



Spatial clustering of metal and metalloid mixtures in unregulated water sources on the Navajo Nation – Arizona, New Mexico, and Utah, USA

Joseph H. Hoover^{a,*}, Eric Coker^b, Yolanda Barney^c, Chris Shuey^d, Johnnye Lewis^a

^a Community Environmental Health Program, College Of Pharmacy, University of New Mexico, 1 University of New Mexico, Albuquerque, NM 87131, USA

^b Center for Environmental Research and Children's Health, School of Public Health, University of California Berkeley, USA

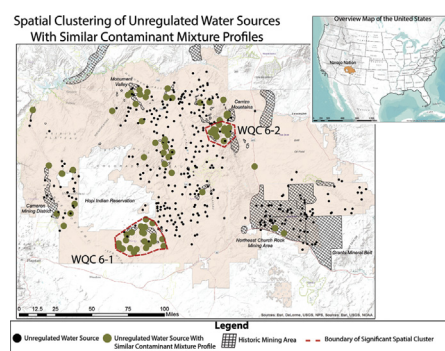
^c Navajo Nation Environmental Protection Agency - Public Water Systems Supervisory Program, PO Box 339, Window Rock, AZ 86515, USA

^d Southwest Research and Information Center, 105 Stanford Drive SE, Albuquerque, NM 87106, USA

HIGHLIGHTS

- Identified contaminant mixtures using Bayesian Profile Regression.
- Arsenic, uranium, lead, and manganese co-occur in some unregulated water sources.
- Spatial clustering of mixtures was observed in the study area.
- Clustering may help identify specific mixtures for future toxicology investigation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 21 November 2017

Received in revised form 23 February 2018

Accepted 24 February 2018

Available online 15 April 2018

Editor: Mae Mae Sexauer Gustin

Keywords:

Unregulated water sources
Metal and metalloid mixtures
Spatial clustering

ABSTRACT

Contaminant mixtures are identified regularly in public and private drinking water supplies throughout the United States; however, the complex and often correlated nature of mixtures makes identification of relevant combinations challenging. This study employed a Bayesian clustering method to identify subgroups of water sources with similar metal and metalloid profiles. Additionally, a spatial scan statistic assessed spatial clustering of these subgroups and a human health metric was applied to investigate potential for human toxicity. These methods were applied to a dataset comprised of metal and metalloid measurements from unregulated water sources located on the Navajo Nation, in the southwest United States. Results indicated distinct subgroups of water sources with similar contaminant profiles and that some of these subgroups were spatially clustered. Several profiles had metal and metalloid concentrations that may have potential for human toxicity including arsenic, uranium, lead, manganese, and selenium. This approach may be useful for identifying mixtures in water sources, spatially evaluating the clusters, and help inform toxicological research investigating mixtures.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Approximately 15% of the US population relies on unregulated water sources for domestic water supply (Maupin et al., 2014). An unregulated water source is a water supply that does not meet the criteria for classification as a public water system: a set of pipes or other

Abbreviations: BPR, Bayesian Profile Regression; UWSs, Unregulated water sources; WQCs, Water quality clusters.

* Corresponding author.

E-mail address: joh Hoover@salud.unm.edu (J.H. Hoover).

conveyance system with 15 or more service connections or provides water to 25 or more people at least 60 days a year. While public water system compliance with the health-based standards of the Safe Drinking Water Act exceeds 90% nationally (US EPA, 2015), unregulated sources are more challenging to evaluate (Backer and Tosta, 2011). Contaminants such as arsenic, uranium, radon, nitrates or others have been identified in private water sources throughout the United States (DeSimone et al., 2009).

Household access to regulated drinking water is more limited for tribal communities than for the US general population (Lewis et al., 2017). On the Navajo Nation, for example, as much as 30% of households lack access to a public water system (Leeper, 2003). As a result, residents may haul water from unregulated water sources. Previous work addressing the quality of unregulated water sources on the Navajo Nation identified arsenic and uranium as contaminants of concern (Corlin et al., 2016; Hoover et al., 2017). Arsenic and uranium were emphasized because previous research indicated their presence in groundwater in the region (Focazio et al., 2000; Ryker, 2003; US EPA, 2006); however, environmental sampling from the region also indicated the presence other contaminants in groundwater sources (US EPA, 2004; US EPA, 2006).

The purpose of this research was to identify, characterize, and assess the spatial patterning of metal and metalloid mixtures measured in a set of unregulated water sources on the Navajo Nation in the southwest United States. A Bayesian clustering method was applied to identify subgroups of unregulated sources with similar water quality profiles. For each subgroup, the potential for human toxicity was evaluated using the Benchmark Index, defined as the summation of contaminant concentrations normalized to a health-based threshold (Toccalino, 2007). Subgroups were also evaluated for spatial clustering to identify locations where unregulated water sources with the same subgroup membership were found at a greater frequency than expected. This approach may be useful for identifying contaminant mixtures for toxicological assessment and informing public health considerations about interventions to reduce adverse human health consequences from consumption of unregulated drinking water.

2. Background

Multivariate statistical techniques have been applied previously to evaluate large, geochemically complex environmental datasets (Swanson et al., 2001). Application of these methods is relevant to the study of metal and metalloid mixtures in water because drinking water is an important environmental exposure pathway in people (Carlin et al., 2013; Carpenter et al., 2002; DeSimone et al., 2009; Hertzberg et al., 2008; Toccalino et al., 2012). Despite previous work that demonstrates the occurrence of multiple contaminants in drinking water sources, there remains limited work using multivariate statistical methods like cluster analysis to identify and evaluate typical mixtures in these sources (Ander et al., 2016; Sanders et al., 2014).

Cluster analysis has proven useful when large, chemically complex datasets are evaluated (Kim et al., 2014). Identification of chemically similar water sources has been accomplished using hierarchical clustering (Flem et al., 2015; Hussain et al., 2008; Swanson et al., 2001; Vandenberg et al., 2015) and partition around medoids methods (Morrison et al., 2011). K-means clustering (Mandel et al., 2015), fuzzy clustering (Gentry, 2013; Güler and Thyne, 2004); and model-based clustering techniques have also been used (Kim et al., 2015; Kim et al., 2014). These methods have known sensitivities, however, that influence clustering outcomes. Partitioning methods are sensitive to the initial number of clusters and it may be challenging to select an appropriate number of clusters using hierarchical methods. Additionally, model-based methods are dependent upon assumptions about the distribution of the geochemical dataset. There is also limited diagnostic information available for assessing how variable selection, based on prior assumptions or expert knowledge, impacts clustering

results. (Güler et al., 2002; Kim et al., 2014; McNeil et al., 2005; Templ et al., 2008). Despite successful applications to other water quality related issues previous research has not used clustering methods to identify contaminant mixtures in drinking water sources (DeSimone et al., 2009; Squillace et al., 2002; Toccalino et al., 2012).

3. Methods

Bayesian Profile Regression was applied to identify subgroups of unregulated water sources with similar metal and metalloid profiles in a predominately rural area of the southwest United States. Subsequently, summary statistics were calculated using Robust Regression on Order Statistics, a left-censored statistical method. A Benchmark Index score was also calculated for each water quality cluster to identify metal and metalloid combinations of potential toxicological interest. Lastly, the spatial distribution of the water quality subgroups was assessed.

3.1. Study area

The Navajo Nation encompasses >70,000 km² in the Four Corners area of Utah, Arizona, and New Mexico in the southwest United States (Fig. 3A). Groundwater is the primary drinking water source for the Navajo Nation, which is supplied by three multi-part aquifer systems (Cooley et al., 1964). As much as 30% of the population lacks household access to a public water supply and may rely on unregulated water sources. While some residents do haul water from public water supplies, which are regulated by national drinking water laws, there remain families who use unregulated sources for domestic water (NN DWR, 2011).

3.2. Water quality data

A water quality dataset for unregulated sources located on the Navajo Nation was queried for this investigation. The collection and analytical procedures used to generate these data are described in detail by Hoover et al. (2017) and the sample collection and laboratory procedures are summarized briefly here. Federal, Tribal, and state agencies (Murphy et al., 2009; US EPA, 2000; US EPA, 2006; US EPA Region IX, 2008; US EPA Region IX, 2010; US EPA Region IX, 2011), and university and non-profit organizations (Hund et al., 2015) worked with Navajo communities to identify and collect water samples from unregulated sources between 1998 and 2010 using similar sample collection methods. The sampled water sources were unregulated and tested, because of community concern about safety for human consumption. Documented use of the unregulated sources was limited to one study (Hund et al., 2015), but the other studies identified and sampled sources thought to be used by residents or in response to residents' concerns. Each source was flushed for 1–2 min before a sample was collected. This collection procedure was designed to match how Navajo Nation residents collect water from unregulated sources (US EPA Region IX, 2008). The samples were subsequently filtered and acidified prior to analysis by a certified drinking water laboratory using Inductively Coupled Plasma Optical Emissions Spectroscopy (USEPA analysis method 200.7/6010B) or Inductively Coupled Plasma Mass Spectroscopy (USEPA analysis method 200.8). For a subset of samples, uranium activity (measured in picocuries per liter), was determined using US EPA method 907.0 or HASL 300 U-02-RC; for this analysis uranium activity was converted to mass using the assumption that 0.67 picocurie uranium was equivalent to 1 µg uranium (Weiner, 2013).

A contaminant was included in this analysis if it was measured in >75% of tested water sources. This threshold was selected so that a set of contaminants measured in unregulated sources dispersed throughout the Navajo Nation was used. Aluminum, arsenic, barium, beryllium, cadmium, chromium, copper, iron, manganese, nickel, lead, selenium, uranium, and zinc met the selection criterion. Other contaminants like antimony, mercury, molybdenum, thallium, and vanadium were

measured, but excluded as they measured in fewer than 75% of tested sources and were not available throughout the Navajo Nation. Different laboratories analyzed the water samples so there were different reporting limits for each contaminant. The highest reporting limit for each contaminant is provided in Table 1. A full list of reporting limits for each element is provided in Supplemental Table 1. Please see Hoover et al. (2017) for more information about the data compilation, quality control, and storage processes.

3.3. Profile regression

Bayesian Profile Regression (BPR), a non-parametric method that uses a Dirichlet process mixture model, was employed to identify subgroups (clusters) of unregulated water sources with similar patterns of co-occurring metals and metalloids. BPR was selected for this analysis due to the large number of correlated variables, no prior information about the optimal number of clusters, and to the need to understand what contaminants might be driving clustering. This method has been used previously for environmental epidemiology research (Coker et al., 2017; Coker et al., 2016; Papathomas et al., 2010; Pirani et al., 2015) as a result of the advantages offered by BPR for clustering analyses.

BPR implements a Dirichlet process mixture model (DPMM) that allows the data to determine the number of subgroups (clusters), hence the number of clusters need not be defined ahead of time (Molitor and Papathomas, 2010). BPR sets the DPMM in a Bayesian framework with Markov chain Monte Carlo (MCMC) estimation that appropriately propagates uncertainty in cluster assignment and the number of clusters and simultaneously links a profile of exposures to a cluster at each iteration. The post-processing of the BPR includes a dissimilarity matrix for assigning optimal cluster membership (or “best clustering”). Uncertainty in cluster membership is ascertained with Bayesian model averaging rather than hard cluster allocation or probabilistic allocation to a cluster. This approach provides greater flexibility than hard cluster assignment and enhances interpretation compared to purely probabilistic clustering methods. Variable selection may also be employed to

identify which variables support clustering (Liverani et al., 2015). Additional information about BPR can be found in other recent publications (Hastie et al., 2013; Liverani et al., 2015; Papathomas et al., 2012).

Briefly, BPR uses a Dirichlet process (DP) as the prior for the mixing distribution, defined as $P \sim DP(P_{00}, \alpha)$ where α is a scale parameter and P_{00} is the base probability distribution (Liverani et al., 2015). For each unregulated water source i , a water quality profile was defined as $x_i = (x_{i1}, \dots, x_{iP})$ where i is the number of unregulated water sources and P is the number of metals and metalloids included in the analysis.

The BPR output produces discreet subgroup (cluster) membership for each water source, referred to as Water Quality Clusters (WQCs). Measured contaminant concentrations were converted to quantiles for clustering and measurements below the limit detection (LOD) were assigned the LOD value before conversion to a quantile value. Similar to Coker et al. (2016), clustering was fit without an outcome and the ‘best’ clustering was used for characterizing and analyzing each cluster and for ease of interpretation. BPR was implemented using the PReMiuM package (v 3.1.3) (Liverani et al., 2015) for R (v 3.3.1).

At each iteration the average subgroup probability was calculated and the collection of these averages represents the posterior distribution of cluster specific parameters. Bayesian model averaging accounts for uncertainty by averaging the posterior distributions of cluster parameters across all iterations (Molitor and Papathomas, 2010). If an observation is consistently placed in the same cluster then the credible interval of cluster specific parameters will be narrower, indicating a higher certainty. Conversely, if an individual observation is placed in different clusters then the credible interval will be relatively wider indicating more uncertainty. In this way uncertainty is accounted for in the ‘best’ clustering assignment. Cluster assignment uncertainty was evaluated by reviewing the MCMC output, as recommended by Molitor and Papathomas (2010).

The probability that each contaminant contributed to the clustering output was evaluated using a variable selection method. The PReMiuM package generates a latent selection weight that ranges between 0 and 1. Values closer to 1 suggest a greater probability of supporting clustering while values closer to 0 suggest less cluster support. There is no

Table 1

Summary statistics for water quality data for each water quality cluster (WQC) and the study area as a whole, Median (25th, 75th percentiles). All units reported in $\mu\text{g/L}$.

	Highest LOD*	Navajo	WQC 1	WQC 2	WQC 3	WQC 4	WQC 5	WQC 6	WQC 7
Primary drinking water contaminants									
As	5.0	1.95 (0.42,5.7)	–	2.05 (0.71,8)	2 (0.54,4)	0.29 (0.1, 0.87)	5.85 (2.23,25.82)	4.5 (2.7,7.5)	3.91 (1.65,6.43)
Cd	1.0	–	–	–	–	–	–	–	–
Cr	10.0	–	–	0.51 (0.29,0.98)	–	–	1.67 (0.63,4.84)	–	–
Pb	1.65	–	–	0.74 (0.28,2.27)	–	–	2 (0.51,6.3)	–	–
Mn	7.5	4.8 (1.2,23.2)	6.95 (2.2,17.25)	5.4 (3.41,14.4)	2 (0.34,4.8)	62 (12.75,225)	18.7 (4.72,39)	0.92 (0.36,2)	12.6 (3.8,39.1)
Ni	20.0	–	1.15 (0.57,1.95)	1.5 (0.62,3)	–	–	2.65 (1.1,7.4)	–	–
Se	2.5	1 (0.22,4)	0.69 (0.16,1.72)	2.2 (0.84,9)	1 (0.21,3)	–	2 (0.4,6.5)	2.53 (1.3,6.5)	–
U	1.25	3.76 (0.51,13)	–	5.76 (1.05,23.5)	3 (0.45,9)	0.09 (0.01,1)	4.44 (1.02,21.58)	8.25 (4.53,16.06)	8.83 (2.41,20.28)
Secondary drinking water contaminants									
Al	25.0	14.78 (2.8,71.95)	8.37 (2.33,37.5)	21.5 (2.15,109.5)	–	–	60.4 (7.1220)	58.45 (30.98,88.45)	71.5 (43.4129)
Cu	10.0	3.4 (1.32,9)	4.55 (1.87,11.25)	7.96 (3.6,19)	3.11 (1.51,7)	1.98 (0.74,6)	4 (1.04,15.4)	3.3 (1.99,5.56)	4.67 (2.32,8.2)
Fe	50.0	160 (40,484)	150 (23.99,500)	256.5 (122.5555)	80 (30,200)	660 (152.15,1200)	570 (219.5,2100)	27.12 (16.77,43.93)	275 (139,616)
Zn	5.0	68 (20.9160)	47.5 (9.97,165)	76 (47,153)	60 (30,120)	255 (36.75,510)	140 (66.6520)	20.9 (6.8,49.7)	94.2 (23.9170)
N		453	53	35	123	43	39	81	79

a “–” indicates that too few observations (<30% of group) were greater than the detection limit and a measure of central tendency was not calculated.

* LOD is limit of detection.

threshold for inclusion/exclusion based on the variable selection weight. This variable is generated to inform interpretation of which covariates might be driving clustering. As suggested by the authors of the PReMiuM package, contaminants were also considered to lack cluster support if the credible intervals for a contaminant did not exhibit significantly higher or lower levels for at least one cluster.

3.4. Summary statistics

Summary statistics were calculated for each WQC using the NADA Package (v 1.5.6) for R (v 3.3.1), because there were observations less than the limit of detection, also known as left-censored data (Antweiler and Taylor, 2008). Left-censored statistical methods are used when a dataset includes observations that are not precisely measured. Environmental datasets commonly contain censored observations that are less than a limit of detection but the precise value cannot be determined (Antweiler and Taylor, 2008; Field, 2011; Helsel, 1990; Helsel, 2006). Robust Regression on Order (ROS) statistics, a left-censored statistical method, was employed for this analysis (Helsel, 2012; Lee, 2013; Lee and Helsel, 2005).

Values for contaminant measurements that were less than the LOD were imputed using the robust ROS method and were not assumed to be zero. The robust ROS method determines the Weibull plotting position of each uncensored and censored (observations less than the limit of detection) observation. Then the plotting position and normal score of the uncensored observations are used to create a linear regression model. Estimates of the censored values are then calculated based on the Weibull plotting position using the regression equation. Lastly, the estimated censored observations are combined with the uncensored observations and summary statistics are calculated using all observations. This method has been used previously in other water quality studies (Helsel, 2005; Levitan et al., 2014) and is appropriate for this dataset because it handles multiple limits of detection, which is present in this dataset.

If fewer than 30% of sources had detectable concentrations then summary statistics were not calculated for that metal or metalloid. Previous research demonstrated that measures of central tendency were biased when fewer than 30% of observations had detectable concentrations (Helsel, 2012).

3.5. Potential for toxicity

The Benchmark Index was employed to assess potential for human health impact (Toccalino et al., 2012), limited to arsenic, cadmium, chromium, lead, nickel, selenium, and uranium. These contaminants have established human-health based thresholds and are regulated in drinking water by the US EPA. Additionally, the Health-Based Screening Level (HBSL) for manganese was selected because of evidence suggesting manganese exposure is associated with negative human health outcomes (Wasserman et al., 2016); manganese is not regulated as a primary drinking water contaminant by the US EPA. The Benchmark Index assumes dose additivity, meaning the overall potential for toxicity of the mixture is a summation of the individual contaminants (Toccalino et al., 2012). There are limitations to the Benchmark Index, such as assuming dose additivity, no synergistic effects, and it can be applied only to contaminants with an established human health benchmark.

For each contaminant and each WQC, a Benchmark Quotient score was calculated by dividing the representative concentration by the respective health-based threshold, as proposed by Toccalino et al. (2012). Subsequently, the individual Benchmark Quotients scores were summed to determine the Benchmark Index for each WQC. An index score >1.0 suggests greater potential for toxicity. The Benchmark Quotient scores were also evaluated and WQCs with two or more scores exceeding 0.1 were identified. This threshold was selected because the US Agency For Toxic Substances and Disease Registry suggested that

mixtures should be considered toxicological study when two or more contaminants are present at concentrations exceeding 10% of the human health based threshold (ATSDR, 2004). Benchmark Quotient and Index scores were calculated for the 50th and 75th percentile concentrations of each metal and metalloid in each WQC. These percentile concentrations were selected to illustrate the potential toxicity for contaminant concentrations that Navajo Nation residents are likely to encounter when using unregulated water sources. Higher percentiles, such as the 90th or 95th, would be encountered by residents at few water sources on the Navajo Nation and likely do not represent common exposures.

3.6. Spatial clustering

While BPR is helpful for identifying subgroups of observations with similar covariate profiles, it is not designed to identify spatial cluster patterns of point data. Therefore, SaTScan, a likelihood-based spatial clustering approach (Kulldorff, 2009; Kulldorff et al., 1997) was employed to identify spatial clusters of unregulated water sources with the same WQC membership. This method enabled identification of areas where unregulated sources with the same WQC membership occurred more frequently than expected. SaTScan has been used for environmental applications such as spatial clustering of microbial contamination of private water wells (Krolik et al., 2014; Krolik et al., 2013).

SaTScan uses a search window to identify spatial clusters. The search window is centered over each location and the number of cases and controls that intersect the window are used to calculate a likelihood ratio. The SaTScan software evaluates clusters at all possible window sizes up to the maximum spatial window size, resulting in thousands of overlapping clusters. The Gini coefficient, rather than a hierarchical approach, was selected to determine the maximum cluster reporting size – the percentage of the population included in the scan window area that produces an optimal collection of spatially distinct clusters (Han et al., 2016; Kim and Jung, 2017). This enabled identification of the most informative collection of non-overlapping clusters without adjusting the maximum spatial window. The cluster with the largest Gini coefficient was considered the best collection of clusters because it balances cluster size with relative risk and limits identification of overly large clusters (Han et al., 2016). No apriori information was available that would lead to selection of a smaller maximum spatial window size so the default maximum spatial window size of up to 50% of the sample population was used. Spatial clusters for each WQC were identified using the SaTScan case-control option (Bernoulli). Using WQC 1 as an example, cases were defined as the sources with WQC membership ($N = 53$) and controls were sources with different cluster membership ($N = 400$). For each cluster a p -value was obtained with Monte Carlo hypothesis testing using 999 random replicates. The spatial clustering results are reported as WQC # -Spatial Cluster #. For example, WQC 2 has two spatial clusters identified as WQC 2-1 and WQC2-2.

Kernel density plots were created to visualize the spatial distribution of each WQC. A kernel density plot visualizes the density of features in a neighborhood area around those features (ESRI, 2016). Kernel Density plots were created using the ESRI Kernel Density tool for ArcGIS Desktop (version 10.3.1).

4. Results

4.1. Correlation

Spearman's ρ correlation values ranged from -0.28 to 0.69 (Fig. 1). Some of the strongest correlations were observed between As-U (0.45), U-Se (0.45), As-Pb (0.52), As-Se (0.57), Al-Cd (0.62), Al-Pb (0.64), Fe-Mn (0.62), Cd-Be (0.64) and Cr-Ni (0.69). Other metals such as Fe (range: -0.2 to 0.39), Ba (range: -0.2 to 0.1), and Zn (range: -0.07 to 0.21) were weakly correlated with other contaminants.

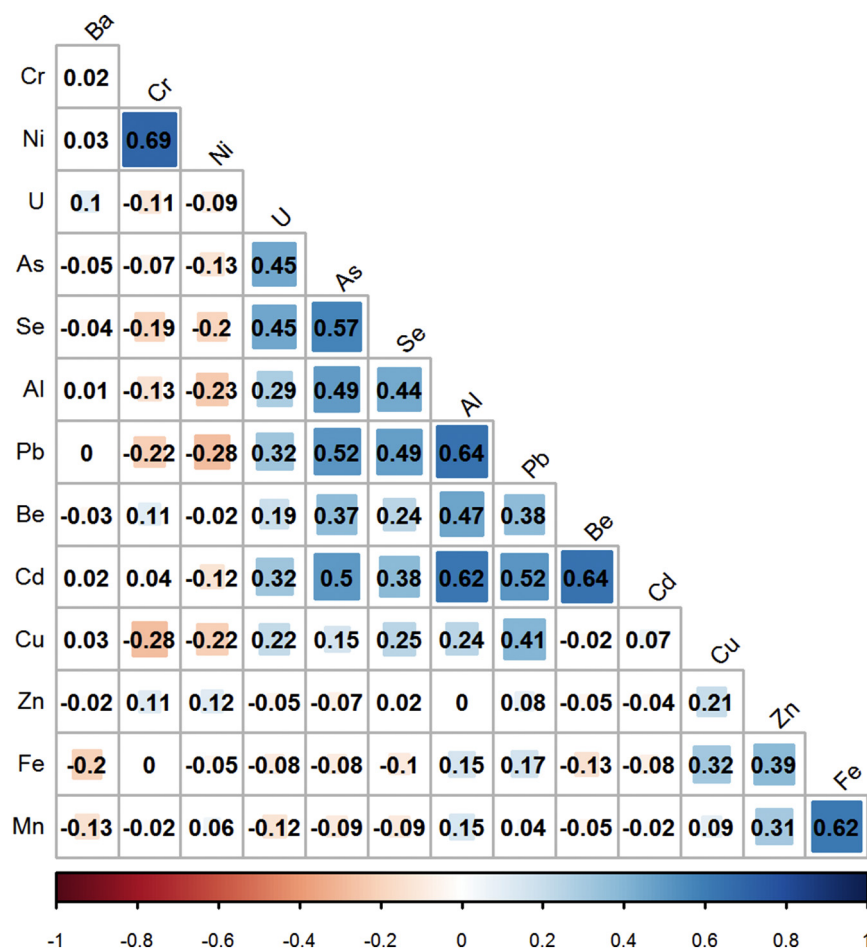


Fig. 1. Matrix illustrating the correlation coefficient (Spearman's ρ) for metals and metalloids measured in unregulated water sources.

4.2. Water quality clusters

Seven water quality clusters (WQC) were identified using Bayesian Profile Regression. Initial clustering runs indicated a lower probability (<0.5) that barium and beryllium were driving clustering. Other contaminants had a greater probability of influencing cluster membership (>0.80). Zinc had had a moderate influence on clustering output (probability 0.67). Upon review of the MCMC post processing visualization (Supplemental Fig. 1), a strong clustering signal was observed with relatively narrow credible intervals, giving confidence in the 'best clustering' allocation. The probability of contaminant concentration categories (quantile scores) varied among the clusters. Some WQCs showed a higher probability for the third or fourth quantile contaminant concentrations (suggesting higher concentrations). Other WQCs had a higher probability for the first or second quantile contaminant concentrations (suggesting lower concentrations).

For WQC 1 ($N = 53$) and WQC 3 ($N = 123$) the median concentration of most tested contaminants was similar to or lower than the median for all tested sources in the study area (Table 1, Fig. 2). Meanwhile the median concentrations of arsenic and uranium in WQC 3 was similar to the median concentrations for the whole study area. Additionally, exceedance of health-based thresholds for arsenic and uranium were similar to the study area as a whole. For WQC 4 ($N = 43$), the median manganese concentration exceeded the median value for the Navajo Nation; the median concentrations of copper, nickel, selenium and uranium were similar to their respective medians for the

Navajo Nation. Approximately 20% of unregulated sources in WQC 4 exceeded the manganese health based screening level of 300 $\mu\text{g/L}$ (Table 2).

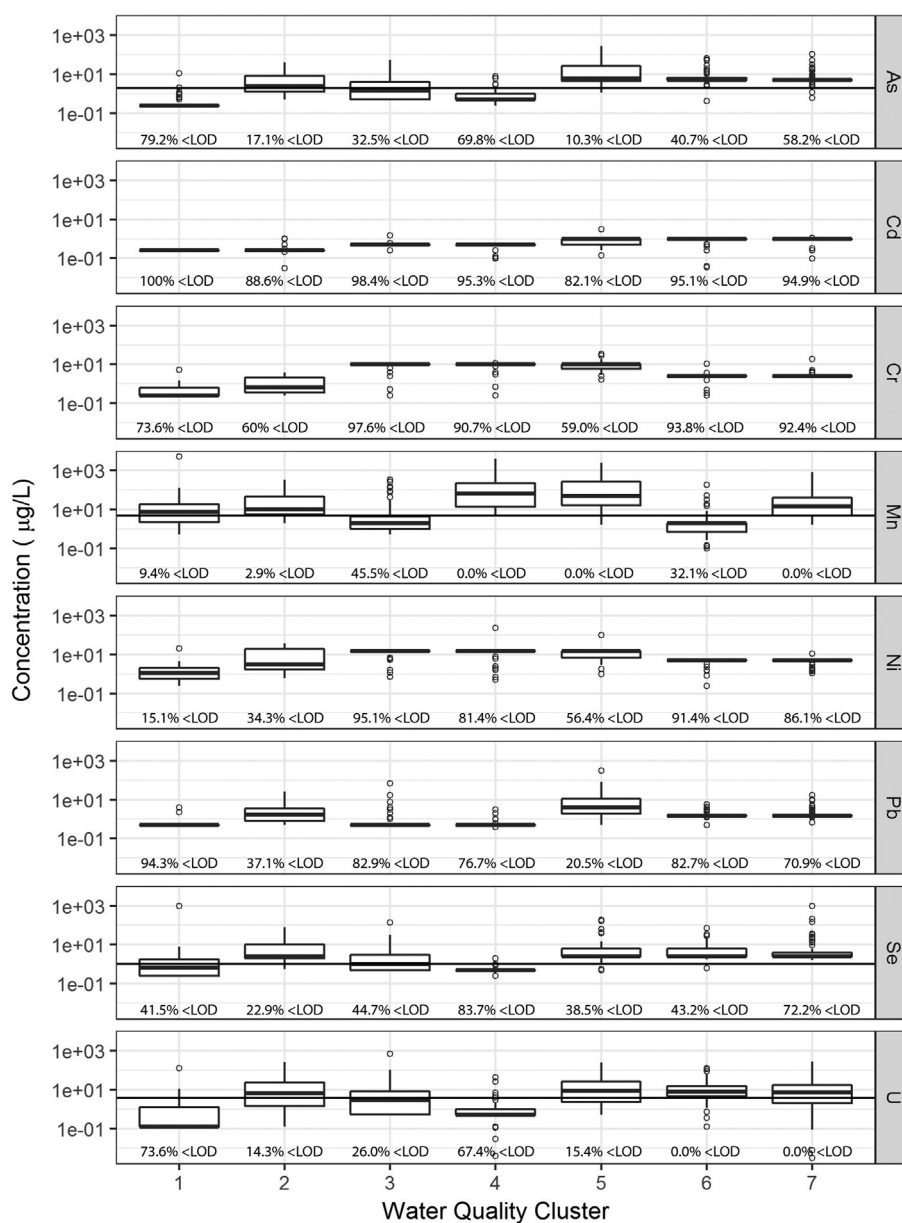
The median concentrations of arsenic and uranium were elevated in WQC 2 ($N = 35$) when compared to the respective study area medians for these contaminants. Furthermore, chromium and lead were detected more frequently in water sources grouped into this WQC, albeit infrequently; exceedance of health based thresholds was more frequent in these water sources for arsenic (~20% of sources) and uranium (~20% of sources) when compared to the study area. Additionally, median concentrations of aluminum and iron were higher.

Water quality clusters 5, 6, and 7 demonstrated elevated concentrations of metals and metalloids and more frequently exceeded health-based thresholds when compared to the entire study area. WQC 5 ($N = 39$) had the highest median concentrations of arsenic. Unregulated sources in this WQC also had higher median concentrations of lead, manganese, nickel, aluminum, iron, and zinc compared to the Navajo Nation. Nearly 44% of sources in WQC5 exceeded the arsenic drinking water standard, 17% exceeded the uranium or manganese standards, and 15% exceeded the lead action level. Sources grouped into WQC 6 ($N = 81$) had elevated concentrations of arsenic, uranium, and aluminum, but lower concentrations of iron when compared to WQC 5. Lastly, WQC 7 ($N = 79$) had the highest median concentration of uranium, and median concentrations of arsenic, manganese, aluminum, and iron were greater than the Navajo Nation medians.

Table 2

Percentage (%) of water sources with contaminant concentrations exceeding National Drinking Water Standards, presented by contaminant and water quality cluster (WQC).

Analyte	Standard ^a (µg/L)	All Navajo UWSs	WQC 1	WQC 2	WQC 3	WQC 4	WQC 5	WQC 6	WQC 7
Primary drinking water contaminants									
As	10	15.0	1.9	20.0	10.6	0.0	41.0	14.8	10.1
Cd	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cr	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pb	15	2.6	0.0	8.6	1.6	0.0	15.4	0.0	1.2
Mn	300 ^b	5.3	1.9	2.9	0.8	23.3	20.5	0.0	3.8
Ni	100	0.4	0.0	0.0	0.0	2.3	2.6	0.0	0.0
Se	50	2.2	1.9	2.9	0.8	0.0	7.7	1.2	3.8
U	30	10.8	1.9	22.9	6.5	2.3	20.5	14.8	13.9
Secondary drinking water contaminants									
Al	50	38.6	20.8	48.6	0.8	7.0	84.6	66.7	70.9
Cu	1000	0.2	0.0	0.0	0.0	2.6	0.0	0.0	0.0
Fe	300	38.6	43.4	71.4	15.4	74.4	89.7	0.0	51.9
Zn	5000	0.4	1.9	0.0	0.0	0.0	2.6	0.0	0.0

^a Reported in µg/L.^b US Geological Survey Health-Based Screening Level.**Fig. 2.** Boxplot of select water contaminants, partitioned into clusters for comparison. The value below each boxplot indicates the percentage of sources with contaminant measurements below the limit of detection. Black horizontal lines for arsenic (As), manganese (Mn), selenium (Se), and uranium (U) indicate the overall median.

4.3. Potential for human health impacts

When the 50th percentile concentration for each contaminant was evaluated two combinations were present that may suggest potential for human health risk: arsenic and uranium (WQCs 2, 3, 6 and 7); and arsenic, uranium, and lead (WQC 5) (Table 3).

When the 75th percentile contaminant concentrations were evaluated five contaminant combinations were present at concentrations that may suggest a need for additional toxicological study: arsenic and uranium (WQC 3); arsenic, uranium, and selenium (WQC 6); arsenic, uranium, and manganese (WQC 7); arsenic, uranium, selenium, and lead (WQC 2); arsenic, uranium, selenium, lead, and manganese (WQC 5).

4.4. Spatial clusters

SaTScan results indicated that 45% of water sources ($N = 204$) were spatially clustered when grouped by WQC (Figs. 3 and 4). Unregulated water sources classified into WQC 1 were spatially clustered in the northern and eastern areas of the Grants Mineral Belt section of the Navajo Nation (Fig. 3B); unregulated water sources classified into WQC 4 were spatially clustered in central Navajo (Fig. 4A). The spatial cluster of WQC4 overlapped with the two spatial clusters for WQC3 (Fig. 3D, Table 4).

Unregulated water sources classified into WQC 2 (Fig. 3C) were spatially clustered near the Northeast Church Rock Mining area of the Grants Mineral Belt (WQC2-1) and a separate group further east in the Grants Mineral Belt (WQC2-2). Unregulated water sources with WQC 6 membership (Fig. 4C) spatially clustered around the Carrizo Mountains (WQC6-2), a former uranium mining area in northeast Arizona, and south of the Hopi Indian Reservation (WQC6-1). Lastly, unregulated water sources with WQC 7 membership (Fig. 4D) spatially clustered in the western part of the study area near the Cameron Mining District (WQC 7-1) and in the northern part of the Carrizo Mountains in northeast Arizona (WQC 7-2). No significant spatial clusters were identified for WQC 5. As illustrated in Fig. 4B, unregulated water sources with WQC 5 membership occurred throughout the study area, but no groups were significantly clustered.

Kernel density plots were also generated for each WQC and compared to the SaTScan results. For WQCs 1, 2, 3, 4, 6, and 7 overlap was observed between the areas with the highest density (from the Kernel density plot) and areas of statistically significant clusters returned by SaTScan (Supplemental Figs. 2 and 3).

Table 3

Summary of Benchmark Index (BI) and Benchmark Quotients (BQ) >0.1 for each water quality cluster (WQC).

WQC	Percentile concentration ^a	BI score	Contaminant of concern (BQ score) ^b
1	50th	0.02	None
1	75th	0.09	None
2	50th	0.53	As (0.21), U (0.19)
2	75th	2.0	As (0.80), U (0.78), Se (0.18), Pb (0.15)
3	50th	0.33	As (0.2), U (0.1)
3	75th	0.78	As (0.40), U (0.30)
4	50th	0.24	Mn (0.21)
4	75th	0.87	Mn (0.75)
5	50th	1.01	As (0.59), U (0.15), Pb (0.13)
5	75th	4.77	As (2.58), U (0.72), Mn (0.80), Pb (0.42), Se (0.13)
6	50th	0.78	As (0.45), U (0.28)
6	75th	1.42	As (0.75), U (0.54), Se (0.13)
7	50th	0.73	A (0.39), U (0.29)
7	75th	1.45	A (0.64), U (0.68), Mn (0.13)

^a Percentile concentration refers to the 50th or 75th percentile of each metal or metalloid for each water quality cluster as reported in Table 1.

^b Contaminants of concern have a benchmark quotient (BQ) >0.1.

5. Discussion

The Bayesian clustering procedure identified distinct profiles of metals and metalloids in unregulated water sources. Description of the water quality clusters using the Benchmark Index enabled identification of clusters with contaminant mixtures occurring at concentrations with potential for human toxicity. Additionally, use of a spatial clustering method identified geospatial groups of unregulated water sources with similar contaminant mixtures.

5.1. Utility of Bayesian profile regression

In this investigation Bayesian Profile Regression was applied to identify metal and metalloid mixtures in unregulated water sources. BPR is designed to identify subgroups with similar joint distributions of a combination of variables in a manner that is mostly data-driven. Unlike other forms of clustering, it does not require pre-specification in the number of subgroups and incorporates uncertainty into clustering. Moreover, BPR can be used to understand associations between joint levels of covariates and an outcome (Molitor and Papathomas, 2010). Alternatively, as done in the present study, BPR may be combined with other analytic approaches to elucidate important relationships (e.g., spatial patterns of clusters or contaminant combinations with potential for human health toxicity).

The use of variable selection enabled identification of contaminants that supported a clustering pattern. Initial runs indicated a lower probability that barium and beryllium supported clustering. Barium may have had limited clustering support because it was weakly correlated with other contaminants, due to no commonly shared contaminant source, as illustrated in Fig. 1. A sensitivity analysis was conducted (results not shown) by including and excluding barium and in cluster assignment and spatial clustering were assessed. Only 2% of unregulated sources ($N = 12$) changed cluster membership, 7 WQCs were identified, and the spatial analysis results were not meaningfully different; spatial clusters of the same WQCs were found in the same parts of the study area.

Similarly, the BPR clustering results were not very sensitive to the inclusion of beryllium, which may be because this contaminant was detected in only 5% of unregulated water sources. A sensitivity analysis indicated that when beryllium was included 7% of unregulated water sources changed WQC membership. There remained seven WQCs in the optimal output despite a small number of sources changing WQC membership. This provides evidence that the number of clusters was stable even without the inclusion or exclusion of barium and beryllium.

BPR also enabled evaluation of cluster assignment uncertainty, which is important because even a noisy dataset will produce a 'best' partition. Ignoring assignment uncertainty could lead to over interpretation of the resulting clusters. The cluster uncertainty for this dataset was relatively low because the best partition, illustrated in Supplemental Fig. 1, demonstrated some clusters with high probabilities for different contaminant categories and very little overlap of credible intervals with expected average probability (e.g., 0.25). If the cluster uncertainty was high then the cluster specific probabilities for many or all categories would not be different than the overall category probability for each cluster (illustrated as a green shaded boxplot). The strong clustering signal may be due to the shared sources of contaminants within clusters.

5.2. Implications for public health and exposure

Despite the well-established co-occurrence of contaminants in unregulated water supplies (Bacquart et al., 2015; DeSimone et al., 2009; Toccalino et al., 2012) there remain few applications of multivariate techniques to identify and characterize mixtures in water sources. On the Navajo Nation for example, previous research focused on the individual occurrence of arsenic or uranium, and few studies addressed co-occurrence of these contaminants (Murphy et al., 2009).

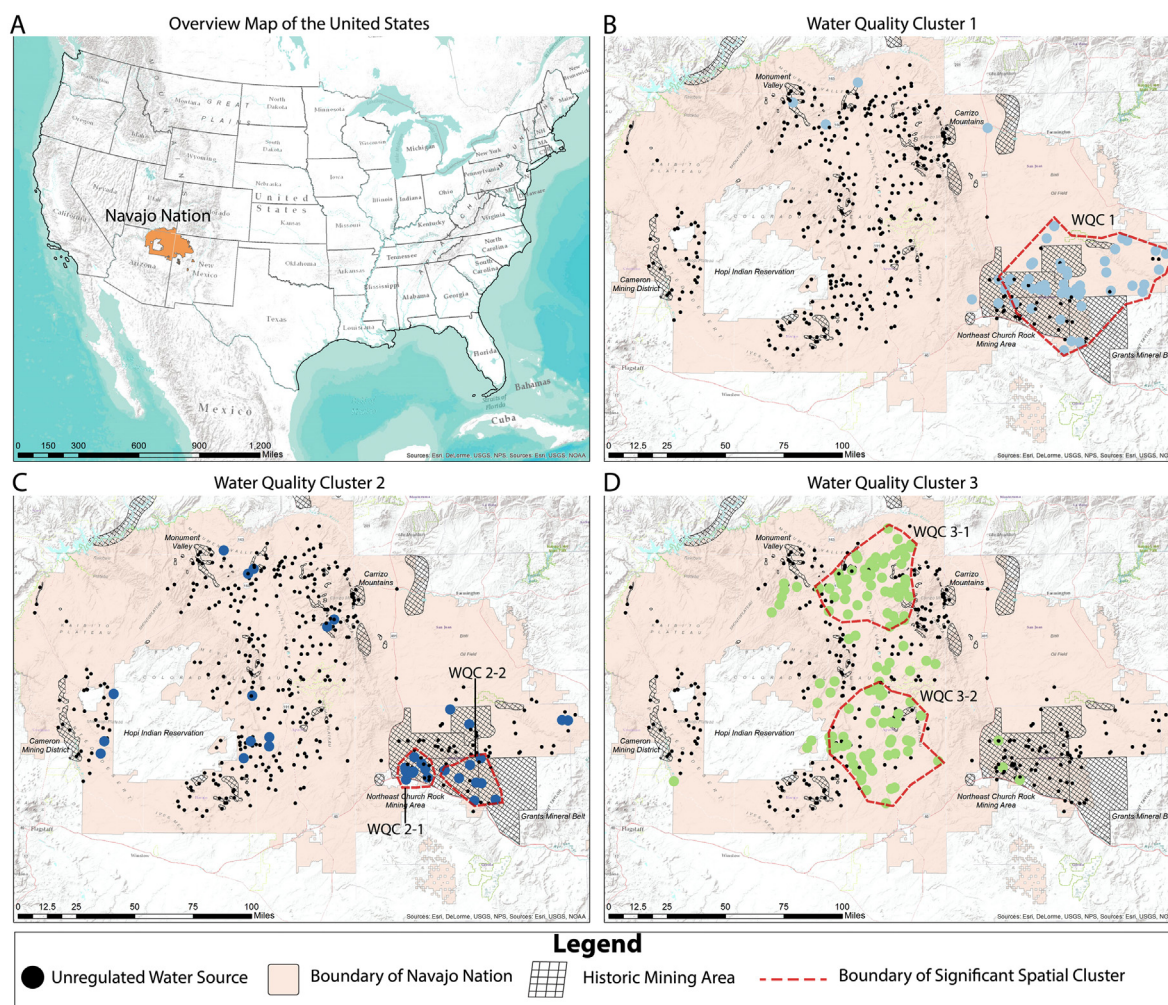


Fig. 3. Overview map (3A) of the Navajo Nation in the southwest United States and locations of unregulated water sources classified into Water Quality Cluster 1, 2, or 3. In panel B, C and D the larger dots are unregulated sources classified into the WQC of interest and the smaller black dots are other tested unregulated sources with different WQC membership. The dashed polygons around subsets of water sources indicate a significant spatial cluster.

Due to the regional focus of this study different contaminant mixtures were expected when compared to national studies. Nationally in the United States, arsenic co-occurs with contaminants such as manganese, molybdenum, strontium, radon, and nitrates (Toccalino et al., 2012); and uranium co-occurs more frequently with manganese and radon. These study results differ from national studies because the list of analyzed contaminants was more limited. Thus, the generalizability of the specific combinations observed in this study may be limited to other water sources in the southwestern United States. Previous water quality research in the study area investigated the presence of arsenic and uranium in unregulated sources (Corlin et al., 2016; Hoover et al., 2017). The work presented here suggests that arsenic, uranium, and other contaminants jointly drive potential toxicity in tested unregulated sources.

For this study, the identified contaminant mixtures are likely representative of unregulated water sources throughout the Navajo Nation, limited to the 14 metals and metalloids included in the analysis. Data included in this analysis came from water samples collected in a similar manner to how Navajo Nation residents collect water from these sources. The water quality data are likely representative of the conditions encountered by Navajo Nation residents who use these unregulated sources. Additionally, the 14 contaminants evaluated in this paper were measured in sources dispersed throughout most of the Navajo Nation. This includes unregulated water sources both close to and far from abandoned uranium mines and other features that might be contaminant sources. However, as noted by Toccalino et al. (2012)

characterization of contaminant mixtures is limited by the availability of occurrence data. For the Navajo Nation there remains a need to test for other contaminants (e.g., pesticides, radionuclides, microbial, etc. ...) that were not available for this analysis.

The spatial clustering analysis indicated specific areas of the Navajo Nation that had higher occurrences of certain WQCs. Small differences between the Kernel Density visualizations and SaTScan results were observed, likely due to how results are visualized by each method. Kernel density plots present a single smoothed surface that represents the density of all points classified as part of the WQC of interest. SaTScan presents the best collection of spatial clusters that are statistically significant. SaTScan results indicate only areas of significant spatial clustering, and not the full geographic extent of members of certain WQCs in the study area. There are some smaller groups of unregulated water sources with the same WQC membership that were not represented due to lack of significance. Overall, the areas with the greatest density of unregulated sources with the same WQC membership overlapped with the areas of statistically significant spatial clustering. These spatially clustered unregulated water sources may be pumping from water bearing geologic units with similar geochemistry or mineral composition. While the geochemistry data to evaluate the drivers of the spatial clustering is unavailable, this observation still has public health relevance. Decision makers may be able to use this information to inform policy choices and focus resources on areas with higher occurrences of deleterious metals or metalloids.

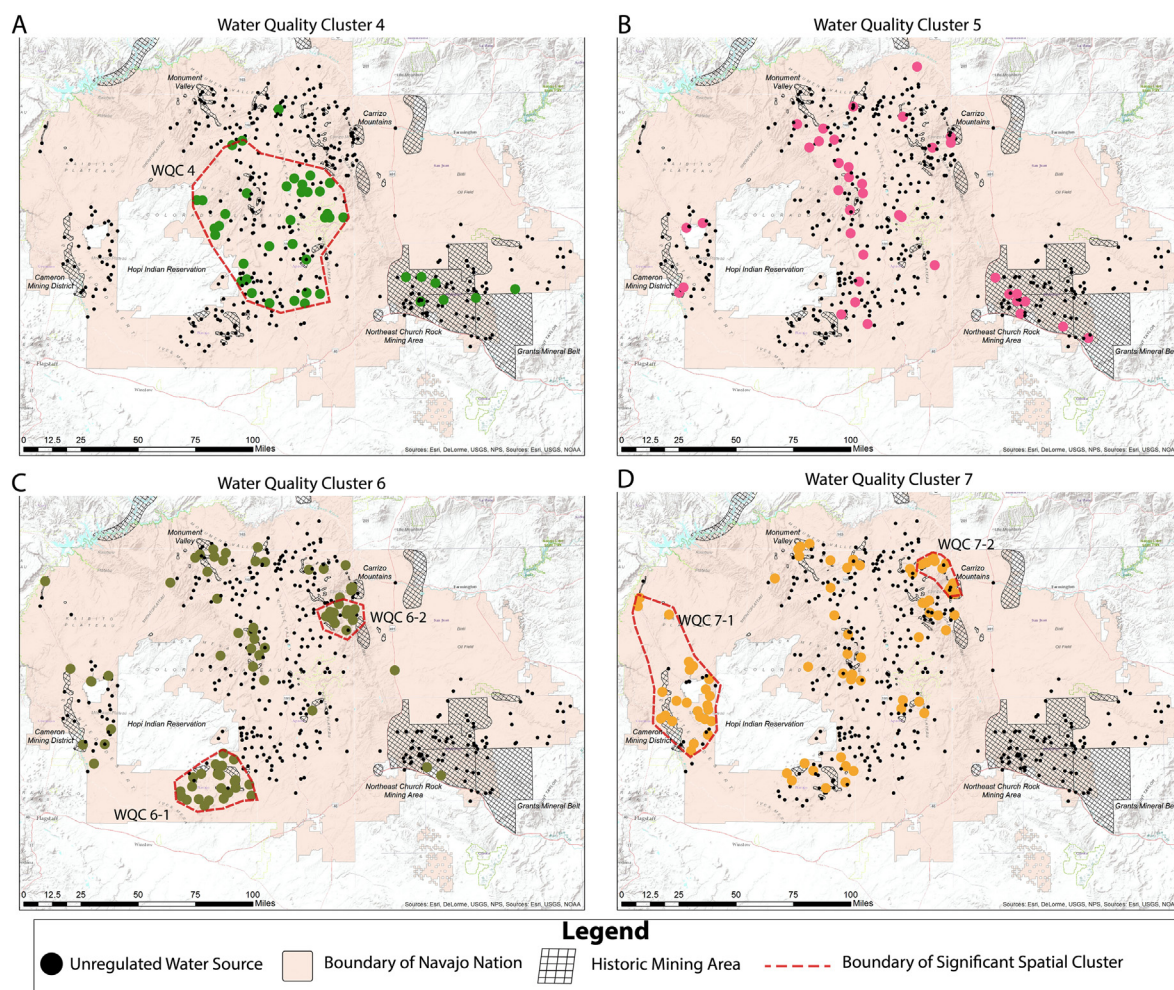


Fig. 4. Locations of unregulated water sources classified into Water Quality Cluster 4, 5, 6 or 7. In each panel, the larger dots are unregulated sources classified into the WQC of interest and the smaller black dots are other tested unregulated sources with different WQC membership. The dashed polygons around subsets of water sources indicate a significant spatial cluster.

The spatial analysis results also indicated that 55% of sources were not part of a spatial cluster. Unregulated water sources from each WQCs occurred outside of a spatial cluster, including all sources classified into WQC 5. This could be an illustration of the complex hydrogeology of the aquifers underlying the Navajo Nation. The C aquifer and N aquifers, named for the Coconino sandstone and Navajo sandstone respectively, generally produce high quality water (Brown and Macy,

2012). The C aquifer is found throughout much of the Navajo Nation, while the N aquifer is found primarily in the central area of Navajo near the Hopi Indian Reservation (Macy et al., 2012). Both aquifers are comprised of multiple water-bearing formations that occur at different depths below surface level (Cooley et al., 1964). The D aquifer is another important water supply, named for Dakota Sandstone, but it produces water with total dissolved solids concentrations exceeding 1000 mg per liter (Truini and Macy, 2006). There are also many small perched and alluvial aquifers that also supply water for domestic, livestock or agricultural uses (NN DWR, 2011). As can be observed in Figs. 3 and 4, unregulated sources with different WQC membership occurred throughout the Navajo Nation, and this may result from water sources tapping different water bearing units at different depths, despite proximity at surface level. For many unregulated water sources on the Navajo Nation there is limited information available about the producing formation, depth to formation, well type or construction materials, or geochemistry for the tested unregulated water sources. Therefore, a link between water quality profiles and specific water-bearing geologic formations could not be made; this remains an opportunity for future research.

5.3. Contaminants of concern

This investigation indicated that lead was detected at a greater frequency in unregulated sources classified into WQCs 2 and 5. The occurrence of lead in unregulated water sources is a potential public health concern; lead exposure is associated with decreased cognitive function

Table 4
Summary of spatial clustering for each water quality cluster (WQC) including the cluster size, total tested sources in the cluster area, and the number of sources with WQC membership matching the spatial cluster membership.

WQC	Max reported cluster size (%) ^a	Totals UWSs (N)	Cases (N)	% Cases in cluster Area
1	20	65	45**	69
2-1	4	17	10**	59
2-2	4	17	8**	47
3-1	20	79	53**	67
3-2	20	73	45**	62
4	40	148	35**	24
5	No significant spatial cluster identified			
6-1	8	30	24**	80
6-2	8	18	13**	72
7-1	8	33	23**	70
7-2	8	19	12*	63

^a Determined by identifying maximum Gini coefficient value.

* p -value < 0.05

** p -value < 0.001.

in children, as well as cardiovascular and renal problems (ATSDR, 2007; Lanphear et al., 2005). Additionally, dissolved lead in drinking water has been associated with elevated blood lead levels (Etchevers et al., 2014). Some of the unregulated water source infrastructure was installed in the 1950s and 60 so the lead detected in select unregulated sources could leach from the pipes used to construct wells or link the water source with a storage tank. Previous research investigating lead water concentrations in homes and schools indicated that during periods of nonuse (hours to days) lead can leach from pipes into stagnant drinking water (Bryant, 2004). In some regulated water systems on the Navajo Nation, federal and tribal agencies have replaced lead pipes in some public buildings and homes; a phosphate-buffer treatment system was also installed to reduce water pH and corrosive activity. Navajo Nation residents may infrequently use some of the sampled unregulated water sources, and therefore, water may stagnate in pipes for days. There exists limited information about well construction history or frequency of use that would help better address the source of lead in these unregulated water sources. A more detailed field analysis could be conducted to better ascertain the possible sources and provide this information to people who may use these water sources.

This investigation also showed that manganese, a neurotoxicant, occurred at concentrations that regularly exceeded health-based screening levels. Consumption of manganese in the range of 0.2–0.3 mg/L (comparable to manganese concentrations in WQCs 4 and 5) has been associated with diminished attention span and lower intellectual growth in children (Wasserman et al., 2006; Wasserman et al., 2016). Unregulated water sources with membership in WQCs 4 or 5 could be further investigated to determine the manganese source and potential intervention methods for reducing human exposure. Similar to ground-water in areas such as Bangladesh and the Mekong Delta, the co-occurrence of manganese with arsenic, uranium, and other contaminants makes contaminant removal more challenging (Bacquart et al., 2015; Frisbie et al., 2009). Water filters developed to remove arsenic via oxidation, such as those developed for use in Bangladesh, may have minimal effect on other contaminants resulting from different oxidation states of contaminants (Bacquart et al., 2015; Frisbie et al., 2009). Other water chemistry factors like pH also play an important role in the removal of metals and metalloids from drinking water (Lakshmanan et al., 2010) and would likely need to be evaluated before a removal system is designed for multiple contaminants. There may also be limited financial and institutional capacity to engage in widespread point-of-extraction filtration, because of high rates of unemployment and large distances between water sources on the Navajo Nation (Navajo Access Workgroup, 2010).

Identified mixtures could be further evaluated for human toxicity or investigated for possible human health links in a population-based study. Additionally, information about common contaminant mixtures could help public health officials prioritize areas for community outreach and intervention. This approach does not lessen the need for comprehensive and regular testing of unregulated sources for a variety of contaminants as recommended by the US Centers for Disease Control and Prevention (CDC, 2009). Additionally, the Navajo Nation is working to educate residents about the health risks of drinking water from an unregulated source.

5.4. Limitations

Water quality data used in this study were retrieved from a dataset comprised of results from seven different studies completed between 1993 and 2010 (Hoover et al., 2017). These studies were conducted by federal, state, tribal, and academic entities using similar collection protocols. While the collection protocol used across these studies was similar, the analysis methods changed over time and became more sensitive. While this evolution likely impacted all measured contaminants, it affects arsenic measurements in particular. The analyses conducted in the 1990s used Inductively Coupled Plasma Optical

Emissions Spectroscopy with a detection limit of 5 µg/L. Later analyses used Inductively Coupled Plasma Mass Spectroscopy that had a limit of detection below 1 µg/L. As a result, 82% of water sources tested for arsenic had a limit of detection at or below 1 µg/L. Clustering water sources using the 5 µg/L detection limit may impact results; however, the Robust ROS method accounts for the higher detection limit when calculating summary statistics.

Water sources tested for these previous studies were identified by community members. A random sampling design was not used and as a result some selection bias may have occurred; however, this was community driven research and a random sampling design would not have been supported by the community. Despite these limitations, these results are likely representative of water quality characteristics Navajo Nation residents may encounter due to the sample collection protocol and the geographic dispersion of tested water sources throughout most of the Navajo Nation.

There are limitations using the Benchmark Index as potential for human toxicity. The Benchmark Index metric can be applied only to contaminants with an established human-health benchmark, such as the Safe Drinking Water Act Maximum Contaminant Level (MCL) or Health-Based Screening Level. MCLs are not strictly health-based and include adjustments for economic feasibility and political considerations, which may result in underestimation of potential for toxicity. Additionally, contaminants without a human health benchmark cannot be assessed using the Benchmark Index. Despite these limitations this metric provides a benchmark for comparing and prioritizing contaminant mixtures for additional toxicological investigation (ATSDR, 2004; Ryker and Small, 2008). Additionally, the metals and metalloids included in this analysis are not the only contaminants that could cause negative human health impacts. Furthermore, low concentrations of these metals and metalloids does not indicate that a water source is safe for all possible contaminants.

There are also limitations to using Bayesian Profile Regression. Similar to other clustering methods, BPR is sensitive to noisy data that lack a strong clustering signal. When the overall clustering signal is weak and the data are noisy, BPR will group together profiles with similar joint levels in covariates or create small clusters, which makes calculating cluster specific parameters difficult (Molitor and Papathomas, 2010). Additionally, it is important to evaluate clustering uncertainty. If the cluster assignment uncertainty is higher than the cluster specific probabilities for many or all categories would not be different than the overall category probability for each cluster (illustrated as a green shaded boxplot in Supplemental Fig. 1). Reviewing the post-processing visualization can inform interpretation, but more developed guidelines to aid interpretation would be beneficial.

6. Conclusions

Drinking water is an important exposure pathway and identifying common contaminant mixtures in water supplies is a critical challenge to address. There are few examples that use multivariate statistical methods to identify representative combinations of metals and metalloids in water sources. To address this gap Bayesian Profile Regression was used to identify subgroups of unregulated water sources that had similar contaminant profiles. A human health toxicity metric was then employed to identify contaminant combinations that might warrant additional investigation. Lastly, a spatial analysis method was applied to evaluate the spatial clustering of unregulated water sources to inform public health officials and policy makers. Five metal combinations that may contribute to negative human health consequences were identified: arsenic and uranium; arsenic, uranium, and selenium; arsenic, uranium, and manganese; arsenic, uranium, selenium, and lead; arsenic, uranium, selenium, lead, and manganese. For a regional study, such as the study presented here, the ability to visualize the spatial distribution of clusters may aid in outreach and public policy decisions. More generally, this investigation demonstrated that a multivariate classification

technique may be used to identify and characterize contaminant mixtures in water supplies. These methods may be applied in the future to inform public policy makers, community outreach, and future research efforts to reduce exposure to multiple contaminants from unregulated water sources.

Acknowledgements

Thank you to the community and non-profit organizations that were instrumental in requesting the monitoring of unregulated water sources on the Navajo Nation. Funding for this work has been provided by the National Institute for Environmental Health Sciences, RO1 ES014565, R25 ES013208 and P30 ES-012072, a NIGMS ASERT IRACDA postdoctoral fellowship (K12 GM088021), the UNM Center for Native Environmental Health Equity Research- A Center of Excellence In Environmental Health Disparities Research- Funded jointly by grants from NIEHS & NIMHD ((1P50ES026102) & USEPA (#83615701), and the National Institute of Environmental Health Sciences Superfund Research Program (Award 1 P42 ES025589). This material was developed in part under Assistance Agreement No. 83615701 awarded by the U.S. Environmental Protection Agency to the University of New Mexico Health Sciences Center. It has not been formally reviewed by EPA. The views expressed are solely those of the speakers and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. We are grateful to the UNM Center for Advanced Research Computing for computational resources.

Appendix A. Supplementary data and figures

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2018.02.288>.

References

- Ander, E.L., Watts, M.J., Smedley, P.L., Hamilton, E.M., Close, R., Crabbe, H., et al., 2016. Variability in the chemistry of private drinking water supplies and the impact of domestic treatment systems on water quality. *Environ. Geochem. Health* 38.6, 1313–1332.
- Antweiler, R.C., Taylor, H.E., 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: 1. Summary statistics. *Environ. Sci. Technol.* 42, 3732–3738.
- ATSDR, 2004. Guidance Manual for the Assessment of Joint Toxic Action of Chemical Mixtures. Agency for Toxic Substances and Disease Registry, Atlanta, GA.
- ATSDR, 2007. Toxicological Profile for Lead. U.S. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry, Atlanta, GA.
- Backer, L.C., Tosta, N., 2011. Unregulated drinking water initiative for environmental surveillance and public health. *J. Environ. Health* 73, 31–32.
- Bacquart, T., Frisbie, S., Mitchell, E., Grigg, L., Cole, C., Small, C., 2015. Multiple inorganic toxic substances contaminating the groundwater of Myingyan Township Myanmar: arsenic, manganese fluoride, iron, and uranium. *Sci. Total Environ.* 517, 232–245.
- Brown, C.R., Macy, J.P., 2012. Groundwater, surface-water, and water-chemistry data from the C-aquifer monitoring program, northeastern Arizona, 2005–2011. Open-File Report 2012-1196. United States Geological Survey, Reston, VA.
- Bryant, S., 2004. Lead-contaminated drinking waters in the public schools of Philadelphia. *J. Toxicol. Clin. Toxicol.* 42, 287–294.
- Carlin, D.J., Rider, C.V., Woychik, R., Birnbaum, L.S., 2013. Unraveling the health effects of environmental mixtures: a NIEHS priority. *Environ. Health Perspect.* 121, a6–a8.
- Carpenter, D.O., Arcaro, K., Spink, D.C., 2002. Understanding the human health effects of chemical mixtures. *Environ. Health Perspect.* 110, 25–42.
- CDC, 2009. Well Testing. 2017. US Centers for Disease Control and Prevention, Atlanta, GA.
- Coker, E., Liverani, S., Ghosh, J.K., Jerrett, M., Beckerman, M., Li, A., et al., 2016. Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environ. Int.* 91, 1–13.
- Coker, E., Gunier, R., Bradman, A., Harley, K., Kogut, K., Molitor, J., et al., 2017. Association between pesticide profiles used on agricultural fields near maternal residences during pregnancy and IQ at age 7 years. *Int. J. Environ. Res. Public Health* 14, 506.
- Cooley, M.E., Harshbarger, J.W., Akers, J.P., Hardt, W.F., 1964. Regional hydrogeology of the Navajo and Hopi Indian reservations, Arizona, New Mexico, and Utah. Open-File Report. United States Geological Survey, Tucson, AZ.
- Corlin, L., Rock, T., Cordova, J., Woodin, M., Durant, J.L., Gute, D.M., et al., 2016. Health effects and environmental justice concerns of exposure to uranium in drinking water. *Curr. Environ. Health Rep.* 3, 434–442.
- DeSimone, L.A., Hamilton, P.A., Billiom, R.J., 2009. The quality of our nation's waters - Quality of water from domestic wells in principal aquifers of the United States, 1991–2004. Overview of Major Findings. U.S. Geological Survey Circular 1332. U.S. Geological Survey, Reston, VA, p. 48.
- ESRI, 2016. How Kernel Density Works. ESRI.
- Etchevers, A., Bretin, P., Lecoffre, C., Bidondo, M.-L., Le Strat, Y., Glorennec, P., et al., 2014. Blood lead levels and risk factors in young children in France, 2008–2009. *Int. J. Hyg. Environ. Health* 217, 528–537.
- Field, M.S., 2011. Application of robust statistical methods to background tracer data characterized by outliers and left-censored data. *Water Res.* 45, 3107–3118.
- Flem, B., Reimann, C., Birke, M., Banks, D., Filzmoser, P., Frengstad, B., 2015. Inorganic chemical quality of European tap-water: 2. Geographical distribution. *Appl. Geochem.* 59, 211–224.
- Focazio, M.J., Welch, A.H., Watkins, S.A., Helsel, D.R., Horn, M.A., 2000. A Retrospective Analysis on the occurrence of arsenic in ground-water resources of the United States and limitations in drinking-water-supply characterizations. Water-Resources Investigations Report 99-4279. US Geological Survey, Reston, VA.
- Frisbie, S., Mitchell, E., Mastera, L.J., Maynard, D.M., Yusuf, A.Z., Siddiq, M.Y., et al., 2009. Public health strategies for western Bangladesh that address arsenic, manganese, uranium, and other toxic elements in drinking water. *Environ. Health Perspect.* 117.
- Gentry, R.W., 2013. Efficacy of fuzzy c-means cluster analysis of naturally occurring radioisotope datasets for improved groundwater resource management under the continued risk of climate change. *Br. J. Environ. Clim. Chang.* 3, 464–479.
- Güler, C., Thyne, G.D., 2004. Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. *Water Resour. Res.* 40, W12503.
- Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol. J.* 10, 455–474.
- Han, J., Zhu, L., Kuldorff, M., Hostovich, S., Stinchcomb, D.G., Tatalovich, Z., et al., 2016. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. *Int. J. Health Geogr.* 15, 27.
- Hastie, D., Liverani, S., Azizi, L., Richardson, S., Stücker, I., 2013. A semi-parametric approach to estimate risk functions associated with multi-dimensional exposure profiles: application to smoking and lung cancer. *BMC Med. Res. Methodol.* 13, 129.
- Helsel, D., 1990. Less than obvious: statistical treatment of data below the detection limit. *Environ. Sci. Technol.* 24, 1767–1774.
- Helsel, D., 2005. Nondetects and data analysis. John Wiley and Sons, New York.
- Helsel, D., 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65, 2434–2439.
- Helsel, D., 2012. Statistics for Censored Environment Data Using Minitab and R. John Wiley & Sons, Inc., Hoboken, NJ.
- Hertzberg, R.C., Rice, G.E., Teuschler, L.K., Wright, J.M., Simmon, J.E., 2008. Health risk assessment of chemical mixtures in drinking water. In: Howd, R.A., Fan, A.M. (Eds.), *Risk Assessment for Chemicals in Drinking Water*. John Wiley & Sons, Inc.
- Hoover, J.H., Gonzales, M., Shuey, C., Barney, Y., Lewis, J., 2017. Elevated arsenic and uranium concentrations in unregulated water sources on the Navajo Nation, USA. *Expo. Health* 9, 113–124.
- Hund, L., Bedrick, E.J., Miller, C., Huerta, G., Nez, T., Ramone, S., et al., 2015. A Bayesian framework for estimating disease risk due to exposure to uranium mine and mill waste on the Navajo Nation. *J. R. Stat. Soc. A. Stat. Soc.* 178, 1069–1091.
- Hussain, M., Ahmed, S.M., Abderrahman, W., 2008. Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *J. Environ. Manag.* 86, 297–307.
- Kim, J., Jung, I., 2017. Evaluation of the Gini coefficient in spatial scan statistics for detecting irregularly shaped clusters. *PLoS One* 12, e0170736.
- Kim, K.-H., Yun, S.-T., Park, S.-S., Joo, Y., Kim, T.-S., 2014. Model-based clustering of hydrochemical data to demarcate natural versus human impacts on bedrock groundwater quality in rural areas, South Korea. *J. Hydrol.* 519, 626–636.
- Kim, K.-H., Yun, S.-T., Kim, H.-K., Kim, J.-W., 2015. Determination of natural backgrounds and thresholds of nitrate in south Korean groundwater using model-based statistical approaches. *J. Geochem. Explor.* 148, 196–205.
- Krolik, J., Maier, A., Evans, A., Belanger, P., Hall, G., Joyce, A., 2013. A spatial analysis of private well water *Escherichia coli* contamination in southern Ontario. *Geospat. Health* 8, 65–75.
- Krolik, J., Evans, G., Belanger, P., Maier, A., Hall, G., Joyce, A., et al., 2014. Microbial source tracking and spatial analysis of *E. coli* contaminated private well waters in southeastern Ontario. *J. Water Health* 12, 348–357.
- Kuldorff, M., 2009. Information Management Services Inc: SaTScan v8.0: Software for the Spatial and Space-time Scan Statistics.
- Kuldorff, M., Feuer, E.J., Miller, B.A., Freedma, L.S., 1997. Breast cancer clusters in the northeast United States: a geographic analysis. *Am. J. Epidemiol.* 146, 161–170.
- Lakshmanan, D., Clifford, D.A., Samanta, G., 2010. Comparative study of arsenic removal by iron using electrocoagulation and chemical coagulation. *Water Res.* 44, 5641–5652.
- Lanphear, B.P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D.C., et al., 2005. Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environ. Health Perspect.* 894–899.
- Lee, L., 2013. Nondetects and Data Analysis for Environmental Data. CRAN.
- Lee, L., Helsel, D., 2005. Baseline models of trace elements in major aquifers of the United States. *Appl. Geochem.* 20, 1560–1570.
- Leeper, J.W., 2003. Navajo Nation plans for their water future. 48th Annual New Mexico Water Conference. New Mexico Water Resources Research Institute, Santa Ana Pueblo.
- Levitán, D.M., Schreiber, M.E., Seall, R.R.L., Bodnar, R.J., Aylor, J.G.J., 2014. Developing protocols for geochemical baseline studies: an example from the Coles Hill uranium deposit, Virginia, USA. *Appl. Geochem.* 43, 88–100.

- Lewis, J., Hoover, J., MacKenzie, D., 2017. Mining and environmental health disparities in native American communities. *Curr. Environ. Health Rep.* 1–12.
- Liverani, S., Hastie, D.L., Azizi, L., Papathomas, M., Richardson, S., 2015. PRemiuM: an R package for profile regression mixture models using dirichlet processes. *J. Stat. Softw.* 64.
- Macy, J.P., Brown, C.R., Anderson, J.R., 2012. Groundwater, surface-water, and water-chemistry data, Black Mesa Area, northeastern Arizona-2010–2011. Open-File Report 2012-1102. United States Geological Survey, Reston, VA.
- Mandel, P., Maurel, M., Chenu, D., 2015. Better understanding of water quality evolution in water distribution networks using data clustering. *Water Res.* 87, 69–78.
- Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., Linsey, K.S., 2014. Estimated Use of Water in the United States in 2010. U.S. Geological Survey, Reston, VA.
- McNeil, V.H., Cox, M.E., Preda, M., 2005. Assessment of chemical water types and their spatial variation using multi-state cluster analysis, Queensland, Australia. *J. Hydrol.* 310, 181–200.
- Molitor, J., Papathomas, M., 2010. Bayesian profile regression with an application to the national survey of children's health. *Biostatistics* 11, 484–498.
- Morrison, J.M., Goldhaber, M.B., Ellefsen, K.J., Mills, C.T., 2011. Cluster analysis of a regional-scale soil geochemical dataset in northern California. *Appl. Geochem.* 26, S105–S107.
- Murphy, M., Lewis, L., Sabogal, R.I., Bell, C., 2009. Survey of unregulated drinking water sources on the Navajo Nation. Annual Meeting of the American Public Health Association, Philadelphia, PA.
- Navajo Access Workgroup, 2010. Mapping of water infrastructure and homes without access to safe drinking water and basic sanitation on the Navajo Nation. Federal Infrastructure Task Force.
- NN DWR, 2011. Draft Water Resource Development Strategy for the Navajo Nation. Navajo Nation Department of Water Resources, Window Rock.
- Papathomas, M., Molitor, M., Richardson, S., Ribald, E., Vines, P., 2010. Examining the joint effect of multiple risk factors using exposure risk profiles: lung cancer in nonsmokers. *Environ. Health Perspect.* 119, 84–91.
- Papathomas, M., Molitor, J., Haggard, C., Hattie, D., Richardson, S., 2012. Exploring data from genetic association studies using Bayesian variable selection and the dirichlet process: application to searching for gene X gene patterns. *Genet. Epidemiol.* 36, 663–674.
- Pirani, M., Best, N., Blangiardo, M., Liverani, S., Atkinson, R., Fuller, G., 2015. Analyzing the health effects of simultaneous exposure to physical and chemical properties of airborne particles. *Environ. Int.* 79, 56–64.
- Ryker, S., 2003. Arsenic in ground water used for drinking water in the United States. In: Welch, A., Stollenwerk, K.G. (Eds.), *Arsenic in Ground Water: Geochemistry and Occurrence*. Kluwer Academic Publishers, Norwell, Massachusetts.
- Ryker, S.J., Small, M.J., 2008. Combining occurrence and toxicity information to identify priorities for drinking-water mixture research. *Risk Anal.* 28, 653–666.
- Sanders, A.P., Desrosiers, T.A., Warren, J.L., Herring, A.H., Enright, D., Olshan, A.F., et al., 2014. Association between arsenic, cadmium, manganese, and lead levels in private wells and birth defects prevalence in North Carolina: a semi-ecological study. *BMC Public Health* 14, 955.
- Squillace, P.J., Scott, J.C., Moran, M.J., Nolan, B.T., Kolpin, D.W., 2002. VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the United States. *Environ. Sci. Technol.* 36, 1923–1930.
- Swanson, S.K., Bahr, J.M., Schwas, M.T., Potter, K.W., 2001. Two-way cluster analysis of geochemical data to constrain spring source waters. *Chem. Geol.* 179, 73–91.
- Templ, M., Filzmoser, P., Reimann, C., 2008. Cluster analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 23, 2198–2213.
- Toccalino, P.L., 2007. Development and Application of Health-Based Screening Levels for Use in Water-Quality Assessments. Citeseer.
- Toccalino, P.L., Norman, J.E., Scott, J.C., 2012. Chemical mixtures in untreated water from public-supply wells in the U.S. - occurrence, composition, and potential toxicity. *Sci. Total Environ.* 431, 262–270.
- Truini, M., Macy, J.P., 2006. Lithology and thickness of the Carmel Formation as related to leakage between the D and N aquifers, Black Mesa, Arizona. Scientific Investigations Report 2005-5187. United States Geological Survey, Reston, VA.
- US EPA, 2000. Abandoned Uranium Mines Project: Arizona, New Mexico, Utah, 1994–2000. US Environmental Protection Agency, Washington DC.
- US EPA, 2004. National Water Quality Inventory: Report to Congress, 2004. US Environmental Protection Agency, Washington DC.
- US EPA, 2006. Abandoned Uranium Mines (AUM) on the Navajo Nation. US EPA: Region 9, CA, San Francisco.
- US EPA, 2015. Providing Safe Drinking Water in America: 2013 National Public Water Systems Compliance Report. US Environmental Protection Agency, Office of Enforcement and Compliance Assurance, Washington, DC.
- US EPA Region IX, 2008. Navajo Nation Drinking Water Source Sampling: February–March 2008. US Environmental Protection Agency, San Francisco, CA.
- US EPA Region IX, 2010. Navajo Nation Unregulated Water Source Sampling Results: October 2009 Sampling Event. US Environmental Protection Agency, San Francisco, CA.
- US EPA Region IX, 2011. Navajo Nation Water Wells Sampling: Church Rock Chapter. US Environmental Protection Agency, San Francisco, CA.
- Vandeberg, G.S., Dixon, C.S., Vose, B., Fisher, M.R., 2015. Spatial assessment of water quality in the vicinity of Lake Alice National Wildlife Refuge, Upper Devils Lake Basin, North Dakota. *Environ. Monit. Assess.* 187, 40.
- Wasserman, G.A., Liu, X., Parvez, F., Ahsan, H., Levy, D., Factor-Litvak, P., et al., 2006. Water manganese exposure and children's intellectual function in Araihaazar, Bangladesh. *Environ. Health Perspect.* 124–129.
- Wasserman, G.A., Liu, X., Parvez, F., Factor-Litvak, P., Kline, J., Siddique, A.B., et al., 2016. Child intelligence and reductions in water arsenic and manganese: a two-year follow-up study in Bangladesh. *Environ. Health Perspect.* 124, 1114.
- Weiner, E.R., 2013. Behavior of radionuclides in the water and soil environment. In: Weiner, E.R. (Ed.), *Applications of Environmental Aquatic Chemistry*. CRC Press, pp. 383–440.