Environmental
Science
Processes & Impacts

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c8em00385h

Assessing changes in groundwater chemistry in landscapes with more than 100 years of oil and gas development†

Tao Wen,‡[a] Amal Agarwal,‡[b] Lingzhou Xue,*[b] Alex Chen,[b] Alison Herman,[c] Zhenhui Li[d] and Susan L. Brantley*[ac]

Received 22nd August 2018
Accepted 20th December 2018

DOI: 10.1039/c8em00385h

rsc.li/espi

With recent improvements in high-volume hydraulic fracturing (HVHF, known to the public as fracking), vast new reservoirs of natural gas and oil are now being tapped. As HVHF has expanded into the populous northeastern USA, some residents have become concerned about impacts on water quality. Scientists have addressed this concern by investigating individual case studies or by statistically assessing the rate of problems. In general, however, lack of access to new or historical water quality data hinders the latter assessments. We introduce a new statistical approach to assess water quality datasets – especially sets that differ in data volume and variance – and apply the technique to one region of intense shale gas development in northeastern Pennsylvania (PA) and one with fewer shale gas wells in northwestern PA. The new analysis for the intensely developed region corroborates an earlier analysis based on a different statistical test: in that area, changes in groundwater chemistry show no degradation despite that area's dense development of shale gas. In contrast, in the region with fewer shale gas wells, we observe slight but statistically significant increases in concentrations in some solutes in groundwaters. One potential explanation for the slight changes in groundwater chemistry in that area (northwestern PA) is that it is the regional focus of the earliest commercial development of conventional oil and gas (O&G) in the USA. Alternate explanations include the use of brines from conventional O&G wells as well as other salt mixtures on roads in that area for dust abatement or de-icing, respectively.

Environmental significance

Intense drilling and high-volume hydraulic fracturing in areas of shale gas development sometimes impact local groundwater. However, both historical and new data are often not available to assess impacts on groundwater quality. Here a comparison of impacts in two areas of the largest shale gas play in the USA reveal decreases in some solute concentrations in the heavily developed region but increases in the less intensely developed region. The latter region is the site of some of the oldest commercial development of conventional oil and gas (O&G) in the world. Historical O&G development or regional differences in the use of salt mixtures and production brines on roads for de-icing or dust abatement might explain the slight concentration changes in groundwater.

Introduction

Hydraulic fracturing has been used to open up the permeability in hydrocarbon reservoirs at least since the 1940s in the USA. Recently, a version of this technique – high volume hydraulic fracturing (HVHF) – has been successful in stimulating gas production from the Marcellus and other shales in the northeastern USA. This development in highly populated areas has increased concerns on the part of the public about possible impacts on water resources. In particular, a debate about the environmental impact of shale gas development activities on water resources has grown to become particularly controversial since the onset of Marcellus drilling in 2004 in Pennsylvania (PA).[1–11] While investigating this controversy, few researchers have documented impacts directly related to or caused by HVHF itself: rather, problems that have been documented generally involve leakage because of casing or cementing issues, faulty impoundments or containers, fluid spills, and well blowouts, all of which have been more common causes of water contamination related to shale gas drilling and production activities.[1,2] To date, most published papers have focused on case studies

---

[a] Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA. E-mail: sxb7@psu.edu

[b] Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA. E-mail: lzxue@psu.edu

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

[d] College of Information Sciences and Technology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8em00385h

‡ These authors contributed equally to this work.

about incidents[8,11–13] and few studies have been able to assess the overall incidence of water contamination over time.

Two reasons for the lack of studies assessing temporal trends of groundwater quality are the lack of publicly released groundwater data both before and after development.[14] In the northeastern USA, only one published study[15] has reported results from a statistical comparison of a moderately large dataset of groundwaters sampled from the same wells before and after shale gas development. That study did not release geographic coordinates along with water quality data for the groundwater wells. Both spatial and temporal comparisons of groundwater quality with respect to shale gas development are needed to answer the public's questions about the frequency and extent of environmental problems related to the activities of the shale gas industry.

One possible approach to look for broad statistical trends is to assess large water quality datasets available for the region of the Marcellus Formation that are collected by oil and gas (O&G) companies to establish the water chemistry pre-drill baseline. These data are now becoming available to the public,[16,17] and increasingly are published with geographic coordinates.[7,8,14,18–20] Studies based on these large water quality datasets (~1000s to ~10000s samples) allow investigation of the temporal and spatial trends of water quality in production areas of gas from unconventional reservoirs (*i.e.*, reservoirs with lower permeability that require HVHF) because, although they are collected as "pre-drill" data for a new well that is planned, they are almost always collected in areas with already-drilled oil or gas wells.[5–7,16–19,21] In such cases, the older wells may be drilled into unconventional or conventional, higher-permeability, reservoirs. Here, these are referred to as unconventional or conventional wells, respectively. In addition, sometimes older groundwater quality datasets are available in the targeted area for comparison. For example, Siegel *et al.*[5] pointed out that groundwater quality in PA was unchanged over time before and after shale gas production based on datasets available from one gas company and other data providers.

Similarly, Wen *et al.*[7] compared water quality documented by datasets in Bradford County in Northeastern PA (NE PA) from before and after the marked increase of shale gas production in that county. They concluded that concentrations decreased for total dissolved solids (TDS), iron (Fe), manganese (Mn), and sulfate (SO_4^{2-}); pH increased; and concentrations of arsenic (As), lead (Pb), barium (Ba), chloride (Cl), and sodium (Na) showed no statistically significant change. This observation thus did not document degradation of groundwater from shale gas development in Bradford County (NE PA), the county with the second highest number of shale gas wells in PA (~1385; including all spudded wells regardless of the well status). Bradford (NE PA) also has 66 conventional oil and gas wells (including all such spudded wells regardless of status).[22]

One observation from statistical and spatial analyses of so-called "pre-drill" data is that groundwater chemistry varies in different sub-areas because of variations in geological faults, folds, topography, lithology, and land use – including the presence of O&G wells.[7,18,19] In this study, we wanted to look at water quality in different parts of PA in a targeted approach to investigate the effect of both conventional and unconventional O&G wells. We took advantage of the fact that PA is the state that hosted the oldest commercially developed oil well in the USA (emplaced in 1859). Specifically, we wanted to test if groundwater quality shows similar or different trends over time for two parts of the Marcellus gas play that have disparate histories of O&G development. We present new data from Mercer County in Northwestern PA (NW PA), an area that has not been a focus for unconventional shale gas drilling but is near the oldest commercial oil well in neighboring Venango County. In Mercer County (NW PA), the state regulator estimates more than 3780 conventional oil or gas wells are either operating, or have been plugged, abandoned, or orphaned.[22,23] The county averages 2.14 conventional O&G wells per km^2 but only 0.03 unconventional O&G wells per km^2. For comparison, we also re-visit NE PA, the location of Bradford County, which is one of the two most intensely developed areas of natural gas development in the Marcellus gas play today. That county averages 0.46 unconventional O&G wells per km^2 but only 0.02 conventional O&G wells per km^2.

We first present a temporal analysis of the data from NW PA.[20] In particular, we compare two groundwater datasets (hereafter "NW PA datasets") that were collected pre-2000 and post-2010 in Mercer County (NW PA). These groundwater quality data, released by the oil and gas regulator in PA (Pennsylvania Department of Environmental Protection or PA DEP) to the public for the first time in this study, are the only data we have found closely related to oil and gas development for Mercer County (NW PA) (data archiving by the PA DEP is variable across the state). We then revisit both the data from the NE and NW PA using a more advanced statistical technique than published previously.[7]

The objectives of this study were to (1) investigate groundwater quality data collected in NW PA (*i.e.*, Mercer County) before and after the onset of unconventional gas drilling in 2012 in that county; (2) compare the temporal trends in groundwater chemistry in NW and NE PA to explore the effects of development of conventional and unconventional hydrocarbon reservoirs; and (3) provide possible explanations for the differences in trends between the two study regions. We also introduce a new statistical test that is useful for comparison of datasets of differing data volume and variance.

Methods and materials

Water quality data

Analysis of groundwater quality data in PA has been made possible following a memorandum of understanding (MOU) signed by Penn State University and the PA DEP in 2013. For this study, PA DEP provided groundwater quality data from oil and gas companies as Excel tables or scanned copies of printed laboratory reports. We manually typed or collated these data into a master database while removing confidential information (*e.g.*, names and addresses of homeowners) and then published the data into the Shale Network online database (DOI: 10.4211/his-data-shalenetwork) as well as the Penn State University Data Commons (DOI: 10.18113/d3967x). Multiple rounds of verification of the data were performed prior to publication as described previously, necessitating up to an hour per laboratory report for the procurement, compilation, cleaning, and management of these water chemistry data.[7] None of these water chemistry data were previously accessible to the public.

NW PA water quality data. For NW PA, the DEP provided the only two datasets of water chemistry from O&G companies that are available. These two batches were collected by O&G companies before 2000 and after 2010 and are hereafter noted as the pre-2000 and post-2010 NW PA datasets (Table 1). The pre-2000 dataset summarizes 1604 groundwater samples for up to 15 analytes (pH, hardness, turbidity, alkalinity, specific conductance, total dissolved solids (TDS), K, Mg, Ca, Cl, Na, SO_4, CH_4, Fe, Mn) in central Mercer County (NW PA) that were mostly collected from 1985 to 1999 (Fig. 1). These pre-2000 data were shared in spreadsheet format without the original laboratory report to cross-check. The post-2010 dataset consisted of 259 pre-drill groundwater samples (244 water wells and 15 springs) collected from 2012 to 2015 in central Mercer County (NW PA) in areas that overlapped with the areas sampled in the pre-2000 dataset (Fig. 1). As many as 43 analytes were reported for each sample.

Hydrocarbon production in Mercer County (NW PA) is mostly dominated by oil or gas extraction from conventional reservoirs (referred to here as conventional wells); no coal mines are present in the study area in central Mercer (NW PA) (see Fig. 1 for coal mining outside of the center part of the county).[24] According to PA DEP records,[22] 3780 conventional wells, including 110 documented orphaned and abandoned wells,[23] had been drilled by 2017 while the first unconventional well was drilled in 2012 and only 61 unconventional wells were drilled in total before 2015.[22] Among all the conventional wells, the PA DEP records show known spudded dates for 3391 while the other 389 wells are not recorded with a spudded date (they are coded with an arbitrary code, *i.e.*, spudded in 1/1/1800).[22] However, since some investigators have estimated that the number of undocumented and unmapped orphaned and abandoned wells overall in PA could be one order of magnitude higher than those that have been mapped as abandoned or orphaned, it is likely that the number of such wells has been underestimated by at least a factor of ten in Mercer County (NW PA).[25]

In this analysis for NW PA, we discuss only the 15 analytes that were reported in both the pre-2000 and post-2010 data.

Table 1 Statistical summary of NW PA pre-2000 and post-2010 datasets

Analyte	Unit	*n*	Number above reporting limit	EPA tandard[a]	Min	Median	95th percentile	Max	Mean	1SD[b]
NW PA pre-2000 dataset										
K	mg L^{-1}	1576	1575	—	0.010	2	7.28	133	3	5
Mg	mg L^{-1}	1603	1602	—	0.002	11	26.4	240	13	1×10^1
Ca	mg L^{-1}	1334	1333	—	0.010	44	105	368	48	3×10^1
Na	mg L^{-1}	1575	1574	—	0.010	10	140	821	32	6×10^1
Fe	mg L^{-1}	1287	1224	0.3	0.002	0.3	7.00	529	2	2×10^1
Mn	mg L^{-1}	1288	1127	0.05	0.010	0.05	0.80	61	0.3	2
Cl	mg L^{-1}	1604	1582	250	0.100	9	106	1214	25	6×10^1
SO_4	mg L^{-1}	1603	1596	250	0.040	34	105	535	43	4×10^1
CH_4	mg L^{-1}	1185	275	—	0.001	0.01	0.500	50	0.4	2
TDS	mg L^{-1}	1604	1604	500	20.0	245	553	2275	274	2×10^2
Total alkalinity	mg L^{-1} as $CaCO_3$	1603	1602	—	0.040	154	319	591	159	9×10^1
pH	pH unit	1604	1604	6.5–8.5	4.9	7.5	8.7	9.1	7.5	0.8
Hardness	mg L^{-1}	1576	1571	—	0.150	168	370	1400	177	1×10^2
Specific conductance	μS cm^{-1}	1604	1604	—	42.6	421	878	4760	457	3×10^2
Turbidity	NTU	1473	1436	—	0.020	2	46	7100	18	2×10^2
NW PA post-2010 dataset										
K	mg L^{-1}	252	226	—	0.470	2	5	13	2	2
Mg	mg L^{-1}	243	230	—	0.030	13	26	54	14	8
Ca	mg L^{-1}	259	252	—	0.130	56	110	180	56	3×10^1
Na	mg L^{-1}	259	259	—	1.32	10	147	351	33	5×10^1
Fe	mg L^{-1}	252	202	0.3	0.010	0.7	8	71	2	7
Mn	mg L^{-1}	259	210	0.05	0.0	0.1	0.6	2.7	0.2	0.3
Cl	mg L^{-1}	259	251	250	0.960	15	159	550	41	7×10^1
SO_4	mg L^{-1}	259	248	250	0.150	31	64	103	33	2×10^1
CH_4	mg L^{-1}	259	91	—	0.0	0.01	0.7	10	0.3	1
TDS	mg L^{-1}	259	259	500	54.5	255	616	1400	301	2×10^2
Total alkalinity	mg L^{-1} as $CaCO_3$	259	259	—	21.7	169	276	372	171	7×10^1
pH	pH unit	219	219	6.5–8.5	6.4	7.4	8.3	8.8	7.4	0.6
Hardness	mg L^{-1}	252	242	—	0.200	185	369	496	187	1×10^2
Specific conductance	μS cm^{-1}	219	219	—	9.90	426	919	2110	485	3×10^2
Turbidity	NTU	259	191	—	0.120	4	42	221	11	2×10^1

[a] For turbidity, EPA primary Maximum Contaminant Level (MCL) is reported. For sodium, EPA drinking water health advisory for individuals on a restricted sodium diet is shown. For other analytes, either EPA secondary MCL or no standard is established. [b] 1SD = 1 standard deviation.

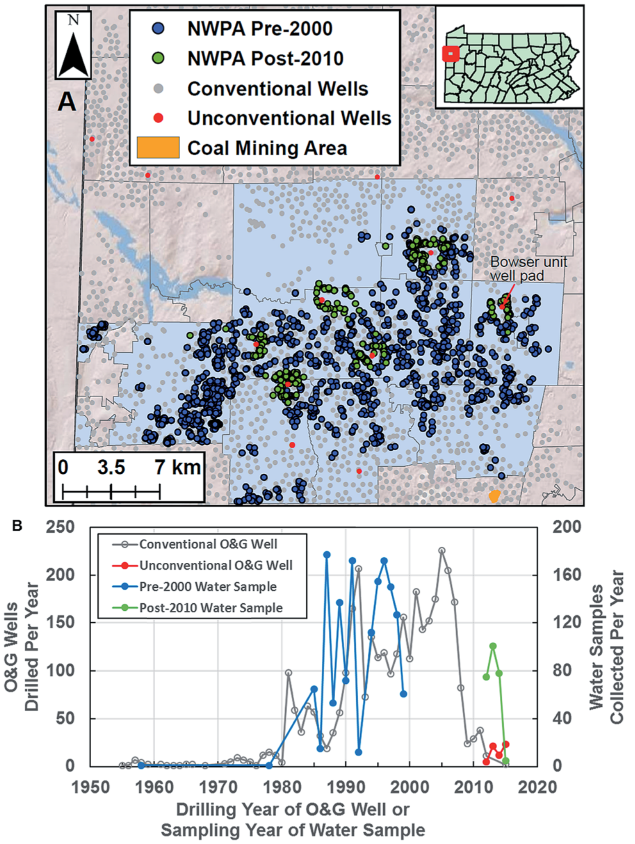

Fig. 1 (A) Locations of the 1863 water samples for the NW PA groundwater datasets in Mercer County (labelled NW PA). Townships in central Mercer (NW PA) are shown in blue: Coolspring, Delaware, East Lackawannock, Fairview, Findley, Fredonia, Hermitage, Jackson, Jackson Center, Jefferson, Lackawannock, and Mercer. Conventional (n = 3391, *i.e.*, conventional wells with known spudded date) and unconventional (n = 61) wells[22,23] and very minor coal mining areas (in the lower right corner) are also indicated on the map.[24] The 389 conventional wells without known spudded dates (coded as '1/1/1800' in PA DEP database[22]) are not plotted here. The Bowser unit well pad (discussed in main text) is indicated by the red arrow. (B) Number of conventional (gray circle) and unconventional (red dot) wells drilled per year along with groundwater samples collected per year in Mercer County (NW PA) and reported in the NW PA dataset are plotted *versus* year. Most conventional wells were drilled between 1980 and 2010. For both (A) and (B), wells reflect the total number of drilled wells regardless of the well status (*e.g.*, active, orphaned, plugged, and abandoned).

Most of the post-2010 data were collected as "pre-drill" samples and were received in the format of the original commercial analytical laboratory report and then were recorded into spreadsheets and checked. Importantly, these samples are only pre-drill with respect to a proposed new well: they can be considered "post-drill" for all other wells already in place in the area.

The only post-2010 data that were not noted as pre-drill were seven groundwater samples collected as a post-drill follow-up for 4 sites. These seven groundwater samples were collected as post-drill samples in Jackson Township around the Bowser unit well pad (Fig. 1A). The laboratory reports of these 7 groundwater samples disclosed no information why these samples were collected. We suspected that these 7 water samples might have been collected for an investigation as a response to a local environmental complaint; however, we could not locate any filed complaint from Jackson Township from 04/01/2014 (10 days before the collection of the oldest post-drill sample among these seven) to 03/01/2015,[26] *i.e.*, the date range when these seven samples were collected. Water chemistry for these 7 post-drill samples as well as associated 4 pre-drill groundwater samples are all listed in Table S1.† The comparison of these pre-drill and post-drill samples indicated no degradation of groundwater quality at these four sites; out of caution and lack of knowledge about the reasons for these extra post-drill samples, they were excluded from analysis.

Statistical tests. The distributions of concentrations are generally skewed; therefore, we often interpret them with respect to medians instead of means and we use non-parametric statistical tests for comparisons of distributions at a significance level = 0.05 (see also ESI†). These tests can determine whether distributions are the same (the null hypothesis) or that the probability is greater than 0.5 that a randomly selected value from one distribution is larger than or smaller than a random value from the second distribution at the 95% confidence level.

Here, we compare the pre-2000 and post-2010 datasets using both Wilcoxon–Mann–Whitney (WMW) rank sum and Brunner-Munzel (BM) tests using the statistical package in R 3.3.3.[27] Both tests are further discussed in ESI† and the Results and discussion sections. The strategy of analysis proceeded according to this workflow: (1) first we tested the null hypothesis of no change between distributions (alternative hypothesis was that distributions were different): if p was greater than 0.05 then we could not reject the null hypothesis; (2) if p was less than 0.05 for (1) then null hypothesis (1) was rejected and we did two more tests. These tests were based on the following null hypotheses: (a) the distributions either increased or were unchanged (alternative hypothesis was that the distribution decreased) and (b) the distributions either decreased or were unchanged (alternative hypothesis was that the distribution increased) between pre-2000 and post-2010. If p was less than 0.05, we rejected the null hypothesis at step (2). Thus, if the distributions did not change, then we calculated a p value greater than 0.05 for null hypothesis at step (1) and stopped. In contrast, if the distributions did change, then we calculated a p value less than 0.05 for rejection of the null hypothesis at step (1) and a p value less than 0.05 for rejection of the hypothesis at step (2).

In addition to a comparison of distributions for the pre-2000 and post-2010 datasets, the temporal trends of groundwater chemistry in NW PA within each of these two time periods (*i.e.*, pre-2000 and post-2010) were also evaluated using the nonparametric Spearman's rank correlation test.

NE PA water quality data. In contrast to Mercer County (NW PA), Bradford County (NE PA) is the site of over a thousand unconventional wells but only 66 conventional wells that have been drilled through 2017.[7,22] Here we briefly summarize the datasets from pre-2000 (108 values) and post-2010 (11 156 values) from Bradford County in NE PA that were reported previously[7] and published online.[20] Hereafter, we refer to these as the "NE PA datasets". Wen *et al.*[7] performed the WMW test on these datasets for 9 analytes: pH, TDS, Fe, Mn, sulfate, Pb, Ba, Cl, and Na. Due to the high fraction of censored data for As both pre-2000 and post-2010, Wen *et al.*[7] did not conduct statistical tests but simply compared the rate that As concentrations failed the EPA standard for As between the two datasets. We summarize the results of the WMW test on the NE PA datasets and present a new analysis using the BM test for comparison.

Results

Central Mercer County (referred to here as NW PA) lies within the Shenango River sub-watershed of the Ohio River Basin. The county is also entirely within the northwestern glaciated Pittsburgh plateau section of the Appalachian Plateaus Province.[28] The study area is generally covered by Wisconsin glacial deposits[29] and underlain by Pennsylvanian and Mississippian sandstone, shale, and limestone bedrock of the Pottsville, Allegheny, and, to a lesser extent, Shenango formations.[6,26,28]

A total of 15 analytes (pH, hardness, turbidity, alkalinity, specific conductance, TDS, K, Mg, Ca, Cl, Na, SO_4, CH_4, Fe, Mn) from this NW PA dataset were grouped into three groups: cation, anion, or other (Table 1). EPA standards (primary Maximum Contaminant Level or MCL, secondary MCL, and health advisory level) were also included in Table 1 for comparison. The data include $n = 1604$ for the pre-2000 and $n = 259$ for post-2010. Table 2 summarizes the results of the WMW tests to determine if the distributions were statistically the same or different for pre-2000 *versus* post-2010.

The WMW test indicated that the distribution of methane values increased with time between the two datasets (p less than 0.05). This can be interpreted as follows: if a value of methane was randomly selected from post-2010 group, it has a greater than 50% chance of being larger than a randomly selected value from the pre-2000 group at the 95% confidence level. This test is not a strong test for methane, however, because the majority (65–77%) of methane data in the compiled (pre-2000 and post-2010) NW PA data, were censored, *i.e.*, below reporting limits and nine different reporting limits were operable among the laboratories. Such a temporal increase in distribution of methane concentrations might thus instead reflect the differences in percentage of censored values among all values at the different times. Following the approach of Wen *et al.*[7] for arsenic, we therefore compared only the fraction of measurements in each dataset that lie above the USA Department of Interior suggested alarm level of 10 mg L^{-1} CH_4:[31] 10 out of 1185 values (0.8%) for the pre-2000 dataset *versus* 1 out of 259 values (0.4%) for the post-2010 dataset. Given the well-known issues in measuring methane at such higher levels[32] and the small discrepancy in values above reporting limits between the two time periods, however, we do not make conclusions about methane concentrations over time.

Most of the other analytes had at least 1100 or 190 reported values above reporting limits (pre-2000 and post-2010, respectively). The distributions of most cation concentrations showed small but statistically significant increases from pre-2000 to post-2010 (p less than 0.05) according to the WMW test (Table 2). A few cations did not show this increase: the distribution for K decreased (p less than 0.05); the distribution for Mn showed no change ($p = 0.70$); and Na showed no change ($p = 0.06$). For the two anions reported in the NW PA data, the WMW test indicated that the distributions of Cl concentrations increased from pre-2000 to post-2010 while sulfate decreased (p less than 0.05; Table 2). Finally, the distributions of TDS, hardness, and specific conductance in NW PA datasets also increased (p less than 0.05), consistent with the increases in distributions of concentrations of some of the major cations and anions, *e.g.*, Ca, Mg, and Cl. For example, the median of the hardness values increased by 17 mg L^{-1} from pre-2000 to post-2010 and over the same time period, the median increased by 11 mg L^{-1} for Ca concentration (Table 1). The distributions of alkalinity in NW PA datasets also increased according to the WMW test (p less than 0.05) while pH showed no change ($p = 0.07$, Table 2).

A more stringent statistical test

Overall, the statistical comparison of groundwater quality data from pre-2000 and post-2010 in central Mercer County (NW PA) was consistent with groundwater quality in this area becoming slightly more saline but remaining well buffered (with slightly increased alkalinity) at a constant pH over the time interval. This conclusion contrasts the earlier conclusion of Wen *et al.*[7] for water quality in Bradford County in NE PA from before and after the marked increase of shale gas production in that county. They concluded that the distributions of concentrations decreased for TDS, Fe, Mn, and sulfate while the distribution of pH values increased between the pre-2000 and post-2010 datasets. They inferred there might have been slight overall improvement in groundwater quality.

However, the pre-2000 and post-2010 data for water quality in both NW PA and NE PA differ in size by a factor of at least 5 (depending upon the analyte, see Table 1 and Wen *et al.* (2018)[7]). The distributions of reported values for many of the analytes in the pre-2000 and post-2010 datasets also have unequal variances (Table 2). Other researchers have shown that the WMW test can fail to be a fair test for distribution[33] for datasets with extremely small size (less than 50) and extreme ratios of variance (*e.g.*, a factor of 10). For such cases, a more sophisticated statistical test (the Brunner-Munzel (BM) test) has been used (see also ESI†).[34,35] Our datasets in NW PA and NE PA do not show such small size or extremely distinct variance: groundwater quality data in NW PA and NE PA all have reasonably large size (greater than 100) and the ratios of variance for many analytes are less than 2 (Table 2). Nonetheless, we decided to apply the stronger BM test on both the NW PA and NE PA datasets to test the WMW test results because of the large discrepancies in data volume (Table 2).

The BM tests for both NW PA and NE PA datasets confirm the temporal trends for all analytes with two exceptions: pH and Na in NW PA data (Table 2). The BM tests for NW PA show that the distributions of Na concentrations increase (p less than 0.05) and that of pH decrease (p less than 0.05), which differ from the

Table 2 Summary of statistical test between pre-2000 and post-2010 datasets in both NW PA and NE PA areas

Analyte	Unit	Equal variance[a]	WMW test[b]	BM test[b]
NW PA				
K	mg L^{-1}	No	↓	↓
Mg	mg L^{-1}	Yes	↑	↑
Ca	mg L^{-1}	Yes	↑	↑
Na	mg L^{-1}	Yes	No change	↑
Fe	mg L^{-1}	No	↑	↑
Mn	mg L^{-1}	No	No change	No change
Cl	mg L^{-1}	Yes	↑	↑
SO$_4$	mg L^{-1}	No	↓	↓
CH$_4$	mg L^{-1}	No	↑	↑
TDS	mg L^{-1}	Yes	↑	↑
Total alkalinity	mg L^{-1} as CaCO$_3$	Yes	↑	↑
pH	pH unit	Yes	No change	↓
Hardness	mg L^{-1}	Yes	↑	↑
Specific conductance	μS cm^{-1}	Yes	↑	↑
Turbidity	NTU	No	↑	↑
NE PA				
pH	pH unit	Yes	↑	↑
TDS	mg L^{-1}	No	↓	↓
Fe	mg L^{-1}	No	↓	↓
Mn	mg L^{-1}	No	↓	↓
SO$_4$	mg L^{-1}	No	↓	↓
As	mg L^{-1}	No	—	—
Pb	mg L^{-1}	No	No change	No change
Ba	mg L^{-1}	No	No change	No change
Cl	mg L^{-1}	No	No change	No change
Na	mg L^{-1}	No	No change	No change

[a] Equal variance means standard deviations between two datasets differ within a factor of $\sqrt{2}$. [b] ↑ indicates the distribution is increasing at significance level of 5% while ↓ indicates the distribution is decreasing at significance level of 5%. When variances are equal, the comparison of distributions can be interpreted as the comparison of medians of two datasets.

WMW test results. However, p values of these two tests are only slightly different for these analytes: $p = 0.0264$ (BM) *vs.* $p = 0.0673$ (WMW) for pH and $p = 0.0496$ (BM) *vs.* $p = 0.0595$ (WMW) for Na concentrations. These results generally confirm that the conclusions are robust from the statistical perspective that the data show slight increases in distributions of salt and some metal concentrations in NW PA groundwaters *versus* slightly lower distributions of concentrations in NE PA groundwaters.

Discussion

Comparing NW and NE PA groundwater

The distributions of sulfate concentrations in both the NW and NE PA datasets decreased from pre-2000 to post-2010 (Table 2).[7] Such decreases in both NE and NW PA are consistent with the state-wide trend of decreasing sulfate concentrations in PA streams.[36] This change has been attributed mostly to the effects of the decline of coal production (and associated acid mine drainage) and the Clean Air Act (CAA) and its amendments since the 1970s that contributed to the amelioration of acid rain in the state.[7,36,37] The similarity between observations of trends in sulfate concentrations in streams in PA and in groundwaters in NW and NE PA may be a good indicator that the trends are temporal indicators of change related to state-wide rather than regionally distinct changes.

While the sulfate changes may be explained by improvements in release of sulfur compounds related to mining or burning coal, the NW PA waters became slightly more saline and Fe-rich while the NE PA waters stayed constant or decreased in concentrations of salts and metal elements. Specifically, the distributions of concentrations of Na, Cl, hardness, and TDS increased in NW PA from pre-2000 to post-2010, while in NE PA changes in Na and Cl distributions were not detected and TDS decreased (Table 2). Similarly, distributions of Fe concentrations increased and Mn showed no change from pre-2000 to post-2010 in NW PA datasets while they decreased in NE PA (Table 2).[7]

Possible explanations for changes in groundwater chemistry in central Mercer County (NW PA)

These discrepant changes in TDS, Na, Cl, Mn, and Fe in NE PA *versus* NW PA might reflect changes in average groundwater chemistry over time. However, because the same water wells are not sampled in each time period, changes that are inferred to indicate temporal trends could instead reflect hidden variables. For example, if the water wells that were sampled in the earlier time period were located in one geological formation and the second set in another formation, then a systematic difference in location could be mis-interpreted as a temporal trend. Alternately, contributions of Cl from natural sources of Appalachian

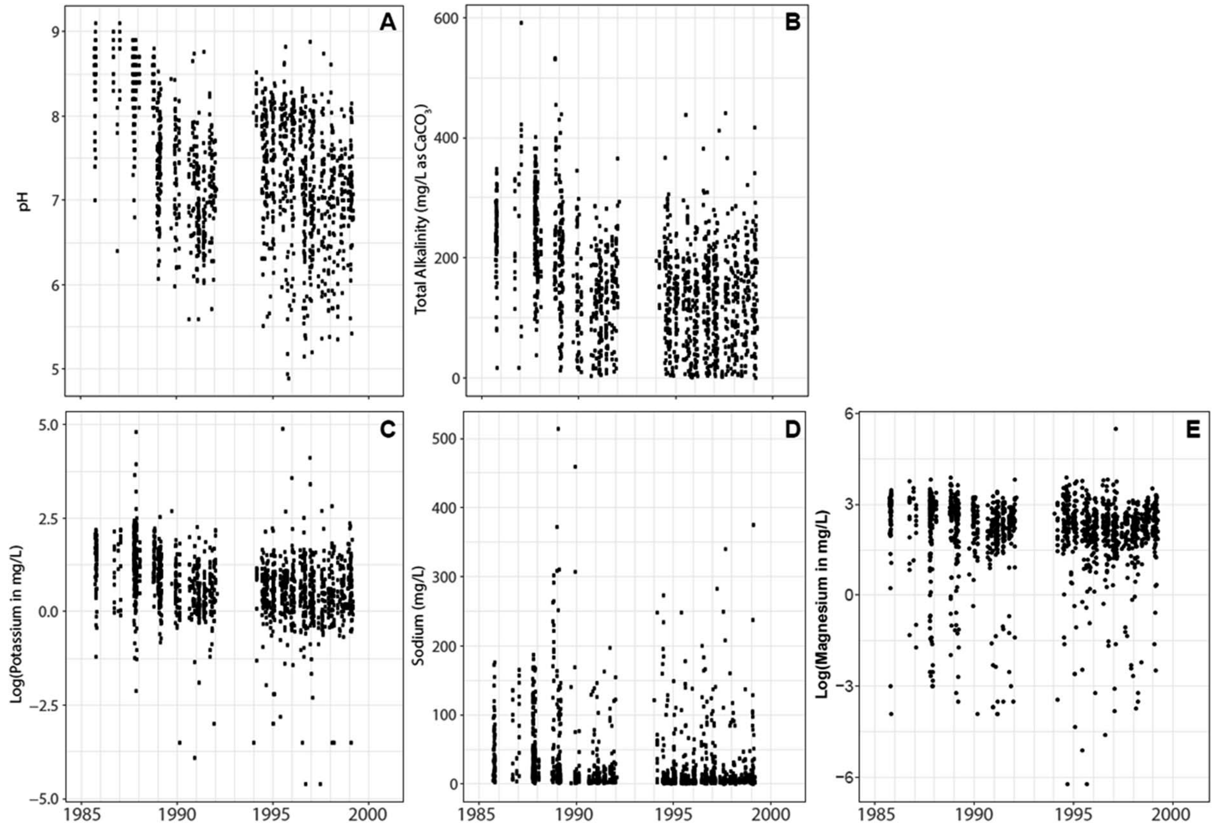

Fig. 2 (A) pH, (B) total alkalinity (in mg L^{-1} as $CaCO_3$), (C) K (in mg L^{-1}), (D) Na (in mg L^{-1}), and (E) Mg (in mg L^{-1}) in NW PA pre-2000 dataset plotted as a function of sampling date. Note that K and Mg are shown on a log scale because of the large range in measured values. All of these five analytes show a decreasing temporal trend according to Spearman's rank correlation test ($p < 0.05$). These five analytes are the only chemical constituents showing statistically significant temporal trends in the pre-2000 dataset. No chemical analytes in the post-2010 dataset show statistically significant temporal changes (significance level = 0.05; see also Fig. S5 and Table S2†).

brine salts[1,5] might be more prevalent in some parts of PA where the density of natural faults and fractures is higher or where the depth to natural brines is shallower.[1,30] Likewise differences in sampling conditions inside a household (*e.g.*, flow velocity during sampling) at different times can also affect parameters such as methane or turbidity.[5] Although we cannot evaluate all of these possible factors, we nonetheless discuss a variety of attributes for the NW PA and NE PA sites as possible explanations in the following sections.

Differences in wetness over time. Larger or lower precipitation in wetter or drier years might dilute or concentrate analytes in groundwaters respectively. To evaluate this possibility, we tested for temporal changes in regional wetness by selecting a USGS stream gauging station in each county: site number 03102850 on the Shenango River in Mercer County (NW PA) and site number 01531500 on the Susquehanna River in Bradford County (NE PA). These two gauging stations are located on the major rivers in these two counties and their discharge data therefore integrate information about precipitation over the regions of interest. Mean values of annual discharge measured from 1985 to 2015 were plotted (Fig. S1†) and checked for a temporal trend using Spearman's nonparametric correlation tests. No significant changes in mean annual discharge were detected at either of the two sites (p greater than 0.05). Therefore, we ruled out changes in precipitation with time as the important factor explaining the observed changes in groundwater chemistry in NW PA *versus* NE PA.

Differences in lithology, topography, or sampling technique. In Pennsylvania, groundwater quality tends to vary with geologic formation and the topographic position where water wells are drilled.[5,38–40] In particular, Siegel *et al.*[5] pointed out that water wells drilled into Allegheny and Pottsville formations tended to have somewhat degraded water quality compared to other formations in western PA. Although we did not have depth information for most water wells in the NW PA datasets in this study and thus could not identify the formations hosting the water wells at the depth of extraction, the vast majority of groundwater samples in both pre-2000 and post-2010 datasets were generally located in areas with identical formations (*i.e.*, Allegheny and Pottsville). This argues against bedrock as an explanation for the inferred temporal trend.

In addition, many researchers have shown that groundwater chemistry in PA is affected by topographic position (*e.g.*, valley or ridge) of water wells.[5,7,41] In NW PA, the relief is generally relatively low compared to the rest of the state, however, and this is consistent with a relatively small effect of topographic position on groundwater chemistry.[5] Therefore, the observed trends in groundwater chemistry in NW PA are also not likely to be explained simply by differences in the topographic positions of the sampling sites in the datasets from the two time periods.

Groundwater samples analyzed in the NW PA datasets – like many such groundwater datasets – were not filtered before analysis. Sediments can be introduced into such samples to cause different magnitudes of turbidity if water wells were pumped at different flow rates during sampling.[5] High turbidity was previously found to be associated with high concentrations of Fe and Mn in Pennsylvania groundwaters sampled largely from homeowner wells.[5] The slight increase in Fe concentrations from pre-2000 to post-2010 in NE PA might therefore be at least partially explained if there were differences in flow rates during sampling that resulted in an increase in turbidity from pre-2000 to post-2010. Although we have no evidence for this, this possibility cannot be eliminated as a potential explanation.

New shale gas wells. The first unconventional well was drilled in 2012 and only 61 of these wells were drilled in total in Mercer County (NW PA). Most groundwater samples in the post-2010 dataset were collected as pre-drill baseline samples around six unconventional well pads: Bowser Unit, Jefferson Mcghee, Jefferson Zigo, Lackawannock James, Mccullough Unit, and Palmer 2082 D. A total of 25 unconventional wells were drilled on these six well pads.

The records of violations issued by PA DEP for these 25 unconventional wells revealed no violations related to cementing/casing failures or fluid spills on- or off-site. Only two wells (API: 085-24642, 085-24669) received violations coded as "SWMA301" meaning "Failure to properly store, transport, process or dispose of a residual waste". Those violations might have allowed the contamination of groundwater only if leakage or spilling of wastewater occurred, but that was not noted by the regulator. Furthermore, no letters were issued to any companies drilling unconventional wells in central Mercer County (NW PA) with a positive determination of possible culpability for impacting surrounding waters.[42,43] Thus, we have no evidence from the regulator that activities at unconventional wells in central Mercer (NW PA) were deemed responsible for nearby water contamination. These observations lead us to infer that the activities of unconventional O&G production in Mercer County (NW PA) might not be the primary cause of the observed slight regional change of groundwater chemistry in the study area. Furthermore, since the county has 61 unconventional wells but 3780 conventional wells, unconventional wells are inferred to be an unlikely cause for the observed slight changes in TDS, hardness, Na, Cl, and Fe distributions.

Conventional wells. Mercer County (NW PA) has a long history of conventional O&G production:[22] the county is located approximately 50 km from Titusville PA, the location of the first commercial oil well in the USA. Fig. 1B presents a summary of the year that each of the more than 3000 conventional wells listed by the regulator with a spud date in Mercer (NW PA) were drilled: most wells were drilled between 1980 and 2010. The 389 conventional wells that are listed by the regulator without a spud date are not shown on the map. Correspondingly, the majority of pre-2000 groundwater samples were collected between 1980 and 2000.

Groundwater contamination by conventional O&G drilling has been repeatedly reported in this area of northwestern PA in the Glaciated Appalachian Plateau area.[44] In particular, Harrison (1983)[44] suggested that the presence of highly permeable surficial sediment, relatively steep hydraulic gradients, and the low to moderate dilution of contaminants along flow paths might render groundwater systems in NW PA prone to contamination from conventional well drilling. Currently, we do not have data for violations related to conventional wells in NW PA. Instead, we use estimates cited for O&G wells in PA in general to estimate violations. Specifically, 0.7–9.1% of the active O&G wells drilled after 2000 in PA have been reported to have had compromised cement and/or casing integrity violations.[10] Therefore, of the 1495 currently active conventional wells that were drilled since 2000 in Mercer County (NW PA), we would expect that 10 to 136 active conventional wells probably had cementing/casing issues. These conventional O&G wells with cementing/casing issues might have caused the changes in groundwater chemistry in Mercer (NW PA) if these well issues were not adequately addressed.

In addition, the PA DEP lists 110 documented orphaned and abandoned conventional O&G wells located in Mercer County (NW PA) in their database[23] (see also Fig. S2†). These orphaned and abandoned wells were not drilled or completed following modern regulations,[25] and are not listed as having been plugged. If these unplugged and possibly poorly constructed O&G wells allow migration of TDS, Na, Cl, and Fe into shallow aquifers, this could explain the trends of increasing distributions of concentrations of those analytes. For example, production waters associated with O&G wells include Ca, Na, Cl, as well as high TDS in Mercer (NW PA). In this regard, the depths and integrity of surface casings and associated cements are of prime importance in protecting shallow groundwater from brines, especially in NW PA where brine waters are known to be present at shallower depths[1] than the rest of PA.

On the other hand, water samples in the post-2010 NW PA dataset are not located in the sub-areas of the county known to host the highest density of documented wells that have been abandoned or orphaned (Fig. S2†). We argue, however, that many such wells may not be marked on the state maps in this part of PA. For example, over all of PA as published by the PA DEP,[23] only ~8700 orphaned and abandoned O&G wells have been documented. However, the total number of abandoned and orphaned O&G wells have been estimated to be as high as 300 000 to 500 000,[25] *i.e.*, a factor of 35 to 55 higher. If we multiply the number (*i.e.*, 110) of orphaned and abandoned O&G wells documented by PA DEP in Mercer County (NW PA) by a conservative estimate of 10 to account for the lack of reporting for many such legacy wells, we estimate that the actual number of orphaned and abandoned O&G wells in Mercer County (NW PA) might be on the order of 1000. Such a large number of old and undocumented conventional O&G wells are likely to be a greater threat to regional groundwater chemistry than the small number of unconventional wells. For example, the first conventional well reported with a spud date in Mercer (NW PA) was drilled in 1955 and the number of conventional wells drilled peaked in 1992 and then again in 2005. In addition, 389 conventional wells are listed by the regulator in central Mercer (NW PA) without a known spud date, *i.e.* they were likely drilled before 1955. Such older conventional wells likely do not comply with modern standards given that no regulations for well completion were imposed in PA until 1984.[45] In comparison, the first unconventional well was not drilled in Mercer (NW PA) until 2012. Therefore, we conclude that conventional O&G wells and legacy wells are potential explanations for the observed slight changes in groundwater chemistry in Mercer County (NW PA).

De-icing salts and brines. Another possible source of Na, Cl, and TDS is the material that is used for de-icing paved roads during winters in PA.[46] De-icing materials are reported to include salt and briny water in the state. We have no evidence that the brines used for de-icing derive from O&G production brines and we therefore infer that these de-icing brines are mined salt dissolved in water. In the U.S., such application of road salt for de-icing became substantial since the 1940s.[47]

Mercer County (NW PA), with an area of 1770 km^2, is relatively more urbanized compared to Bradford County in NE PA. Specifically, Mercer (NW PA) has a population density of 63 per km^2 in 2015, and 12.4% of its land use was categorized as "developed" in 2011.[48,49] The population in Mercer (NW PA) decreased by 7.8% from 1985 to 2015 (Fig. S3†)[49] while land use did not change significantly compared to 2001.[50] The road density in Mercer (NW PA) was 1.85 km km^{-2} in 2015 (ref. 51) and public data shows little change in total length of public roads since 2010.[52] Although earlier data of road density at the county level is not available online to our knowledge, state-level data[53] show that the total length of public roads in Pennsylvania increased by only ~6437 km from 186 142 km from 1985 to 2000 and about 644 km from 2000 to 2015.

In comparison to Mercer County (NW PA), Bradford County (NE PA) is more rural: 5% of the county was reported as "developed" in both 2011 (ref. 48) and 2001.[50] Bradford County (NE PA) has a larger area of 3000 km^2 but a much smaller population density of 20 km^{-2} in 2015 than Mercer (NW PA).[48,49] The population density in Bradford (NE PA) decreased by 0.3% from 1985 to 2015 (Fig. S3†).[49] Bradford (NE PA) has smaller road density of 1.34 km km^{-2} in 2015 (ref. 51) and has not changed since 2010.[52]

Given these land use data, both urbanization and paved road density are higher in Mercer County (NW PA) than Bradford County (NE PA). In the winter of 2016–17, for example, a total of 15 200 tons salt and 1.6 million liters brine were used in Mercer County (NW PA) for de-icing.[46] In contrast, less salt (*i.e.*, 12 500 tons salt and 0.27 million liters brine) were used in Bradford (NE PA) in the same winter even though Bradford (NE PA) is almost twice the area of Mercer (NW PA).[46] Although the information of the amount of road salt and brine applied for de-icing in earlier years is not available, we assume the amount is generally increasing each year. These brines used for de-icing are likely to enter surface or groundwaters, especially since they are used in time periods with large precipitation when temperatures fluctuate in PA above and below freezing.

Such contamination effects related to road salt that is dominantly NaCl have been noted throughout the USA.[54] In addition, increases in Na and Cl concentrations may impact dissolution and ion exchange reactions between soil and water, releasing Ca, Mg, and bicarbonate into groundwater.[54–56] For multiple reasons, therefore, road salt might contribute to the observed changes in groundwater chemistry in Mercer (NW PA), where slight evidence was observed for salinization.

Spreading wastewaters from conventional O&G production on roads. Given the extremely long history of oil and gas extraction in NW PA, the region also has a long history of dealing with briny wastewaters that return to the land surface with the oil and gas. Among other disposal mechanisms, some counties in the NW PA spread brines on dirt roads for dust suppression in the dry summertime months.[57] Specifically, published data has documented that briny production waters from conventional O&G wells were spread on roads for dust suppression in Mercer County (NW PA) from 2010 to 2017.[57,58] In contrast, spreading of brine wastes on roads in NE PA is not reported.[57] Instead, most of the brine wastes that are recovered at shale gas wells in NE PA are re-injected for hydraulic fracturing of new wells.[59] As of 2018, road spreading of brines from conventional drilling for dust abatement has been terminated in PA. When briny wastewater is used for dust abatement of dirt roads, it can recharge into aquifers. For example, based on previously published calculations, solutes (for example, from brines) could flow vertically into groundwater through depths of approximately ~70 meters within one year.[7]

PA DEP has reported road spreading in Mercer County (NW PA) after 2010 (Fig. S4†). Wastes might have been spread on roads before 2010 in NW PA but waste reports were not available online.[58] Based on the temporal trend of volume of wastewater spread on Mercer roads (NW PA) post-2010 (Fig. S4†), the extrapolated volume of spread wastewater before 2010 might not be as large as that after 2010. All the post-2010 water samples were collected in 2012–2015, corresponding to the time period when an average of 0.30 million liters of wastewater were spread on Mercer County (NW PA) dirt roads annually. In contrast, from 2010 to 2017, an average of 0.62 million liters of wastewater were spread annually in Mercer (NW PA). These two average annual volumes were higher than the 0.29 million liters per county reported for Ohio[57] where two incidents of groundwater contamination and salinization have been reported and attributed to road spreading.[60,61] In both cases, Cl and specific conductance increased in groundwater following road spreading.

After spreading of brine on dirt roads to suppress dust, the road material retains dissolved solutes from the brine. Leaching experiments on PA road aggregate mixed with wastewater have shown that nearly all these waste solutes (*i.e.*, Cl, Br, Na, Mg, Ca, and Sr) can then leach readily from the road material and could thus potentially discharge into groundwater.[57] However, to our knowledge, no such incidents of groundwater contamination have been reported for PA. Most of the species for which distributions increased with time in the NW PA datasets (Ca, Mg, Na, Cl) are components that Tasker *et al.* (2018)[57] observed to leach easily from experimental road material after washing with brine.

Temporal trends within the pre-2000 and post-2010 datasets in NW PA. To summarize the discussion so far, we observed changes in groundwater chemistry in Mercer County (NW PA) from pre-2000 to post-2010. We consider it unlikely that unconventional O&G development has been a major impact on the water since no groundwater contamination related violations or environmental complaints were reported in the PA DEP databases[26,42,43] and only 61 such wells were drilled in the county. On the other hand, leakage of brine salts from older and less well constructed conventional and legacy O&G wells, mixtures of salts used for road de-icing, and production brines (from conventional wells) spread on roads for dust suppression are potential explanations. Given that Mercer County (NW PA) is more urbanized than Bradford County (NE PA), has used production brines for dust abatement, and has historically been the location of much early O&G development, any of these explanations could explain the differences between the two counties. Strictly on the basis of volume of salt used, the practice of de-icing may be the largest source.

Given that road spreading of produced wastewater for dust suppression in Mercer County (NW PA) was not substantial until 2010 (Fig. S4†) while the application of mixtures of salts used for road de-icing and the drilling of conventional reservoirs have both been in place in Mercer (NW PA) since before 2000 (Fig. 1B), we decided to also look at temporal trends within the two NW PA datasets. Specifically, to investigate which of these activities might have had more of an impact on groundwater chemistry in Mercer County (NW PA), we applied the nonparametric Spearman's rank correlation test on time series for all of the 15 analytes in the NW PA datasets. Such a test was performed for pre-2000 and post-2010 datasets separately to assess the temporal trend of groundwater chemistry within each of these two time periods. No significant change was observed for any analyte in the post-2010 dataset (Table S2†). Only five analytes (pH, total alkalinity, K, Na, and Mg) in the pre-2000 NW PA dataset show statistically significant temporal trends and all these trends are decreasing (Table S2,† Fig. 2).

Such pH decreases were also previously observed at some stream sites in the USA following the CAA amendments in 1990.[54] This generally decreasing trend in concentrations of four analytes in NW PA pre-2000 as compared to the overall change between pre-2000 and post-2010 datasets could be consistent with something different occurring between 2000 and 2010. For example, more conventional and unconventional wells were drilled later in the 2000s and the volume of wastewater spread on roads increased annually since 2010. However, considering the number of data points and the lack of data for application of salts and brine for road de-icing pre-2000 and post-2010, we cannot fully distinguish any of these possibilities.

Conclusions

In this study, we compiled and presented groundwater quality data from central Mercer County in NW PA to document temporal trends in that area: we compared 259 groundwater samples collected from pre-2000 and 1604 samples from post-2010. A total of 15 analytes (*i.e.*, pH, hardness, turbidity, alkalinity, specific conductance, CH_4, TDS, K, Mg, Ca, Cl, Na, SO_4, Fe, Mn) that were reported in both pre-2000 and post-2010 datasets were compared using both the simpler Wilcoxon–Mann–Whitney (WMW) rank sum and the more rigorous Brunner-Munzel (BM) tests. We saw little difference in the results from these two tests. Although the BM test might not have been required in this study, for datasets with smaller size (*e.g.*, less than 50) and larger ratio of variance (*e.g.*, greater than 10), the BM test should be considered.

These statistical tests indicated that solute concentrations in groundwater in central Mercer County in NW PA could have slightly increased over the 30 year interval between datasets. In particular, the distributions of concentrations of Mg, Ca, Na, Fe, Cl, TDS, total alkalinity, hardness, specific conductance, and turbidity increased slightly from pre-2000 to post-2010. This conclusion was compared to observations for NE PA where a similar study revealed some evidence for slight decreases in solute concentrations, even though NE PA is one of the two most heavily developed areas for shale gas in the state.

Potential explanations for the slight changes in groundwater chemistry detected in Mercer County (NW PA) were the use of mixtures of salts to de-ice roads, the spreading of brines from conventional O&G wells for dust suppression on roads, or leakage of brine salts from older conventional O&G wells (including orphaned and abandoned) with cementing/casing issues. The issues related to de-icing and dust abatement of roads might be the most important factors. The impact of unconventional O&G drilling and production activities was considered to be less significant largely because no groundwater contamination related to violations or environmental complaints were found in PA DEP databases.[26,42,43] Furthermore, the number of new shale gas wells ($n = 61$) is small compared to the number of active conventional ($n = 3670$) or legacy (*i.e.*, orphaned and abandoned wells; estimated between 110 and greater than 1100) or to the overall density of roads. All of these possible explanations for the changes in groundwater chemistry in NW PA provide working hypotheses for future studies. The data presented here show the possibilities of historic dataset analysis in the context of HVHF operations.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Funding derived from National Science Foundation IIS-16-39150 and USGS 104b award G16AP00079 through the Pennsylvania Water Resource Research Center to SLB and ZL. We thank Xianzeng Niu for technical support, and Seth Pelepko and Stew Beattie (PA DEP) for access and help with data. AA, LX, and AC are partially supported by the National Science Foundation grants DMS-1505256 and DMS-1811552 and the National Institutes of Health P50 DA039838. Matthew Gonzales, supported by the Penn State Earth and Environmental Systems Institute and Institutes of Energy and Environment, is acknowledged for data management.

References

1 S. L. Brantley, D. Yoxtheimer, S. Arjmand, P. Grieve, R. Vidic, J. Pollak, G. T. Llewellyn, J. Abad and C. Simon, Water Resource Impacts during Unconventional Shale Gas Development: The Pennsylvania Experience, *Int. J. Coal Geol.*, 2014, **126**, 140–156.

2 R. D. Vidic, S. L. Brantley, J. M. Vandenbossche, D. Yoxtheimer and J. D. Abad, Impact of Shale Gas Development on Regional Water Quality, *Science*, 2013, **340**, 1235009.

3 R. B. Jackson, A. Vengosh, T. H. Darrah, N. R. Warner, A. Down, R. J. Poreda, S. G. Osborn, K. Zhao and J. D. Karr, Increased Stray Gas Abundance in a Subset of Drinking Water Wells near Marcellus Shale Gas Extraction, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 11250–11255.

4 S. M. Olmstead, L. A. Muehlenbachs, J.-S. Shih, Z. Chu and A. J. Krupnick, Shale Gas Development Impacts on Surface Water Quality in Pennsylvania, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 4962–4967.

5 D. I. Siegel, B. Smith, E. Perry, R. Bothun and M. Hollingsworth, Pre-Drilling Water-Quality Data of Groundwater Prior to Shale Gas Drilling in the Appalachian Basin: Analysis of the Chesapeake Energy Corporation Dataset, *Appl. Geochem.*, 2015, **63**, 37–57.

6 D. I. Siegel, N. A. Azzolina, B. J. Smith, A. E. Perry and R. L. Bothun, Methane Concentrations in Water Wells Unrelated to Proximity to Existing Oil and Gas Wells in Northeastern Pennsylvania, *Environ. Sci. Technol.*, 2015, **49**, 4106–4112.

7 T. Wen, X. Niu, M. Gonzales, G. Zheng, Z. Li and S. L. Brantley, Big Groundwater Data Sets Reveal Possible Rare Contamination Amid Otherwise Improved Water Quality for Some Analytes in a Region of Marcellus Shale Development, *Environ. Sci. Technol.*, 2018, **52**, 7149–7159.

8 J. Woda, T. Wen, D. Oakley, D. Yoxtheimer, T. Engelder, M. C. Castro and S. L. Brantley, Detecting and Explaining Why Aquifers Occasionally Become Degraded near Hydraulically Fractured Shale Gas Wells, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**(49), 12349–12358.

9 T. H. Darrah, A. Vengosh, R. B. Jackson, N. R. Warner and R. J. Poreda, Noble Gases Identify the Mechanisms of Fugitive Gas Contamination in Drinking-Water Wells Overlying the Marcellus and Barnett Shales, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14076–14081.

10 A. R. Ingraffea, M. T. Wells, R. L. Santoro and S. B. C. Shonkoff, Assessment and Risk Analysis of Casing and Cement Impairment in Oil and Gas Wells In, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 10955–10960.

11 S. G. Osborn, A. Vengosh, N. R. Warner and R. B. Jackson, Methane Contamination of Drinking Water Accompanying Gas-Well Drilling and Hydraulic Fracturing, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 8172–8176.

12 G. T. Llewellyn, F. Dorman, J. L. Westland, D. Yoxtheimer, P. Grieve, T. Sowers, E. Humston-Fulmer and S. L. Brantley, Evaluating a Groundwater Supply Contamination Incident Attributed to Marcellus Shale Gas Development, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 6325–6330.

13 G. Schout, N. Hartog, S. M. Hassanizadeh and J. Griffioen, Impact of an Historic Underground Gas Well Blowout on the Current Methane Chemistry in a Shallow Groundwater System, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 296–301.

14 S. L. Brantley, R. D. Vidic, K. Brasier, D. Yoxtheimer, J. Pollak, C. Wilderman and T. Wen, Engaging over Data on Fracking and Water Quality, *Science*, 2018, **359**, 395–397.

15 E. W. Boyer, B. R. Swistock, J. Clark, M. Madden and D. E. Rizzo, *The impact of Marcellus gas drilling on rural drinking water supplies*, Center for Rural Pennsylvania, 2012.

16 L. J. Molofsky, J. A. Connor, A. S. Wylie, T. Wagner and S. K. Farhat, Evaluation of Methane Sources in Groundwater in Northeastern Pennsylvania, *Groundwater*, 2013, **51**, 333–349.

17 L. J. Molofsky, J. A. Connor, T. E. McHugh, S. D. Richardson, C. Woroszylo and P. J. Alvarez, Environmental Factors Associated With Natural Methane Occurrence in the Appalachian Basin, *Groundwater*, 2016, **54**, 656–668.

18 Z. Li, C. You, M. Gonzales, A. K. Wendt, F. Wu and S. L. Brantley, Searching for Anomalous Methane in Shallow Groundwater near Shale Gas Wells, *J. Contam. Hydrol.*, 2016, **195**, 23–30.

19 Z. Li, C. You, M. Gonzales, A. K. Wendt, F. Wu and S. L. Brantley, Corrigendum to "Searching for Anomalous Methane in Shallow Groundwater near Shale Gas Wells", *J. Contam. Hydrol.*, 2016, **195**, 23–30; *J. Contam. Hydrol.*, 2017, **207**, 50–51, DOI: 10.1016/j.jconhyd.2016.10.005, S0169772216300985.

20 S. L. Brantley, *Shale Network Database*, Consortium for Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI), 2018, DOI: 10.4211/his-data-shalenetwork.

21 G. Zheng, F. Wu, M. Gonzales, S. L. Brantley, T. Lauvaux and Z. Li, Assessing Environmental Impacts of Shale-Gas Development in an Area of Hydraulic Fracturing, KDD 2017 DATA Sci. Intell. food, energy, water work.

22 PADEP, PA Oil and Gas Mapping, http://www.depgis.state.pa.us/PaOilAndGasMapping, accessed 1 January 2018.

23 PADEP, Abandoned, Orphan & DEP Plugged Wells Listing, http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Abandoned_Orphan_Web, accessed 3 August 2018.

24 PADEP, Pennsylvania Digitized Mined Areas, http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=257, accessed 13 July 2018.

25 M. Kang, C. M. Kanno, M. C. Reid, X. Zhang, D. L. Mauzerall, M. A. Celia, Y. Chen and T. C. Onstott, Direct Measurements of Methane Emissions from Abandoned Oil and Gas Wells in Pennsylvania, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 18173–18177.

26 PADEP, Water Supply Resolved Complaints, http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Complaints/CTS_WS_Complaints, accessed 27 July 2018.

27 R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2017.

28 C. Poth, *Geology and hydrology of the Mercer Quadrangle, Mercer, Lawrence, and Butler Counties*, Pennsylvania, Harrisburg, PA, 1963.

29 PA Department of Conservation and Natural Resources, Glacial deposits of Pennsylvania, 1997.

30 T. M. Berg, W. E. Edmunds, A. R. Geyer, A. D. Glover, D. M. Hoskins, D. B. MacLachlan, S. I. Root, W. D. Sevon, A. A. Socolow and C. E. Miles, *Geologic map of Pennsylvania*, Map, Harrisburg, PA, 4th edn, 1980.

31 K. K. Eltschlager, J. W. Hawkins, W. C. Ehler, F. Baldassare and P. Dep, *Technical measures for the investigation and mitigation of fugitive methane hazards in areas of coal mining*, Department of the Interior Office of Surface Mining, 2001.

32 L. J. Molofsky, S. D. Richardson, A. W. Gorody, F. Baldassare, J. A. Black, T. E. McHugh and J. A. Connor, Effect of Different Sampling Methodologies on Measured Methane Concentrations in Groundwater Samples, *Groundwater*, 2016, **54**, 669–680.

33 J. W. Pratt, Robustness of Some Procedures for the Two-Sample Location Problem, *J. Am. Stat. Assoc.*, 1964, **59**, 665.

34 M. Chen, A Nonparametric Procedure Associated with a Clinically Meaningful Efficacy Measure, *Biostatistics*, 2000, **1**, 293–298.

35 E. Brunner and U. Munzel, The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation, *Biom. J.*, 2000, **42**, 17–25.

36 X. Niu, A. Wendt, Z. Li, A. Agarwal, L. Xue, M. Gonzales and S. L. Brantley, Detecting the Effects of Coal Mining, Acid Rain, and Natural Gas Extraction in Appalachian Basin Streams in Pennsylvania (USA) through Analysis of Barium and Sulfate Concentrations, *Environ. Geochem. Health*, 2018, **40**, 865–885.

37 P. A. Raymond and N. H. Oh, Long Term Changes of Chemical Weathering Products in Rivers Heavily Impacted from Acid Mine Drainage: Insights on the Impact of Coal Mining on Regional and Global Carbon and Sulfur Budgets, *Earth Planet. Sci. Lett.*, 2009, **284**, 50–56.

38 J. H. Williams, L. E. Taylor and D. J. Low, *Hydrogeology and groundwater quality of the glaciated valleys of Bradford, Tioga, and Potter Counties*, Pennsylvania Geological Survey, Pennsylvania, Harrisburg, PA, 1998.

39 B. Yan, M. Stute, R. A. Panettieri, J. Ross, B. Mailloux, M. J. Neidell, L. Soares, M. Howarth, X. Liu, P. Saberi and S. N. Chillrud, Association of Groundwater Constituents with Topography and Distance to Unconventional Gas Wells in NE Pennsylvania, *Sci. Total Environ.*, 2017, **577**, 195–201.

40 J. D. Stoner, D. R. Williams, T. F. Buckwalter, J. K. Felbinger and K. L. Pattison, *Water Resources and the Effect of Coal Mining*, Greene County, Pennsylvania, Harrisburg, PA, 1987.

41 E. L. Gross and C. A. Cravotta, *Groundwater Quality for 75 Domestic Wells in Lycoming County, Pennsylvania, 2014 Scientific Investigations Report 2016 – 5143*, U.S. Geological Survey, Reston, VA, 2017.

42 PADEP, PA DEP Water Supply Determination Letters, http://files.dep.state.pa.us/OilGas/BOGM/BOGMPortalFiles/OilGasReports/Determination_Letters/Regional_Determination_Letters.pdf, accessed 1 January 2018.

43 PADEP, PA DEP Oil & Gas Compliance Report, http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/OG_Compliance, accessed 1 January 2018.

44 S. S. Harrison, Evaluating System for Ground-Water Contamination Hazards Due to Gas-Well Drilling on the Glaciated Appalachian Plateau, *Ground Water*, 1983, **21**, 689–700.

45 R. M. Dilmore, J. I. Sams, D. Glosser, K. M. Carter and D. J. Bain, Spatial and Temporal Characteristics of Historical Oil and Gas Wells in Pennsylvania: Implications for New Shale Gas Resources, *Environ. Sci. Technol.*, 2015, **49**, 12015–12023.

46 Pennsylvania Department of Transportation, Winter Service Guide 2017-18, https://www.penndot.gov/TravelInPA/Winter/Pages/default.aspx, accessed 1 May 2018.

47 S. R. Corsi, L. A. De Cicco, M. A. Lutz and R. M. Hirsch, River Chloride Trends in Snow-Affected Urban Watersheds: Increasing Concentrations Outpace Urban Growth Rate and Are Common among All Seasons, *Sci. Total Environ.*, 2015, **508**, 488–497.

48 U.S. Geological Survey, *National Land Cover, Version 2*, https://gapanalysis.usgs.gov/gaplandcover/data/.

49 United States Census Bureau, https://www.census.gov/, accessed 21 November 2018.

50 C. Homer, J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel and J. Wickham, Completion of the 2001 National Land Cover Database for the Conterminous United States, *Photogramm. Eng. Remote Sens.*, 2007, **73**, 337–341.

51 Pennsylvania Department of Transportation, 2015 Highway Statistics, http://www.dot7.state.pa.us/BPR_PDF_FILES/Documents/Traffic/Highway_Statistics/Annual_Report/2016/9_10_Mileage_Jurisdiction_2015.pdf, accessed 24 November 2018.

52 Pennsylvania Department of Transportation, 2010 Highway Statistics, http://www.dot7.state.pa.us/BPR_PDF_FILES/Documents/Traffic/Highway_Statistics/Annual_Report/2011/2010_Highway_StatisticsPub60010-11.pdf, accessed 24 November 2018.

53 U.S. Department of Transportation Federal Highway Administration, Highway Statistics Series, https://www.fhwa.dot.gov/policyinformation/statistics.cfm, accessed 24 November 2018.

54 S. S. Kaushal, G. E. Likens, M. L. Pace, R. M. Utz, S. Haq, J. Gorman and M. Grese, Freshwater Salinization Syndrome on a Continental Scale, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E574–E583.

55 J. B. Shanley, Effects of Ion Exchange on Stream Solute Fluxes in a Basin Receiving Highway Deicing Salts, *J. Environ. Qual.*, 1994, **23**, 977.

56 D. W. Ostendorf, B. Xing and N. Kallergis, Cation Exchange in a Glacial till Drumlin at a Road Salt Storage Facility, *J. Contam. Hydrol.*, 2009, **106**, 118–130.

57 T. L. Tasker, W. D. Burgos, P. Piotrowski, L. Castillo-Meza, T. A. Blewett, K. B. Ganow, A. Stallworth, P. L. M. Delompré, G. G. Goss, L. B. Fowler, J. P. Vanden Heuvel, F. Dorman and N. R. Warner, Environmental and Human Health Impacts of Spreading Oil and Gas Wastewater on Roads, *Environ. Sci. Technol.*, 2018, **52**, 7081–7091.

58 PADEP, PA DEP Waste Report, http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?%2FOil_Gas%2FOil_Gas_Well_Waste, accessed 1 January 2018.

59 K. O. Maloney and D. A. Yoxtheimer, Production and Disposal of Waste Materials from Gas and Oil Extraction from the Marcellus Shale Play in Pennsylvania, *Environ. Pract.*, 2012, **14**, 278–287.

60 Y. Eckstein, Is Use of Oil-Field Brine as a Dust-Abating Agent Really Benign? Tracing the Source and Flowpath of Contamination by Oil Brine in a Shallow Phreatic Aquifer, *Environ. Earth Sci.*, 2011, **63**, 201–214.

61 E. S. Bair and R. K. Digel, Subsurface Transport of Inorganic and Organic Solutes from Experimental Road Spreading of Oil-Field Brine, *Groundwater Monit. Rem.*, 1990, **10**, 94–105.