A spatial autologistic model to predict the presence of arsenic in private wells across Gaston County, North Carolina using geology, well-depth, and pH

Abstract

Chronic exposure to arsenic-contaminated drinking water is detrimental to human health. We develop an autologistic regression model to evaluate if the geology, pH, and well depth can improve our ability to predict the presence of arsenic at and above detectable levels (≥ 5 μg/L) found in private wells. We use arsenic samples measured in private well water across Gaston County, North Carolina, from 2011 to 2017. We use kriging to map the probability of arsenic at detectable levels across Gaston County. Arsenic at detectable levels was reported at 78 private wells and the median pH for all the samples was 7.1. Our spatial autologistic model suggests that arsenic at detectable levels is positively associated with pH. In addition, private wells set in Mica schist (€Zms) were associated with arsenic, suggesting a local-scale geologic source influence of arsenic in the county. Our kriging map shows that the northwestern section of the county has more than a 50 percent probability to have arsenic at detectable levels. In conclusion, the results of our model provide evidence to warrant testing of wells in the Mica schist unit, and those using wells with higher arsenic levels could take action to reduce their risk. The map of probability of arsenic at and above detectable levels can be used to implement cost-effective targeted interventions.

Keywords: *Arsenic, Autologistic Regression, Geology, GIS, Private Wells, Water*

1. Background and Rationale

Chronic exposure to elevated arsenic levels (>10 µg/L) in drinking water has been associated with several types of cancers including prostate (Benbrahim-Tallaa and Waalkes 2008), lung (Heck et al. 2009; Dauphiné et al. 2013), bladder (Steinmaus et al. 2003), kidney (Yuan et al. 2010), and skin (Karagas et al. 2015). Recent studies have suggested that even low levels of arsenic (<10 µg/L) in drinking water may impact fetal development (Bloom et al. 2016; Almberg et al. 2017), increase odds of diabetes (Mahram et al. 2013), and cause heart diseases (Bräuner et al. 2014; James et al. 2015).

In the United States (U.S) alone, 2.1 million people out of 44.1 million Americans are relying on private wells for water consumption, and do so at unsafe arsenic concentration levels above the public drinking water standard (10 µg/L) set by the U.S. Environmental Protection Agency (USEPA) (Ayotte et al. 2017). Yet, private wells are not regulated in the U.S (MacDonald Gibson and Pieper 2017). In Gaston County, North Carolina (the focus of our study, Fig 1), nearly 42% of the residents rely on private well water (Centers for Disease Control and Prevention (CDC) 2019). The accurate prediction of the spatial variation of arsenic in groundwater is critical to water supply management.

Arsenic has been found at elevated levels in groundwater aquifers across North Carolina, US (Sanders et al. 2012), China (He et al. 2020a), Bangladesh (Hossain and Sivakumar 2006), Nepal (Gurung et al. 2005) and many other countries. He et al. (2020b) determined that elevated arsenic levels in groundwater in Datong Basin, China was associated with geological and climatic conditions characterized by active water–rock interactions, strong evaporation, and low groundwater flow rate. In North Carolina, some studies have underlined possible associations between elevated arsenic concentrations and metavolcanic or metavolcaniclastic rocks (Pippin 2005; Harden et al. 2009), and metamorphosed clastic sedimentary rocks (Chapman et al. 2013). Reid et al. (2005) have suggested that the occurrence of arsenic in groundwater in the Piedmont of North Carolina could be related to fracture coatings in iron-manganese filled borehole cores from oxidized zones. The northwestern part of Gaston County, North Carolina (Fig 1) is within the area described as the physiographic and general geologic Piedmont of North Carolina. Chapman et al. (2013) have suggested that elevated arsenic concentrations in groundwater from rock units are positively correlated with pH of 7.2 or greater in the Piedmont of North Carolina. At a high pH, the formation of soluble ions can increase arsenic mobilization through desorption processes (Ayotte et al. 2003; Ayotte et al. 2006).

Most private wells in the Piedmont of North Carolina obtain water by drilling into bedrock, but a few wells tap water from the regolith at shallow depth (Daniel and Dahlen 2002). Two studies have examined the relationship between arsenic concentration and well depth in bedrock aquifers in the Piedmont of North Carolina. Kim et al. (2011) found associations between elevated arsenic levels and deep wells within welded tuffs and quartz units that were close to the transition zones between primarily pyroclastic and primarily volcaniclastic sedimentary rocks. Chapman et al. (2013) found that arsenic concentrations in crystalline lithologies were positively correlated with well depth. However, to date there is no predictive model of the presence of arsenic in Gaston County, even though the county is ranked among the top counties in North Carolina, with the most private wells with arsenic concentrations exceeding the United States EPA drinking water standards (Sanders et al. 2012). The complexity and spatial distribution of geologic formations make it difficult to assume that arsenic concentration would be evenly distributed in the county.

Spatial modeling and geostatistics have received considerable attention for the prediction of arsenic in groundwater (Gaus et al. 2003; Goovaerts et al. 2005; Meliker et al. 2008; Kim et al. 2011; Dummer et al. 2015). When most of the data contain arsenic values that are reported as below the limit, researchers have relied on geostatistical techniques such as indicator kriging to estimate the occurrence of arsenic (Goovaerts et al. 2005; Lee et al. 2008; Goovaerts 2009; Hassan and Atkins 2011; Antunes and Albuquerque 2013), yet these approaches typically do not incorporate predictor variables. Some studies have used logistic regression with various predictors (geologic and anthropogenic sources of arsenic, geochemical processes, hydrogeologic, and land-use factors) to model the occurrence of arsenic ≥ 5 μg/L (Ayotte et al. 2006; Bretzler et al. 2017). The ordinary logistic regression is based on the assumption that the relationship between the presence of arsenic and potential confounding factors would not change across a region. However, spatial autocorrelation, defined as a measure of the similarity in values for nearby observations (Griffith 1987), is frequently present in environmental data. For example, there are different geologic regions in Gaston County, and samples taken from those distinct regions may exhibit strong similarities, violating the assumption of spatial stationarity. Therefore, ignoring spatial effects in ordinary logistic regression could result in a biased and under-performing model (Bo et al. 2014). Autologistic regression could be used to alleviate this problem (Griffith 2004; Dormann 2007; Fu et al. 2013; Bo et al. 2014; Seeley et al. 2019).

The autologistic regression is a spatial model that incorporates a spatial autocorrelation (autocovariate) variable into a logistic regression model to obtain robust inference (Griffith 2004; Dormann 2007; Fu et al. 2013; Bo et al. 2014; Liu et al. 2018). The autocovariate variable introduced in an autologistic regression reflects the first law of geography that near things are more related than distant things (Tobler 1970; Tobler 1979; Miller 2004). In this study, we assumed that the probability of arsenic occurrence in a private well is higher if it is also present in nearby private wells. The autologistic regression has gained attention in ecological studies (Wu and Huffer 1997; Dormann 2007; Tsuyuki 2008), transportation research (Liu and Sharma 2019), but have not been applied to model the occurrence of arsenic.

We develop a spatial autologistic regression model to evaluate if the geology, pH, and well depth can improve our ability to predict the presence of arsenic at and above detectable levels (≥5 µg/L) in private wells. We used this threshold because all arsenic concentration data in our study used EPA method 200.8 that has a detection limit of 5 µg/L (USEPA 1994; North Carolina Department of Health and Human Services 2020). Also, we used this threshold because lifetime exposure to even relatively low arsenic concentration can have adverse health effects.

2. Study Area and Geologic Setting

Gaston County, North Carolina (364 mi^2 or 942.756km^2) is a fast-growing county of nearly 225,000 residents (2019) in the South-Central Piedmont section of North Carolina, with the city of Gastonia serving as the county seat (35.2621° N, 81.1873° W). The county is bounded on the east by the Catawba River and Mecklenburg County, on the west by Cleveland County, on the north by Lincoln County and on the south by York

County, South Carolina. Gaston County enjoys a temperate climate with moderate temperature variations and humidity. The topography of the County is gently rolling to hilly, with several pronounced ridges. Elevations above sea level range from 587 feet (179 meters) in the southeast corner to 1,705 feet (520 meters).

Gaston County, North Carolina, is composed of Inner Piedmont (1), Kings Mountain (2), and Charlotte (3) geologic belts (Fig 1) (North Carolina Department of Environmental Quality 2020). In contrast with the historic terminology of (east to west) Charlotte belt, Kings Mountain belt, and Inner Piedmont, Gaston County sits astride the Central Piedmont suture zone (a complex tectonic boundary) that joins the Carolina terrane to the Cat Square terrane in the Inner Piedmont (Huebner et al. 2017). Huebner et al. (2017) described the Cat Square terrane as a remnant of an early Paleozoic ocean basin.

<Fig 1 around here>

Fig 1 Spatial distribution of arsenic concentrations in well water samples in Gaston County, North Carolina. (Geologic data source: North Carolina Department of Environmental Quality, 2020)

The geologic belts have bedrock of varying ages and formations. A geologic formation is a fundamental unit in the classification of rocks based on similar characteristics in mineral composition, grain size, and color (Carter et al. 2002). The geologic formations in the Inner Piedmont consist of amphibolite and biotite gneiss, Cherryville granite, metamorphosed granitic rock, and mica schist (Fig 1). The mica schist formation consists of both €Zs and €Zms. €Zs is a “white-mica schist” that, depending on locality, contains layers of biotite gneiss, quartz-mica schist, micaceous quartzite, and rare amphibolite. In Gaston County, much of the €Zms unit appears to be a country-rock to the Mississippian Cherryville Granite (Mc in Fig 1) (Goldsmith et al. 1988). Following the interpretation of Goldsmith et al. (1988), the ЄZms unit in this study forms a suite of mainly stratified groups of similar age and thus related source environments. The Cherryville Granite is a late- to post-metamorphic two-mica granite that is associated with elevated radon (Waldron et al. 2007; Werner et al. 2009).

The geologic formations in the Charlotte belt consist of granitic rock, metamorphosed granitic rock, metamorphosed quartz diorite, gabbro of concord plutonic suite, and felsic metavolcanic rock. The Kings Mountain belt consists of metamorphic rocks in the Battleground Formation, Blacksburg Formation, foliated to massive granitic rock, Cherryville granite, and metamorphosed quartz diorite (Goldsmith et al. 1988). The Battleground Formation consists of protoliths formed during the Late Proterozoic and later metamorphosed to form a combination of quartz-sericite schist with metavolcanic rocks, quartz-pebble metaconglomerate, and kyanite-sillimanite quartzite. The Blacksburg Formation consists of sericite schist with graphite, phyllite, amphibolite, and calc-silicate rocks formed in the late Proterozoic-Cambrian.

3. Material and Methods

3.1. Arsenic Concentration in Well Water

Arsenic data for private wells was obtained from the Gaston County Department of Health and Human Services (GC-DHHS) for 2011 through 2017. The data also contained information on the permit number, owner’s name, residential address, collection date, sampling point, pH, and other inorganic chemicals. We used a GIS to geocode residential addresses to determine their geographic coordinates (Owusu et al. 2017). Some of the records represent repeated sampling of the same well – e.g., when water samples are taken from the kitchen sink and at the well. We therefore retained only the maximum recorded value from the location with the multiple tests to reflect potential groundwater concentration, which reduced our samples to 1082. This method has been used in similar studies to preserve the number of samples above the reporting limit (Ayotte et al. 2006; Kim et al. 2011; Gross and Low 2013; Ayotte et al. 2017; VanDerwerker et al. 2018). We also excluded 92 records because the pH values were missing, which reduced our final samples set to 990.

3.2. **Estimating Well Depth and Geologic Information**

We obtained a digital copy of Gaston County's private wells permit data from GC-DHHS to get well depth information to associate with the arsenic data (Fig 2). The well depth does not accurately reflect the depth of the water sample, because geologic changes, groundwater flow, weather cycles and precipitation patterns can affect the level of the water table (United States Geological Survey 2020). In this study, we relied on the well depth because the actual depth of the water sample was not available.

Out of the 990 samples, we were able to merge 509 arsenic samples to the permit data using either the permit numbers, residential address, or name to extract the well depth information (Fig 2). For the remaining 481 sampled wells that were not merged to the permit data due to missing data, we imputed the well depth information using an inverse distance weighting (IDW), an interpolation technique (Fig 2). The IDW surface was developed from 7837 well depths in the permit data.

Ethan and Xiao-Ming (2018) have suggested that the depth from the regolith to the bedrock aquifer frequently tapped by shallow wells ranges from 0 to 150 feet in Orange County, which is also in the Piedmont region. We assumed this could also be the case in Gaston County and classified the well depths into three groups; 1) shallow (≤ 150 feet), 2) moderate (151 – 300 feet), and deep (≥ 301 feet) evaluate the differences in risk of arsenic in private wells. It was appropriate to categorize well depth into three groups (shallow, moderate, and deep) because the relationship between probability of arsenic concentration ≥ 5 μg/L and well depth may not be linear due to differences in well construction. The well depth groups can therefore help to understand the real relationships and differences in the probability of arsenic concentration ≥ 5 μg/L considering that the characteristics of the sampled private wells are not provided in the data.

<Fig. 2 around here>

Fig 2 Workflow to estimate well depth for private wells

We obtained the geologic data for Gaston County from the North Carolina online GIS Portal. The NC Department of Environmental Quality Division of Land Resources, NC Geological Survey, and NC Center for GIS developed the digital data at a scale of 1: 250,000 miles. We spatially joined the sampled arsenic locations to the geologic data using a GIS.

3.3. **Development of the Autologistic Regression Model**

Similar to Ayotte et al. (2006), we converted the arsenic concentration to 1 if ≥ 5 μg/L and 0 if < 5 μg/L because 912 samples were marked as '< 5 μg/L' and 78 samples were reported arsenic concentrations. Because of the small number of samples with arsenic concentration ≥ 5 μg/L, we did not split the datasets into train and validation data. Instead, we used all the data in the model development to allow for a better model. We used an autologistic regression model to predict locations where the presence of arsenic concentration is ≥ 5 μg/L in private wells. The assumption for the autologistic regression is that relationships between the presence of arsenic and the explanatory factors are similar for nearby private wells than distant wells. We estimated the probability of elevated arsenic concentration at a location *i* using the autologistic function (Tsuyuki 2008).

$$p_i = \frac{1}{1 + exp\left[-\left(\beta_0 + \beta_1 x_{1,i} + .. \beta_n x_{n,i} + C(auto\ cov_i)\right)\right]} \quad (1)$$

i is the location of the private well, $x_1 \ldots x_n$ are the covariates, $\beta_0, \beta_1, \beta_n\ and\ C$ are the estimated coefficients. The introduction of the autocovariate variable in the autologistic regression penalizes the regression constant and reduces the contribution of the residuals to produce robust predictions (Griffith 2004; Dormann 2007; Fu et al. 2013; Bo et al. 2014). The autocovariate variable for a location *i* is calculated using Equation 2.

$$Auto\ cov_i = \frac{\sum_{j=1}^{k} w_{ij} \hat{P}_j}{\sum_{j=1}^{k} w_{ij}} \quad (2)$$

The autocovariate variable ($auto\ cov$) is a weighted average of the probabilities of arsenic concentration ≥ 5 μg/L of a set of nearby private wells *j* ($j = 1 \ldots k$) to the private well at *i*. The weight between private wells *i and j* is $w_{ij} = \frac{1}{d_{ij}}$, where d_{ij} is the Euclidean distance between private wells *i* and *j*, and $\hat{P}_j$ probability of arsenic concentration ≥ 5 μg/L at *j*. We determined that the minimum Euclidean distance (bandwidth) at which no private well had zero neighbors was 1976 meters and used this value (d_{ij}) in the analysis.

We used the “spatialEco” package in R/R Studio version 3.6 (Evans and Ram 2020) to implement the spatial autologistic regression model. We assessed the overall model performance by computing the receiver operating characteristic (ROC) area under curve (AUC) value. This value is a ratio of the true positive rate to the false positive rate, integrated over a range of probability thresholds, and indicates model fit (Hamel 2009). AUC values range from 0.5 to 1; where 0.5 means that the model is no better than predicting the outcome by a random chance, 0.7 is a good model; 0.8 is a robust model, and 1 is a perfect model (Hamel 2009). We also report the percentage of the correctly classified and the Chi-Square test for goodness of model fit.

3.4. **Development of an Interpolated Probability Surface**

Our model results return the probability of arsenic concentrations ≥ 5 μg/L that we mapped to reveal spatial patterns throughout Gaston County, along with the residuals using Kriging. Kriging is an interpolation method to estimate the values of a variable at unsampled locations using observations from known sites (Hengl 2009; Li and Heap 2011; Li and Heap 2014). The interpolation surface allows for delineating areas with a high probability of arsenic concentration ≥ 5 μg/L in well water in Gaston County. The kriging interpolation was developed with the Gstat R statistical package (Pebesma et al. 2019).

4. Results

4.1. Distribution of Arsenic Concentration

Out of the 990 arsenic measurements, a total of 912 samples contained arsenic concentrations < 5 μg/L; 78 samples were ≥ 5 μg/L. Out of 78 samples with detectable levels of arsenic (≥ 5 μg/), 42 samples had concentrations from 5 to 6 μg/L (Fig 3). The maximum reported arsenic concentration in well water was 81 μg/L in the Kings mountain geological belt (Fig 1). The pH in well water samples ranged from 5.1 to 9.7. The median pH in well water samples was 7.1. Sampled wells that contained arsenic concentrations ≥ 5 μg/L had an average pH of 7.3, while those at lower levels (< 5 μg/L) had a pH of 7.1. The minimum and maximum well depth were 30 ft and 1205 ft respectively. The average prediction error for the IDW was -3.1 ft and the RMSE of 125 ft.

<Fig. 3 around>

Fig 3 Distribution of arsenic for the 78 samples at and above detectable levels (5 μg/L) (samples marked as '< 5 μg/L' in the data had a frequency of 912- not included in the histogram)

As shown in Fig 1, the spatial distribution of the presence of arsenic and the geologic units in Gaston County suggest that most of the samples with arsenic concentration ≥ 5 μg/L were in the northwestern part of the county, which is an area within the €Zms - Mica schist geologic unit. Specifically, within the €Zms - Mica schist unit, 28% (n = 26) of the samples in that unit exhibited arsenic concentration ≥ 5 μg/L (Table 1). Noteworthy, 15.1% (n = 8) of samples with arsenic concentration ≥ 5 μg/L were found in private wells located on the Mc - Cherryville Granite.

Table 1 Samples with arsenic concentration (≥ 5 μg/L) for geologic units in Gaston County, North Carolina

Geologic unit	**Total (N)**	***(n)* *(≥ 5 μg/L)***	***%* *(≥ 5 μg/L)***
€Zab - Amphibolite and biotite gneiss	7	1	14.3
€Zbg - Mica schist	4	0	0
€Zbl - Blacksburg Formation	57	5	8.8
€Zfv - Felsic metavolcanic rock	58	3	5.2
€Zg - Metamorphosed granitic rock	39	2	5.1
€Zms - Mica schist	93	26	28
DOg - Granitic rock	52	4	7.7
Mc - Cherryville Granite	53	8	15.1

O€g - Metamorphosed granitic rock	2	0	0
PPmg - Foliated to massive granitic rock	199	13	6.5
PzZq - Metamorphosed quartz diorite	279	10	3.6
Zbt - Battleground Formation	147	6	4.1

We summarized the number and percent of samples with arsenic concentration ≥ 5 µg/L for the different geologic belts in Gaston County (Table 2). Overall, 22.1% (n = 34) of the sampled wells in the Inner Piedmont belt had arsenic concentrations ≥ 5 µg/L. We found 6.3% (n = 27) of the sampled wells in the Kings Mountain belt had arsenic concentration ≥ 5 µg/L. The Charlotte belt had 4.2% (n = 17) sampled wells with arsenic concentration ≥ 5 µg/L, respectively.

Table 2 Samples with arsenic concentration (≥ 5 µg/L) for geologic belts in Gaston County, North Carolina

Geologic belt	**Total (N)**	***(n)*** (≥ 5 µg/L)	***%*** (≥ 5 µg/L)
Charlotte	409	17	4.2
Inner Piedmont	154	34	22.1
Kings Mountain	427	27	6.3

We also examined the number and percent of samples with ≥ 5 µg/L arsenic concentration by different well depths (Table 3). We found that 5.7% (n = 7) of the sampled wells with depth ≤ 150 feet contained arsenic concentrations ≥ 5 µg/L. Also, 6.9% (n = 40) of the sampled wells with depth from 151 to 300 feet contained arsenic concentrations ≥ 5 µg/L. Sampled wells with depth ≥ 301 feet had 10.6% (n = 32) with arsenic concentration ≥ 5 µg/L.

Table 3 Samples with arsenic concentration (≥ 5 µg/L) for different well depths in Gaston County, North Carolina

Depth (feet)	**Total (N)**	***(n)*** **(≥ 5 µg/L)**	***%*** **(≥ 5 µg/L)**

≤ 150	105	6	5.7
151 to 300	582	40	6.9
≥ 301	303	32	10.6

4.2. **Model Results for Arsenic Concentration ≥ 5 µg/L**

The results of the autologistic regression model adjusted for confounding factors suggests that the €Zms - Mica schist and pH are associated with the presence of arsenic (≥ 5 µg/L) in well water (Table 4). The presence of arsenic ≥ 5 µg/L is significantly associated with private wells located in €Zms - Mica schist formation, (OR = 2.99, with 95% confidence interval: (1.37 - 6.52). When adjusted for potential confounding variables, the odds of arsenic > 5 µg/L in wells on €Zms - Mica schist was 2.99 times that of other wells. We found that one unit increase in pH in well water, the log odds of arsenic concentrations ≥ 5 µg/L increased by 0.75, when adjusted for other confounding factors. An OR= 2.11 with 95% CI: (1.31 – 3.38), indicated that arsenic concentration significantly increased with pH levels. The positive autocovariate coefficient (C= 2.80) indicates an inherent spatial effect in the model residuals. The spatial effect in the model increase or decrease by a factor of 2.80 in the regression. This residual spatial autocorrelation term (autocovariate) in the spatial autologistic regression reduced the spatial bias in the residual errors to produce robust estimates.

Table 4 Results of significant ($p < 0.05$) variables in the spatial autologistic regression model. Positive coefficient suggests an increased probability of arsenic ≥ 5 µg/L

Variable	Coefficient (β)	Odds Ratio (OR)	95% CI of OR
€Zms - Mica schist	1.09	2.99	1.37 - 6.52
pH	0.75	2.11	1.31 - 3.38
Autocovariate constant (C)	2.80	16.46	4.58 – 59.15

The model correctly classified 90.1% of the arsenic concentrations ≥ 5 µg/L (sensitivity) and 93.1% of the arsenic concentration < 5 µg/L (specificity). Overall, our model classification accuracy was 93.0%. The chi-square goodness of fit was significant ($p < 0.05$), which indicates that the model was better than a null model. The model AUC was 0.8, which indicates the model had classification capability with accuracy 80% of the time in predicting the presence of arsenic concentrations ≥ 5 µg/L across Gaston County, North Carolina.

4.3. **Spatial Autocorrelation (Autocovariate) of Arsenic Concentration ≥ 5 µg/L**

The spatial distribution of the autocovariate variable represents the residual spatial autocorrelation in the autologistic regression model. The values indicate the strength of the correlation between with arsenic concentrations ≥ 5 µg/L as a function of the distance separating the samples. These values range from -1 to 1, with 1 indicating areas with strong positive autocorrelation (spatial clustering), -1 indicating areas with strong negative autocorrelation, and 0 indicating a random spatial pattern with no spatial autocorrelation. As shown in Fig 4, areas with negative values can be observed in the central part of the county (dispersion of arsenic concentrations ≥ 5 µg/L), and a large proportion of the county with zero values (random spatial pattern). We observed areas with positive spatial autocorrelation ≥ 0.41 in the northwest, northeast, and southeast areas in the county indicating samples with arsenic concentrations ≥ 5 µg/L are near each other. Having many samples with arsenic concentrations ≥ 5 µg/L near each other in the northwest, northeast, and southeast areas in the county may suggest a possible common contamination source is within the area. These areas may have a poor groundwater quality compared to other parts of the county with spatial autocorrelation <0.41. The areas with spatial autocorrelation ≥ 0.41 are consistent with the pattern in Fig 1 showing locations with arsenic concentrations ≥ 5 μg/L particularly in the northwest of the county.

<Fig. 4 around here>

Fig 4 Distribution of the spatial autocorrelation (autocovariate variable)

4.4. **Spatial Probability of Arsenic Concentration ≥ 5 μg/L**

Using the model, we generated a kriging map of the probability of arsenic concentrations ≥ 5 μg/L (Fig 5). A probability higher than 0.5 indicated that well water was predicted to have arsenic concentration ≥ 5 μg/L, considering the combined effects of geology, pH, and well depth. Although the map shows that most places in the county have a low likelihood of arsenic concentration ≥ 5 μg/L, we can observe a high probability (> 0.5) in the northwest section of the county (Fig 5). This area covers about 8.4 $km^{2,}$ and our model predicts that wells contained within the area are highly susceptible to arsenic concentration ≥ 5 μg/L.

<Fig. 5 around here>

Fig 5 Spatial distribution of the probability of arsenic concentration ≥ 5 μg/L in private wells

5. Discussion

Our results suggest the presence of arsenic ≥ 5 μg/L concentration in well water is related to the geologic formation and pH. We found 26.5% of the sampled wells in the Mica schist (€Zms) formation contained arsenic concentrations ≥ 5 μg/L. This high percentage of samples with arsenic concentration ≥ 5 μg/L supports our results of the spatial autologistic regression model; private wells located in Mica schist (€Zms) were predicted to have a threefold likelihood of having arsenic concentrations ≥ 5 μg/L after controlling for other confounding factors. Mica schist (€Zms) formation consists of metamorphic rocks including quartz schist, micaceous quartzite, phyllite, and calc-silicate rock (Goldsmith et al. 1988; North Carolina Department of Environmental Quality 2020). Previous studies have identified high arsenic levels in these rocks with similar assemblages of silicate rock-forming minerals (Smedley and Kinniburgh 2002; Garelick et al. 2009). The Mica schist (€Zms) formation is also part of the Inner Piedmont belt of North Carolina, a region that has been found to contain elevated arsenic concentrations (≥ 10 μg/L) in groundwater supplies due to geologic sources (Pippin 2005; Reid et al. 2005; Harden et al. 2009; Chapman et al. 2013). Our study corroborates these findings.

The 8.4 km^2 area with a probability ≥ 0.5 for the presence of arsenic concentration ≥ 5 μg/L (Fig 5), coincides with the Mica schist (€Zms) formation. Further, we observed a positive spatial autocorrelation ≥ 0.41 in the northwest (Fig 4) to support evidence of a possible common contamination source related to the geology in the area.From the GIS permit database, we found that there were 75 private wells within that area, and 12 were sampled during this study period. Out of the 12 sampled private wells in the area, 75% (n=9) contained arsenic concentration ≥ 5 μg/L. The average arsenic concentration for the 9 sampled private wells was 16 μg/L. Given that lifetime exposure to even lower levels of arsenic concentration can be detrimental to human health (Mahram et al. 2013; Bräuner et al. 2014; James et al. 2015; Bloom et al. 2016; Almberg et al. 2017), well users in this area should be encouraged to monitor well water quality. Future epidemiological studies are needed to determine whether residents have arsenic-related health outcomes, including cancer.

We found evidence that sampled wells with arsenic concentration ≥ 5 μg/L in the water had an average pH of 7.3, which may indicate alkaline conditions that could increase arsenic mobilization in well water. Also, our model results indicated a positive association between pH and increased probability of arsenic ≥ 5 μg/L. These findings suggest that arsenic concentration ≥ 5 μg/L occur on higher pH (≥ 7.3). The potential for arsenic mobilization to occur as a result of ion exchange-related increases in pH could be due to interactions between geologic minerals and aquifer waters (Ayotte et al. 2003; Ayotte et al. 2006). The pH values greater than 7.2 have closely been related to high arsenic concentrations in groundwater aquifers in the Piedmont of North Carolina (Chapman et al. 2013). Our findings corroborate these studies.

Our model results indicate no statistically significant relationship between the presence of arsenic and well depth after adjusting for other confounding factors. Similarly, we found no significant relationship between well depth and probability of arsenic concentration ≥ 5 μg/L when using data for the 509 samples with known well depth. Previous studies have found an association between deeper wells and elevated arsenic levels (Sun 2004; Focazio et al. 2006; Kim et al. 2011; Chapman et al. 2013). Yet, our model results did not corroborate findings from these studies. We found 10.6% of sampled wells with depth ≥ 301 feet had arsenic concentration ≥ 5 μg/L. Subsequently, 7 out of 12 sampled wells in the northwestern area with an estimated 50% chance of having arsenic concentration ≥ 5 μg/L had well depth ≥ 301 feet. Given the small sample size in the most affected area, we recommend that future studies obtain more samples to determine whether there could be a relationship between arsenic concentration ≥ 5 μg/L and deeper wells.

We used publicly available data in the analysis. Thus, our approach can be applied to other areas where geologic data is available and with existing data on private wells' water quality. Also, we used a spatial autologistic regression model rather than the commonly used ordinary logistic regression model because our dependent and predictor variables were inherently spatial. The spatial autologistic regression model was used because it adjusts for spatial autocorrelation in prediction residuals due to spatial effects, which is not rectified in the non-spatial ordinary logistic regression (Tsuyuki 2008; Bo et al. 2014). A limitation of our study is that we imputed well depth information for 481 sampled wells from an IDW interpolated surface of all wells in the county. Interpolated values may not reflect the actual well depth, but we selected this approach because excluding these samples would have reduced our sample size by 49 percent. This would have affected the model statistical power and reduced our ability to find spatial patterns of the probability of arsenic concentration ≥ 5 μg/L. If we ignored the wells with interpolated depth, we would have in fact removed 31 samples with arsenic concentration ≥ 5 μg/L. Also, no significant relationship was found between the probability of arsenic concentration ≥ 5 μg/L and any of the geologic rocks from using the 509 samples with known well depth. Another weakness of our model is that we are not able to validate the results through field investigation and the number of samples that had arsenic concentration ≥ 5 μg/L was only 78 to split the data into training and testing during the model development. We recommend more sampling of arsenic data in the future could be useful in validating our results.

Further, our model could be improved by the addition of other variables, including distance to potential arsenic sources (e.g., coal ash, landfills), geochemical, and hydrogeological conditions. Also, a detailed geologic map such as that produced by Goldsmith et al. (1988), could allow for incorporating finer geologic information and improve the model. However, this map could not be used in this study because it was unavailable in a GIS usable format.

6. Conclusions

Out of 990 sampled wells, 78 contained arsenic concentration ≥ 5 µg/L, and the highest reported level was 81 µg/L. The pH of well water ranged between 5.1 to 9.7, and private wells with arsenic concentration ≥ 5 µg/L had an average pH of 7.3. The pH value of well water was positively associated with an increased probability of arsenic concentration ≥ 5 µg/L after controlling for confounding factors. Furthermore, the presence of arsenic ≥ 5 µg/L in well water was primarily related to private wells located on Mica schist (ЄZms) formation after controlling for other confounding factors.

The model results can be used to explain questions related to "why," "where," and "what" factors are influencing arsenic occurrence at and above detectable levels. For example, the model results were utilized to investigate "where are the risk areas of arsenic at or above detectable levels?" To answer this question, kriging was used to estimate probabilities of arsenic at and above detectable levels at unsampled locations across Gaston County. From the kriging map, we identified an area in the northwestern part of Gaston County has a 50% chance of having arsenic at and above the detection limit. Further, we found a positive spatial autocorrelation ≥ 0.41 suggesting a spatial clustering of samples with arsenic concentration ≥ 5 µg/L in the northwest, northeast and southeast parts of the county which may suggest a possible common contamination source within these areas.

Our analysis further reveals that, the northwest area with spatial clustering arsenic concentration ≥ 5 µg/L and with a 50% chance of reporting elevated levels of arsenic coincided with the Mica schist (€Zms) formation. Our maps offer two relevant practical use cases - 1) private wells in the "hot spot" area can be targeted for interventions, and 2) the map can be shared with the community so well owners can take action to reduce their risk of drinking unsafe water. The model results improve our ability to predict the presence of arsenic because the area we identified as a hotspot coincide with the Mica schist and 9 out of the 12 samples in the area were at and above 5 µg/L. The model results provide evidence to warrant testing of wells for arsenic across Gaston County.

References

Almberg KS, Turyk ME, Jones RM, Rankin K, Freels S, Graber JM, Stayner LT (2017) Arsenic in drinking water and adverse birth outcomes in Ohio. Environmental Research 157:52-59. doi:10.1016/j.envres.2017.05.010

Antunes IM, Albuquerque MT (2013) Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). Science of the Total Environment 442:545-552. doi:10.1016/j.scitotenv.2012.10.010

Ayotte JD, Medalie L, Qi SL, Backer LC, Nolan BT (2017) Estimating the high-arsenic domestic-well population in the conterminous United States. Environmental science & technology 51(21):12443-12454. doi:10.1021/acs.est.7b02881

Ayotte JD, Montgomery DL, Flanagan SM, Robinson KW (2003) Arsenic in groundwater in eastern New England: occurrence, controls, and human health implications. Environmental science & technology 37(10):2075-2083. doi:10.1021/es026211g

Ayotte JD, Nolan BT, Nuckols JR, Cantor KP, Robinson GR, Baris D, Hayes L, Karagas M, Bress W et al. (2006) Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment. Environmental science & technology 40(11):3578-3585. doi:10.1021/es051972f

Benbrahim-Tallaa L, Waalkes MP (2008) Inorganic arsenic and human prostate cancer. Environmental health perspectives 116(2):158-164. doi:10.1289/ehp.10423

Bloom MS, Neamtiu IA, Surdu S, Pop C, Anastasiu D, Appleton AA, Fitzgerald EF, Gurzau ES (2016) Low level arsenic contaminated water consumption and birth outcomes in Romania—An exploratory study. Reproductive Toxicology 59:8-16. doi:10.1016/j.reprotox.2015.10.012

Bo YC, Song C, Wang JF, Li XW (2014) Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in Mainland China. BMC Public Health 14(1):358. doi:10.1186/1471-2458-14-358

Bräuner EV, Nordsborg RB, Andersen ZJ, Tjønneland A, Loft S, Raaschou-Nielsen O (2014) Long-term exposure to low-level arsenic in drinking water and diabetes incidence: a prospective study of the diet, cancer and health cohort. Environmental health perspectives 122(10):1059-1065. doi:10.1289/ehp.1408198

Bretzler A, Lalanne F, Nikiema J, Podgorski J, Pfenninger N, Berg M, Schirmer M (2017) Groundwater arsenic contamination in Burkina Faso, West Africa: Predicting and verifying regions at risk. Science of the Total Environment 584:958-970. doi:10.1016/j.scitotenv.2017.01.147

Carter JM, Driscoll DG, Williamson JE, Lindquist VA (2002) Atlas of water resources in the Black Hills area, South Dakota. https://pubs.usgs.gov/ha/ha747/pdf/definition.pdf. Accessed: February 9, 2020

Centers for Disease Control and Prevention (CDC) (2019) Safer Well Water through Stronger Public Health Programs. https://blogs.cdc.gov/yourhealthyourenvironment/2019/11/25/safer-well-water-through-stronger-public-health-programs/. Accessed: December 23, 2019

Chapman MJ, Cravotta C, Szabo Z, Lindsey BD (2013) Naturally Occurring Contaminants in the Piedmont and Blue Ridge Cystalline-rock Aquifers and Piedmont Early Mesozoic Basin Siliciclastic-rock Aquifers, Eastern United States, 1994-2008. US Department of the Interior, US Geological Survey.

Daniel CC, Dahlen PR (2002) Preliminary hydrogeologic assessment and study plan for a regional ground-water resource investigation of the Blue Ridge and Piedmont Provinces of North Carolina. vol 2. vol 4105. US Department of the Interior, US Geological Survey.

Dauphiné DC, Smith AH, Yuan Y, Balmes JR, Bates MN, Steinmaus C (2013) Case-control study of arsenic in drinking water and lung cancer in California and Nevada. International journal of environmental research and public health 10(8):3310-3324. doi:10.3390/ijerph10083310

Dormann CF (2007) Assessing the validity of autologistic regression. ecological modelling 207(2-4):234-242.

Dummer T, Yu Z, Nauta L, Murimboh J, Parker L (2015) Geostatistical modelling of arsenic in drinking water wells and related toenail arsenic concentrations across Nova Scotia, Canada. Science of the Total Environment 505:1248-1258. doi:10.1016/j.scitotenv.2014.02.055

Ethan D, Xiao-Ming L (2018) Examining the Geologic Link of Arsenic Contamination in Groundwater in Orange County, North Carolina. Frontiers in Earth Science 6. doi:10.3389/feart.2018.00111

Evans JS, Ram K (2020) Spatial Analysis and Modelling Utilities. https://github.com/jeffreyevans/spatialEco. Accessed: April 2, 2020 2020

Focazio MJ, Tipton D, Dunkle Shapiro S, Geiger LH (2006) The chemical quality of self-supplied domestic well water in the United States. Groundwater Monitoring & Remediation 26(3):92-104. doi:10.1111/j.1745-6592.2006.00089.x

Fu R, Thurman AL, Chu T, Steen-Adams MM, Zhu J (2013) On estimation and selection of autologistic regression models via penalized pseudolikelihood. Journal of agricultural, biological, and environmental statistics 18(3):429-449. doi:10.1007/s13253-013-0144-z

Garelick H, Jones H, Dybowska A, Valsami-Jones E (2009) Arsenic pollution sources. In: Reviews of Environmental Contamination Volume 197. Springer, New York, NY, pp 17-60.

Gaus I, Kinniburgh D, Talbot J, Webster R (2003) Geostatistical analysis of arsenic concentration in groundwater in Bangladesh using disjunctive kriging. Environmental Geology 44(8):939-948. doi:10.1007/s00254-003-0837-7

Goldsmith R, Milton DJ, Horton Jr JW (1988) Geologic map of the Charlotte 1 degrees by 2 degrees Quadrangle, North Carolina and South Carolina. https://pubs.usgs.gov/imap/1251e/report.pdf. Accessed: February 20, 2020

Goovaerts P (2009) AUTO-IK: a 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. Computers & geosciences 35(6):1255-1270. doi:10.1016/j.cageo.2008.08.014

Goovaerts P, AvRuskin G, Meliker J, Slotnick M, Jacquez G, Nriagu J (2005) Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. Water Resources Research 41(7). doi:10.1029/2004WR003705

Griffith DA (1987) Spatial autocorrelation. Resource publications in geography.

Griffith DA (2004) A spatial filtering specification for the autologistic model. Environment and Planning A : Economy and Space 36(10):1791-1811. doi:doi.org/10.1068/a36247

Gross EL, Low DJ (2013) Arsenic concentrations, related environmental factors, and the predicted probability of elevated arsenic in groundwater in Pennsylvania. US Department of the Interior, US Geological Survey.

Gurung JK, Ishiga H, Khadka MSJEG (2005) Geological and geochemical examination of arsenic contamination in groundwater in the Holocene Terai Basin, Nepal. 49(1):98-113.

Hamel L (2009) Model assessment with ROC curves. In: Encyclopedia of Data Warehousing and Mining, Second Edition. IGI Global, pp 1316-1323.

Harden SL, Chapman MJ, Harned DA (2009) Characterization of Groundwater Quality Based on Regional Geologic Setting in the Piedmont and Blue Ridge Physiographic Provinces, North Carolina. US Geological Survey.

Hassan MM, Atkins PJ (2011) Application of geostatistics with Indicator Kriging for analyzing spatial variability of groundwater arsenic concentrations in Southwest Bangladesh. Journal of Environmental Science and Health, Part A 46(11):1185-1196. doi:10.1080/10934529.2011.598771

He X, Li P, Ji Y, Wang Y, Su Z, Elumalai V (2020a) Groundwater arsenic and fluoride and associated arsenicosis and fluorosis in China: occurrence, distribution and management. Exposure and Health. doi:10.1007/s12403-020-00347-8

He X, Li P, Wu J, Wei M, Ren X, Wang D (2020b) Poor groundwater quality and high potential health risks in the Datong Basin, northern China: research from published data. Environmental Geochemistry and Health. doi:10.1007/s10653-020-00520-7

Heck JE, Andrew AS, Onega T, Rigas JR, Jackson BP, Karagas MR, Duell EJ (2009) Lung cancer in a US population with low to moderate arsenic exposure. Environmental health perspectives 117(11):1718-1723. doi:10.1289/ehp.0900566

Hengl T (2009) A practical guide to geostatistical mapping. 2nd edn. Office for Official Publications of the European Communities, Luxembourg.

Hossain F, Sivakumar B (2006) Spatial pattern of arsenic contamination in shallow wells of Bangladesh: regional geology and nonlinear dynamics. Stochastic Environmental Research and Risk Assessment 20(1):66-76. doi:10.1007/s00477-005-0012-7

Huebner MT, Hatcher RD, Jr., Merschat AJ (2017) Confirmation of the southwest continuation of the Cat Square terrane, southern Appalachian Inner Piedmont, with implications for middle Paleozoic collisional orogenesis. American Journal of Science 317(2):95-176. doi:10.2475/02.2017.01

James KA, Byers T, Hokanson JE, Meliker JR, Zerbe GO, Marshall JA (2015) Association between lifetime exposure to inorganic arsenic in drinking water and coronary heart disease in Colorado residents. Environmental health perspectives 123(2):128-134. doi:10.1289/ehp.1307839

Karagas MR, Gossai A, Pierce B, Ahsan H (2015) Drinking water arsenic contamination, skin lesions, and malignancies: a systematic review of the global evidence. Current environmental health reports 2(1):52-68. doi:10.1007/s40572-014-0040-x

Kim D, Miranda ML, Tootoo J, Bradley P, Gelfand AE (2011) Spatial modeling for groundwater arsenic levels in North Carolina. Environmental science & technology 45(11):4824-4831. doi:10.1021/es103336s

Lee JJ, Liu CW, Jang CS, Liang CP (2008) Zonal management of multi-purpose use of water from arsenic-affected aquifers by using a multi-variable indicator kriging approach. Journal of Hydrology 359(3):260-273. doi:10.1016/j.jhydrol.2008.07.015

Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecological Informatics 6(3-4):228-241. doi:10.1016/j.ecoinf.2010.12.003

Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling & Software 53:173-189. doi:10.1016/j.envsoft.2013.12.008

Liu C, Sharma A (2019) Are you going to get a ticket or a warning for speeding? An autologistic regression analysis in Burlington, VT. Transportation research interdisciplinary perspectives 1:100001. doi:10.1016/j.trip.2019.100001

Liu J, Piegorsch WW, Grant Schissler A, Cutter SL (2018) Autologistic models for benchmark risk or vulnerability assessment of urban terrorism outcomes. Journal of the Royal Statistical Society: Series A (Statistics in Society) 181(3):803-823. doi:10.1111/rssa.12323

MacDonald Gibson J, Pieper KJ (2017) Strategies to improve private-well water quality: a North Carolina perspective. Environmental health perspectives 125(7):076001-076009. doi:10.1289/EHP890.

Mahram M, Shahsavari D, Oveisi S, Jalilolghadr S (2013) Comparison of hypertension and diabetes mellitus prevalence in areas with and without water arsenic contamination. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences 18(5):408-412.

Meliker JR, AvRuskin GA, Slotnick MJ, Goovaerts P, Schottenfeld D, Jacquez GM, Nriagu JO (2008) Validity of spatial models of arsenic concentrations in private well water. Environmental research 106(1):42-50. doi:10.1016/j.envres.2007.09.001

Miller HJ (2004) Tobler's first law and spatial analysis. Annals of the Association of American Geographers 94(2):284-289. doi:10.1111/j.1467-8306.2004.09402005.x

North Carolina Department of Environmental Quality (2020) NC Geological Survey. https://deq.nc.gov/about/divisions/energy-mineral-land-resources/north-carolina-geological-survey. Accessed: February 22, 2020 2020

North Carolina Department of Health and Human Services (2020) Minimum detection level for arsenic in NC Division of State Laboratory of Public Health Raleigh, NC.

Owusu C, Lan Y, Zheng M, Tang W, Delmelle E (2017) Geocoding fundamentals and associated challenges. In: Karimi HA, Karimi B (eds) Geospatial Data Science Techniques and Applications. CRC Press, Boca Raton, FL, pp 41-62.

Pebesma E, Graeler B, Pebesma ME (2019) Package ‘gstat’.

Pippin CG (2005) Distribution of total arsenic in groundwater in the North Carolina Piedmont, paper presented at 2005 NGWA Naturally Occurring Contaminants Conference—Arsenic, Radium, Radon, and Uranium,. Charleston, SC, p. 89–102. http://h2o.enr.state.nc.us/gwp/Arsenic_Studies.htm. Accessed: May 20, 2019

Reid JC, Pippin CG, Haven WT, Wooten R (2005) Assessing the Source for Arsenic in Groundwater, North Carolina Piedmont. Raleigh.

Sanders AP, Messier KP, Shehee M, Rudo K, Serre ML, Fry RC (2012) Arsenic in North Carolina: public health implications. Environment international 38(1):10-16. doi:10.1016/j.envint.2011.08.005

Seeley M, Goring S, Williams JW (2019) Assessing the environmental and dispersal controls on Fagus grandifolia distributions in the Great Lakes region. Journal of biogeography 46(2):405-419. doi:10.1111/jbi.13491

Smedley PL, Kinniburgh D (2002) A review of the source, behaviour and distribution of arsenic in natural waters. J Applied geochemistry 17(5):517-568. doi:10.1016/S0883-2927(02)00018-5

Steinmaus C, Yuan Y, Bates MN, Smith AH (2003) Case-control study of bladder cancer and drinking water arsenic in the western United States. American journal of epidemiology 158(12):1193-1201. doi:10.1093/aje/kwg281

Sun G (2004) Arsenic contamination and arsenicosis in China. Toxicology and applied pharmacology 198(3):268-271. doi:10.1016/j.taap.2003.10.017

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Economic geography 46(1):234-240. doi:10.2307/143141

Tobler WR (1979) Cellular geography. In: Gale S, Olsson G (eds) Philosophy in geography Theory and Decision Library (An International Series in the Philosophy and Methodology of the Social and Behavioral Sciences), vol 20. Springer, Dordrecht, pp 379-386.

Tsuyuki S (2008) GIS-based modeling of Javan Hawk-Eagle distribution using logistic and autologistic regression models. Biological Conservation 141(3):756-769.

United States Geological Survey (2020) Aquifers and Groundwater. https://www.usgs.gov/special-topic/water-science-school/science/aquifers-and-groundwater?qt-science_center_objects=0#qt-science_center_objects. Accessed: June 12, 2020

USEPA (1994) Method 200.8: Determination of Trace Elements in Waters and Wastes by Inductively Coupled Plasma-Mass Spectrometry. https://www.epa.gov/sites/production/files/2015-06/documents/epa-200.8.pdf. Accessed: February 25, 2020 2020

VanDerwerker T, Zhang L, Ling E, Benham B, Schreiber M (2018) Evaluating geologic sources of arsenic in well water in Virginia (USA). International journal of environmental research and public health 15(4):787. doi:10.3390/ijerph15040787

Waldron AJ, Bobyarchick AR, Diemer J, Eppes MC, Meentemeyer R (2007) Spatial analysis of factors affecting home radon levels around Moss Lake, NC. Geological Society of America Abstracts with Programs 39(2):28-29.

Werner CK, Bender JF, Bobyarchick AR, Diemer JA, Eppes MC, Waldron AS (2009) Investigation of the source of high radon levels in Cleveland County, North Carolina. Geological Society of America Abstracts with Programs 41(1):18.

Wu H, Huffer FRW (1997) Modelling the distribution of plant species using the autologistic regression model. Environmental and ecological Statistics 4(1):31-48. doi:10.1023/A:1018553807765

Yuan Y, Marshall G, Ferreccio C, Steinmaus C, Liaw J, Bates M, Smith AH (2010) Kidney cancer mortality: fifty-year latency patterns related to arsenic exposure. J Epidemiology:103-108. doi:10.1097/EDE.ObO.13e3181.c21.e46

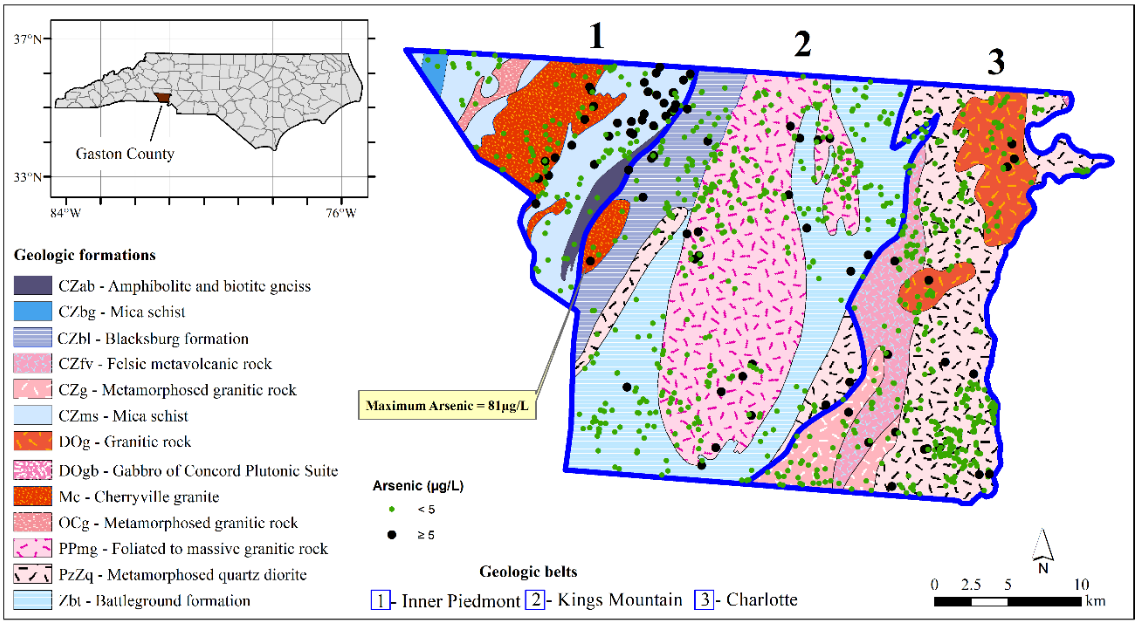

Fig 6 Spatial distribution of arsenic concentrations in well water samples in Gaston County, North Carolina. (Geologic data source: North Carolina Department of Environmental Quality, 2020)

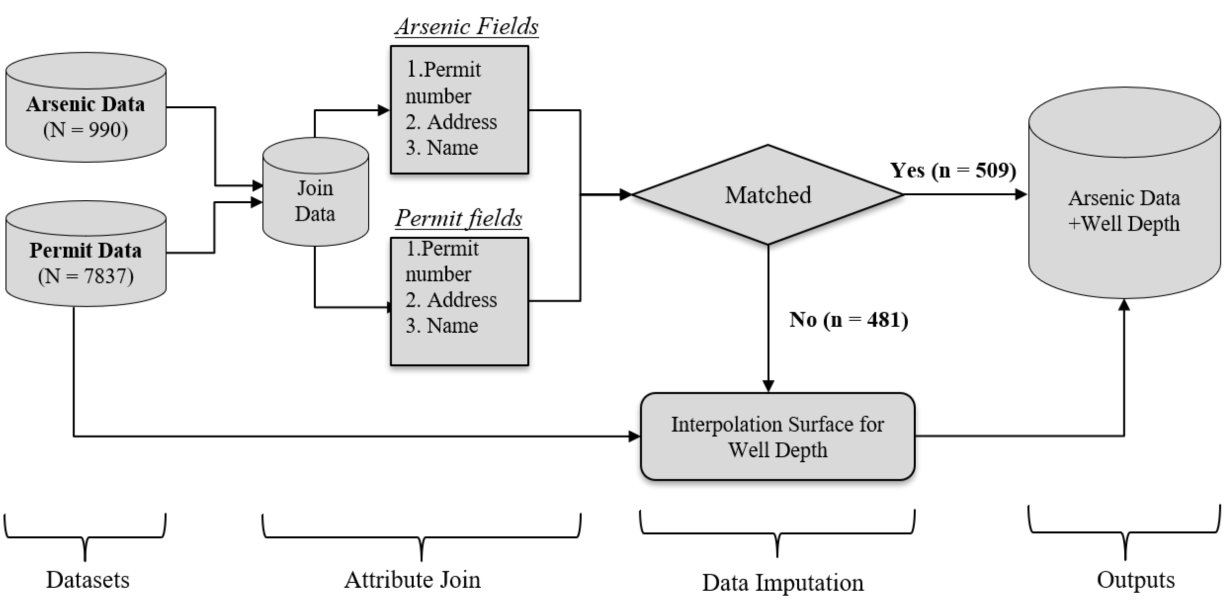

Fig 7 Workflow to estimate well depth for private wells

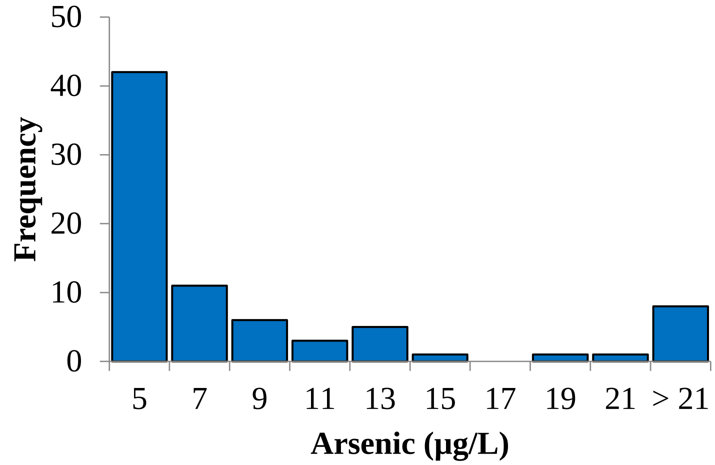

Fig 8 Distribution of arsenic for the 78 samples at and above detectable levels (5 µg/L) (samples marked as '< 5 µg/L' in the data had a frequency of 912- not included in the histogram)

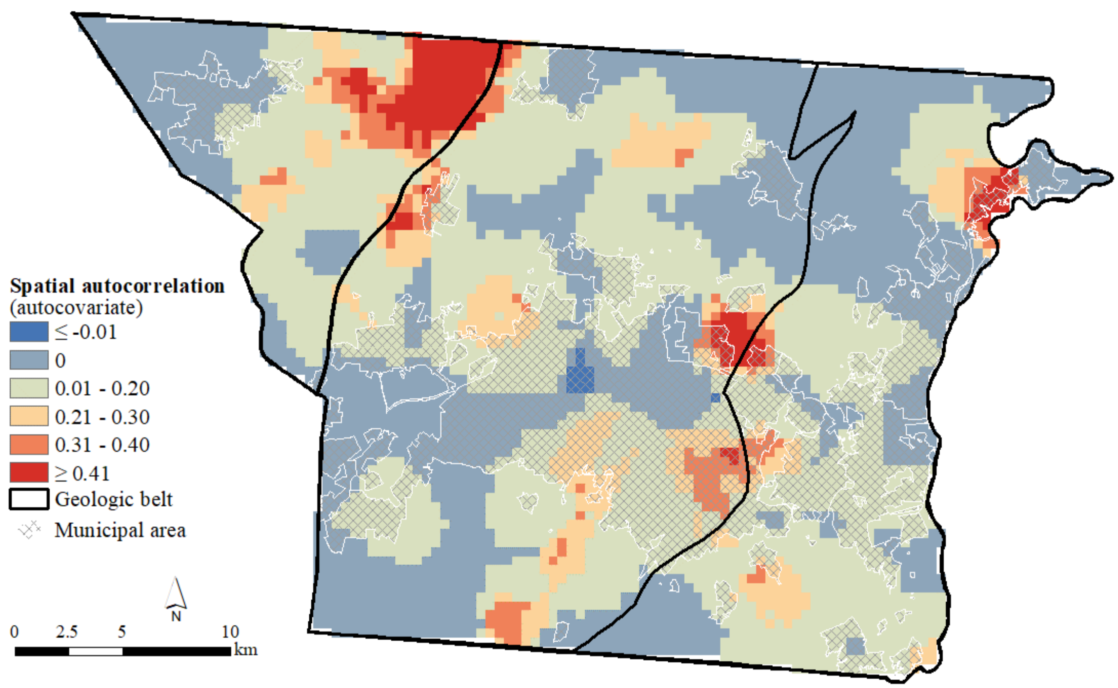

Fig 9 Distribution of the spatial autocorrelation (autocovariate variable)

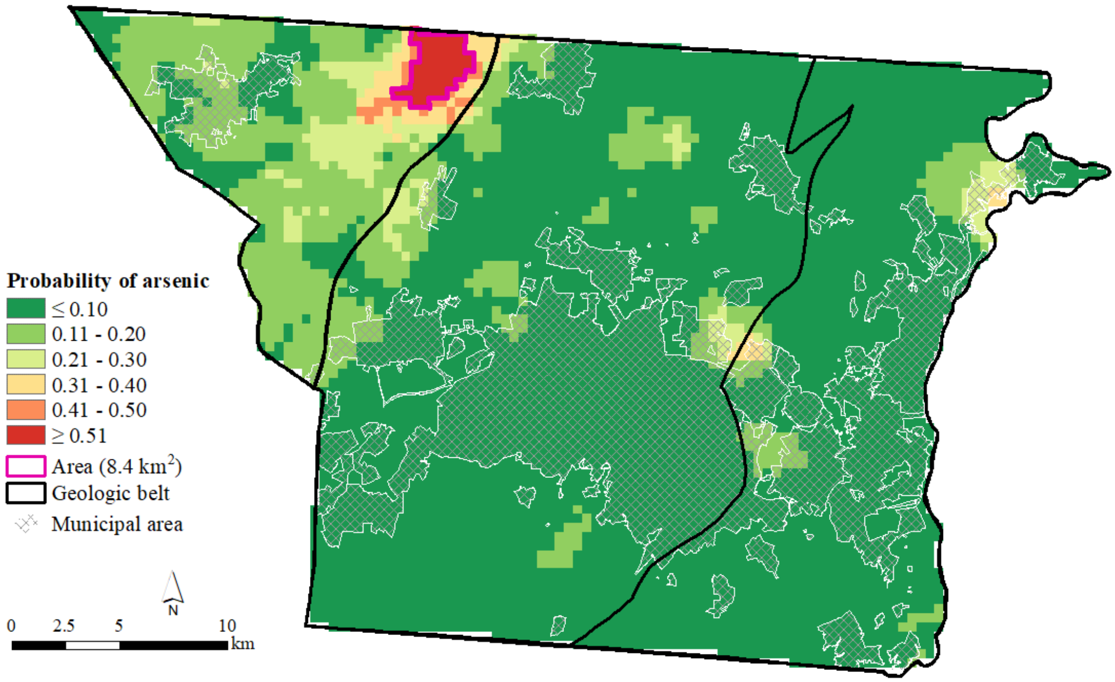

Fig 10 Spatial distribution of the probability of arsenic concentration ≥ 5 μg/L in private wells