1 **Classification:** Physical Sciences: Environmental Sciences

2

3 **Elevated levels of diesel range organic compounds in**

4 **groundwater near Marcellus gas operations are derived from**

5 **surface activities**

6

7 Brian D. Drollette[a], Kathrin Hoelzer[b], Nathaniel R. Warner[c], Thomas H. Darrah[d], Osman

8 Karatum[e], Megan P. O'Connor[e], Robert K. Nelson[f], Loretta A. Fernandez[g], Christopher

9 M. Reddy[f] Avner Vengosh[h], Robert B. Jackson[i], Martin Elsner[b], Desiree L. Plata[a]

10
11 [a] Department of Chemical and Environmental Engineering, Yale University, New Haven,
12 CT 06511
13 [b] Institute of Groundwater Ecology, Helmholtz Zentrum München, Germany
14 [c] Department of Civil and Environmental Engineering, Pennsylvania State University,
15 University Park, PA 16802
16 [d] School of Earth Sciences, The Ohio State University, Columbus, OH 43210
17 [e] Department of Civil and Environmental Engineering, Duke University, Durham, NC
18 27708
19 [f] Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic
20 Institution, Woods Hole, MA 02543
21 [g] Departments of Civil and Environmental Engineering and Marine and Environmental
22 Sciences, Northeastern University, Boston, MA 02115
23 [h] Division of Earth and Ocean Sciences, Duke University, Durham, NC 27708
24 [i] School of Earth, Energy, and Environmental Sciences, Woods Institute for the
25 Environment, and Precourt Institute for Energy, Stanford University, Stanford, CA 94305
26
27 **Corresponding Author:** Desiree L. Plata; desiree.plata@yale.edu

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.

\body

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),

76   or deep transport routes? Furthermore, are organic compounds present in groundwater

77   derived from naturally occurring, geogenic sources or associated with industrial activities,

78   such as HVHF? Finally, what are the chemical fingerprints that enable one to make this

79   distinction?

80   Although few studies have examined the occurrence and origin of organic

81   contaminants in groundwater [7], the presence of light hydrocarbon gases (i.e., methane

82   and ethane) and inorganic constituents has been investigated frequently. Osborn et al.

83   (2011) and Jackson et al. (2013) demonstrated elevated methane levels within 1 km of

84   unconventional gas wells over the Marcellus Shale [8, 9]. Further, Darrah et al. (2014)

85   showed that stray gas contamination in a subset of groundwater wells likely resulted from

86   poor well integrity (i.e., casing and cementing issues) [10]. In contrast, inorganic

87   chemical constituents (e.g., Cl^- and Br^-) in groundwater over the Marcellus Shale seem to

88   reflect geogenic sources and provide evidence of hydraulic connectivity between shallow

89   groundwaters and deeper formation brines on geological timescales in some areas in

90   northeastern PA [11, 12]. This deep-origin, saline groundwater has a chemical and

91   isotopic fingerprint similar to the Marcellus brines [11] but distinct from Marcellus

92   flowback water [13]. However, it is unknown if this deep saline water carries a unique

93   organic chemical fingerprint of either geogenic or anthropogenic origin.

94   The same mechanistic approaches taken to source apportion methane and

95   inorganic compounds have not yet been applied to organic compounds in groundwater.

96   While Gross et al. (2013) reported surface spills of hydraulic fracturing fluids and wastes

97   that could impact groundwater with organic chemicals in Colorado [14], and the

98   Pennsylvania Department of Environmental Protection (PA DEP) has assessed multiple

99   instances of local groundwater pollution by gas drilling operations [15], these studies

100   relied on voluntary industry accident disclosure and did not probe for possible alternative

101   exposure paths (i.e., through a broad geospatial sample set). In addition, they did not

102   provide a detailed characterization of the water that could identify the dominant transport

103   processes associated with the contamination (i.e., organic and inorganic markers, along

104   with hydrocarbon composition, noble gas isotopes, and spatial distribution analysis).

105   Recently, Llewellyn et al. (2015) investigated a localized incident of stray gas

106   groundwater pollution in Pennsylvania that exposed the potential for groundwater

107   contamination from natural gas extraction practices, where the authors attributed the

108   contamination to flow from HVHF wells through shallow subsurface pathways [7]. While

109   critical and detailed, the targeted nature of this case study (i.e., sampling wells with

110   documented contamination or close-to-contamination sites) precludes the identification of

111   geologic transport mechanisms that may be occurring in the region. Thus, the

112   mechanisms of organic chemical transport in groundwater associated with HVHF regions

113   remain unclear.

114   To address this research gap, we sampled 64 private residential groundwater wells,

115   ranging from 9 – 213m deep, over a three-year period (2012-2014) in northeastern

116   Pennsylvania ($n = 62$) and in southern New York ($n = 2$) for analyses of gas

117   chromatography (GC)-amenable organic compounds (Fig. 1). Fifty-nine samples were

118   analyzed for volatile organic compounds (VOCs) and gasoline range organic compounds

119   (GRO; defined as the hydrocarbons eluting between 2-methylpentane and 1,2,4-

120   trimethylbenzene; approximately between $n\text{C}_6$ and $n\text{C}_{10}$), and 41 were also analyzed for

121   diesel range organic compounds (DRO; defined as the hydrocarbons eluting between

122    $n\mathrm{C}_{10}$ and $n\mathrm{C}_{28}$) [16]. Analytical details are provided in the Supporting Information (SI)

123   Appendix. (Note that compounds included in the Environmental Protection Agency

124   (EPA)-designated definitions of GRO and DRO are not necessarily gasoline or diesel

125   derived). A subset of these samples was analyzed using comprehensive two-dimensional

126   gas chromatography (GC×GC) to evaluate whether compound-specific organic chemical

127   fingerprints were associated with either HVHF activities or natural geologic processes.

128   Complementary analyses of the inorganic chemical, methane stable isotope, and helium

129   composition (i.e., $[^4\mathrm{He}]$ and $^3\mathrm{He}/^4\mathrm{He}$) were conducted to evaluate potential transport

130   mechanisms for organic compounds into the shallow groundwater (i.e., from surface

131   spills, leaky well casings, or communication with deep shale formations). Finally, we

132   investigated the spatial distribution of disclosed surface spills, active shale gas wells, and

133   groundwater samples with elevated GRO and DRO to determine if there is an increased

134   risk associated with the co-location of natural gas extraction activities with drinking-

135   water supplies over the Marcellus Shale.

136   **Results and Discussion**

137   Trace levels of GRO and DRO compounds were detected in 9 of 59 (0 – 8.8 ppb

138   total GRO) and 23 of 41 (0 - 157.6 ppb total DRO) groundwater samples, respectively.

139   While the highest concentrations of GRO and DRO were always detected within 1 km of

140   active shale gas operations, this difference in concentration within 1 km ($n = 21$) and

141   beyond 1 km ($n = 20$) from shale gas wells was only significantly higher in the case of

142   DRO (Figure 2; $p = 0.01$, Mann-Whitney U test, SI Appendix Fig. S1). A sensitivity

143   analysis indicated that the statistically significant difference in DRO levels persisted at

144   0.75 – 3.0 km (at 0.25 km intervals) from a gas well, as well as at 0.76 km away (the PA

145   DEP's suggested distance of evaluation; $p < 0.01$) [17]. Notably, although the highest

146   GRO and highest DRO containing samples all occurred within 1 km of a shale gas well,

147   the elevated GRO and DRO were not co-occurring (Figure 2, inset; discussed below).

148   Finally, trace levels (<1 ppb) of VOCs, including BTEX compounds (benzene, toluene,

149   ethylbenzene, and xylenes), were detected in six samples (10%) and well below the

150   EPA's drinking water maximum contaminant level (MCL) recommendations [18].

151   In this region, there are multiple potential sources of elevated DRO in

152   groundwater, including: (a) upward migration of naturally occurring, formation derived

153   organic compounds over geologic time, (b) lateral transport of drilling muds, flowback,

154   or produced fluids from faulty wells, (c) leaking oil and gas waste containment ponds, (d)

155   input of organic contaminants from surface spills of either raw chemicals or residual

156   fracturing fluids, and (e) leaking underground storage containers or local traffic. To

157   evaluate these sources systematically, we employed geochemical fingerprinting using

158   inorganic constituents (i.e., Br/Cl ratios) [11], groundwater residence times (i.e., ^{4}He

159   concentration) [10], and dissolved methane concentrations [8, 9], coupled with our GRO

160   and DRO analysis and geospatial analysis.

161   *Upward Migration from Deep Formation.* Warner et al. (2012) demonstrated that

162   shallow groundwater in some areas in northeastern PA is saline with molar Br/Cl ratios

163   similar to deeper Marcellus Formation water (designated "Type D" water), suggesting

164   natural upward migration of deep saline water over geologic timescales [11]. Using this

165   inorganic fingerprinting approach, we found no statistical difference in GRO or DRO

166   contents based on water type (SI Appendix, Fig. S2; GRO, $p > 0.05$; DRO, $p > 0.05$;

167   Kruskal-Wallis test), suggesting that the increased GRO and DRO signals were not a

168   result of upward migration of deeper naturally occurring formation fluids. Furthermore,

169   samples with elevated GRO (>5 ppb), but lower DRO (<50 ppb), which might be

170   considered geogenic and shale-derived considering the distinct transport rates of each

171   (i.e., retarded transport of higher molecular weight compounds due to slower diffusion

172   through and higher sorptivity to porous media) were not found in Type D waters

173   uniformly (2 of 3 were not Type D).

174   Water migration from the Marcellus formation to shallow groundwater would also

175   lead to significant enrichments in ^{4}He and fractionation of air-saturated water noble gases

176   (i.e., ^{20}Ne/^{36}Ar) [10]. In contrast, we find that the highest concentration DRO and GRO

177   samples occur in tritium-active groundwater (i.e., relatively young) and have the lowest

178   ^{4}He abundance (an integrated proxy for residence time, SI Appendix, Fig. S3) and,

179   therefore, the lowest apparent crustal residence times [10]. This suggests that

180   contamination occurs in the younger groundwater systems. Consequently, our results

181   indicate that connectivity with deep subsurface brines is not a dominant source of organic

182   compounds in the shallow groundwater.

183   *Lateral Transport from Faulty Wells.* An alternate source of organic compounds

184   to shallow aquifers could be faulty gas well casings, as poor well integrity has been

185   documented in gas wells targeting the Marcellus Shale [6, 9, 10, 19, 20]. Llewellyn et al.

186   (2015) recently reported groundwater contamination of 2-n-butoxyethanol [7], while

187   others documented stray gas contamination of light hydrocarbons (e.g., nC$_1$-nC$_3$) from

188   poor well integrity [8-10]. Previous studies suggested that hydraulic fracturing fluids and

189   denser formation brines could also migrate laterally, but on timescales longer than typical

190   of methane transport [5, 21, 22]. Such migration through porous media from a well

191   casing would result in elevated GRO with lower levels of DRO (due to the higher

192   diffusivities and generally lower sorptivities of the lower molecular weight compounds),

193   along with higher levels of methane and a salinity signature similar to that of flowback or

194   produced waters. Methane abundance from paired samples or previous sampling

195   campaigns showed no correlation with GRO or DRO (SI Appendix, Fig. S4), and the

196   noble gas analysis provided no evidence for fugitive gas contamination in the elevated

197   GRO and DRO samples (e.g., low air-saturated water abundances ($[^{36}Ar]$, $[N_2]$), or

198    $^{4}He/CH_4$ [10]). Furthermore, samples with elevated GRO (>5 ppb) had relatively low

199   methane and Br^- (<1 ppm for both). Thus, leaky well casings are an unlikely source of

200   GRO compounds.

201   It is possible that a leaky well casing during slickwater injection could be a source

202   of elevated GRO or DRO without commensurate brine or methane inputs in a relatively

203   young well. If so, then one might expect some relationship between GRO or DRO

204   occurrence in groundwater and the age of the nearest HVHF well *if and only if* chemical

205   or bulk fluid transport times were fast relative to or on the same order as the well ages.

206   To entertain the possibility of a well age effect, we calculated time from the "spud date"

207   (the drilling date of the nearest HVHF well) to sample collection. Well ages ranged

208   broadly from 10 days to over 5 years with a fairly even temporal distribution, and levels

209   of GRO and DRO were not correlated with the age of the well (DRO, $p > 0.05$; GRO, $p >$

210   0.05; SI Appendix, Fig. S5).

211   Since the preceding well age argument relies on rapid fluid transport relative to

212   the well ages, we note that typical bulk groundwater velocities are highly variable in the

213   sampled aquifers and on the order of 0.1-8.2 km yr^{-1} (spanning velocities in alluvium to

214   fractured bedrock aquifers) [23], and sorption-retarded transport velocities of the

215   chemicals we detected (described below) would be on the order of 0.02-7.53 km yr^{-1}

216   (approximately 2 mo - 50 yr to migrate a 1 km distance; see SI Appendix). However,

217   depending on the topography, hydraulic connectivity, and large pressure gradients

218   experienced during injection [5], transport times could be faster than predicted by simple

219   porous media transport models. For example, Llwellyn et al. [7] argue that fracturing

220   fluids could be driven 1-3 km in a 2-3 yr timeframe, which is reasonable for the fractured

221   bedrock case. While it is not possible to put an exact timeframe on the fluid transport

222   under the hypothetical condition of leaky casing during slickwater injection, two

223   conditions emerge: either (1) chemical transport is slow and could not give rise to the

224   elevated DRO compounds observed here (i.e., within 1 km and less than 2 years), or (2)

225   transport is faster and a relationship between DRO and well age could have been

226   observed, which it was not. In either case, our data suggest that leaky well is not a source

227   of DRO to nearby groundwater wells.

228   *Leaking Oil and Gas Waste Containment Ponds.* Following hydraulic fracturing,

229   the flowback and produced waters are often stored in polymer-lined, open waste

230   containment pits, which are demonstrated sources of contamination to surface water and

231   groundwater in cases where the liner integrity was compromised (e.g., torn, ripped,

232   folded, or other failure due to a physical breach that allowed fluid to pass unrestricted)

233   [24]. While many of these pits have been phased out voluntarily, many were still in use at

234   the time of our study. Unfortunately, Pennsylvania does not maintain a publically

235   available database of the location of the polymer-lined containment pits, and no spatial

236   analysis between elevated DRO or GRO levels and containment pits is possible.

237   Nonetheless, these pits were designed to allow volatile compounds to outgas and

238   particles to settle [25], and the residual wastes are often highly saline with a high organic

239   content [24, 26]. Thus, leakage into groundwater from such containment basins would

240   result in low GRO levels (due to volatile out gassing) and elevated DRO, such as

241   observed in our samples (see SI Appendix for discussion of potential GRO/DRO

242   fingerprints in groundwater and flowback water, SI Appendix, Fig. S8). Leaking from

243   pits with compromised liners would give rise to elevated chloride and bromide in the high

244   DRO samples, which was not observed (SI Appendix, Fig. S6). Therefore, diffusive

245   transport of DRO through uncompromised liners could give rise to the observed chemical

246   composition of the groundwater. However, the types of compounds revealed in our

247   compound-specific analysis (detailed below) have very long transport times through

248   model polymers characteristic of such liners. Considering the fastest-possible transport,

249   the compound would not migrate through a 4-mm liner to the soil interface after 4 yr

250   (only 2×10^{-27}% of the water-side content would migrate to 1 mm depth after 4 yr; model

251   details provided in SI Appendix, Fig. S7). Over these long timescales, transport through

252   intact pit liners could not have given rise the DRO observed in our samples. This implies

253   that organic chemical transport through the liners was not the primary source of material

254   in our samples.

255   *Surface Spills of Hydraulic Fracturing Chemicals*. State databases are maintained

256   for disclosed releases of hydraulic fracturing components at the surface, as these present a

257   direct route for surface water and groundwater contamination. Surface releases could

258   result in low GRO due to volatilization within weeks [27, 28], with elevated levels of

259   DRO leaching to groundwater. Such elevated DRO with low GRO was observed in our

260   samples, with higher DRO close to active shale gas wells (< 1km). In order to further

261   evaluate the possible link between elevated DRO and proximity to the nearest well, we

262   analyzed PA DEP violation reports. According to the PA DEP Oil and Gas Compliance

263   Report, 5,791 violations were reported associated with 1,729 unconventional gas wells

264   throughout the state between January 1, 2007 and June 1, 2014 [29] and classified as

265   either "Administrative" or "Environmental Health & Safety" (EHS) violations. DRO

266   concentrations were elevated significantly in groundwater samples in close proximity to

267   EHS violations ($p = 0.03$, Spearman correlation; Figure 3), but GRO concentrations were

268   not ($p = 0.36$). Furthermore, DRO occurrence in samples within 2 km of an

269   unconventional well with an EHS violation ($n = 20$) was statistically higher than in

270   samples further away than 2 km ($n = 21$; $p = 0.03$, Mann-Whitney U test), whereas GRO

271   did not show the same relationship ($p = 0.92$). Neither DRO nor GRO levels were

272   significant at the 1 or 0.76 km cutoff distances, perhaps due to the fact that the

273   distribution of shale gas wells with an EHS violation is spatially diffuse compared to

274   individual shale gas well locations.

275   Groundwater well depth could also provide information on the nature of the flow

276   path of the compounds detected in our samples. For example, since vertical transport

277   times are long, a deep or shallow source might give rise depth-dependent concentration

278   gradient. There was no statistically significant difference between DRO or GRO

279   concentrations in the shallowest (<100m) or deepest (>100m) sampled wells (DRO, $p =$

280   0.57; GRO, $p = 0.89$; Mann-Whitney U test), or at any other depth cutoff (50m, 75m,

281   125m), and neither GRO or DRO were correlated with depth (SI Appendix, Fig. S9). This

282   could be an artifact of the scale and spatial resolution of the sampling effort. Designed to

283   cover a large area (approximately 7,400 km^2) and constrained by well access, the

284   groundwater samples were separated by widely varying lateral distances (7 ± 15 and 11 ±

285   18 km for groundwater wells containing detectable DRO and GRO, respectively). As a

286   result, any point source or spatially constrained "plume" of organic material could

287   conceivably affect only a small population of groundwater wells, obfuscating any effect

288   of well depth on the GRO or DRO concentration. Nevertheless, the samples with the

289   highest GRO and DRO are found in groundwater wells less than 100m deep.

290   *Leaking Underground Storage Tanks or Local Traffic.* We also explored the

291   hypothesis that leaking underground storage tanks that typically contain gasoline, diesel,

292   or fuel oil for both domestic and industrial use could provide a significant source of GRO

293   and/or DRO. Leaking tank incident data obtained from PA DEP [30] showed no spatial

294   correlations with DRO (p = 0.95, Spearman correlation) or GRO (p = 0.81) in the

295   groundwater samples (SI Appendix, Fig. S10). In addition, the chemicals identified in the

296   compound-specific analysis are not commonly stored in underground storage containers

297   and are distinct from the chromatographic fingerprints of gasoline, diesel, or hydraulic

298   fluids (SI Appendix, Fig S11). Indeed, were these materials present, they would be

299   readily obvious via our analytical methods (detection limits near 100 pg L^{-1} or parts per

300   quadrillion). Their absence implies that leaking underground storage tanks were not a

301   source of material to the groundwater. Similarly, if local truck traffic were a source, then

302   one might expect a distinct chemical fingerprint and correlation with distance to the

303   nearest road. No such fingerprints (SI Appendix, Fig. S11) or correlations existed (DRO,

304   p = 0.78; GRO, p = 0.63), suggesting that traffic was not responsible for the DRO

305   observed in the studied groundwater.

306   *Organic Chemical Fingerprinting via GC×GC-TOF-MS.* In order to further

307   evaluate the source of elevated DRO compounds, we conducted a compound-specific

308   investigation using GC×GC with time-of-flight mass spectrometry (GC×GC-TOFMS). In

309   particular, a subset ($n = 12$) of groundwater liquid-liquid extracts was analyzed. including

310   those with the highest DRO levels, along with a field blank and one natural salt spring in

311   Susquehanna County, PA, which is typically used to indicate the composition of natural

312   gas and brine-rich fluids derived from shale rock sources in the northern Appalachian

313   Basin [7-12]. The samples with the highest DRO ($n = 2$) contained bis(2-ethylhexyl)

314   phthalate (Figure 4), whereas the salt spring, lower-DRO samples ($n = 8$), and field blank

315   did not (SI Appendix, Fig. S12). Fatty acid phthalate esters (including bis(2-ethylhexyl)

316   phthalate) are used in drilling and in hydraulic fracturing fluids [31], and bis(2-

317   ethylhexyl) phthalate has been reported in Marcellus Shale, Barnett Shale, and Denver-

318   Julesburg basin flowback waters [26, 31-33], in runoff and surface waters following an

319   incident at a gas well site [34], and in multiple residential groundwater wells in Dimock,

320   PA where the EPA has identified contamination directly from hydraulic fracturing

321   operations (discerned by the timing onset of the contamination) [35].

322   Bis(2-ethylhexyl) phthalate is a ubiquitous chemical that is used in many

323   industrial practices and materials, and it is difficult to attribute its presence solely to

324   hydraulic fracturing activities. However, we present several lines of evidence that this

325   particular phthalate is likely to be derived from HVHF activities. First, only our highest

326   DRO samples contained bis(2-ethylhexyl) phthalate, suggesting that the compound was

327   not derived from any step of our own analytical procedure. Second, if polyvinylchloride

328   (PVC) pipes (known to contain phthalates and to be pervasive in water distribution

329   systems) were a source of the bis(2-ethylhexyl)phthalate, then one would expect a

330   widespread presence in the analyzed samples ($n = 12$). In contrast, it was only detected in

331   the highest DRO samples ($n = 2$). Third, compound-specific analysis of the natural salt

332   spring did not contain the phthalate. Thus, the presence of bis(2-ethylhexyl) phthalate

333   likely reflects its presence in the contamination source and is not an artifact of our

334   sampling or preparation protocol.

335   Curiously, this particular phthalate has relatively low aqueous solubility. In

336   chemical disclosure databases, bis(2-ethylhexyl) phthalate is reportedly used in "perfball"

337   form (i.e., it is transported and injected as a solid). A solubilized form of a phthalate

338   could be derived from surface spills of flowback or produced waters, or transport through

339   containment pit liners. However, the former would carry a brine signature, which was not

340   observed in the high DRO groundwater samples (SI Appendix, Fig. S6), and the latter has

341   prohibitive transport timescales and could not give rise to the phthalate observed here (SI

342   Appendix, Fig. S7). Consequently, our data suggests that some solubilized form of the

343   phthalate (e.g., perfballs placed in a liquid carrier) is responsible for their appearance in

344   shallow aquifers sampled in this study. Indeed, in all cases where the phthalate was

345   detected, toluene was present as a co-contaminant. Further, since bis(2-ethylhexyl)

346   phthalate is a disclosed additive in fracturing fluids, it is both (a) plausible that its

347   presence in these samples is due to accidental surface releases of the parent fluids in the

348   Marcellus region and (b) reasonable given our statistical spatial analysis using the

349   disclosed spill database, as well as the complementary inorganic, methane, and helium

350   abundance measurements. Nevertheless, one can not rule out the possibility that the

351   bis(2-ethylhexyl) phthalate is derived from some non-HVHF source and just

352   coincidentally correlated with proximity to disclosed HVHF EHS violations.

353   Bis(2-ethylhexyl) phthalate is "reasonably anticipated to be a human carcinogen"

354   [36]. Due to the analytical challenges of obtaining clean blanks and ubiquitous industrial

355   use, the environmental fate of phthalates has been understudied since their presumable

356   first appearances after the advent of plastics in the 1970s [37-40]. Detection of phthalates

357   in environmentally derived samples, as well as their source apportionment, may have

358   been overlooked out of fear of cross-contamination from other sources. However, if

359   HVHF practices are employing phthalates (which are disclosed, but not with great

360   frequency), the environmental geochemistry community is challenged to develop robust

361   methods to track and source apportion these materials. Careful efforts to avoid

362   contamination (i.e., the use of pre-combusted, all-glass or metal materials) and

363   accountability for all other potential local industrial sources will be critical.

364   *Implications*. This is the first study of its kind to evaluate, on a regional scale,

365   different possible mechanistic sources of organic compounds detected in drinking water

366   wells in the Marcellus region using complementary inorganic chemical analyses and

367   residence time approximations. Based on the evaluation of different possible mechanisms,

368   our data are consistent with a surface-derived source of organic compounds in the study

369   area, possibly from releases of hydraulic fracturing materials near drill sites. The

370   emergent question arises: is the spill rate associated with unconventional shale gas

371   development worse than any other industrial chemical or energy extraction activity?

372   Unfortunately, a quantitative comparison can not be made due to the construct of the PA

373   DEP disclosed violation reports, for which details are limited [41]. Often, ambiguous

374 language is used to describe the nature of the violation and volume estimates of reported

375 releases are not provided. If volume data were available, an appropriate comparison of

376 the environmental impacts of these releases could be calculated with the ratio of volume-

377 of-spill to total-transported volume, as is done with crude oil (in 2014, 0.00007% of all

378 oil transported by sea was spilled [42]). Incorporating volumetric data on releases from

379 natural gas operations would allow direct comparisons to other energy industries. Clearly,

380 such a report relies on accurate self-reporting or more robust monitoring [43]. Further, if

381 the PA DEP required volume and chemical identity estimates, a more accurate

382 assessment of the relative risks due to domestic energy extraction could be constructed.

383　　　Irrespective of the reporting nuances, it is clear that surface releases of fracturing

384 fluids are usually accidental. Therefore, it is not necessarily the hydraulic fracturing

385 process (i.e., the fluid injection) that can lead to groundwater contamination, but rather,

386 the existence of the operation itself (i.e., the inherent risk associated with mechanical

387 failure and human error in industrial practice). Domestic natural gas production

388 necessitates co-location of residential areas with extraction facilities, and, like any

389 industrial activity, the economic benefits come with some level of environmental and

390 public health risk.

391　　　In summary, we show that some private residential groundwater wells contained

392 trace concentrations of organic compounds (<200 ppb DRO) in close proximity to active

393 shale gas wells and disclosed EHS violations. Surface sources are consistent with the

394 presence of DRO compounds in groundwater with the lowest apparent groundwater

395 residence times. We found no evidence for direct communication of deeper formation

396 water or injected fracturing fluids with shallow drinking water wells due to upward

397 migration from shale horizons. This result is encouraging, as it implies there is some

398 degree of temporal and spatial separation between injected fluids and drinking-water

399 supply. However, shallow groundwater should be monitored over longer timescales [44]

400 in areas of enhanced fracturing activities (e.g., where preferential faults could enhance

401 deep-to-surface communication [45]). Future research should also focus on investigating

402 chemical fingerprints of shale-derived organic matter via a careful comparison of raw

403 fracturing fluids, flowback water, and geologic formation waters.

404 **Materials and Methods**

405   Shallow groundwater samples were collected in pre-combusted glass vials over

406 three sampling campaigns from private residential groundwater wells. Wells were purged

407 of stagnant water until stable readings of conductivity, pH, and temperature were

408 recorded, upstream of any treatment system. The samples were fixed with acid, then

409 stored on ice until analysis within 14-28 days. For the organic compound analysis, light

410 hydrocarbons were analyzed using standard purging and preconcentration techniques (see

411 SI Appendix for details) whereas heavier hydrocarbons were concentrated via liquid-

412 liquid extraction into organic solvents. Compounds and compound classes were

413 quantified via gas chromatography with flame ionization detection (GC-FID) and

414 qualitatively identified with confirmed standards using GC-mass spectrometry (MS). A

415 subset of liquid-liquid extracts was interrogated using comprehensive two-dimensional

416 gas chromatography with time of flight mass spectrometry (GC×GC-TOF-MS). Inorganic

417 constituents were analyzed by methods detailed in Warner et al. (2012) [11]. Methane

418 was analyzed by methods detailed in Jackson et al. (2013) [9]. Noble gases were analyzed

419 by methods detailed in Darrah et al. (2014) [10]. Maps and spatial data analysis were

420   prepared with ArcMap^{TM} and all statistics were analyzed with the R statistical computing

421   platform.

422

427

References

1. Kerr RA (2010) Natural Gas From Shale Bursts Onto the Scene. *Science* 328(5986):1624-1626.
2. Malakoff D (2014) The gas surge. *Science* 344(6191):1464-1467.
3. US Energy Information Administration (2015) *Drilling Productivity Report, March 2015*. (USEIA, Washington, DC).
4. Howarth RW, Ingraffea A, & Engelder T (2011) Natural gas: Should fracking stop? *Nature* 477(7364):271-275.
5. Vengosh A, Jackson RB, Warner N, Darrah TH, & Kondash A (2014) A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. *Environ Sci Technol* 48(15):8334-8348.
6. Vidic RD, Brantley SL, Vandenbossche JM, Yoxtheimer D, & Abad JD (2013) Impact of Shale Gas Development on Regional Water Quality. *Science* 340(6134).
7. Llewellyn GT, et al. (2015) Evaluating a groundwater supply contamination incident due to Marcellus Shale gas development. *Proc Natl Acad Sci USA* 112(20):6325-6330.
8. Osborn SG, Vengosh A, Warner NR, & Jackson RB (2011) Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc Natl Acad Sci USA* 108(20):8172-8176.
9. Jackson RB, et al. (2013) Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proc Natl Acad Sci USA* 110(28):11250-11255.
10. Darrah TH, Vengosh A, Jackson RB, Warner NR, & Poreda RJ (2014) Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proc Natl Acad Sci USA* 111(39):14076-14081.
11. Warner NR, et al. (2012) Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania. *Proc Natl Acad Sci USA* 109(30):11961-11966.
12. Llewellyn GT (2014) Evidence and mechanisms for Appalachian Basin brine migration into shallow aquifers in NE Pennsylvania, USA. *Hydrogeol J* 22(5):1055-1066.
13. Warner NR, et al. (2014) New tracers of hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environ Sci Technol* 48(21): 12552-12560.
14. Gross SA, et al. (2013) Analysis of BTEX groundwater concentrations from surface spills associated with hydraulic fracturing operations. *J Air Waste Manage Assoc* 63(4):424-432.
15. Office of Oil and Gas Management (2015) *Oil and Gas Reports: Water Supply Determination Letters*. (PA DEP, Harrisburg PA).
16. US Environmental Protection Agency (2003) *Method 8015D Nonhalogenated Organics Using GC/FID*. (US EPA, Washington DC).
17. 25 Pa. Code § 38 *Oil and Gas Wells*.

473 18. US Environmental Protection Agency (2009) *National Primary Drinking Water*
474   *Regulations*. (US EPA, Washington DC).
475 19. Davies RJ et al. (2014) Oil and gas wells and their integrity: Implications for
476   shale and unconventional resource exploration. *Marine Petrol Geol* 56(0):239-
477   254.
478 20. Considine TJ, Watson RW, Considine NB, Martin JP (2013) Environmental
479   regulation and compliance of Marcellus Shale gas drilling. *Environ Geosci*
480   20(1):1-16.
481 21. Harrison SS (1983) Evaluating System for Ground-Water Contamination Hazards
482   Due to Gas-Well Drilling on the Glaciated Appalachian Plateau. *Groundwater*
483   21(6):689-700.
484 22. Harrison SS (1985) Contamination of Aquifers by Overpressuring the Annulus of
485   Oil and Gas Wells. *Groundwater* 23(3):317-324.
486 23. Gelhar LW, Welty C, & Rehfeldt KR (1992) A Critical Review of Data on Field-
487   Scale Dispersion in Aquifers. *Water Resour Res* 28(7):1955-1974.
488 24. Ziemkiewicz PF, Quaranta JD, Darnell A, & Wise R (2014) Exposure pathways
489   related to shale gas development and procedures for reducing environmental and
490   public risk. *J Nat Gas Sci Eng* 16(0):77-84.
491 25. Rich AL & Crosby EC (2013) Analysis of Reserve Pit Sludge from
492   Unconventional Natural Gas Hydraulic Fracturing and Drilling Operations for the
493   Presence of Technologically Enhanced Naturally Occurring Radioactive Material
494   (TENORM). *New Solut* 23(1):117-135.
495 26. Hayes T (2009) *Sampling and Analysis of Water Streams Associated with the*
496   *Development of Marcellus Shale Gas.* (Gas Technology Institute, Des Plaines IL).
497 27. Serrano A, Gallego M, Gonzalez JL & Tejada M (2008) Natural attenuation of
498   diesel aliphatic hydrocarbons in contaminated agricultural soil. *Environ Pollut*
499   151(3):494-502.
500 28. Serrano A, Gallego M & Gonzales JL (2006) Assessment of natural attenuation of
501   volatile aromatic hydrocarbons in agricultural soil contaminated with diesel fuel.
502   *Environ Pollut* 144(1): 203-209.
503 29. Office of Oil and Gas Management (2015) *Oil and Gas Reports: Oil and Gas*
504   *Compliance Report.* (PA DEP, Harrisburg PA). *Available at* http://www.portal.
505   state.pa.us/portal/server.pt/community/oil_and_gas_reports/20297. Accessed
506   January 20, 2015.
507 30. Department of Environmental Protection (2015) *Storage Tank Cleanup Locations.*
508   (PA DEP, Harrisburg PA) *Available at* http://www.portal.state.pa.us/portal/server.
509   pt/community/storage_tank_cleanup_program/20605/storage_tank_cleanup_locati
510   ons/1053538. Accessed January 22, 2015.
511 31. Maguire-Boyle SJ & Barron AR (2014) Organic compounds in produced waters
512   from shale gas wells. *Environ Sci Process Impacts* 16(10):2237-2248.
513 32. Ferrer I & Thurman, EM (2015) Chemical constituents and analytical approaches
514   for hydraulic fracturing waters. *Trends Analyt Chem* 5:18-25.
515 33. Lester Y, et al. (2015) Characterization of hydraulic fracturing flowback water in
516   Colorado: Implications for water treatment. *Sci Total Environ* 512-513(0):637-
517   644.

518   34. Justice J (2014) *U.S. Environmental Protection Agency Pollution/Situation Report,*
519   *Statoil Eisenbarth Well Response – Removal Polrep.* (US EPA, Washington DC).
520   35. US Environmental Protection Agency (2015) *Dimock Residential Groundwater*
521   *Site.* (US EPA, Washington DC). *Available at* http://www.epa.gov/reg3hscd/npl/
522   PAN000306785.htm. Accessed January 22, 2015.
523   36. NTP (National Toxicology Program). 2014. *Report on Carcinogens, Thirteenth*
524   *Edition.* Research Triangle Park, NC: U.S. Department of Health and Human
525   Services, Public Health Services. *Available at* http://ntp.niehs.nih.gov/pubhealth/
526   roc/roc13/
527   37. Wade TL (1974) Measurements of hydrocarbons, phthalic acid, and phthalic acid
528   ester concentrations in environmental samples from the north Atlantic. *M.S.*
529   *Thesis, University of Rhode Island.*
530   38. Sanborn JR, Metcalf RL, Yu CC & Lu PY (1975) Plasticizers in the environment:
531   the fate of di-N-octyl phthalate (DOP) in two model ecosystems and uptake and
532   metabolism of DOP by aquatic organisms. *Arch Environ Contam Toxicol*
533   3(2):244-255.
534   39. Hites RA & Biemann K (1972) Water pollution: organic compounds in the
535   Charles River, Boston. *Science* 178(4057):158-60.
536   40. Mayer FL, Stalling DL & Johnson JJ (1972) Phthalate Esters as Environmental
537   Contaminants. *Nature* 238(5364):411-413.
538   41. Brantley SL, et al. (2014) Water resource impacts during unconventional shale
539   gas development: The Pennsylvania experience. *Int J Coal Geol* 126(0):140-156.
540   42. US Energy Information Administration (2014) *World Oil Transit Chokepoints*
541   (USEIA, Washington DC).
542   43. Manda AK, Heath JL, Klein WA, Griffin MT, & Montz BE (2014) Evolution of
543   multi-well pad development and influence of well pads on environmental
544   violations and wastewater volumes in the Marcellus shale (USA). *J Environ*
545   *Manage* 142(0):36-45.
546   44. Myers T (2012) Potential Contaminant Pathways from Hydraulically Fractured
547   Shale to Aquifers. *Groundwater* 50(6):872-882.
548   45. Davies RJ, Mathias SA, Moss J, Hustoft S, & Newport L (2012) Hydraulic
549   fractures: How far can they go? *Marine Petrol Geol* 37(1):1-6.
550

551

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.
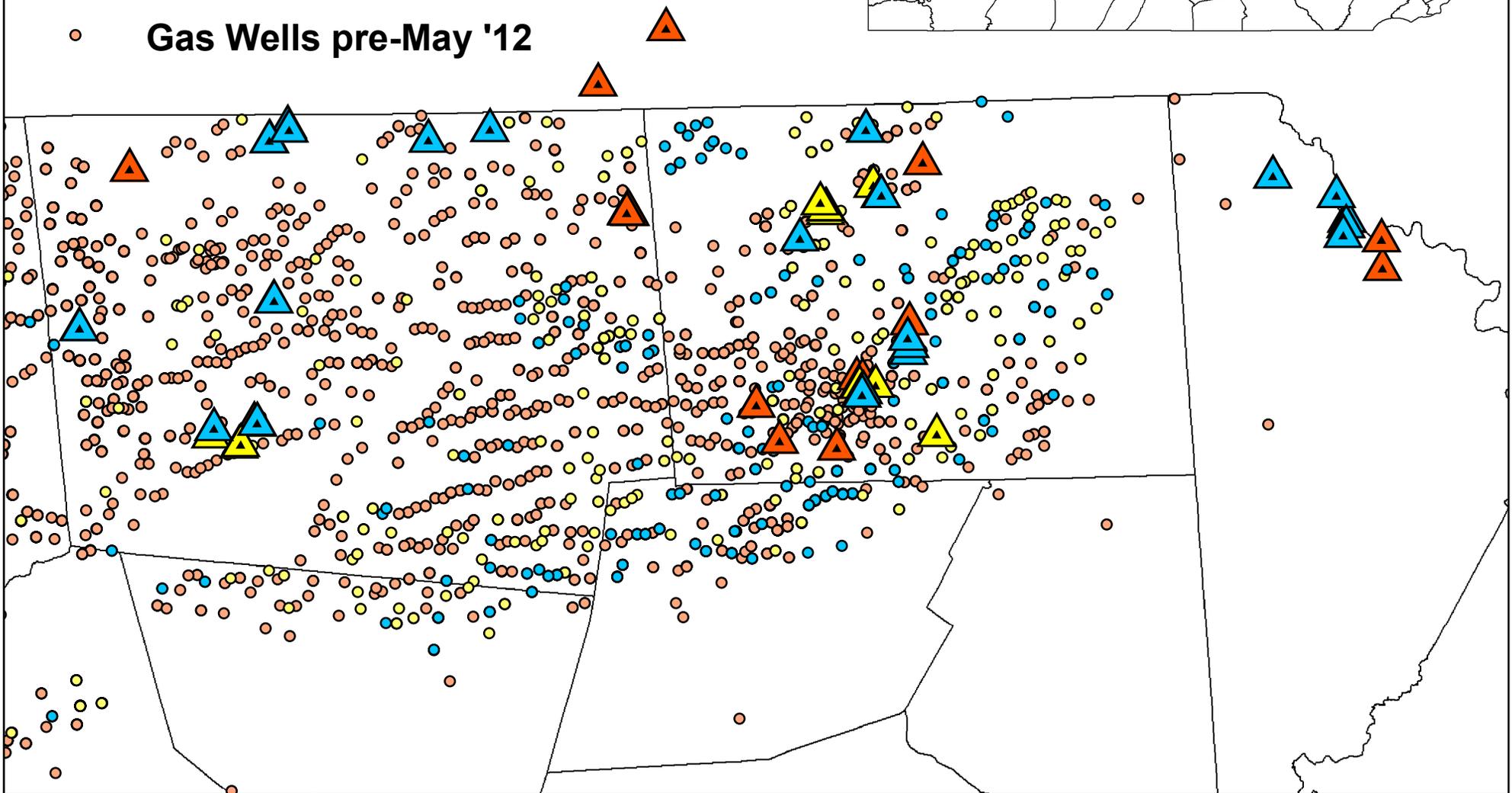
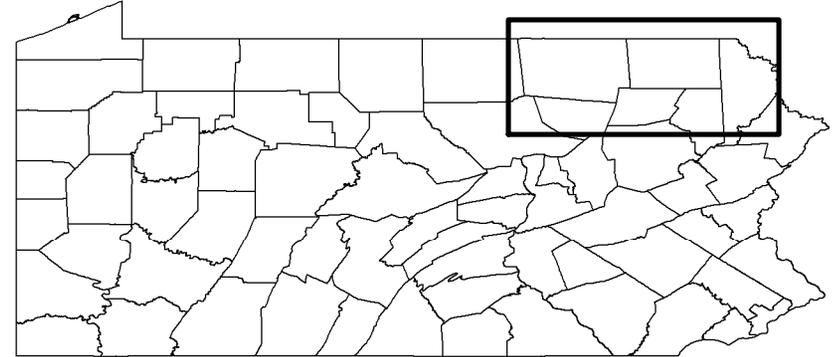Groundwater Samples June 2014 (n = 25)
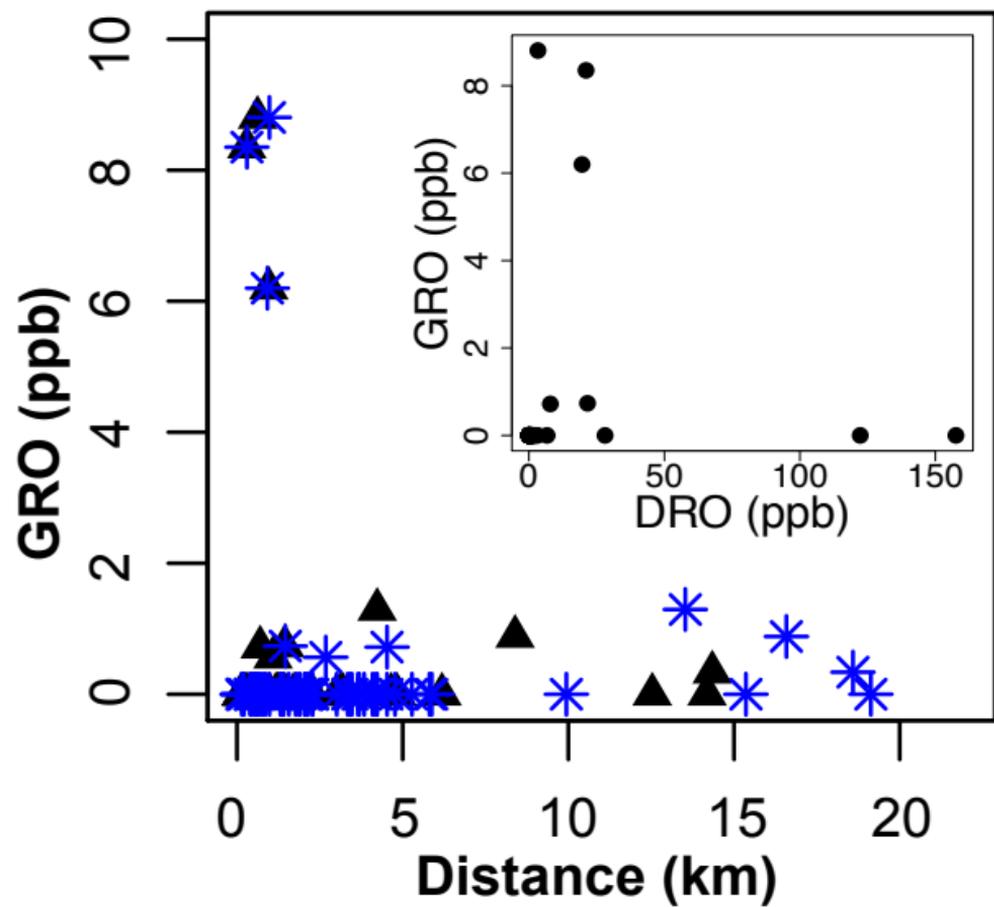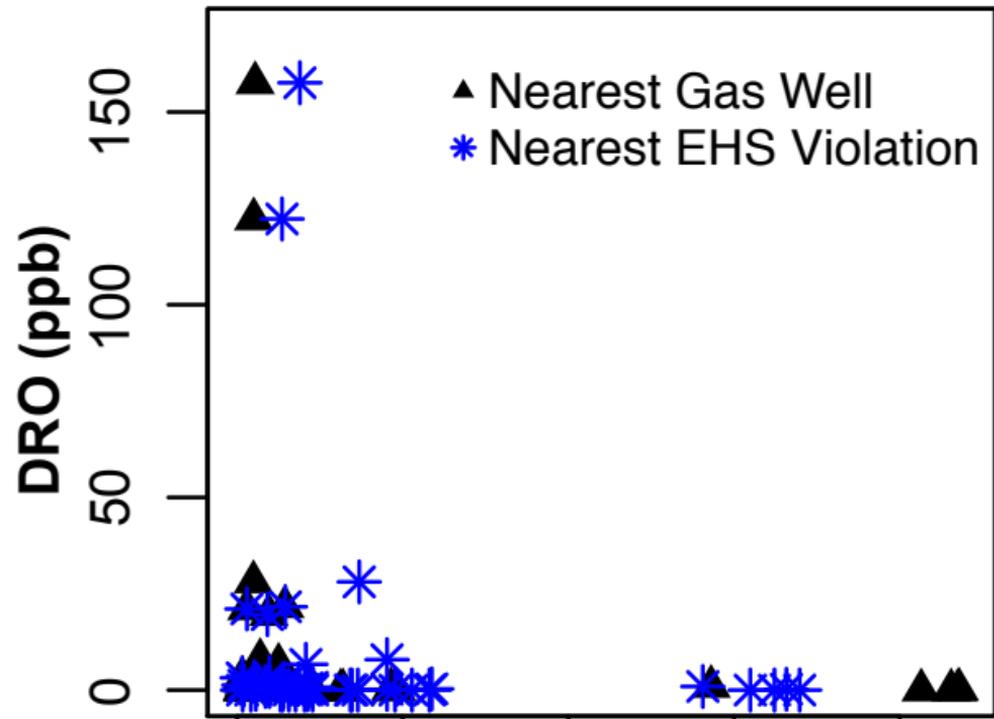
Groundwater Samples July 2013 (n = 16)

Groundwater Samples May 2012 (n = 23)

Gas Wells July '13 - June '14

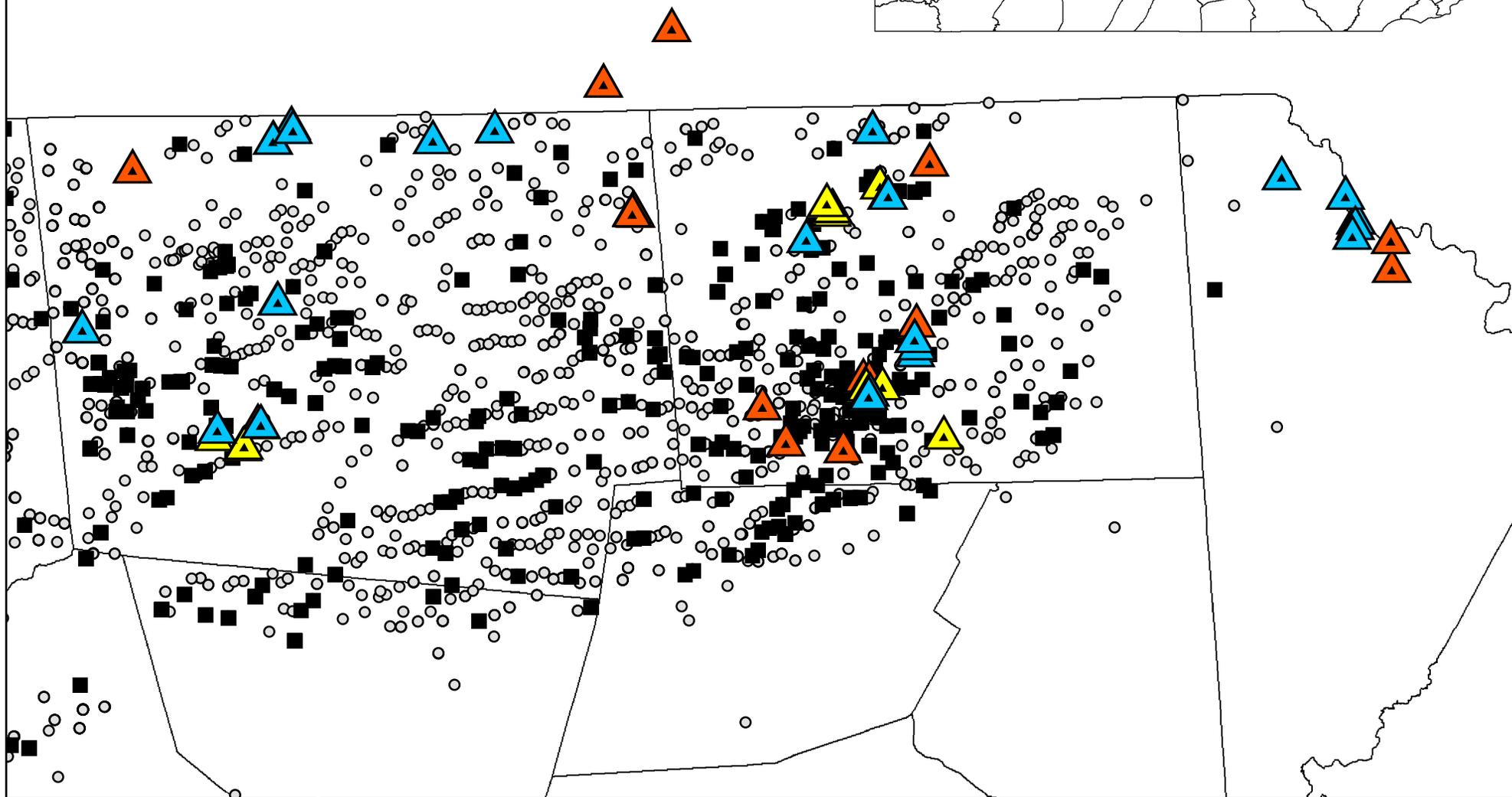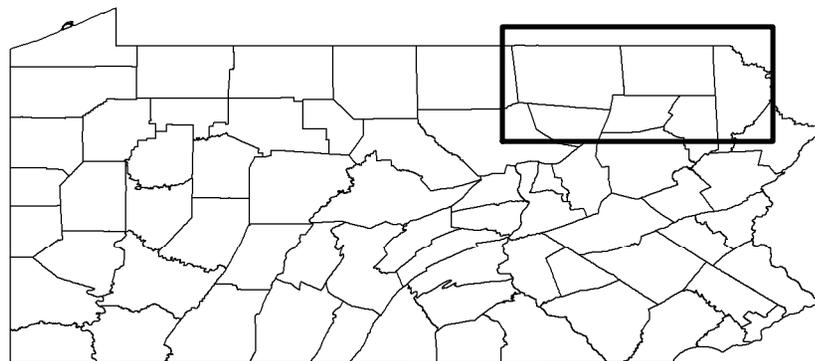Gas Wells May '12 - July '13

Gas Wells pre-May '12

Groundwater Samples June 2014 (n = 25)
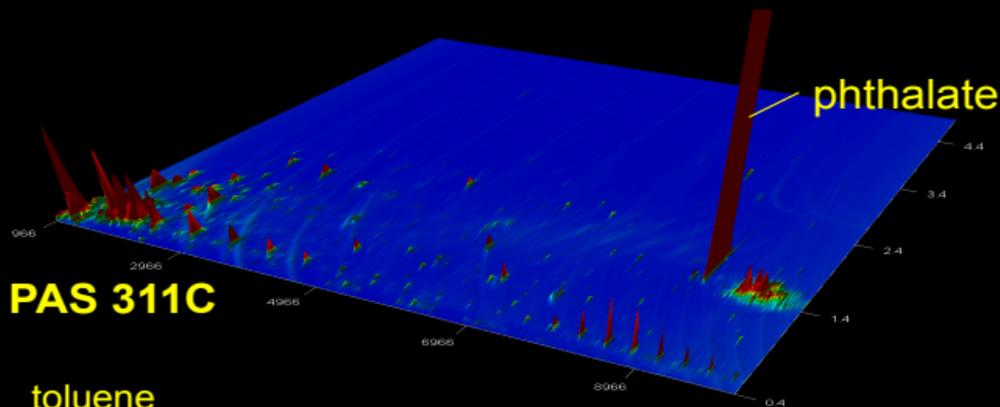
Groundwater Samples July 2013 (n = 16)

Groundwater Samples May 2012 (n = 23)
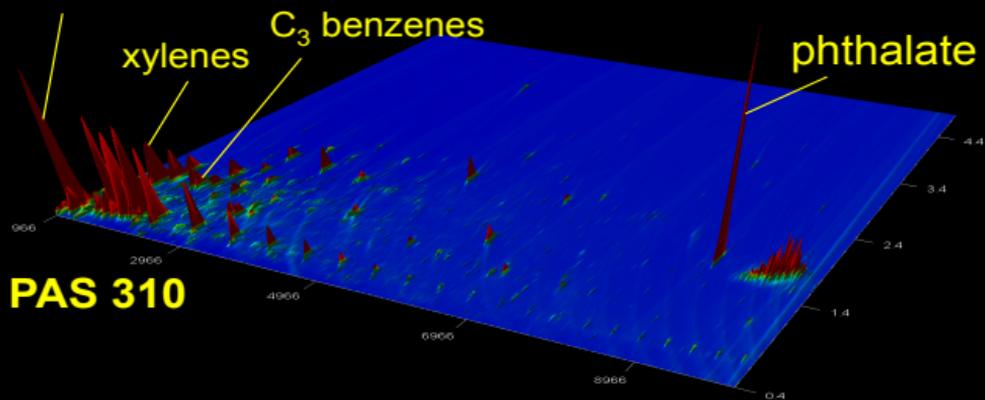
Unconventional Gas Wells

EHS Violations

phthalate

PAS 311C

toluene

xylenes

C$_3$ benzenes

phthalate

PAS 310

Salt Spring