

Published in final edited form as:

Environ Sci Technol. 2012 February 21; 46(4): 2080–2087. doi:10.1021/es203793x.

Can arsenic occurrence rates in bedrock aquifers be predicted?

Qiang Yang^{a,b}, Hun Bok Jung^a, Robert G. Marvinney^c, Charles W. Culbertson^d, and Yan Zheng^{a,b,*}

^aCity University of New York, Queens College and Graduate Center, 6530 Kissena Blvd., Flushing, NY 11367

^bColumbia University, Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10964

^cMaine Geological Survey, 22 State House Station, Augusta, ME 04333

^dUSGS Maine Water Science Center, 196 Whitten Road, Augusta, ME 04330

Abstract

A high percentage (31%) of groundwater samples from bedrock aquifers in the greater Augusta area, Maine was found to contain greater than 10 $\mu\text{g L}^{-1}$ of arsenic. Elevated arsenic concentrations are associated with bedrock geology, and more frequently observed in samples with high pH, low dissolved oxygen, and low nitrate. These associations were quantitatively compared by statistical analysis. Stepwise logistic regression models using bedrock geology and/or water chemistry parameters are developed and tested with external data sets to explore the feasibility of predicting groundwater arsenic occurrence rates (the percentages of arsenic concentrations higher than 10 $\mu\text{g L}^{-1}$) in bedrock aquifers. Despite the under-prediction of high arsenic occurrence rates, models including groundwater geochemistry parameters predict arsenic occurrence rates better than those with bedrock geology only. Such simple models with very few parameters can be applied to obtain a preliminary arsenic risk assessment in bedrock aquifers at local to intermediate scales at other localities with similar geology.

Introduction

Elevated arsenic in groundwater has been recognized as a worldwide public health problem. In the United States, millions of people are subjected to increased cancer risks from arsenic in drinking water from private wells ¹. In New England, the danger is exacerbated because most rural residents rely solely on private wells completed in bedrock aquifers. Arsenic concentrations in these well waters can exceed 10 $\mu\text{g L}^{-1}$, the EPA Maximum Contaminant Level (MCL). The problem is acute for New Hampshire and Maine, where approximately 25% and 43% of the population use private wells built in bedrock aquifers ², and where 12% and 18% of these wells were estimated to have > 10 $\mu\text{g L}^{-1}$ of arsenic in well water, respectively ^{3, 4}.

While the majority of documented arsenic affected aquifers around the world are sedimentary, more attention has recently been turned to elevated arsenic in bedrock aquifers,

*Corresponding author: Yan Zheng, City University of New York, Queens College and Graduate Center, 6530 Kissena Blvd., Flushing, NY 11367; Phone: 718-997-3329; Fax: 718-997-3299, yan.zheng@qc.cuny.edu; yzheng@ldeo.columbia.edu.

Supporting Information Available

To protect the privacy of the homeowners, the tabulated arsenic concentration data are not included but available to researchers upon request for usage in accordance of IRB rules. Correlations between groundwater arsenic and hydrogeochemical parameters in the greater Augusta area, groundwater arsenic, pH, dissolved oxygen, nitrate and sulfate distributions in 2006 sampling, distribution maps of arsenic in 2007 sampling, temporal variation of groundwater arsenic, results of logistic regression models, and a summary of documented bedrock aquifers with natural arsenic problems are available free of charge via the Internet at <http://pubs.acs.org>.

such as New England^{5–20}, Pennsylvania²¹, North Carolina²², Michigan^{23, 24} of U.S., British Columbia of Canada²⁵, Zimapan Valley of Mexico²⁶, Chattisgarh of central India²⁷, Northern Bavaria of Germany²⁸, Ghana^{29–31} and Burkina Faso³². Although each region has distinct bedrock formations, there are similarities in their tectonic setting and geologic history (Table S2). The bedrock units started either as marine basin deposits of fine sediments or volcanic sequences with high arsenic concentrations, followed by regional metamorphism resulting from episodes of tectonic collisions and uplift. These episodes were often accompanied by contact metamorphism from granite-diorite intrusions resulting in the formation of quartz-sulfide veins in host rocks. They were sometimes overlain by younger glacial drift with various spatial coverage and thickness. Studies have related the spatial distribution of high arsenic occurrence rates (the percentages of arsenic concentrations higher than $10 \mu\text{g L}^{-1}$) with meta-sedimentary or meta-volcanic formations^{12–16, 19, 21, 23, 30, 32}, pegmatite zones contacting the granitic intrusions^{6, 11, 22, 27}, as well as overlying glacial deposits²⁴. The major ion composition of groundwater varies among the aquifers, but high arsenic groundwater tends to associate with high pH and bicarbonate, as well as low dissolved oxygen and Eh in bedrock aquifers^{9, 10, 19, 21, 23, 26, 31, 32}. However, research is still needed to ascertain the extent to which the associations between these parameters and arsenic concentrations exist among various types of bedrock.

In New England, significant progress has been made to understand the spatial distribution of groundwater arsenic at local^{6, 10–13, 15, 16, 19} to regional scales^{7, 9, 14, 17}, as well as the source and mobilization mechanisms of releasing arsenic^{6, 9–11, 13, 15, 17, 19}. At the very local scale of 10^0 km, arsenic distribution was largely affected by fracture characteristics, hydrogeology and water chemistry¹⁰, although no quantitative relationships with any of these factors could be established in part due to the complexity of water flow and chemical transport in bedrock aquifers³³. Studies at the intermediate scales of 10^0 – 10^2 km^{6, 11, 12, 15–17} defined the spatial pattern of arsenic distribution, and its associations with bedrock geology and major water chemistry parameters, but no quantitative estimation or prediction model based on these associations have been made. Using a logistic regression model on the regional scale of $>10^2$ km, groundwater arsenic occurrence has been shown to depend on factors such as bedrock geology, arsenic levels in stream sediments, proximity to granitic intrusions, and landscape variables related to groundwater residence time, and these parameters have been used to predict arsenic occurrence⁷. This regional model however did not include groundwater chemistry. Risk assessment models based on association of arsenic occurrence and a set of by-and-large similar parameters including climate, sedimentary depositional environments, soil and elevation have been developed for sedimentary aquifers on global and regional scales^{34, 35}. To verify and to refine model prediction of arsenic occurrence rates at sub-regional scales of 10^0 – 10^2 km requires further quantitative description of the association between arsenic and groundwater chemistry.

Our recent study in the greater Augusta area, Maine has highlighted the health risk from groundwater arsenic exposure¹³. Arsenic occurrence rates are different among the major bedrock formations and are the highest in the meta-sedimentary sequences. Questions remain regarding whether these differences are also regulated by groundwater chemistry and other environmental factors, and if so, to what extent. A stepwise logistic regression model is used to examine the correlations between groundwater arsenic and hydrogeochemical parameters including groundwater pH, dissolved oxygen, major anions and cations. Multivariate logistic regression models are developed based on the most statistically significant parameters and tested with a separate data set also from the greater Augusta area and two external data sets from other areas in Maine. Our approach of establishing a prediction model based on bedrock type and a few water chemistry parameters can be used to improve risk assessment of groundwater arsenic in bedrock aquifers around the world that

share similar tectonic history. It is also applicable to other New England states for targeting risk reduction in communities.

Materials and Methods

Study Area

In 13 towns of Kennebec County, central Maine, 13,100 (or 54%) of the total 24,500 households are estimated to be self-supplied by private wells⁴. Five major geologic formations that compose the study area (Figure S1) displayed high to low groundwater arsenic occurrence rates from west to east in the order of Ss (Sangerville Formation, 42%), Sw (Waterville Formation, 39%), SOv (Vassalboro Formation, 24%), D (Granite Intrusions, 15%) and OZc (Cushing Formation, 9%)¹³. Sedimentary facies suggested that the sediment of three meta-sedimentary formations (Ss, Sw, SOv) was deposited in anoxic marine basins during the Silurian and may contain abundant sulfide minerals³⁶.

Groundwater

Unfiltered groundwater samples were collected throughout the study area in 2006 (n=789), and then from four clusters in 2007 (n=331) (Figure S2), with 36 wells sampled in 2007 being re-visits to assess temporal variability. Sampling protocol and analytical methods for arsenic have been reported¹³, noting that the instrumental method also collected data for 32 other elements simultaneously³⁷. Groundwater pH was recorded after the temperature, electrical conductivity and pH readings stabilized in a flow-through cell, usually after 20 min of pumping. Dissolved oxygen was measured in the field using CHEMets Kit. Anions including fluoride, bromide, chloride, nitrate, nitrite, phosphate and sulfate were measured by ion-exchange chromatography (IC) following U.S. EPA method 300.0. A subset of groundwater samples from 2006 (n=105) evenly distributed in the study area, and all samples from 2007 were analyzed for alkalinity using Gran titration. The estimation of charge balance between the sums of major cations and anions was conducted to ensure the quality of measurements (Supporting Information).

pearman's Correlation

Non-parametric Spearman's rank correlation coefficient (ρ) between groundwater arsenic and other hydrogeochemical parameters was calculated because the data did not follow normal distributions. The test significance (p value) was determined using two-tailed student's *t* test with n-2 degrees of freedom.

Logistic Regression Model

A stepwise logistic regression method was conducted to examine the association between each parameter and groundwater arsenic occurrence using the 2006 data set. The groundwater arsenic concentrations were converted into binary data using a threshold of 10 $\mu\text{g L}^{-1}$ and then used as the response variable. All the available hydrogeochemical parameters were used as explanatory variables, including well depth, yield, elevation, water temperature, pH, dissolved oxygen level, major cation and anion concentrations, alkalinity, [Fe], [Mn], and bedrock unit (categorical variable). Each was individually tested to determine whether it could be used as an indicator of groundwater arsenic exceeding 10 $\mu\text{g L}^{-1}$ by a univariate logistic regression, and those with statistical significance (p value <0.05) were included in the multivariate logistic regression model. The Hosmer and Lemeshow goodness-of-fit test was used to verify the model. Only significant parameters from the multivariate model were retained in the three prediction models (using bedrock geology only, water geochemistry parameters only, and all significant parameters, respectively) to estimate arsenic occurrence rates. The odds ratio of each variable is calculated as $\exp(\beta_i)$,

where β_i stands for the “estimate” value of each variable in the regression model results (Supporting Information). The prediction models were verified by the Hosmer and Lemeshow goodness-of-fit test. Deviance residuals were used to detect ill-fitting predictions. Model discrimination was estimated by the c statistic. For binary outcomes, c varies between 0.5 and 1.0 for sensible models; the higher the better. The best prediction model was then applied in other areas. All the regression models and tests were performed in the R program.

A predicted probability of 0.5 is used as the cutoff value, so the prediction for an individual sample is considered as less than $10 \mu\text{g L}^{-1}$ if the predicted probability is < 0.5 , and exceeding the MCL if the predicted probability is > 0.5 . The overall correct classification rate, model specificity (the percentage of correct classification of samples with $[\text{As}] < 10 \mu\text{g L}^{-1}$), and sensitivity (the percentage of correct classification of samples with $[\text{As}] > 10 \mu\text{g L}^{-1}$) were calculated. The 2007 dataset from four clusters in the study area, as well as two external data sets from Buxton and Hollis, southern Maine¹² and from 5 drainage basins throughout the state from the Maine Geological Survey’s Ambient Bedrock Water Quality program, were used to test the prediction model.

Results and Discussions

Correlations

The correlations between groundwater arsenic and hydrogeochemical parameters are established for the entire study area by statistically significant Spearman’s ρ value ($p < 0.05$), with higher values suggesting stronger correlations (Table S1). There is a moderate positive correlation with pH ($\rho = 0.54$) and a weak negative correlation with dissolved oxygen ($\rho = -0.35$) (Figure 1 and Table S1). Previous studies in New England^{9, 10}, Pennsylvania²¹ and Ghana³¹ have also shown that elevated arsenic concentrations were more frequent in samples with higher pH. That high groundwater arsenic concentrations are associated with low dissolved oxygen is a new result for New England, but this has been noted in bedrock aquifers of Burkina Faso³², Mexico²⁶, Pennsylvania²¹ and Michigan²³. A weak negative correlation is observed for nitrate ($\rho = -0.31$) (Figure 1 and Table S1), consistent with the correlations found for other aquifers from New England⁸ and for surface water³⁸. Weak positive correlations are found for fluoride ($\rho = 0.39$) and molybdenum ($\rho = 0.38$), consistent with the Mo-As association observed in Burkina Faso³². There is a weak negative correlation between arsenic and chloride ($\rho = -0.23$) (Figure 1 and Table S1), also observed in Burkina Faso³² and Pennsylvania²¹. An even weaker but positive correlation is observed with the molar sum of cations (with sodium subtracted in equivalent molar amount to chloride which is likely attributable to road salt), a parameter more likely to reflect the extent of water-rock interaction. Instead of using scatter plots of arsenic vs. hydrogeochemical parameters by previous studies to show correlations, the Spearman’s ρ value presents a better and more quantitative description of associations because arsenic and some of these hydrogeochemical parameters do not follow normal distributions.

Within each unit of meta-sedimentary, granitic and volcanic rocks, the above correlations exist with similar strength and significance (Table S1). Moderate positive correlations with pH ($\rho = 0.44 \sim 0.65$), weak to moderate negative correlations with dissolved oxygen ($\rho = -0.44 \sim -0.29$) and nitrate ($\rho = -0.36 \sim -0.20$), weak to moderate positive correlations with fluoride ($\rho = 0.36 \sim 0.51$) and molybdenum ($\rho = 0.29 \sim 0.48$), weak negative correlations with chloride ($\rho = -0.21 \sim -0.26$), and very weak to moderate correlations with the sum of cations (minus chloride) ($\rho = 0.14 \sim 0.44$) are observed.

Similar correlations between arsenic and aforementioned parameters are also found in four clusters of wells sampled in high-density in 2007 representative of granite (Chelsea), meta-sedimentary rocks (Sidney), and their combinations (Litchfield and Manchester) (Table S1).

Spatial Distribution of Parameters Correlated to Arsenic

The spatial distribution of groundwater arsenic displays a decreasing trend from west (meta-sedimentary rocks) to east (granites and volcanic rocks). The key parameters (groundwater pH, dissolved oxygen, nitrate and sulfate) that are correlated with arsenic are also different in the western and eastern parts of the study area (Figure S1).

The median pH values are statistically significantly higher in meta-sedimentary or meta-volcanic rocks, such as Ss (pH 8.0), Sw (pH 7.9), SOv (pH 7.3) and OZc (pH 7.3), than in granitic rocks, D (pH 6.9) (Table 1). The median pH value (6.8) in groundwater from the Chelsea cluster within the granite intrusions (D) according to the most updated bedrock geology map³⁹ and also showing the lowest arsenic occurrence, is statistically significantly lower than those in other three clusters (8.2, 7.8, 8.3, respectively) that are mostly in meta-sedimentary rocks (Table 1).

The median dissolved oxygen concentrations in meta-sedimentary or volcanic rocks ($1.0\text{--}2.0\text{ mg L}^{-1}$) are much less than that in granites (4.0 mg L^{-1}) (Table 1). The median dissolved oxygen level (2.0 mg L^{-1}) of groundwater in the Chelsea cluster (granitic rocks) is also higher than those in meta-sedimentary rocks (0.6, 1.0, 1.0 mg L^{-1} , respectively) (Table 1). There is a slightly higher percentage of low dissolved oxygen groundwater ($<1\text{ mg L}^{-1}$) in the western part than in the eastern part (Figure S1).

Comparison of hydrogeochemical parameters within different bedrock geologic units (Table 1) suggests that the meta-sedimentary rocks have a higher arsenic source and tend to form favorable water chemistry for releasing arsenic, i.e. higher pH and more reducing conditions. Meta-volcanic rocks could develop favorable groundwater environment for arsenic mobilization, but are limited by low arsenic source in the bedrock. Granitic rocks may or may not contain high arsenic source, but do not tend to form favorable chemical conditions to release arsenic into groundwater, probably due to the lack of carbonate minerals. This finding is consistent with previous findings (Table S2) that high arsenic occurrence rates were mostly found in meta-sedimentary formations^{6, 12–15, 21, 23, 24, 32}, but were rare in meta-volcanic formations³⁰ or in granite intrusions, granodiorite, and pegmatite^{6, 11, 17}. This correlation between bedrock geology and groundwater geochemistry does not reduce the predictive power of the logistic regression model, but may affect the calculations regarding individual predictors.

Temporal Pattern

No significant difference was found when arsenic concentrations were compared for the 36 wells sampled in 2006 and 2007 (Figure S3). In addition to arsenic, pH, dissolved oxygen, and major ions concentrations are also comparable. This suggests that groundwater chemistry in the study area is relatively stable, possibly due to low water withdrawal rates since water is used only for domestic consumption in this rural area.

Logistic Regression Model

Univariate regressions found that twelve hydrogeochemical parameters showed significant associations with groundwater arsenic occurrence in the 2006 data set from 789 wells. Six parameters (well depth, pH, dissolved oxygen, chloride, fluoride, and nitrate) besides bedrock geology type also displayed significant Spearman's ρ value of > 0.2 (Table S1 column 2), while five others (sulfate, alkalinity, Mg, Ca, and Si) did not. The multivariate

logistic regression model using 11 parameters mentioned above (alkalinity excluded) had a p-value of 0.62 by the Hosmer and Lemeshow goodness-of-fit test, much higher than the critical value of 0.05, indicating the robustness of the model (greater p-values indicate better fit because the null hypothesis in the Hosmer and Lemeshow test is that the model fits the data). Alkalinity was excluded because the data points were too few for the model to converge. Five parameters, bedrock geology, groundwater pH, dissolved oxygen, nitrate and sulfate were found to be statistically significant in this multivariate model. In the final prediction model with these 5 parameters, the *p*-value of Hosmer and Lemeshow goodness-of-fit test decreased to 0.38 (Supporting Information). The model output is that Ss, Sw, SOv and D rock units show 5.22-fold, 4.84-fold, 3.67-fold and 1.77-fold higher odds ($\exp(\beta)$) of containing elevated groundwater arsenic than rock unit OZc, respectively (Table 2), same as the order of observed arsenic occurrence rates for each rock unit¹³. Among the water chemistry parameters, pH has a higher odds ratio of 2.93, while dissolved oxygen and nitrate have lower odds ratio of 0.89 and 0.79, respectively. Sulfate shows an odds ratio (0.97) not too different from 1, suggesting a weaker influence. Therefore, the most important water chemistry factor is pH, followed by redox conditions as indicated by dissolved oxygen, nitrate and sulfate. The high value of the intercept in the model outcome indicates there are unknown parameters or processes that have impact on groundwater arsenic.

The final prediction model of groundwater arsenic occurrence rates in the greater Augusta area is as follows.

$$\frac{\text{Probability}([As] > 10 \mu\text{g L}^{-1})}{1 - \text{Probability}([As] > 10 \mu\text{g L}^{-1})} = \exp(-9.52 + 1.075\text{pH} - 0.117\text{DO} - 0.239\text{NO}_3 - 0.029\text{SO}_4 + 1.652\text{Ss} + 1.577\text{Sw} + 1.229\text{SOv} + 0.570\text{D})$$

where Ss is Sangerville formation, Sw is Waterville formation, SOv is Vassalboro formation, D is granite intrusion, pH is groundwater pH values, DO is dissolved oxygen concentration in mg L^{-1} , NO_3 is nitrate concentration as of N in mg L^{-1} , SO_4 is sulfate concentration in mg L^{-1} .

The c statistic of the final prediction model with five parameters is 0.81, similar to that (0.80) of the model including only water chemistry parameters, but much improved from 0.64 of the model with only bedrock geology (Supporting Information). These suggest that groundwater chemistry, e.g. pH and redox conditions, are better predictors for groundwater arsenic occurrence at intermediate spatial scales than bedrock type, because bedrock geology is not mapped to a spatial resolution that fully represents subsurface heterogeneity. These results should not be used to suggest that bedrock type is irrelevant; rather it is more likely that water chemistry data obtained at sufficiently high spatial density (1 per 1.5 km^2) represented the aquifer heterogeneity better than bedrock type. Actually, the lithochemistry of aquifer material varies even within the same bedrock type, thus arsenic content and geochemical composition of source rocks could be better parameters to use in arsenic prediction models. However, there is no adequate data of arsenic concentration or mineral composition of bedrock in the study area.

Model Test

The percentage of overall correct prediction on all the 789 samples using the model with all 5 parameters is 76%, with a correct prediction rate of 85% on the samples with $[As] < 10 \mu\text{g L}^{-1}$ (specificity) and that of 55% on the samples with $[As] > 10 \mu\text{g L}^{-1}$ (sensitivity) (Table 3). The simplified model with only water geochemical parameters has a specificity of 85% and a sensitivity of 50%. The simplified model with only bedrock geology has a perfect

specificity of 100%, but 0% of sensitivity, which means the model completely mis-predicts those samples with $[As] > 10 \mu\text{g L}^{-1}$. The regional model for New England without water geochemical parameters⁷ also showed a low sensitivity (37%) on samples with $[As] > 5 \mu\text{g L}^{-1}$, which was used as the threshold value. Together, they suggest that including water geochemical parameters can improve the prediction of arsenic occurrence.

In Buxton-Hollis, Maine, the major bedrock units are Silurian meta-sedimentary rocks with Devonian or younger granite intrusions. Median pH value of groundwater samples ($n=94$) is 6.95, median dissolved oxygen is 1.82 mg L^{-1} , similar to those in Devonian granites in the greater Augusta area. Application of logistic regression models with groundwater geochemical parameters results only 14% correct prediction on the samples with $[As] > 10 \mu\text{g L}^{-1}$, which might be due to the highly clustered sampling targeting high arsenic samples. But nevertheless the correct prediction rate is improved from 0% of the prediction model with bedrock geology only. The major bedrock units in 5 areas included in the Maine Geological Survey's Ambient Bedrock Water Quality program are meta-sedimentary rocks and granite/pegmatite, comparable to those in the greater Augusta area. The correct prediction rates on the samples with greater than $10 \mu\text{g L}^{-1}$ of arsenic are 20–30% using models including water geochemical parameters, while the model with bedrock geology only results no correct prediction. These low percentages of correct prediction could be influenced by the much lower sampling density of geochemical predictors (Table 3), and also indicate the difficulty of applying prediction models based on one local area to others, because the significance of influence of bedrock geology and groundwater geochemistry on arsenic might vary in different areas. However, models including these basic groundwater geochemistry parameters, which are often available or easy to obtain, can apparently improve the correct prediction rate.

For the four clusters representing Sw, SOv, and D rock units where 331 groundwater samples were obtained in 2007 (Figure S1), the correct prediction rates on samples with $[As] > 10 \mu\text{g L}^{-1}$ are 62% and 69% using models with all parameters and geochemical parameters only, respectively. Higher density groundwater geochemistry data may have contributed to these better prediction rates. The overall occurrence rates predicted by the model with all 5 parameters were 12% in Chelsea, 46% in Sidney, 39% in Litchfield, and 45% in Manchester. The overall occurrence rates predicted by the model with geochemical parameters only were 19% in Chelsea, 41% in Sidney, 36% in Litchfield, and 48% in Manchester. The observed arsenic occurrence rates were 7% in Chelsea, 29% in Sidney, 57% in Litchfield, and 78% in Manchester. Both models predicted that the arsenic occurrence rate in Chelsea (granitic rocks) is lower than those in other clusters (metamorphic and granitic rocks), but underestimated the occurrence rates in high arsenic clusters. This is not unusual in models developed to predict arsenic occurrence rates³⁵, but it nevertheless underscores difficulties in predicting arsenic occurrence rates quantitatively even when the model was developed based on the data collected from the same area. The underestimation of arsenic occurrence rates in Litchfield and Manchester may be due to: 1) the contact zone between granite intrusions and Silurian meta-sedimentary formations in these two cluster areas may cause higher arsenic occurrence rates¹⁷, 2) the lower sampling density in these two areas (5–6 samples per km^2) compared with higher sampling density in Chelsea and Sidney (42 and 24 samples per km^2 , respectively) may lead to larger uncertainties in the observed arsenic occurrence rates.

Implications on Predicting Arsenic Occurrence in Bedrock Aquifers

Despite the limitations, the value of such a predictive logistic regression model is that it can be applied to obtain a preliminary arsenic risk assessment based on surprisingly few water chemistry parameters and/or bedrock types. Improvement in reliability of prediction is shown to be possible by increasing sampling density of wells, and may be possible with

higher resolution bedrock geology maps and bedrock lithochemistry data. The models are more reliable in identifying areas with low arsenic risk ($< 10 \mu\text{g L}^{-1}$) by correctly predicting 70% to 90% of the cases if only water geochemistry parameters are used. The accuracy is improved to 85% to 92% when all 5 parameters are used. The models tend to under-predict the occurrence rates in areas with high arsenic risk ($> 10 \mu\text{g L}^{-1}$). Thus a caveat in interpreting model prediction is that if any occurrence rate of arsenic is predicted by the model, the observed arsenic occurrence rate is likely to be higher. The reason for this under-prediction is not well understood, but it might be because the median arsenic concentration is usually much lower than the MCL, which was used as the threshold value in the models. As a comparison, the model sensitivity on high arsenic samples could be improved if the threshold of $5 \mu\text{g L}^{-1}$ is used (model results are not shown), which is more comparable with the median arsenic concentration of $3.8 \mu\text{g L}^{-1}$ in the greater Augusta area. Furthermore, the regression method inherently tends to weaken the impact of extreme values, especially when the arsenic data do not follow normal distributions. The analytical uncertainty of those samples with arsenic around threshold values can also add error to the model performance. Lastly, the spatial heterogeneity of arsenic distribution among individual wells within meters or tens of meters is not addressed here (this suggests that the use of models as a predictive tool in no way negates the need for well owners to test for arsenic).

There are reasons to believe that the approach established here in the greater Augusta area, Maine can be applied to predict arsenic occurrence rates in fractured bedrock aquifers at local to intermediate scales at other localities. This is because the bedrock geology, major water geochemical parameters, and arsenic mobilization processes show similarities to a large extent (Table S2). However, it is recommended that locally appropriate logistic regression model be developed and calibrated to improve the accuracy of prediction. On a practical level for the New England region, towns with similar bedrock geology can use our model to preliminarily identify potential risk areas for arsenic using the data of pH, dissolved oxygen, nitrate and sulfate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by the U.S. National Institute of Environmental Health Sciences Superfund Research Program 2 P42 ES10349 to Y. Zheng. We thank A. Smith and his team at Maine Bureau of Health for providing consultation to the families in the study area; Z. Cheng of Brooklyn College for assistance with ICP-MS analysis. The manuscript benefited from reviews of S. Peters and an anonymous reviewer. Finally, we thank the families in the study area for their cooperation.

Literature Cited

1. Kumar A, Adak P, Gurian PL, Lockwood JR. Arsenic exposure in US public and domestic drinking water supplies: A comparative risk assessment. *Journal of Exposure Science and Environmental Epidemiology*. 2010; 20:245–254. [PubMed: 19401722]
2. Hutson SS, Barber NL, Kenny JF, Linsey KS, Lumia DS, Maupin MA. Estimated use of water in the united states in 2000. U.S. Geological Survey Circular. 2004; 1268
3. Peters SC, Blum JD, Karagas MR, Chamberlain CP, Sjostrom DJ. Sources and exposure of the New Hampshire population to arsenic in public and private drinking water supplies. *Chemical Geology*. 2006; 228(1–3):72–84.
4. Nielsen MG, Lombard PJ, Schalk LF. Assessment of Arsenic Concentrations in Domestic Well Water, by Town, in Maine, 2005–09. U.S. Geological Survey Scientific Investigations Report. 2010; 5199:68.

5. Sidle WC, Fischer RA. Detection of H-3 and Kr-85 in groundwater from arsenic-bearing crystalline bedrock of the Goose River basin, Maine. *Environmental Geology*. 2003; 44(7):781–789.
6. Sidle WC, Wotten B, Murphy E. Provenance of geogenic arsenic in the Goose River basin, Maine, USA. *Environmental Geology*. 2001; 41(1–2):62–73.
7. Ayotte JD, Nolan BT, Nuckols JR, Cantor KP, Robinson GR, Baris D, Hayes L, Karagas M, Bress W, Silverman DT, Lubin JH. Modeling the Probability of Arsenic in Groundwater in New England as a Tool for Exposure Assessment. *Environmental Science and Technology*. 2006; 40(11):3578–3585. [PubMed: 16786697]
8. Peters SC. Arsenic in groundwaters in the Northern Appalachian Mountain belt: A review of patterns and processes. *Journal of Contaminant Hydrology*. 2008; 99(1–4):8–21. [PubMed: 18571283]
9. Ayotte JD, Montgomery DL, Flanagan SM, Robinson KW. Arsenic in groundwater in eastern New England: occurrence, controls, and human health implications. *Environmental Science and Technology*. 2003; 37(10):2075–2083. [PubMed: 12785510]
10. Lipfert G, Reeve AS, Sidle WC, Marvinney R. Geochemical patterns of arsenic-enriched ground water in fractured, crystalline bedrock, Northport, Maine, USA. *Applied Geochemistry*. 2006; 21(3):528–545.
11. Peters SC, Blum JD. The source and transport of arsenic in a bedrock aquifer, New Hampshire, USA. *Applied Geochemistry*. 2003; 18(11):1773–1787.
12. Marvinney, RG.; Loiselle, MC.; Hopeck, JT.; Braley, D.; Krueger, JA. Arsenic in Maine ground water - An example from Buxton, Maine; the proceeding of the 1994 Focus Conference on Eastern Regional Ground Water Issues, Burlington, Vermont; 1994. p. 701-715.
13. Yang Q, Jung HB, Culbertson CW, Marvinney RG, Loiselle MC, Locke DB, Cheek H, Thibodeau H, Zheng Y. Spatial Pattern of Groundwater Arsenic Occurrence and Association with Bedrock Geology in Greater Augusta, Maine. *Environmental Science and Technology*. 2009; 43(8):2714–2719. [PubMed: 19475939]
14. Ayotte JD, Nielsen MG, Robinson GR, Moore RB. Relation of Arsenic, Iron, and Manganese in Ground Water to Aquifer Type, Bedrock Lithogeochemistry, and Land Use in the New England Coastal Basins. U.S. Geological Survey, Water-Resources Investigations Report 99-4162. 1999:70.
15. Pagach, JJ. Distribution of Arsenic in Bedrock Wells in the Eastern Highlands of Connecticut; the proceeding of New England Private Well Symposium, 2009, Portland, ME; 2009.
16. Montgomery DL, Ayotte JD, Carroll PR, Hamlin P. Arsenic concentrations in private bedrock wells in southeastern New Hampshire. U.S. Geological Survey Fact Sheet 051-03. 2003
17. Peters SC, Blum JD, Klaue B, Karagas MR. Arsenic Occurrence in New Hampshire Drinking Water. *Environmental Science and Technology*. 1999; 33(9):1328–1333.
18. Sidle W. 18OSO4 and 18OH2O as prospective indicators of elevated arsenic in the Goose River ground-watershed, Maine. *Environmental Geology*. 2002; 42(4):350–359.
19. Ryan PC, Kim J, Wall AJ, Moen JC, Corenthal LG, Chow DR, Sullivan CM, Bright KS. Ultramafic-derived arsenic in a fractured bedrock aquifer. *Applied Geochemistry*. 2011; 26(4): 444–457.
20. Colman JA. Arsenic and uranium in water from private wells completed in bedrock of east-central Massachusetts—Concentrations, correlations with bedrock units, and estimated probability maps. U.S. Geological Survey Scientific Investigations Report 5013. 2011:113.
21. Peters SC, Burkert L. The occurrence and geochemistry of arsenic in groundwaters of the Newark basin of Pennsylvania. *Applied Geochemistry*. 2008; 23(1):85–98.
22. Kim D, Miranda ML, Tootoo J, Bradley P, Gelfand AE. Spatial Modeling for Groundwater Arsenic Levels in North Carolina. *Environmental Science and Technology*. 2011; 45(11):4824–4831. [PubMed: 21528844]
23. Kim M-J, Nriagu J, Haack S. Arsenic species and chemistry in groundwater of southeast Michigan. *Environmental Pollution*. 2002; 120(2):379–390. [PubMed: 12395852]
24. Meliker JR, Slotnick MJ, AvRuskin GA, Haack SK, Nriagu JO. Influence of groundwater recharge and well characteristics on dissolved arsenic concentrations in southeastern Michigan groundwater. *Environmental Geochemistry and Health*. 2009; 31:147–157. [PubMed: 18496757]

25. Boyle DR, Turner RJW, Hall GEM. Anomalous arsenic concentrations in groundwaters of an island community, Bowen Island, British Columbia. *Environmental Geochemistry and Health*. 1998; 20(4):199–212.
26. Armienta MA, Villasenor G, Rodriguez R, Ongley LK, Mango H. The role of arsenic-bearing rocks in groundwater pollution at Zimapan Valley, Mexico. *Environmental Geology*. 2001; 40(4–5):571–581.
27. Shukla DP, Dubey CS, Singh NP, Tajbakhsh M, Chaudhry M. Sources and controls of Arsenic contamination in groundwater of Rajnandgaon and Kanker District, Chattisgarh Central India. *Journal of Hydrology*. 2010; 395:49–66.
28. Heinrichs G, Udluft P. Natural arsenic in Triassic rocks: A source of drinking-water contamination in Bavaria, Germany. *Hydrogeology Journal*. 1999; 7(5):468–476.
29. Kortatsi B. Hydrochemical framework of groundwater in the Ankobra Basin, Ghana. *Aquatic Geochemistry*. 2007; 13(1):41–74.
30. Buamah R, Petrusevski B, Schippers JC. Presence of arsenic, iron and manganese in groundwater within the gold-belt zone of Ghana. *Journal of Water Supply: Research & Technology-AQUA*. 2008; 57(7):519–529.
31. Smedley PL. Arsenic in rural groundwater in Ghana. *Journal of African Earth Sciences*. 1996; 22(4):459–470.
32. Smedley PL, Knudsen J, Maiga D. Arsenic in groundwater from mineralised Proterozoic basement rocks of Burkina Faso. *Applied Geochemistry*. 2007; 22:1074–1092.
33. Day-Lewis FD, Hsieh PA, Gorelick SM. Identifying fracture-zone geometry using simulated annealing and hydraulic-connection data. *Water Resources Research*. 2000; 36(7):1707–1721.
34. Amini M, Abbaspour KC, Berg M, Winkel L, Hug SJ, Hoehn E, Yang H, Johnson CA. Statistical Modeling of Global Geogenic Arsenic Contamination in Groundwater. *Environmental Science and Technology*. 2008; 42(10):3669–3675. [PubMed: 18546706]
35. Winkel L, Berg M, Amini M, Hug SJ, Annette Johnson C. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. *Nature Geosciences*. 2008; 1(8):536–542.
36. Hanson, LS.; Bradley, DC. Sedimentary facies and tectonic interpretation of the Lower Devonian Carrabassett Formation, north-central Maine. In: Tucker, RD.; Marvinney, RG., editors. *Studies in Maine Geology: Volume 2 - Structure and Stratigraphy*. Maine Geological Survey; 1989. p. 101–126.
37. Cheng Z, Zheng Y, Mortlock R, van Geen A. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. *Analytical and Bioanalytical Chemistry*. 2004; 379:512–518. [PubMed: 15098084]
38. Senn DB, Hemond HF. Nitrate Controls on Iron and Arsenic in an Urban Lake. *Science*. 2002; 296(5577):2373. [PubMed: 12089437]
39. Marvinney RG, Barker DS. Bedrock Geology Map of Augusta Quadrangle, Maine. Maine Geological Survey Open File No 08-83. 2008

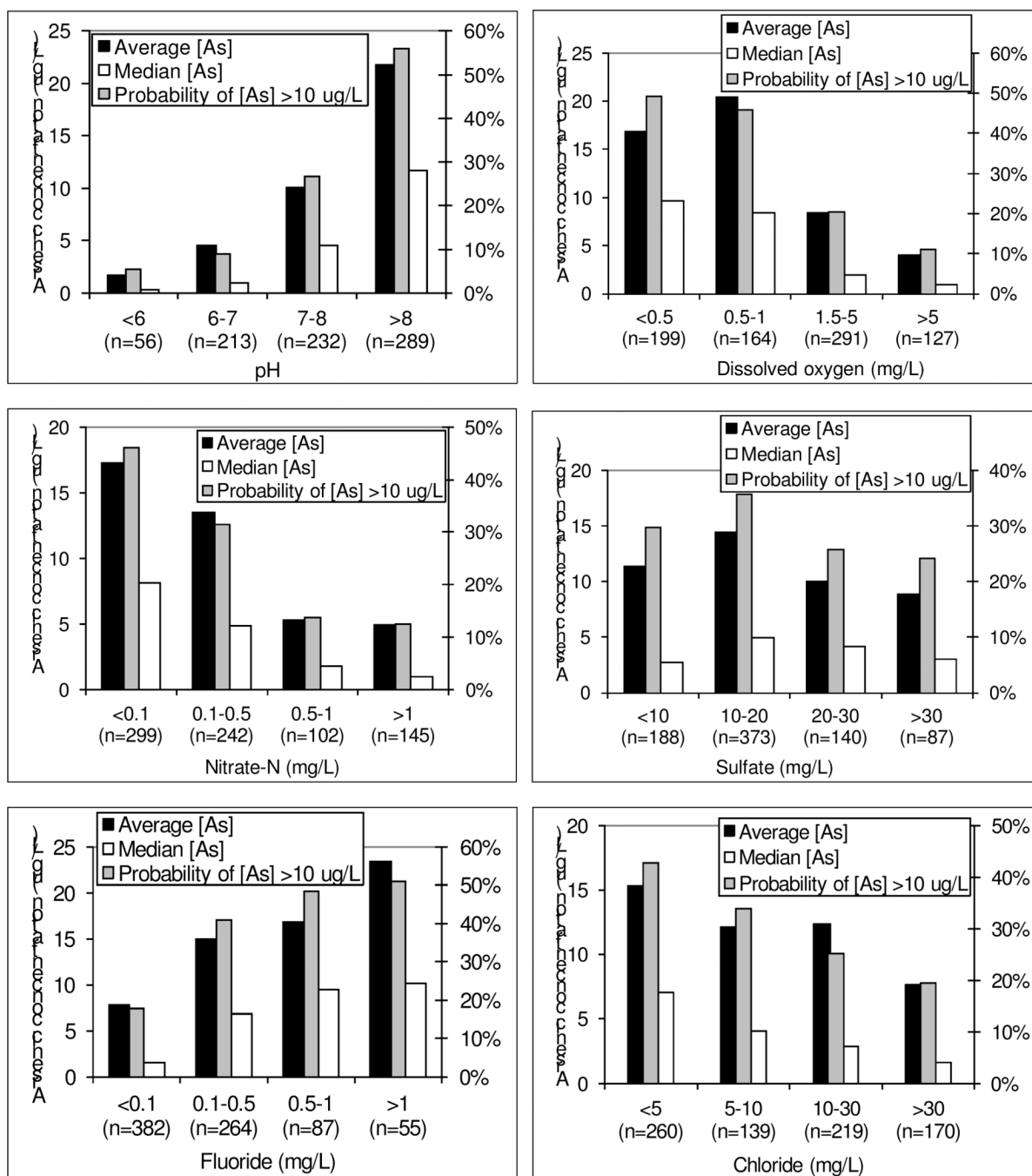


Figure 1.

Arsenic concentrations and occurrence rates vs. pH, dissolved oxygen, nitrate as of N, sulfate, fluoride, and chloride in groundwater of Greater Augusta aquifers sampled in 2006. Parameters are shown in the order of their association with groundwater arsenic: pH > DO > nitrate > sulfate.

Table 1

Median values of hydrogeochemical parameters of groundwater sampled in 2006 and 2007 from Greater Augusta, Maine

Parameter	Unit	2006						2007					
		Greater Augusta			meta-sedimentary rocks			volcanics			Chelsea		
		Ss	Sw	SOv, D, OZc	Ss	Sw	SOv	granite	D	OZc	D	Sw	SOv, D
bedrock unit													
n		790	259		161	259	248	86	30		107	62	49
well depth	m	59	57		61	57	53	61	40		56	61	65
temperature deg. C		10.5	10.5		10.5	10.5	10.5	10.5	10.4		10.7	10.6	10.6
As	ug/L	3.8	6.1		6.8	6.1	2.0	2.5	0.2		0.9	4.1	14.6
pH		7.59	7.92		7.99	7.92	7.30	6.94	7.35		6.80	8.17	7.78
dO ₂	mg/L	1.5	1.5		1.0	1.5	2.0	4.0	1.3		2.0	0.6	1.0
Na	mg/L	10.9	11.1		9.1	11.1	11.8	10.0	12.5		13.2	8.4	8.3
Mg	mg/L	3.2	2.8		2.7	2.8	3.9	2.9	3.4		4.3	2.0	1.4
Si	mg/L	8.7	8.3		8.3	8.3	9.4	8.5	9.9		7.4	6.8	6.4
K	mg/L	2.3	2.5		1.8	2.5	2.7	2.1	3.0		3.0	1.5	2.1
Ca	mg/L	30.8	30.9		26.7	30.9	37.7	28.0	23.7		36.9	18.2	25.8
alkalinity	mmol/L	1.28									1.45	1.17	1.27
Cl	mg/L	9.7	8.9		6.2	8.9	14.5	12.1	6.3		44.0	8.1	11.8
SO ₄	mg/L	14.6	14.9		11.9	14.9	17.4	10.6	14.2		17.5	19.4	15.4
F	mg/L	0.10	0.13		0.22	0.13	0.04	0.12	0.13		0.04	0.39	0.09
NO ₃ -N	mg/L	0.21	0.14		0.15	0.14	0.26	0.24	0.12		1.02	0.00	0.00
Fe	ug/L	113.2	0.11		0.14	0.11	0.10	0.07	0.17		47.2	125.9	92.0
Mn	ug/L	10.4	9.5		13.8	9.5	10.4	5.8	51.5		2.9	23.5	12.5
Mo	ug/L	1.19	1.17		2.09	1.17	0.86	1.78	1.31		0.49	2.49	0.55
U	ug/L	1.06	0.73		0.58	0.73	1.31	4.88	0.98		7.71	0.55	2.01
													6.97

Explanatory variables in the final logistic regression model of groundwater arsenic in bedrock aquifers of Greater Augusta, Maine

Table 2

Variable	Type	β	$\exp(\beta)$	p-value
Bedrock geology				
Intercept		-9.520	0.00	<0.0001
Sangerville formation (Ss)	categorical	1.652	5.22	0.0118
Waterville formation (Sw)	categorical	1.577	4.84	0.0148
Vassalboro formation (SOv)	categorical	1.299	3.67	0.0459
Granite intrusion (D)	categorical	0.570	1.77	0.4243
Cushing formation (OZe, comparison varibale)	categorical			
Water geochemistry				
pH	continuous	1.075	2.93	<0.0001
Dissolved Oxygen	continuous	-0.117	0.89	0.0201
Nitrate	continuous	-0.239	0.79	0.0652
Sulfate	continuous	-0.029	0.97	0.0022

* β - the relative importance of a variable (the larger in absolute values the more important a variable)

* p-value - if a variable is important by chance or not (a value closer to zero is more desirable)

Table 3

Summary of model prediction of groundwater arsenic in bedrock aquifers in Maine

observation				correct prediction rate from models using					
n	n <10 µg L ⁻¹	n >10 µg L ⁻¹	exceedance	bedrock geology		water geochemistry		all parameters	
				specificity ^a	sensitivity ^b	specificity ^a	sensitivity ^b	specificity ^a	sensitivity ^b
2006 sampling in Greater Augusta area (sampling density = 1 per 1.4 km ²)									
all	789	543	246	31%	100%	0%	0%	85%	50%
Ss	159	91	68	43%	100%	0%	0%	75%	59%
Sw	263	159	104	40%	100%	0%	0%	80%	57%
SOv	248	190	58	23%	100%	0%	0%	92%	29%
D	87	74	13	15%	100%	0%	0%	86%	54%
OZc	32	29	3	9%	100%	0%	0%	90%	33%
2007 sampling in Greater Augusta area (sampling density = 1 per 0.1 km ²)									
all	331	190	141	43%	100%	0%	0%	88%	69%
Chelsea	107	100	7	7%	100%	0%	0%	94%	71%
Sidney	62	44	18	29%	100%	0%	0%	73%	72%
Litchfield	49	21	28	57%	100%	0%	0%	90%	46%
Manchester	113	25	88	78%	100%	0%	0%	92%	75%
Buxton and Hollis (sampling density = 1 per 2 km ²)									
	94	59	35	37%	100%	0%	0%	90%	14%
Ambient Bedrock Water Quality program (sampling density = 1 per 14 km ²)									
	228	218	10	4%	100%	0%	0%	70%	20%
								86%	30%

^a Specificity is defined as the correct prediction rate of the model on samples with [As] < 10 µg L⁻¹.

^b Sensitivity is defined as the correct prediction rate of the model on samples with [As] > 10 µg L⁻¹.