101

Methodological details: GRO and VOCs were analyzed using an Agilent 7890B with a split/splitless injector, DB-624 column (Agilent Technologies, 60 m × 0.32 mm × 1.80 μm) and a flame ionization detector (FID). VOCs were calibrated using Restek MegaMix™ standard mixture (Restek 502.2) and CalMix 1(Restek 624). The GC oven was set to have an initial temperature of 40 °C with 6 min hold time, followed b 6 °C/min to 158 °C , 10 °C/min to 190 °C, and 20 °C/min to 240℃. DRO compounds were analyzed using a HP-5 column (30 m × 0.32 mm × 0.25 μm), and the GC oven was set to have an initial temperature of 40°C and 10 °C/min to 300 °C with 10 min hold time.

Table S3. Concentrations of seven most frequently detected gasoline range compounds detected were below Maximum Contaminant Levels determined by EPA.

Compound name	EPA Maximum Contaminant Level (ppb)	Maximum concentration detected (ppb)	Limit of Quantification (LOQ, ppb)	% Occurrence (based on LOQ)
Bromochloromethane	N/A	1.18	0.17	99
Trichloroethene	5	4.11	0.05	54
Chloroform	70	0.63	0.18	35
Toluene	1000	0.26	0.05	28
Bromomethane	N/A	0.29	0.08	26
1,2-Dichloroethane & Benzene*	5	0.10	0.06	20

N/A: not available

LOQ: limit of quantification. This limit of quantification is the average of 27 daily limit of quantification numbers.

*1,2-Dichloroethane and Benzene were co-eluting.

Some halogenated materials are disclosed for use in hydraulic fracturing, while others are thought to form through subsurface transformation processes. These are detailed in Hoelzer and Sumner *et al.* [4]and a series of articles by Sumner *et al*.[5-7].

Solid phase extraction (SPE) and liquid chromatography-mass spectrometry

Analytes were extracted using 1 g Bondesil PPL polymeric sorbents (Agilent, USA) packed in pre-combusted glass cartridges (Sigma-Aldrich, USA). The SPE cartridges were sequentially preconditioned using 6 mL of methanol and 6 mL of ultrapure water. The sample (10 mL) was then loaded onto the cartridge at 1 mL/min, after which the cartridges were rinsed with 15 mL ultrapure water and then dried under vacuum for 15 min. The cartridges were eluted with 1-2 mL methanol into pre-combusted glass vials.

An Agilent 1290 Infinity II liquid chromatography system equipped with 20 μL sample loop and Agilent ZORBAX Eclipse Plus C18 column (3x50mm, 1.8 μm) was used for the analysis. A linear gradient from 5 – 95% acetonitrile/50 mM ammonium formate (pH = 3.7) at a flow rate of 0.2 mL/min was applied over 20 min followed by a 3 min hold. Mass spectrometry was performed using an Agilent 6495 iFunnel triple quadrupole system. Spectral data were acquired in MS2 scan mode from 50 – 2000 m/z. Resulting total ion chromatograms were analyzed for prominent peaks using MassHunter Qualitative Analysis software. Three samples with DRO ranging from 31.36-101.58 ppb were analyzed using the method described above and no significant peak or mass spectra trace was found, suggesting limited hydrophilic compound present in the water sample with high DRO.

Inorganic constituents analysis

Inorganic compounds were analyzed from samples collected simultaneously with organic compounds. Major cation and Fe analyses were conducted using inductively coupled plasma - optical emission spectrometry (ICP-OES) at the Cary Institute for Environmental Sciences. Major anions were determined using ion chromatography (IC) and trace metals by inductively coupled plasma mass spectrometry (ICP-MS) at the Yale Analytical and Stable Isotope Center (YASIC). Dissolved inorganic content (DIC) and dissolved organic content (DOC) were measured using a Shimadzu TOC analyzer at Yale University. Alkalinity was calculated from DIC measurements using stable pH and temperature measurements

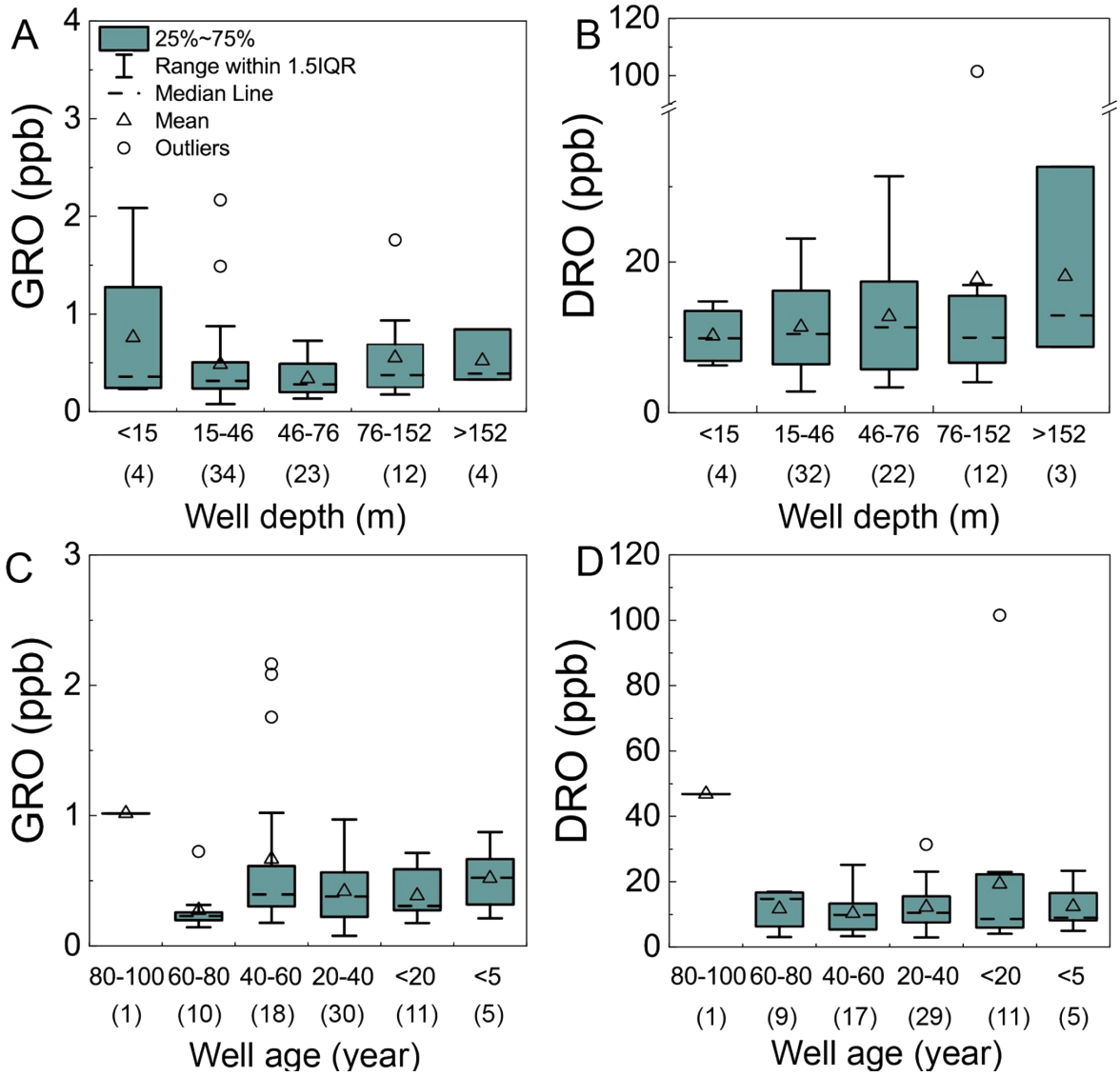

Figure S1. A) GRO and B) DRO with resident-reported drinking water well depth. C) GRO and D) DRO with self-reported groundwater well age. There are no statistically significant differences among the GRO or DRO levels across different well ages or depths (Well age: p= 0.077, GRO; p= 0.561, DRO; well depth: p= 0.289, GRO; p= 0.781, DRO, Kruskal-Wallis test).

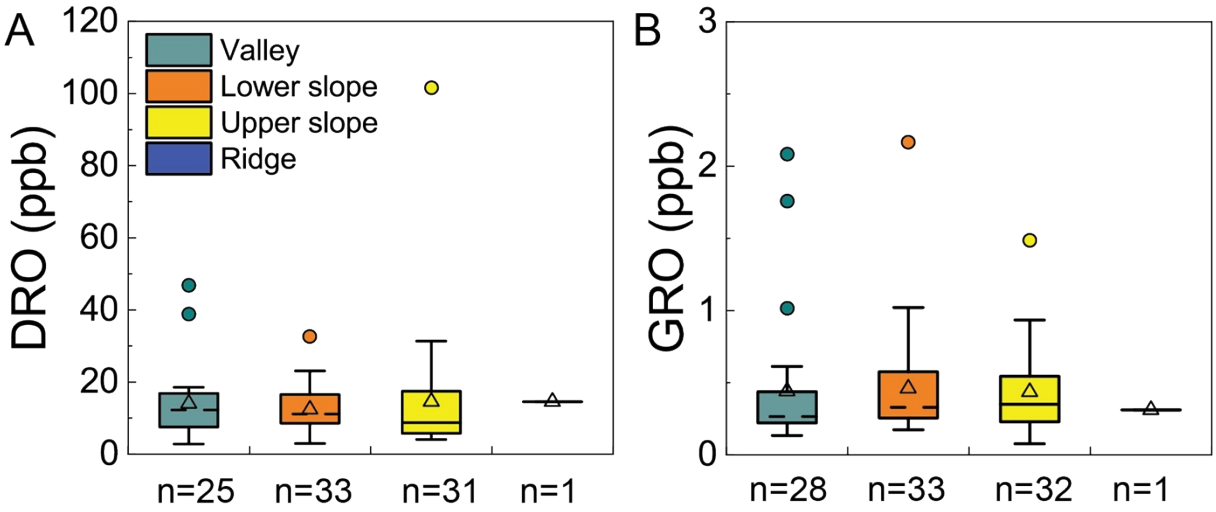

Figure S2. GRO or DRO are not correlated with the sampling-location topography (p= 0.525, GRO; p= 0.891, DRO, Kruskal-Wallis test)

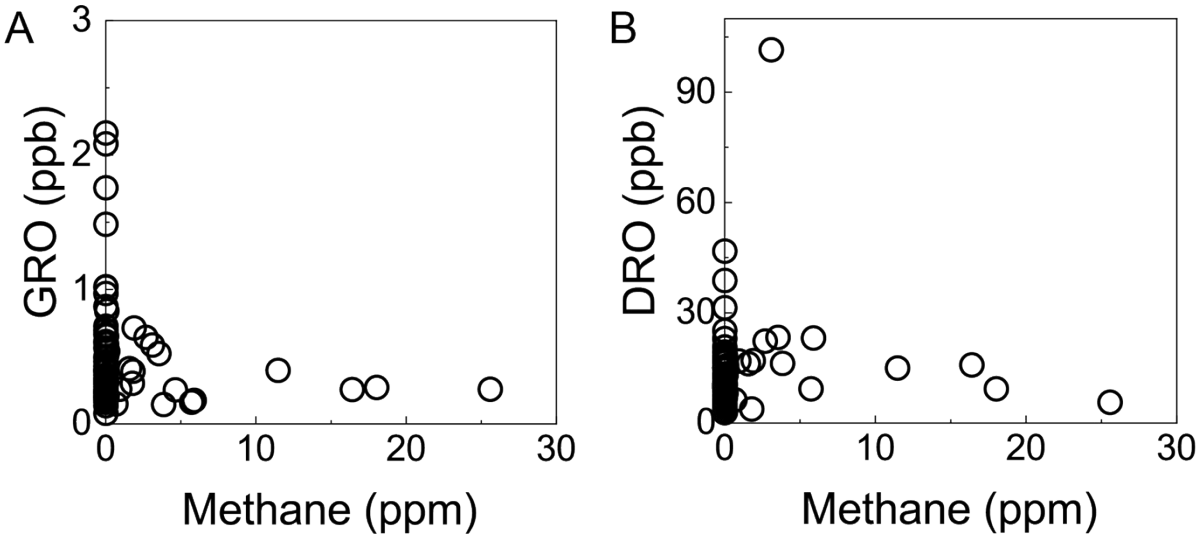

Figure S3. DRO have statistically significant correlation with methane level in the groundwater sample, and GRO do not have such correlation. (p= 0.901, ρ=0.0137, GRO; p= 0.029, ρ=0.242, DRO, Spearman correlation)

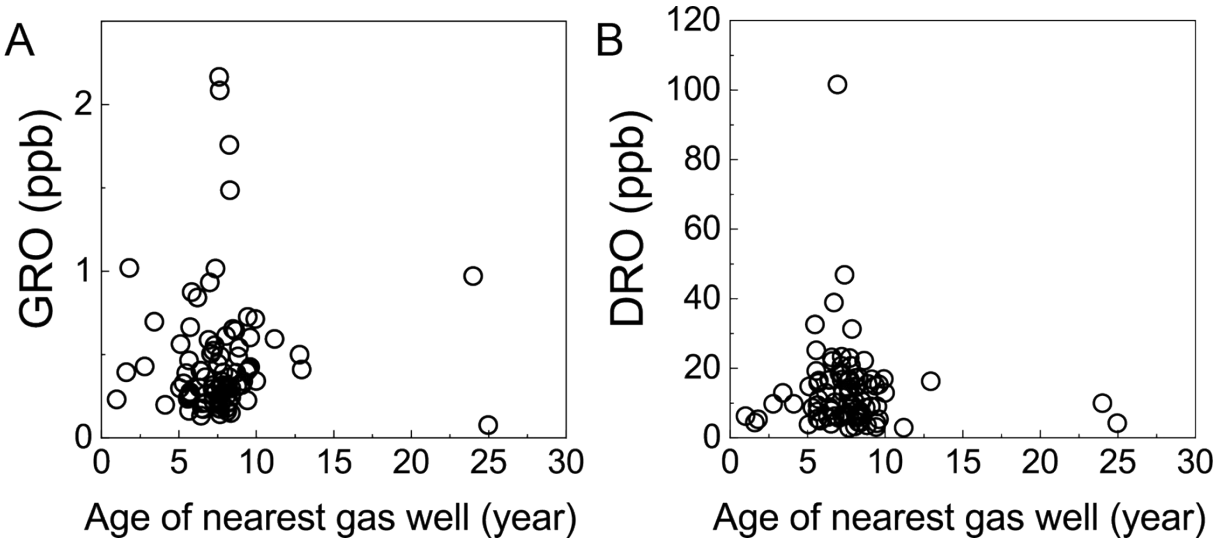

Figure S4. Days post drill of nearest oil and gas wells (0.1-4.8 km) to the drinking water wells with DRO and GRO levels. (p= 0.221, ρ=0.127, GRO; p= 0.356, ρ=-0.098, DRO, Spearman correlation)

Compound identification using GC×GC time-of-flight MS and LC Triple-Q MS

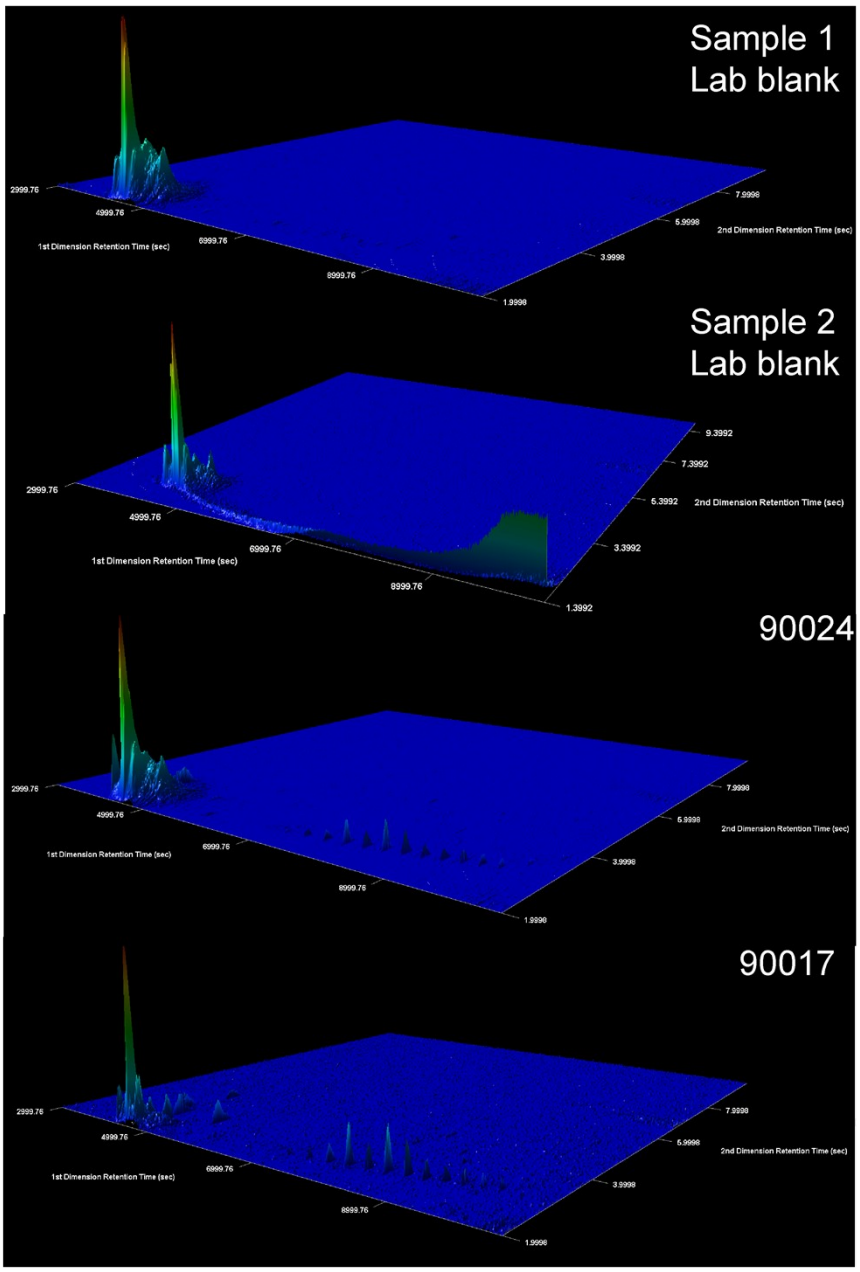

Figure S5. GC × GC time-of-flight MS extracted ion (*m/z* 41) chromatographs of two lab blanks that were extracted on the same date as sample 1 and 2 shown in Figure 3 (top), two field blanks that were collected at the same time as sample 1 and 2 (bottom). Note that the color map indicates the extracted ion intensity and each image is scaled to the same maximum height. This results in many small peaks in the foreground that are colored similarly to the baseline (i.e., royal blue) due to their sparingly small abundance.

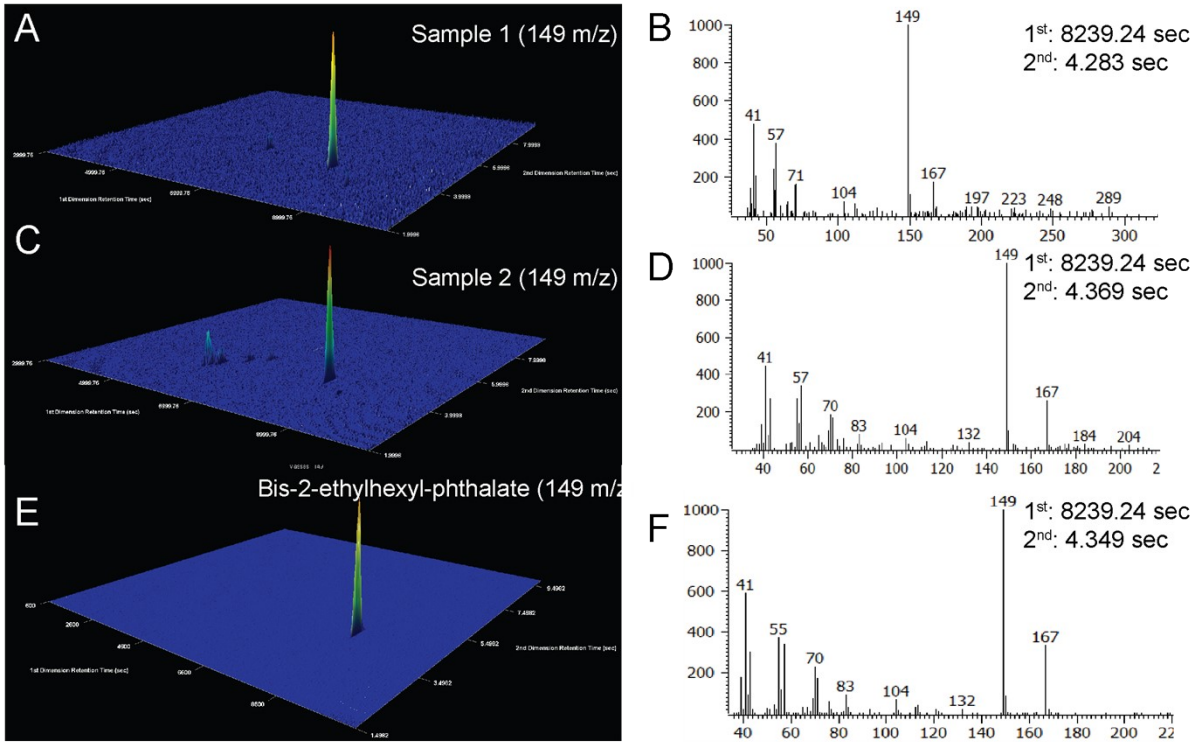

Figure S6. GC × GC time-of-flight MS extracted ion (*m/z* 149) chromatographs and ion spectra of sample 1 (A and B), 2 (C and D) and bis-2-ethylhexyl phthalate standard (E and F, Sigma Aldrich). Z-axis is scaled automatically to the highest intensity peak at given X and Y range. Top right corner of ion spectra indicates the retention time at the first and second dimension during GC. Note that the color map indicates the extracted ion intensity and each image is scaled to the same maximum height (where the maximum height is colored red). This results in many small peaks in the foreground that are colored similarly to the baseline (i.e., royal blue) due to their sparingly small abundance.

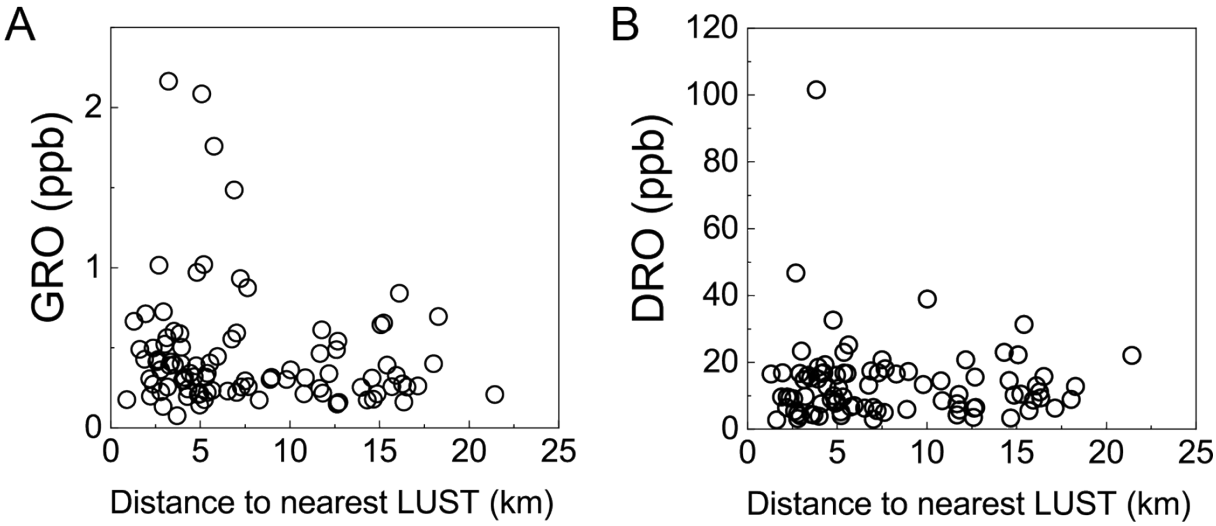

Figure S7. GRO level correlated significantly with the distance to nearest leaking underground storage tank, and DRO did not have such correlation. (p= 0.110, ρ=-0.165, GRO; p= 0.785, ρ=0.0291, DRO, Spearman correlation)

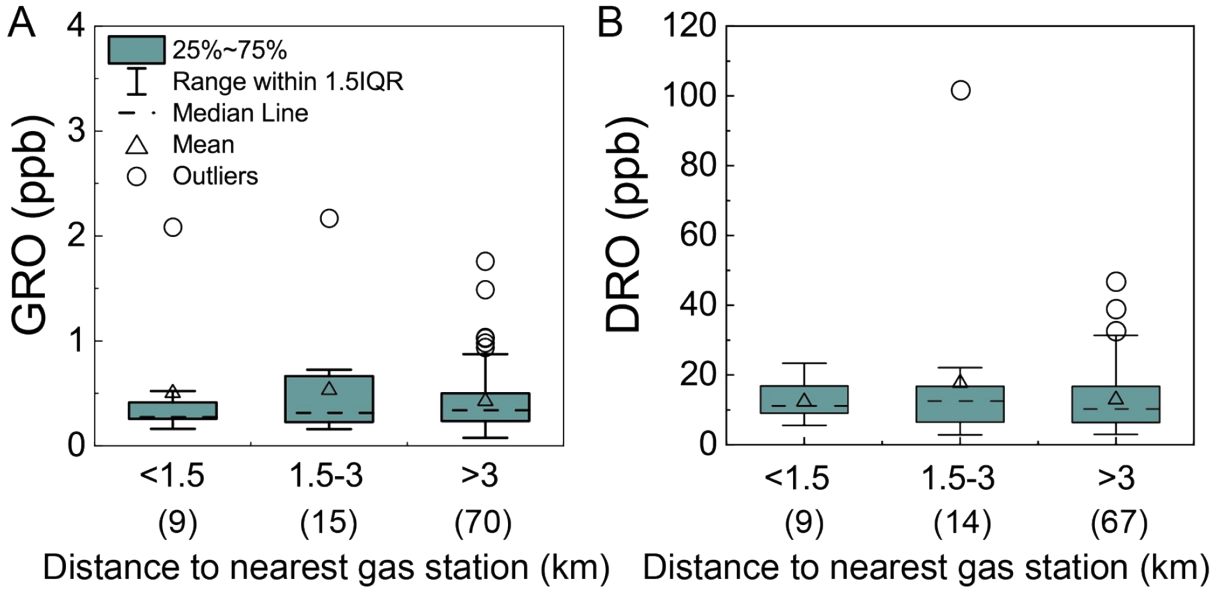

Figure S8. Elevated GRO and DRO are not correlated with the distance to nearest gas station, indicating the gas stations were not a major source of organic contaminants. The distance to nearest gas station was self-reported by homeowners. (p= 0.782, GRO; p= 0.950, DRO, Kruskal-Wallis test)

Groundwater model development

A three-dimensional flow and transport model for a 190-km^2 area in southeastern Bradford county was constructed to estimate the spatiotemporal distribution of hydraulic heads and groundwater velocities. A control volume finite element hydrologic simulator Hydrogeosphere[8] was used to solve the groundwater flow equation and the advection-dispersion equation for solute transport. The model domain was areally discretized into triangular elements with a maximum side length of 30 meters, and vertically discretized into 21 layers with a thickness ranging from 0.5 to 50 m. Finer horizontal discretization was used in proximity to wells, streams, and external boundaries. The model was parameterized using pilot points with a Tikhonov regularization scheme for preferred homogeneity[9, 10]. Model calibration was accomplished through a Gauss-Marquardt-Levenberg optimization scheme as implemented in PEST++[11]. The calibrated model successfully replicated hydraulic head observations extracted from the USGS National Water Information System[12] and the Pennsylvania Groundwater Information System databases[13] as well as groundwater discharge estimated from streamflow regression equations developed by Stuckey [14]. All computations were carried out at the Yale Center for Research Computing's GRACE high performance cluster.

Table S4. Parameters used in calculating the distribution of solute transport distance in **Fig. 5**. K_{oc} value was obtained by EPA-Estimation Program Interface Suite program. A lognormal distribution of q_w was estimated by fitting the lognormal probability density function to the 2 x 10^7 calculations of q_w by groundwater model domain. A maximum and minimum value of ρ_s, ϕ, and f_{oc} were available and obtained from various sources [15-18]. The retardation factor (10,000 values each) for acrylamide and phthalate were calculated separately using **Eqn 1** with random sampled values from uniform distributions of ϕ, ρ_b, f_{oc} and respected fixed K_{oc}.

Parameter	Range	Distribution
Specific discharge, q_w, km/yr	1.6 x 10^{-5} to 2.3 x 10^0 [a]	Lognormal[b]
Soil bulk density, ρ_b, kg/L	1.6 to 2.4	Uniform
Porosity, ϕ	0.08 to 0.25	Uniform
Organic carbon partition coefficient, L/kg		
$K_{oc,acrylamide}$	3.55	Constant
$K_{oc,bis\text{-}2\text{-}ethylhexyl\ phthalate}$	9.90 x 10^4	Constant
Fraction of organic carbon, f_{oc} (kg_{oc}/kg_{sed})	0.001 to 0.1	Uniform

[a] Determined from the calibrated groundwater model

[b] Fitted to the groundwater model output

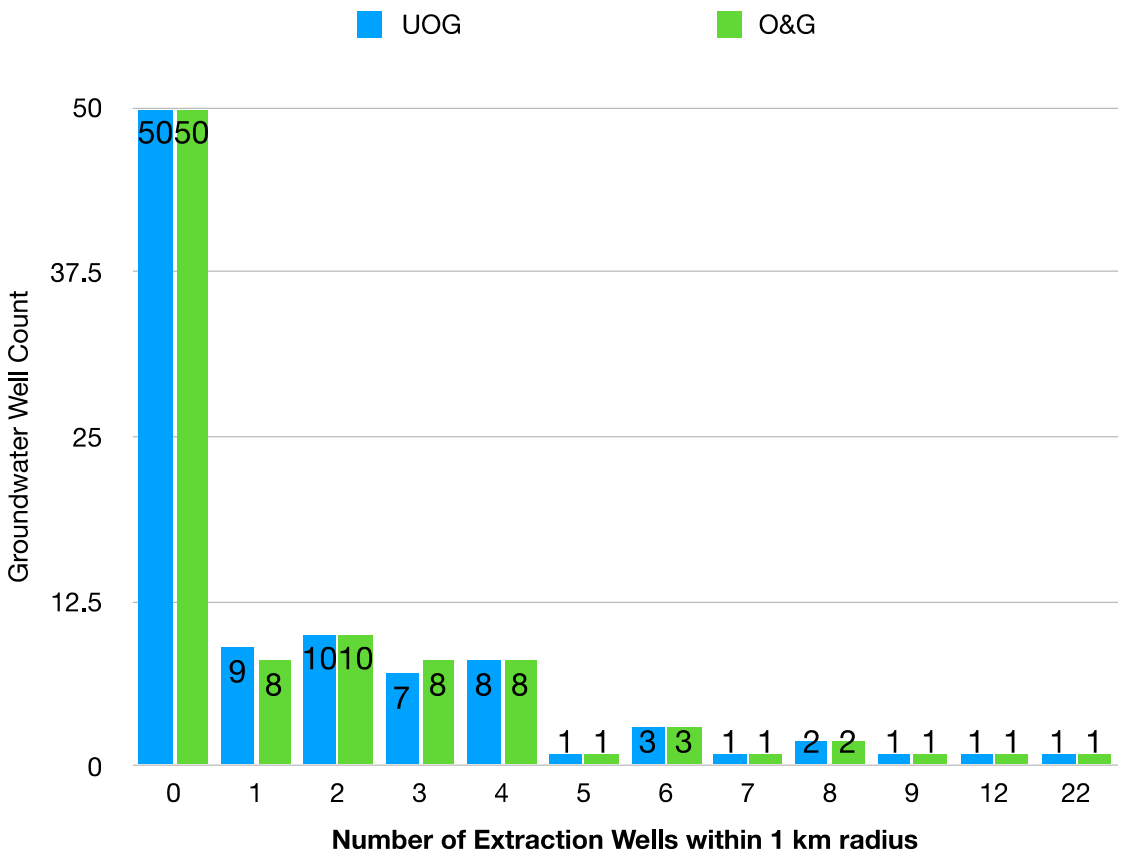

Figure S9. Number count of groundwater wells with a particular "density" of extraction wells within 1 km (unconventional oil and gas (UOG, blue) or all oil and gas wells (O&G, green)). For example, most groundwater wells (n=50) had zero gas extraction wells within a 1 km radius, whereas 9 groundwater wells had 1 UOG well within a 1 km radius and 8 groundwater wells had 1 O&G well within a 1 km radius. One groundwater well had 22 UOG wells within 1 km.

The statistical distribution of log K_{oc} values of disclosed fracturing chemical and their frequencies of use are calculated using data organized by Elsner and Hoelzer 2016[19]. The organized data is primarily based on disclosure until November 2014 from FracFocus Chemical Disclosure Registry that includes states Colorado, North Dakota, Pennsylvania, and Texas that have been summarized by EPA and Rogers *et al.*[20]. In total, 959 entries of chemicals that had distinct names, 508 of which had an estimated K_{oc} value, among which 246 indicated their frequency of use. The 508 log K_{oc} values had a 25th-percentile value of -0.1, a 75th-percentile value of 3.1, and a median value of 1.4, which is within the range of log K_{oc} from acrylamide (log K_{oc} = 0.5) to phthalate (log K_{oc} = 4.9). This distribution is identical to the 246 chemicals that also indicated frequency of use (**Fig. S8)**. Frequency of use is the percentage of reports that

included the use of compound out of total reports[19]. For the 93 chemicals with higher frequency of use (>1), the 25%-75% of their log K_{oc} ranges from -0.4 to 3, which has good overlap with the range of log K_{oc} of acrylamide and phthalate. Thus, the selection of acrylamide and phthalate represents the majority of the hydraulic fracturing organic compounds with high frequency use.

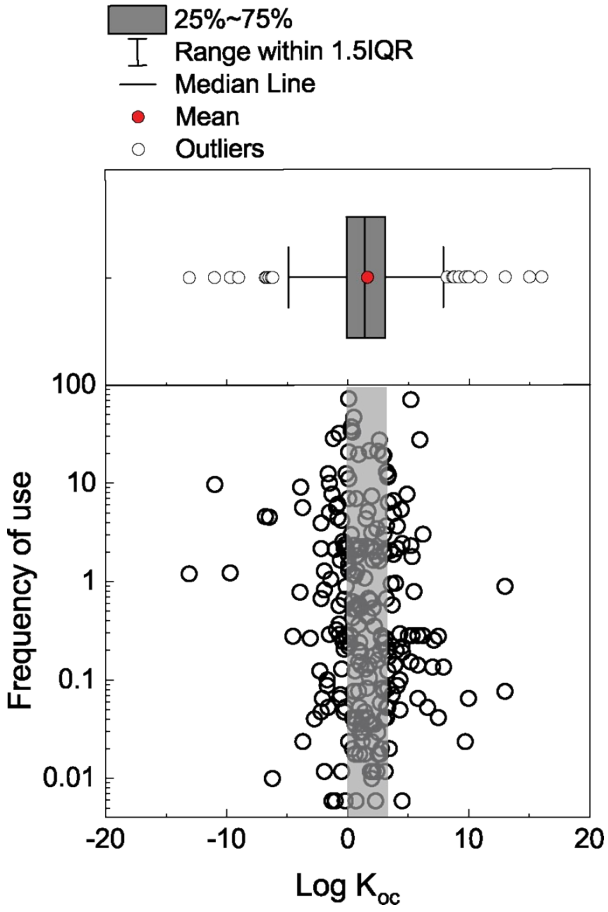

Figure S10. Top panel: the distribution of Log K_{oc} of 508 disclosed fracturing chemicals that have known K_{oc} values. Bottom panel: A subset of those chemicals (246) contained frequency use data, where frequency (%) is in percentage of appearance in fracking reports out of total reviewed reports. The grey shaded area indicates the 25-75-percentile range of the log K_{oc} values shown.

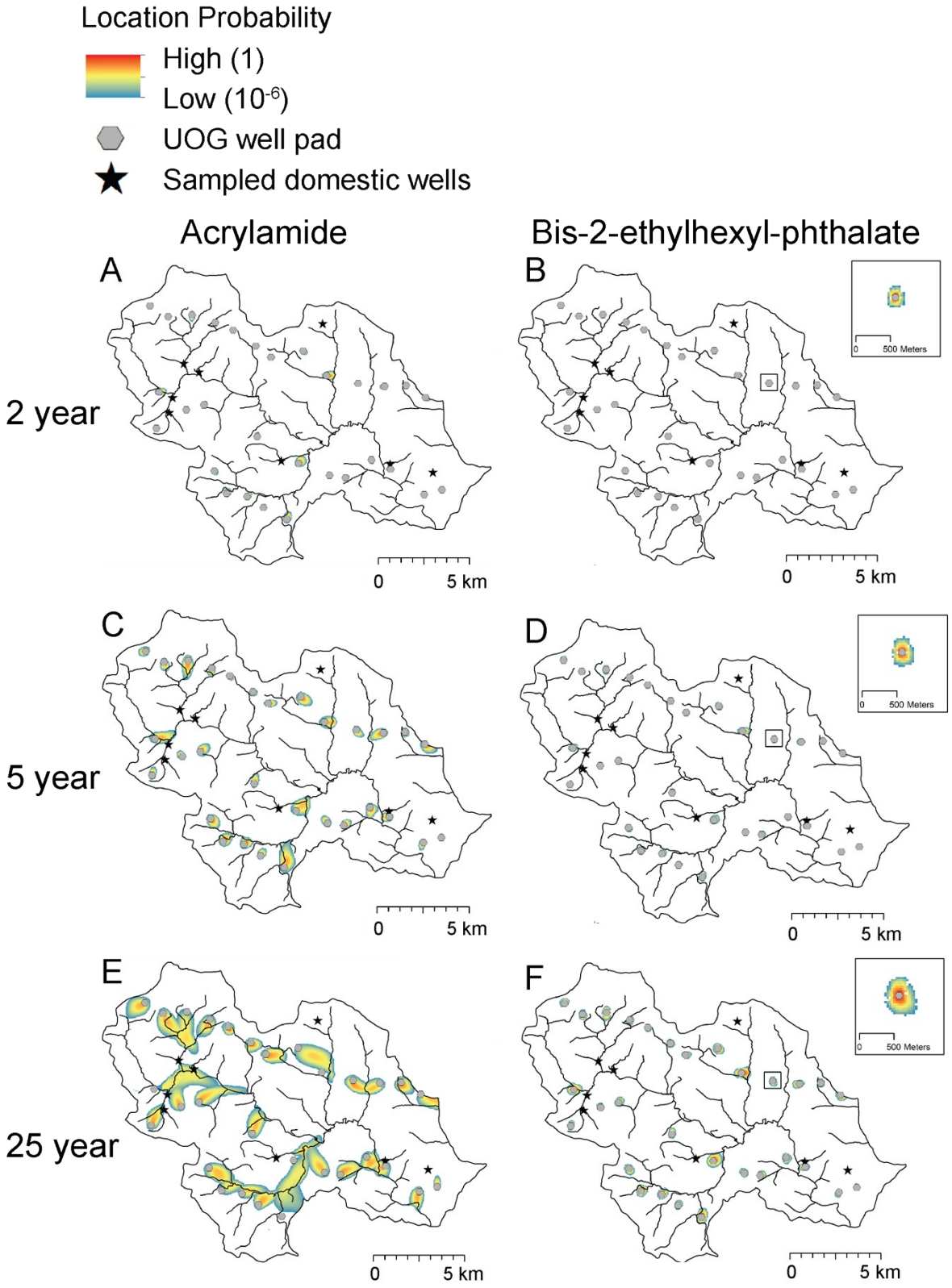

Figure S11. Location probability for a pulse injection of acrylamide and bis-2-ethylhexyl phthalate from gas well for transport times of 1 (A and B), 5 (C and D), and 25 (E and F) years.

The low probability is always set as 10^{-7}. The high probability is a function of the size of the plume in order to keep the entire area of the plume the same as the injected unit probability, therefore varies as a function of time and the retardation factor of the compounds. The high probability for acrylamide transport in A, C, E is defined as $10^{-0.1}$ (0.8), $10^{-0.3}$ (0.5), and $10^{-1.2}$ (0.06), respectively. Due to the slow transport of bis-2-ethylhexyl phthalate and the size of the plume does not vary significantly; the high probabilities of B, D, F are all $10^{-0.004}$ (0.99).

Velocity distribution justification given literature context: We are not aware of any direct groundwater velocity measurements in our study area to which we can compare our estimates. Water well records in the region report volumetric yields (usually in gallons/minute), which are a bit ambiguous to translate to velocities. Models or approximations are needed. For example, Wen *et al.*[21] reported a range of 0.2-4 m/day for vertical velocities. Their calculation is from velocity = K*(dh/dl)/ϕ, where they used a range of K (hydraulic conductivity) of 0.6-12 m/day reported by Yager[22]. Yager's values were determined from model calibration by trial-and-error, and were for an alluvial valley aquifer in Broome county, NY, which is immediately north of Bradford. Wen *et al.*[21] then assumed a "typical" value of 0.15 for ϕ (porosity) and 0.05 for dh/dl (hydraulic gradient) to come up with their velocity estimates. Horizontal velocity is typically assumed to be greater than vertical velocity, so an upper value of 4 m/day for vertical velocity would result in a horizontal velocity that is more than a factor of 2 outside of our range. However, the spatial extent of alluvial valley aquifers in our study area is quite limited.

Data from Barth-Naftilan *et al.*[23] (Table S1) reported horizontal hydraulic conductivity of 0.004-1.3 m/day for fracture intervals in Susquehanna county, PA, immediately east of Bradford. These values were measured via straddle-packer tests. Using the same assumptions of ϕ= 0.15 and dh/dl = 0.05, we would get velocities ranging from 0.001-0.4 m/day, which is covered by our modeled velocity range. We believe most of the aquifer in our study area is fractured rock, so our velocity range is reasonable.

Lastly, Llewellyn *et al.*[24] hypothesized that a contamination event in Bradford county could be explained by contaminants migrating horizontally 1-3 km within 2.5 years, translating to horizontal velocities of 1.1-3.3 m/day. The upper values (>1.8 m/d) are outside our range, but not extremely erroneous. All of these points of comparison based on an interpretation of migration pathways and not a direct measurement of velocity.

Parameters	Statistical result	GRO (ppb)	DRO (ppb)	Statistical test
Water type	P value	0.73325	0.83035	Kruskal-Wallis ANOVA
Topography	P value	0.52535	0.89094	Kruskal-Wallis ANOVA
Well Depth (m)	P value	0.28897	0.78081	Kruskal-Wallis ANOVA
Well age (year)	P value	0.07705	0.56152	Kruskal-Wallis ANOVA
Distance to nearest gas station (km)	P value	0.78184	0.94974	Kruskal-Wallis ANOVA
GRO	P value	NA	0.62831	Spearman correlations
	Spearman coefficient	NA	0.05172	
Methane(ppm)	P value	0.90125	*0.04646*	Spearman correlations
	Spearman coefficient	0.01366	*0.22473*	
Nearest well distance (km)	P value	0.91288	0.43327	Spearman correlations
	Spearman coefficient	0.01144	0.08362	
Nearest violated well distance (km)	P value	0.1243	0.77207	Spearman correlations
	Spearman coefficient	-0.15964	0.03096	
LUST distance (km)	P value	0.11035	0.78543	Spearman correlations
	Spearman coefficient	-0.16576	0.0291	
Gas well age	P value	0.22092	0.35641	Spearman correlations
	Spearman coefficient	0.12744	-0.09835	
As (ppb)	Spearman coefficient	0.35046	0.35984	Spearman correlations
	P value	0.09738	0.09765	
Ba (ppb)	Spearman coefficient	0.95497	*0.01324*	Spearman correlations
	P value	0.0059	*0.26025*	
U (ppb)	Spearman coefficient	0.4527	0.91237	Spearman correlations
	P value	0.07838	0.01176	
Pb (ppb)	Spearman coefficient	0.43591	0.18617	Spearman correlations
	P value	0.08132	-0.14062	
Sr (ppb)	Spearman coefficient	0.44365	*0.0139*	Spearman correlations
	P value	0.07996	*0.25848*	
Mn (ppb)	Spearman coefficient	0.25022	*0.00975*	Spearman correlations
	Spearman coefficient	1.20E-01	*0.27109*	
Li (ppb)	P value	0.25255	0.07686	Spearman correlations
	Spearman coefficient	0.11919	0.18746	
F (ppb)	P value	0.15895	0.45433	Spearman correlations
	Spearman coefficient	0.14646	0.07986	
Cl (ppb)	P value	0.36418	0.64206	Spearman correlations
	Spearman coefficient	0.09465	0.04966	

Table S5. Summary of all statistical analyses for correlation between DRO and GRO and other water parameters, distances, inorganic, and organic indicators.

NO3 (ppb)	P value	0.33594	*0.04393*	Spearman correlations
	Spearman coefficient	-0.10034	*-0.21291*	
SO4 (ppb)	P value	0.60393	0.3825	Spearman correlations
	Spearman coefficient	0.05419	-0.09316	
Br (ppb)	Spearman coefficient	0.30291	0.22756	Spearman correlations
	P value	0.10739	0.12846	
Na (ppb)	Spearman coefficient	0.6994	0.18754	Spearman correlations
	P value	0.04035	0.14019	
Fe (ppb)	Spearman coefficient	0.81964	*0.02706*	Spearman correlations
	P value	0.02383	*0.23306*	
Ca (ppb)	Spearman coefficient	0.75223	0.90249	Spearman correlations
	P value	-0.03299	-0.0131	
K (ppb)	Spearman coefficient	0.55715	*0.04113*	Spearman correlations
	P value	0.06132	*0.21574*	
Mg (ppb)	Spearman coefficient	0.88968	0.46867	Spearman correlations
	Spearman coefficient	-0.0145	0.07735	
HCO_3 (ppb)	P value	0.29941	0.14855	Spearman correlations
	Spearman coefficient	0.10997	0.15619	
Carbonate m (ppb)	P value	0.34635	0.13091	Spearman correlations
	Spearman coefficient	0.09985	0.16322	
DOC (ppb)	P value	0.46405	*0.006*	Spearman correlations
	Spearman coefficient	-0.07643	*-0.28754*	
Nearest well, Distance cut off 1 km	P value	0.59482	0.74446	Mann-Whitney U test
Nearest well, Distance cut off 1.5 km	P value	0.65307	0.79413	Mann-Whitney U test
Nearest well, Distance cut off 2 km	P value	0.86573	0.11512	Mann-Whitney U test
Nearest violated well, Distance cut off 1 km	P value	0.19133	0.77377	Mann-Whitney U test
Nearest violated well, Distance cut off 1.5 km	P value	0.39599	0.54772	Mann-Whitney U test
Nearest violated well, Distance cut off 2 km	P value	*0.02101*	0.73382	Mann-Whitney U test
Nearest violated well, Distance cut off 3 km	P value	0.141	0.86449	Mann-Whitney U test

References

1. Clark, C. X., Boya; Soriano, Mario; Gutchess, Kristina; Siegel, Helen; Ryan, Emma; Johnson, Nicholaus; Cassell, Kelsie; Elliott, Elise; Li, Yunpo; Cox, Austin; Bugher, Nicolette; Glist, Lukas; Brenneis, Rebecca; Sorrentino, Keli; Plano, Julie; Ma, Xiaomei; Warren, Joshua; Plata, Desiree; Saiers, James; Deziel, Nicole, Assessing unconventional oil and gas exposure in the appalachian basin: Comparison of exposure surrogates and residential drinking water measurements. . *Environmental Science & Technology* **2021,** *In revision*.
2. Li, Y. T., Nathalie ; Siegel, Helen ; Clark, Cassandra; Ryan, Emma; Brenneis, Rebecca; Gutchess, Kristina; Soriano, Mario; Xiong, Boya; Deziel, Nicole; Saiers, James; Plata, Desiree, Groundwater methane in northeastern pennsylvania attributable to thermogenic sources and hydrogeomorphologic migration pathways. *Environmental Science and Technology* **2021,** *In revision*.
3. Rahm, B. G.; Vedachalam, S.; Bertoia, L. R.; Mehta, D.; Vanka, V. S.; Riha, S. J., Shale gas operator violations in the marcellus and what they tell us about water resource risks. *Energy Policy* **2015,** *82*, 1-11.
4. Hoelzer, K.; Sumner, A. J.; Karatum, O.; Nelson, R. K.; Drollette, B. D.; O'Connor, M. P.; D'Ambro, E. L.; Getzinger, G. J.; Ferguson, P. L.; Reddy, C. M.; Elsner, M.; Plata, D. L., Indications of transformation troducts from hydraulic fracturing additives in shale-gas wastewater. *Environmental Science & Technology* **2016,** *50*, (15), 8036-8048.
5. Sumner, A. J.; Plata, D. L., Halogenation chemistry of hydraulic fracturing additives under highly saline simulated subsurface conditions. *Environmental science & technology* **2018,** *52*, (16), 9097-9107.
6. Sumner, A. J.; Plata, D. L., Oxidative breakers can stimulate halogenation and competitive oxidation in guar-gelled hydraulic fracturing fluids. *Environmental science & technology* **2019,** *53*, (14), 8216-8226.
7. Sumner, A. J.; Plata, D. L., A geospatially resolved database of hydraulic fracturing wells for chemical transformation assessment. *Environmental Science: Processes & Impacts* **2020,** *22*, (4), 945-955.
8. Brunner, P.; Simmons, C. T., Hydrogeosphere: A fully integrated, physically based hydrological model. *Ground Water* **2012,** *50*, (2), 170-176.
9. Alcolea, A.; Carrera, J.; Medina, A., Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources* **2006,** *29*, (11), 1678-1689.
10. Tonkin, M. J.; Doherty, J., A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research* **2005,** *41*, (10).
11. Welter, D. E.; White, J. T.; Hunt, R. J.; Doherty, J. E., *Approaches in highly parameterized inversion— pest++ version 3, a parameter estimation and uncertainty analysis software suite optimized for large environmental models*; Reston, VA, **2015**; p 54.
12. U.S. Geological Survey, Usgs water data for the nation. 2019. (Accessed by 8/17/2019).
13. Department of Conservation and Natural Resources, Pennsylvania groundwater information system databases (Accessed by (06/01/2020)).
14. Stuckey, M. H., *Low-flow, base-flow, and mean-flow regression equations for pennsylvania streams*; Reston, VA, **2006**; p 84.
15. Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M., *Environmental organic chemistry*. John Wiley & Sons: 2016.
16. Todd, D. K.; Mays, L. W., Groundwater hydrology edition. *Welly Inte* **2005**.
17. Gelhar, L. W.; Welty, C.; Rehfeldt, K. R., A critical review of data on field-scale dispersion in aquifers. *Water resources research* **1992,** *28*, (7), 1955-1974.

18. Drollette, B. D.; Hoelzer, K.; Warner, N. R.; Darrah, T. H.; Karatum, O.; O'Connor, M. P.; Nelson, R. K.; Fernandez, L. A.; Reddy, C. M.; Vengosh, A.; Jackson, R. B.; Elsner, M.; Plata, D. L., Elevated levels of diesel range organic compounds in groundwater near marcellus gas operations are derived from surface activities. *Proceedings of the National Academy of Sciences* **2015,** *112*, (43), 13184-13189.
19. Elsner, M.; Hoelzer, K., Quantitative survey and structural classification of hydraulic fracturing chemicals reported in unconventional gas production. *Environmental science & technology* **2016,** *50*, (7), 3290-3314.
20. Rogers, J. D.; Burke, T. L.; Osborn, S. G.; Ryan, J. N., A framework for identifying organic compounds of concern in hydraulic fracturing fluids based on their mobility and persistence in groundwater. *Environmental Science & Technology Letters* **2015,** *2*, (6), 158-164.
21. 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. *Environmental science & technology* **2018,** *52*, (12), 7149-7159.
22. Yager, R. M., *Estimation of hydraulic conductivity of a riverbed and aquifer system on the susquehanna river in broome county, new york*. US Government Printing Office: 1993.
23. Barth-Naftilan, E.; Sohng, J.; Saiers, J. E., Methane in groundwater before, during, and after hydraulic fracturing of the marcellus shale. *Proceedings of the National Academy of Sciences* **2018,** *115*, (27), 6970-6975.
24. Llewellyn, G. T.; Dorman, F.; Westland, J.; Yoxtheimer, D.; Grieve, P.; Sowers, T.; Humston-Fulmer, E.; Brantley, S. L., Evaluating a groundwater supply contamination incident attributed to marcellus shale gas development. *Proceedings of the National Academy of Sciences* **2015,** *112*, (20), 6325-6330.