



# Geostatistical model of the spatial distribution of arsenic in groundwaters in Gujarat State, India

Ruohan Wu · Joel Podgorski · Michael Berg · David A. Polya 

Received: 15 December 2019 / Accepted: 24 June 2020 / Published online: 11 July 2020  
© The Author(s) 2020

**Abstract** Geogenic arsenic contamination in groundwaters poses a severe health risk to hundreds of millions of people globally. Notwithstanding the particular risks to exposed populations in the Indian sub-continent, at the time of writing, there was a paucity of geostatistically based models of the spatial distribution of groundwater hazard in India. In this study, we used logistic regression models of secondary groundwater arsenic data with research-informed secondary soil, climate and topographic variables as principal predictors generate hazard and risk maps of groundwater arsenic at a resolution of 1 km across Gujarat State. By combining models based on different arsenic concentrations, we have generated a pseudo-contour map of groundwater arsenic concentrations, which indicates greater arsenic hazard (> 10 µg/L) in the northwest, northeast and south-

east parts of Kachchh District as well as northwest and southwest Banas Kantha District. The total number of people living in areas in Gujarat with groundwater arsenic concentration exceeding 10 µg/L is estimated to be around 122,000, of which we estimate approximately 49,000 people consume groundwater exceeding 10 µg/L. Using simple previously published dose–response relationships, this is estimated to have given rise to 700 (prevalence) cases of skin cancer and around 10 cases of premature avoidable mortality/annum from internal (lung, liver, bladder) cancers—that latter value is on the order of just 0.001% of internal cancers in Gujarat, reflecting the relative low groundwater arsenic hazard in Gujarat State.

**Keywords** Groundwater · Arsenic · Health impacts · Gujarat · Logistic regression · Geostatistics

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10653-020-00655-7>) contains supplementary material, which is available to authorized users.

R. Wu · J. Podgorski · D. A. Polya (✉)  
Department of Earth and Environmental Sciences, School of Natural Sciences and Williamson Research Centre for Molecular Environmental Science, University of Manchester, Manchester M13 9PL, UK  
e-mail: david.polya@manchester.ac.uk

J. Podgorski · M. Berg  
Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

## Introduction

Arsenic (As) is a toxic element, found in more than 200 minerals in nature (Thornton and Farago 1997; Ravenscroft et al. 2009) with arsenic being released into groundwater under specific biogeochemical and hydrogeological conditions (Islam et al. 2004; Guo et al. 2011). In many parts of the world, arsenic-contaminated groundwater is used for drinking water and irrigation (Nickson et al. 2005; Rahman and

Hasegawa 2011). The long-term consumption of arsenic may greatly increase the risk of skin cancers, bladder cancers, lung cancers, cardiovascular disease and other detrimental health outcomes (Chen and Ahsan 2004; Chowdhury et al. 2000). The provisional guideline value of arsenic in drinking water established by the World Health Organization (WHO) is 10 µg/L (WHO/UNICEF 2018); however, an increasing number of studies have pointed to detrimental health outcomes for exposure at lower arsenic concentrations (Medrano et al. 2010; García-Esquinas et al. 2013; Monrad et al. 2017; Moon et al. 2017; Polya et al. 2019b; Ahmad et al. 2020).

Groundwater arsenic contamination (WHO/UNICEF 2018; Bhattacharya et al. 2017; Bretzler and Johnson 2015) is the most substantive contributor to preventable detrimental health outcomes arising from chemicals such as F, Mn, Pb, pesticides in drinking water (Smith et al. 2000). As many as 100,000 preventable deaths may arise each year from exposure to such groundwater arsenic across the globe (Polya et al. 2019a, b; Smith et al. 2000), particularly in densely populated (van Geen 2008) areas in south and south-east Asia (Polya and Charlet 2009; Fendorf et al. 2010), Bangladesh (Argos et al. 2010, Flanagan et al. 2012), Pakistan (Podgorski et al. 2017) and India (Chakraborti et al. 2004).

Although there have been numerous studies of arsenic-contaminated groundwaters utilized for domestic consumption (e.g. Chatterjee et al. 1995; Chowdhury et al. 1999, 2000; Chakraborti et al. 2003), relatively few studies have produced hazard prediction maps indicating the spatial distribution of groundwater arsenic for whole districts or states in India (Buragohain and Sarma 2012; Ghosh et al. 2004, 2019). However, such maps have been generated for other regions (Amini et al. 2008; Winkel et al. 2008) and countries (Lado et al. 2008; Sovann and Polya 2014; Bretzler et al. 2017), notably including Bangladesh (Kinniburgh and Smedley 2001) and Pakistan (Podgorski et al. 2017).

Spatial geostatistical models used to predict the distribution of groundwater contaminants include logistic regression (Winkel et al. 2008; Ayotte et al. 2017; Podgorski et al. 2017; Bretzler et al. 2017) Tyson polygons (Ghosh et al. 2019), ordinary Kriging (Ghosh et al. 2019; Sovann and Polya 2014), regression Kriging (Sovann and Polya 2014), and random forest models (Podgorski et al. 2018). Methods such as

logistic regression and random forest find statistical relationships between a target variable and predictor variables in order to make predictions (Winkel et al. 2008; Ayotte et al. 2017; Podgorski et al. 2017; Bretzler et al. 2017; Podgorski et al. 2018). Such methods can be used to consider a variety of environmental factors that may act as proxies or have a direct relationship to the release and accumulation of arsenic in groundwaters. Due to the often highly heterogeneous distribution of groundwater arsenic in sedimentary aquifers, modelling based on a binary target variable to produce probabilities, such as logistic regression, is often performed rather than attempting to predict a continuous variable.

As a preliminary step to developing a comprehensive model of the spatial distribution of arsenic in groundwaters across India, we (1) present logistic regression-based geostatistical models of the distribution of arsenic in groundwaters in the state of Gujarat, (2) outline methods permitting model results to be rendered as a pseudo-contour map of likely concentrations and (3) combine the modelled arsenic hazard with simple exposure route and dose–response models to provide plausible estimates of detrimental health outcomes in Gujarat that can be attributed to arsenic in drinking water.

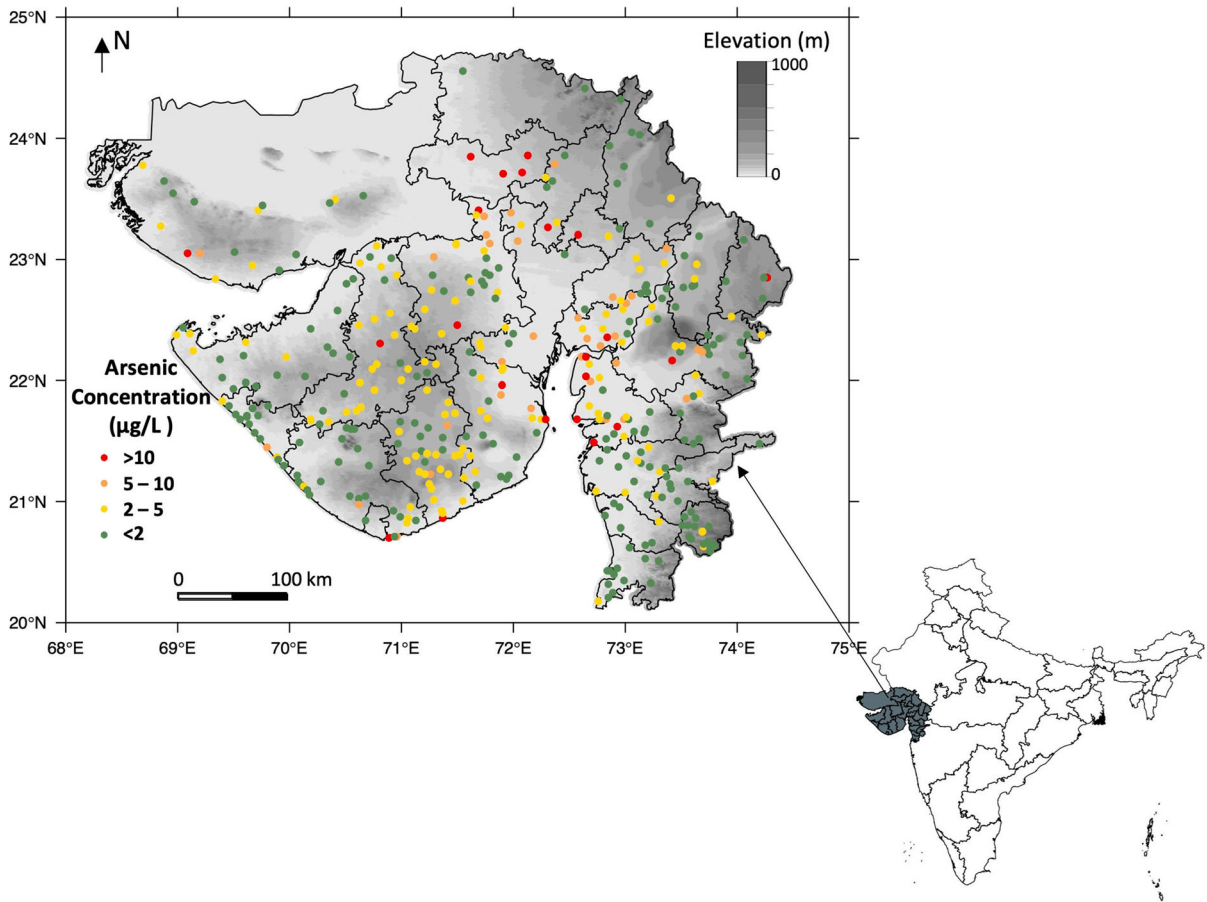
## Materials and methods

### Study area

Gujarat is located between 20° 06' and 24° 42' north latitude and 68° 10' to 74° 28' east longitude, with an area of 196,024 sq. km (CGWB 2016) (Fig. 1). The population of Gujarat State is 70,445,000 (Chandramouli 2011). Gujarat has nearly 1600 km of coastline which is the longest coastline in India (CGWB 2016). Diverse climatic, topographic and geological and physiographic conditions result in diversification of groundwater conditions in different parts of Gujarat State (Sharma and Kumar 2008).

### Dataset compilation

Groundwater arsenic data from throughout Gujarat were obtained from surveys conducted by the Central Ground Water Board of India (CGWB) in 2015 (CGWB 2016). The CGWB collected groundwater



**Fig. 1** The location of Gujarat State and distribution of groundwater arsenic concentrations used in modelling

samples from dug wells, tube wells and bore wells during May 2015, which is the end of dry season shortly before the onset of monsoon and analysed for arsenic by a colorimetric method using a visible spectrophotometer with an implied detection limit of around 1 µg/L. Of the 599 samples reported, (1) 183 samples for which the arsenic concentration was recorded as “nd” we have taken to have not been analysed and have excluded from the dataset; (2) a further 18 samples for which arsenic concentrations were reported without location data were also excluded from the dataset, leaving 398 datapoints with both groundwater arsenic and location data (Fig. 1): of these only 6% showed arsenic concentrations greater than 10 µg/L, with the maximum reported arsenic concentration being 26 µg/L. The frequency distribution of groundwater arsenic concentrations is shown in Fig. S1.

Potential independent variables ( $n = 28$ ) related to geology, hydrology, soil properties, climate, and topography were compiled from a variety of sources, many based on or relying upon remote sensing (Table S1). These variables were initially chosen based on established and proposed relationships with the release and enrichment of groundwater arsenic (Smedley and Kinniburgh 2002; Islam et al. 2004; McArthur et al. 2004; Charlet and Polya 2006; Polya and Charlet 2009; Rodríguez-Lado et al. 2013; Polya and Middleton 2017; Podgorski et al. 2018; Polya et al. 2019a, b) and prepared to predict the distribution of groundwater arsenic in Gujarat State. The resolution and sources (Trabucco and Zomer 2009, 2010; ISRIC 2017; Hijmans et al. 2005; Hengl 2018; Fan et al. 2013; The World Bank 2017; Pelletier et al. 2016; Hartmann and Moosdorf 2012) of the independent variables dataset are shown in Table S1.

## Dataset preparation

The six thresholds of 10 µg/L, 5 µg/L, 4 µg/L, 3 µg/L, 2 µg/L and 1 µg/L were used to create binary datasets for creating six different geostatistical models. These values were chosen based on being the WHO provisional guideline of 10 µg/L and due to 85% of arsenic concentrations in the dataset being in the range of 1 to 5 µg/L. Of the 398 groundwater arsenic concentrations, 24 (6%), 57 (14%), 78 (20%), 124 (31%), 185 (46%) and 301 (76%) arsenic concentrations exceeded 10 µg/L, 5 µg/L, 4 µg/L, 3 µg/L, 2 µg/L and 1 µg/L, respectively. The dataset was converted into high and low classes by assigning one to all arsenic concentrations > threshold concentrations and zero to all arsenic concentrations ≤ the threshold concentrations. The converted dataset was randomly divided into training (80%) and testing (20%) datasets maintaining the same ratio of low to high values as in the entire dataset.

## Statistical modelling

In this study, we used the logistic regression models to predict arsenic contamination in Gujarat groundwaters. Logistic regression uses a logistic function to predict a binary dependent variable with the probability between 0 and 1 (Hosmer et al. 2013). In this case, the binary dependent variable represents whether or not groundwater arsenic concentration exceeds a given threshold. The logistic function is as follows:

$$\log\left(\frac{P(y=1)}{1-P(y=1)}\right) = \log\left(\frac{P(y=1)}{P(y=0)}\right) = \beta_0 + \beta_1x_1 + \dots + \beta_nx_n$$

$$\frac{P(y=1)}{1-P(y=1)} = \text{odds} = \exp(\beta_0 + \beta_1x_1 + \dots + \beta_nx_n)$$

$$P(y=1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1x_1 + \dots + \beta_nx_n))}$$

where  $P(y=1)$  and  $P(y=0)$  are the probability of the dependent variable being 1 or 0;  $x_1 \dots x_n$  are the independent variables;  $\beta_0 \dots \beta_n$  are the regression intercept and other coefficients.

Multicollinearity is a statistical phenomenon in which predictor variables of a logistic regression model are highly correlated. The existence of

collinearity increases the variances of parameter estimates and thus leads to erroneous inferences about the relationship between dependent and independent variables (Midi et al. 2010). Variance inflation factor (VIF) quantifies the severity of multicollinearity of independent variables (predictors) in regression analysis (Franke 2010). It was used for independent variable selection in this study.

$$VIF = \frac{1}{1 - R^2}$$

where  $R^2$  is the coefficient of determination,  $R^2 = 1 - e^{-\frac{D}{n}}$  ( $D$  is the test statistic of the likelihood ratio test,  $n$  is the sample size.)

The empirical judgment method is that if  $VIF > 10$  then multicollinearity is high (Franke 2010).

We used stepwise variable selection in which Akaike information criterion (AIC) is used as criterion for removing or adding variables to determine final logistic regression models. AIC is an estimator of the complexity and goodness of fit of statistical models (Akaike 1974).

$$AIC = 2k - 2 \ln(L)$$

where  $k$  is the number of parameters;  $L$  is the Likelihood of the model.

The objectively preferred variable combination in stepwise selection was the one with the lowest AIC value, providing the best combination of performance and complexity.

## Variable selection

Based on their known or potential relationships to arsenic occurrence in groundwater, twenty-eight independent variables (see Table S1), including twenty-four continuous variables and four categorical variables, were considered for potential use in logistic regression modelling. In order to help identify effective independent variables, univariate logistic regressions were run for each of six thresholds on the training dataset which is consistent with the dataset used for logistic regression analysis. The significance of each independent variable was assessed through its  $p$  value tested by the analysis of variance (ANOVA) type II test (Pearce and Ferrier 2000). Independent variables with  $p$  values < 0.05 (within the 95% confidence interval) were retained for further selection. Multicollinearity of the continuous variables

following the univariate analysis was then calculated on the training dataset at each threshold. Predictor variables with a variance inflation factor (VIF) > 10 were removed on the basis of strong multicollinearity. The univariate regression and multicollinearity analysis were repeated 1000 times in order to avoid the random bias produced by specific splitting of training and testing datasets at one time. The averaged *p* value and VIF were used to determine the addition or removal of variables during variable selection.

### Logistic regression analysis

Logistic regression analysis was run on the training dataset for each of six thresholds using a stepwise selection of variables (both directions), which removes or adds variables according to their improvement to the Akaike information criterion (AIC). The Hosmer–Lemeshow goodness-of-fit test (Hosmer et al. 2013) was also used on the testing dataset to determine the accuracy of regressions at the 95% confidence level, such that there is no significant difference between the fitted values and observed values if the *p* value is > 0.05. In order to avoid introducing bias to the model by performing only a single split of training and testing datasets, logistic regressions were performed 1000 times with the Hosmer–Lemeshow goodness-of-fit test. The logistic regression models passing the Hosmer–Lemeshow goodness-of-fit test (*p* value is > 0.05) provided various variable combinations determined by AIC values. The different combinations of variables of the logistic regressions passing the Hosmer–Lemeshow goodness-of-fit test were counted. The mean of coefficients of each combination passing the Hosmer–Lemeshow goodness-of-fit test were utilized as the coefficients of the model.

The true-positive rate (sensitivity) and true-negative rate (specificity) were calculated on both the entire dataset and testing datasets passing the Hosmer–Lemeshow goodness-of-fit test for each of the six thresholds. Plotting sensitivity against specificity for the range of probability cut-off values from 0 to 1 on the entire dataset produced a receiver operating characteristic (ROC) curve and the associated area under the ROC curve (AUC), which generally ranges from 0.5 (no predictive capability) to 1 (perfect predictive capability) (Fawcett 2006). Mean AUC values were also calculated on the test dataset of each

logistic regression. The largest AUC value among the variable combinations passing the Hosmer–Lemeshow goodness-of-fit test was used to select the final model.

### Hazard and potential exposure maps

The final logistic regression models were utilized to calculate the probability of groundwater arsenic concentration exceeding each of the threshold concentrations. The sensitivity, accuracy, and specificity of the final models were plotted against cut-offs. The cut-off values at which sensitivity and specificity are equal were used to classify whether arsenic concentrations exceed the given thresholds (Podgorski et al. 2017). These were then used to generate a pseudo-contour map of groundwater arsenic concentrations, which was combined with population density (Pages et al. 2018) to generate a potential exposure map.

### Health risk estimation

Based on the potential exposure map, we used dose response functions for arsenic-induced cancers to evaluate the health effects of exposure to groundwater arsenic in Gujarat.

- (1) Prevalence ratio of arsenic-induced skin cancer as a function of arsenic concentration, *c*, and age, *t* (Brown et al. 1989).

$$p(c, t) = 1 - \exp\left(- (q_1c + q_2c^2)(t - m)^k H(t - m)\right)$$

where *p*(*c*, *t*) denotes prevalence ratio of the gender with arsenic-induced skin cancer; *c* denotes arsenic concentration, µg/L; *t* denotes age, year; *q*<sub>1</sub>, *q*<sub>2</sub>, *k*, *m* are the nonnegative parameters, listed in Table S2;

*H*(*t* − *m*) denotes the Heaviside function with *H*(*t* − *m*) = 0 for *t* < *m* and *H*(*t* − *m*) = 1 for *t* ≥ *m*.

- (2) Incidence rate of arsenic-induced internal cancer (lung cancer, bladder cancer, liver cancer) as a function of arsenic concentration, *c*, and age, *t* (NRC 1999, 2001; Yu et al. 2003).

$$h(c, t) = k(q_1c + q_2c^2)(t - m)^{k-1} H(t - m)$$

where *h*(*c*, *t*) denotes incidence rate of the gender with arsenic-induced internal cancer, per year; *c* denotes arsenic concentration, µg/L; *t* denotes age, year;

$q_1, q_2, k, m$  are the nonnegative parameters, listed in Table S2;

$H(t - m)$  denotes the Heaviside function with  $H(t - m) = 0$  for  $t < m$  and  $H(t - m) = 1$  for  $t \geq m$ .

## Results and discussion

### Logistic regression models

The univariate regression and multicollinearity analysis retained 10, 15, 16, 13, 9 and 2 independent variables for models with thresholds of 10  $\mu\text{g/L}$ , 5  $\mu\text{g/L}$ , 4  $\mu\text{g/L}$ , 3  $\mu\text{g/L}$ , 2  $\mu\text{g/L}$ , and 1  $\mu\text{g/L}$ , respectively (Table 1). Of the 1000 logistic regression iterations performed, 707, 535, 679, 736, 858 and 473 regression runs passed the Hosmer–Lemeshow goodness-of-fit test of models using the thresholds of 10  $\mu\text{g/L}$ , 5  $\mu\text{g/L}$ , 4  $\mu\text{g/L}$ , 3  $\mu\text{g/L}$ , 2  $\mu\text{g/L}$ , and 1  $\mu\text{g/L}$ , respectively. The variables appearing in the regressions passing the Hosmer–Lemeshow goodness-of-fit test for each of six thresholds are listed in Table S3.

The optimum combinations of independent variables in the final model for each threshold were determined using the areas under the ROC curve (AUC). Both the AUC calculated using entire dataset and testing datasets of regressions passing the Hosmer–Lemeshow goodness-of-fit test were very similar. Six variable combinations with highest AUC values (Table 2) were selected as final models. The coefficients and intercepts of normalized variables and their standard deviations in final models are summarized in Table 3. However, many other variable combinations not selected for various thresholds may also have good predictive capabilities, as evidenced by high AUC values, see Tables S4–S8.

The AUC values indicate that models with thresholds of 10  $\mu\text{g/L}$  (Fig. 2), 5  $\mu\text{g/L}$  (Fig. 2), 4  $\mu\text{g/L}$  (Fig. S2), 3  $\mu\text{g/L}$  (Fig. S2), and 2  $\mu\text{g/L}$  (Fig. 2) perform well (AUC 0.71–0.83), whereas the classification performance of the 1  $\mu\text{g/L}$  model is not satisfactory (AUC 0.60, Fig. S2), which may be due to detection limits of the arsenic analysis. The 1  $\mu\text{g/L}$  model was therefore excluded from the further consideration. The crossover between sensitivity (true-positive rate) and specificity (true-negative rate) against cut-offs (Figs. 2 and S3) were utilized to determine high-risk areas of groundwater arsenic concentrations.

### Predictor variables

Eight predictor variables were included in the final models to predict the distribution of groundwater arsenic in Gujarat and can be grouped into three categories: (1) climate variables, (2) geological variables and (3) topographic variables (Fig. S5). Positive coefficients were found for fluvisols, soil and sedimentary deposit thickness, potential evapotranspiration, temperature, and topographic wetness index, whereas negative coefficients were found for aridity, slope, and soil water capacity.

Climate variables (temperature, potential evapotranspiration, and aridity) in final models relate to arsenic accumulation in aquifers significantly. High temperature promotes the evapotranspiration and can increase drought. The combination of high temperature, high evapotranspiration and low aridity index (average precipitation/potential evapotranspiration) can increase the evaporative concentration of groundwater and hence increase arsenic concentrations, particularly in inland and/or enclosed basins in arid or semi-arid climates (Smedley and Kinniburgh 2002; Ravenscroft et al. 2009; Alarcón-Herrera et al. 2013).

Fluvisols and soil and sedimentary deposit thickness are also conducive to the enrichment of arsenic in groundwaters. Fluvisols are genetically young soils in alluvial deposits (IUSS 2015). Previous studies (Ahmed et al. 2004; Chakraborti et al. 2013; McArthur et al. 2001) have shown that arsenic pollution occurs dominantly in the alluvial deposits of major rivers which flow south and east from the Himalayas and Tibetan plateau, where rivers flow through the highest mountains with the largest rainfall and generate the greatest sedimentary deposit worldwide. The widely accepted mechanism of arsenic release into groundwaters in alluvial aquifers is the microbially mediated dissimilatory reductive dissolution of arsenic-bearing Fe oxides (Fe oxyhydroxides, hydroxides, and oxides) (Islam et al. 2004; Berg et al. 2007). The abundance of relatively young reactive organic matter in sedimentary deposits is plausibly causally linked to the occurrence of high arsenic concentrations in groundwaters (Rowland et al. 2007, 2011; Mukherjee et al. 2019). Hence, increased fluvisols, soil and sedimentary deposit thickness promote arsenic accumulation in groundwaters.

Low slope can be regarded as a proxy for slow groundwater flow, which suppresses the flushing of

**Table 1** Univariate logistic regressions and multicollinearity analysis for models with 1–10 µg/L as thresholds for groundwater arsenic

Variables	Thresholds											
	10 µg/L		5 µg/L		4 µg/L		3 µg/L		2 µg/L		1 µg/L	
	<i>p</i> value	VIF	<i>p</i> value	VIF	<i>p</i> value	VIF	<i>p</i> value	VIF	<i>p</i> value	VIF	<i>p</i> value	VIF
Actual evapotranspiration	$3.80 \times 10^{-1}$		$4.69 \times 10^{-1}$		<b><math>1.38 \times 10^{-2}</math></b>	103	$1.81 \times 10^{-3}$	94.11	<b><math>3.34 \times 10^{-4}</math></b>	81.8	$6.94 \times 10^{-1}$	
Aridity	<b><math>2.35 \times 10^{-2}</math></b>	<b>2.43</b>	<b><math>4.11 \times 10^{-3}</math></b>	<b>2.91</b>	<b><math>9.97 \times 10^{-6}</math></b>	12	<b><math>1.95 \times 10^{-7}</math></b>	<b>7.13</b>	<b><math>1.43 \times 10^{-7}</math></b>	<b>7.36</b>	$7.45 \times 10^{-1}$	
Calcsols	$1.78 \times 10^{-1}$		$1.10 \times 10^{-1}$		<b><math>3.78 \times 10^{-2}</math></b>	<b>2.12</b>	$1.76 \times 10^{-1}$		$3.96 \times 10^{-1}$		$1.61 \times 10^{-1}$	
Fluvisols	$6.85 \times 10^{-2}$		<b><math>1.43 \times 10^{-3}</math></b>	<b>1.26</b>	<b><math>1.62 \times 10^{-4}</math></b>	<b>1.54</b>	<b><math>1.86 \times 10^{-3}</math></b>	<b>1.56</b>	$1.02 \times 10^{-1}$		$5.04 \times 10^{-1}$	
Gleysols	$5.10 \times 10^{-1}$		$1.35 \times 10^{-1}$		$2.97 \times 10^{-1}$		$6.73 \times 10^{-1}$		$7.26 \times 10^{-1}$		$7.36 \times 10^{-1}$	
Potential evapotranspiration	<b><math>6.71 \times 10^{-4}</math></b>	<b>2.87</b>	<b><math>2.57 \times 10^{-6}</math></b>	<b>2.95</b>	<b><math>3.49 \times 10^{-6}</math></b>	<b>4.66</b>	<b><math>5.98 \times 10^{-8}</math></b>	<b>3.89</b>	<b><math>6.54 \times 10^{-8}</math></b>	<b>4.96</b>	<b><math>2.19 \times 10^{-2}</math></b>	<b>1.36</b>
Precipitation	$2.61 \times 10^{-1}$		$3.13 \times 10^{-1}$		<b><math>6.88 \times 10^{-3}</math></b>	87.94	<b><math>5.72 \times 10^{-4}</math></b>	71.56	<b><math>1.43 \times 10^{-4}</math></b>	54.75	$5.99 \times 10^{-1}$	
Priestly–Taylor alpha coefficients	$8.57 \times 10^{-2}$		$5.24 \times 10^{-2}$		<b><math>1.02 \times 10^{-3}</math></b>	40.22	<b><math>1.56 \times 10^{-5}</math></b>	33.96	<b><math>1.62 \times 10^{-6}</math></b>	33.92	$1.48 \times 10^{-1}$	
Slope	<b><math>4.79 \times 10^{-5}</math></b>	<b>1.69</b>	<b><math>1.99 \times 10^{-5}</math></b>	<b>2.04</b>	<b><math>9.14 \times 10^{-6}</math></b>	<b>1.91</b>	<b><math>5.86 \times 10^{-6}</math></b>	<b>1.90</b>	<b><math>5.56 \times 10^{-4}</math></b>	<b>2.16</b>	$1.02 \times 10^{-1}$	
Soil and sedimentary deposit thickness	<b><math>3.00 \times 10^{-4}</math></b>	<b>1.66</b>	<b><math>1.00 \times 10^{-6}</math></b>	<b>2.07</b>	<b><math>2.10 \times 10^{-6}</math></b>	<b>2.41</b>	<b><math>3.74 \times 10^{-4}</math></b>	<b>2.11</b>	<b><math>4.11 \times 10^{-2}</math></b>	<b>1.88</b>	$5.15 \times 10^{-1}$	
Soil cation exchange capacity	$6.95 \times 10^{-2}$		<b><math>7.79 \times 10^{-3}</math></b>	<b>4.13</b>	<b><math>3.12 \times 10^{-2}</math></b>	<b>4.75</b>	$3.08 \times 10^{-1}$		$4.82 \times 10^{-1}$		$1.22 \times 10^{-1}$	
Soil organic carbon content	$6.50 \times 10^{-1}$		$1.90 \times 10^{-1}$		$2.76 \times 10^{-1}$		$7.43 \times 10^{-1}$		$7.38 \times 10^{-1}$		$1.66 \times 10^{-1}$	
Soil organic carbon density	$5.91 \times 10^{-1}$		$6.27 \times 10^{-1}$		$3.78 \times 10^{-1}$		$5.97 \times 10^{-1}$		$5.52 \times 10^{-1}$		$6.49 \times 10^{-1}$	
Soil organic carbon stock	$5.65 \times 10^{-1}$		$9.22 \times 10^{-2}$		$7.15 \times 10^{-2}$		$5.59 \times 10^{-1}$		$4.57 \times 10^{-1}$		$2.03 \times 10^{-1}$	
Soil pH	$1.11 \times 10^{-1}$		$1.88 \times 10^{-1}$		<b><math>3.54 \times 10^{-2}</math></b>	<b>2.46</b>	<b><math>7.03 \times 10^{-3}</math></b>	<b>2.36</b>	<b><math>8.03 \times 10^{-4}</math></b>	<b>2.69</b>	$2.79 \times 10^{-1}$	
Soil water capacity	<b><math>1.89 \times 10^{-2}</math></b>	<b>1.34</b>	<b><math>2.27 \times 10^{-3}</math></b>	<b>5.28</b>	<b><math>1.61 \times 10^{-2}</math></b>	<b>6.36</b>	$1.56 \times 10^{-1}$		$6.95 \times 10^{-1}$		$8.73 \times 10^{-2}$	
Solonchaks	$5.01 \times 10^{-1}$		$5.56 \times 10^{-1}$		$7.75 \times 10^{-1}$		$4.30 \times 10^{-1}$		<b><math>4.68 \times 10^{-2}</math></b>	<b>2.35</b>	$6.09 \times 10^{-1}$	
Temperature	<b><math>2.98 \times 10^{-3}</math></b>	<b>3.50</b>	<b><math>1.35 \times 10^{-5}</math></b>	<b>3.40</b>	<b><math>5.42 \times 10^{-4}</math></b>	<b>8.65</b>	<b><math>4.23 \times 10^{-4}</math></b>	<b>7.01</b>	<b><math>2.85 \times 10^{-2}</math></b>	<b>7.34</b>	<b><math>2.33 \times 10^{-2}</math></b>	<b>1.36</b>
Topographic wetness index	<b><math>2.45 \times 10^{-3}</math></b>	<b>2.29</b>	<b><math>1.19 \times 10^{-5}</math></b>	<b>2.55</b>	<b><math>9.52 \times 10^{-7}</math></b>	<b>2.46</b>	<b><math>4.20 \times 10^{-7}</math></b>	<b>2.76</b>	<b><math>1.39 \times 10^{-4}</math></b>	<b>3.10</b>	$4.65 \times 10^{-1}$	
Volume percentage of coarse fragments	$1.92 \times 10^{-1}$		<b><math>7.90 \times 10^{-3}</math></b>	<b>1.25</b>	<b><math>2.48 \times 10^{-3}</math></b>	<b>1.52</b>	<b><math>2.20 \times 10^{-2}</math></b>	<b>1.49</b>	<b><math>4.51 \times 10^{-2}</math></b>	<b>1.50</b>	$1.96 \times 10^{-1}$	
Water table depth	<b><math>1.50 \times 10^{-2}</math></b>	<b>1.76</b>	<b><math>6.02 \times 10^{-3}</math></b>	<b>1.76</b>	<b><math>3.15 \times 10^{-3}</math></b>	<b>1.83</b>	<b><math>2.30 \times 10^{-2}</math></b>	<b>1.90</b>	$1.70 \times 10^{-1}$		$2.46 \times 10^{-1}$	
Weight percentage of clay particles	$7.88 \times 10^{-2}$		<b><math>1.97 \times 10^{-2}</math></b>	<b>4.55</b>	<b><math>3.11 \times 10^{-2}</math></b>	<b>5.46</b>	$2.41 \times 10^{-1}$		$7.11 \times 10^{-1}$		$1.49 \times 10^{-1}$	
Weight percentage of sand particles	$2.22 \times 10^{-1}$		<b><math>2.77 \times 10^{-2}</math></b>	<b>4.70</b>	<b><math>1.59 \times 10^{-2}</math></b>	<b>5.75</b>	$1.11 \times 10^{-1}$		$6.19 \times 10^{-1}$		$1.73 \times 10^{-1}$	
Weight percentage of silt particles	$5.51 \times 10^{-1}$		$6.63 \times 10^{-1}$		$2.67 \times 10^{-1}$		$2.01 \times 10^{-1}$		$4.49 \times 10^{-1}$		$7.06 \times 10^{-1}$	
Acid plutonic rocks	$4.10 \times 10^{-1}$		$1.95 \times 10^{-1}$		$1.23 \times 10^{-1}$		<b><math>4.50 \times 10^{-2}</math></b>		$1.27 \times 10^{-1}$		$3.65 \times 10^{-1}$	
Basic volcanic rocks	<b><math>3.03 \times 10^{-2}</math></b>		<b><math>6.02 \times 10^{-3}</math></b>		<b><math>1.39 \times 10^{-2}</math></b>		<b><math>2.46 \times 10^{-2}</math></b>		$3.57 \times 10^{-1}$		$6.77 \times 10^{-1}$	
Metamorphic rocks	$5.67 \times 10^{-1}$		$5.37 \times 10^{-1}$		$3.06 \times 10^{-1}$		$2.75 \times 10^{-1}$		$3.30 \times 10^{-1}$		$3.38 \times 10^{-1}$	
Sedimentary rocks	<b><math>2.89 \times 10^{-2}</math></b>		<b><math>1.28 \times 10^{-4}</math></b>		<b><math>2.24 \times 10^{-3}</math></b>		<b><math>2.86 \times 10^{-3}</math></b>		$1.03 \times 10^{-1}$		$6.58 \times 10^{-1}$	

Bold shows where *p* values are less than 0.05 and variance inflation factor (VIF) values are less than 10

**Table 2** Variable combinations and AUC values of final models in 1000 logistic regressions using thresholds of 1–10 µg/L for groundwater arsenic

No.	Threshold (µg/L)	Variable combination	AUC value
1	10	Potential evapotranspiration, slope, soil water capacity	0.83
2	5	Soil and sedimentary deposit thickness, potential evapotranspiration, slope	0.79
3	4	Topographic wetness index, potential evapotranspiration, fluvisols	0.77
4	3	Aridity, fluvisols, temperature	0.76
5	2	Aridity, temperature	0.71
6	1	Evapotranspiration	0.60

arsenic from groundwater systems. The gentle slope facilitates the accumulation of abundant organic matter within floodplains and alluvial deposits, arsenic-bearing Fe-oxyhydroxide minerals, and finer sediments (Shamsudduha and Uddin 2007; Shamsudduha et al. 2009). Then, arsenic is released into groundwaters by microbial activities, resulting in groundwater arsenic occurring in flat, low-lying areas where groundwater flows are sluggish; such areas include low-lying deltaic and floodplain areas (Shamsudduha et al. 2009).

#### Hazard maps

The probabilities of arsenic concentration exceeding 10 µg/L, 5 µg/L, 4 µg/L, 3 µg/L and 2 µg/L were calculated by the 5 final models, and the probability maps of arsenic concentrations are shown in Figs. 3 and S4. The cut-offs where sensitivity equals specificity in the respective 5 final models (shown in Figs. 2 and S3) were 0.69 (10 µg/L), 0.66 (5 µg/L), 0.61 (4 µg/L), 0.57 (3 µg/L) and 0.50 (2 µg/L), which were used to create maps of the occurrence of arsenic concentration exceeding each of the thresholds. Figure 4 contains the pseudo-contour map of various concentrations of groundwater arsenic, combined from the individual hazard map of each of the thresholds. Of the 26 districts of Gujarat State as defined by the 2011 Indian Census (Chandramouli 2011), our map predicts that groundwater arsenic exceeds 10 µg/L in the northwest, northeast and south-east parts of Kachchh district and the north-western and south-western part of Banas Kantha district. In comparison, a pseudo-contour map of groundwater arsenic determined using a fixed cut-off of 0.50

indicates more widely varying higher concentrations (Fig. S6).

The pseudo-contour map of arsenic concentrations (Fig. 4) shows a similar spatial pattern to the distribution map of soil organic carbon content (Fig. S7), which is not one of the predictor variables. Dissolved organic matter is the main driver of microbe-mediated reductive dissolution of arsenic-bearing Fe-oxyhydroxide (Fendorf et al. 2010). Other processes, including complexation of arsenic by dissolved humic substances, competitive sorption and electron shuttling reactions mediated by humic substances may also influence arsenic mobility in groundwaters (Guo et al. 2011; Mladenov et al. 2015). The amount and availability of organic carbon in sediments and soil affect the spatial variability of groundwater arsenic concentrations (McArthur et al. 2004; McArthur et al. 2011).

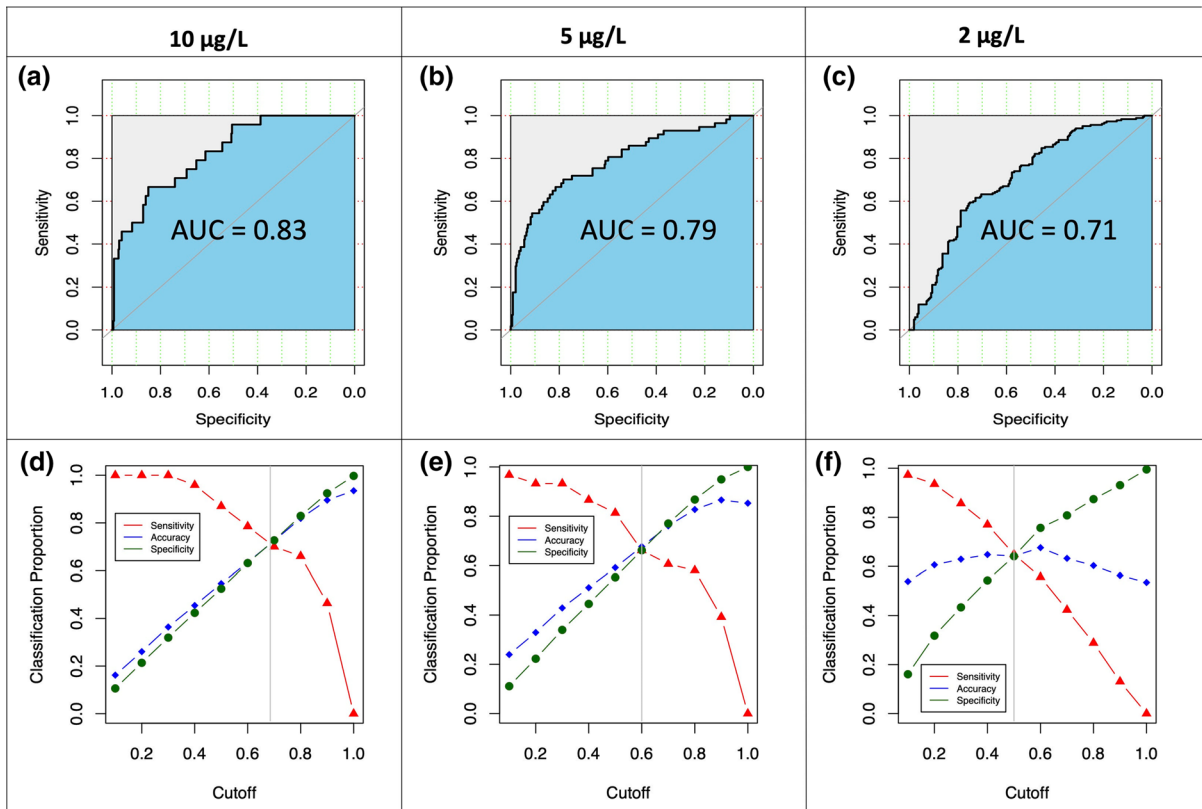
#### Potential exposure map

We combined the pseudo-contour map of varying arsenic concentrations with projected 2020 population density (Pages et al. 2018) to produce a potential exposure map showing the population living in areas with different groundwater arsenic concentrations (Fig. 5). Of a projected total population of Gujarat of 70,445,000 (Chandramouli 2011), approximately 122,000 (i.e. about 0.17% of total Gujarat population) live in areas where groundwater arsenic concentrations exceed 10 µg/L. The number of people living in areas with other groundwater arsenic concentrations is summarized in Table 4. In Gujarat State, only a low percentage of people (0.07%) were exposed to high arsenic groundwaters, and most people are likely to be exposed to low groundwater concentrations of arsenic.



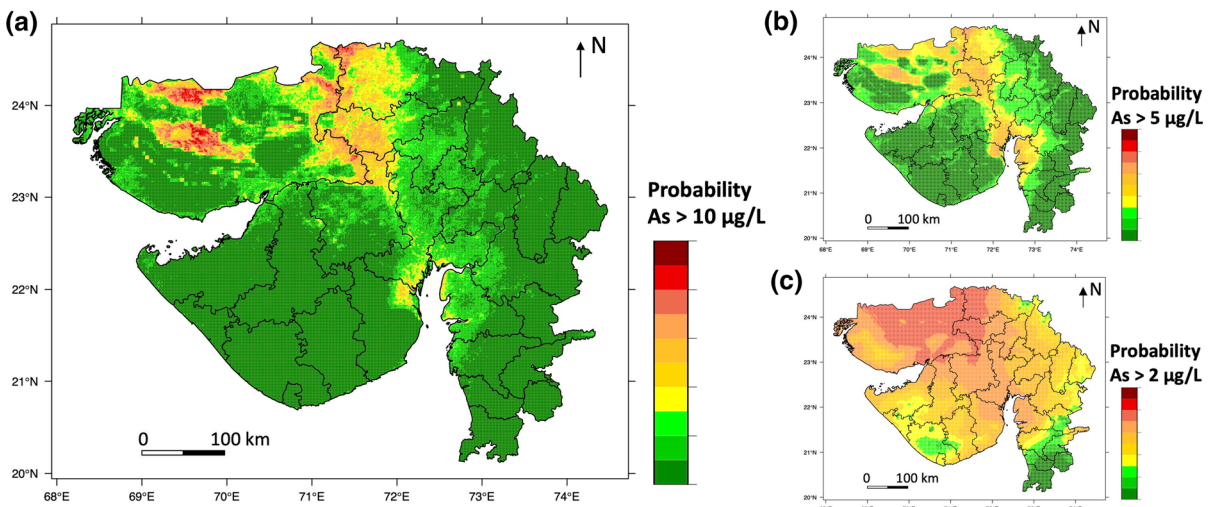
**Table 3** Intercepts, coefficients, and their standard deviations of normalized predictor variables of final models in 1000 logistic regression runs using thresholds of 1–10 µg/L for groundwater arsenic

Variables	Thresholds											
	10 µg/L		5 µg/L		4 µg/L		3 µg/L		2 µg/L		1 µg/L	
	Coefficient	Standard deviation	Coefficient	Standard deviation	Coefficient	Standard deviation	Coefficient	Standard deviation	Coefficient	Standard deviation	Coefficient	Standard deviation
Intercept	-1.75	0.36	-3.36	0.10	-7.94	0.52	-2.14	0.21	-0.22	0.13	0.34	0.07
Aridity							-6.99	0.72	-5.99	0.49		
Fluvisols					2.58	2.79	1.97	0.11				
Potential evapotranspiration	3.47	0.39	2.67	0.22	2.76	0.33					1.68	0.15
Slope	-25.19	3.6	-10.51	0.21								
Soil and sedimentary deposit thickness			1.18	0.05								
Soil water capacity	-3.55	0.27										
Temperature							2.94	0.2	1.95	0.23		
Topographic wetness index					4.83	0.58						

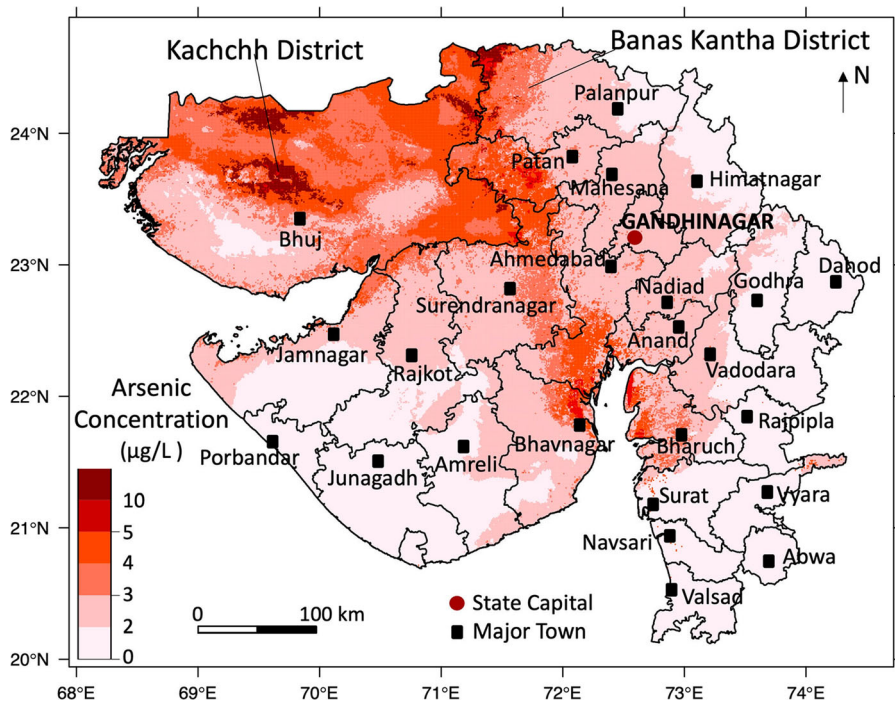


**Fig. 2** ROC curves of final logistic regression models with **a** 10 µg/L, **b** 5 µg/L, and **c** 2 µg/L as thresholds for groundwater arsenic in Gujarat State, India. Plots of sensitivity (true-positive rate), specificity (true-negative rate) and accuracy against cut-

offs of the final logistic regression models with **d** 10 µg/L, **e** 5 µg/L, and **f** 2 µg/L as thresholds for CGWB (2016) dataset for groundwater arsenic in Gujarat State, India



**Fig. 3** Hazard maps showing the probability of the geospatially modelled occurrences of groundwater arsenic concentration exceeding thresholds of **a** 10 µg/L, **b** 5 µg/L, and **c** 2 µg/L in Gujarat State, India



**Fig. 4** Pseudo-contour map of geospatially modelled groundwater arsenic hazard distribution in Gujarat. Contour boundaries surround the regions in which the modelled probability of groundwater arsenic exceeding the contour value is equal to the

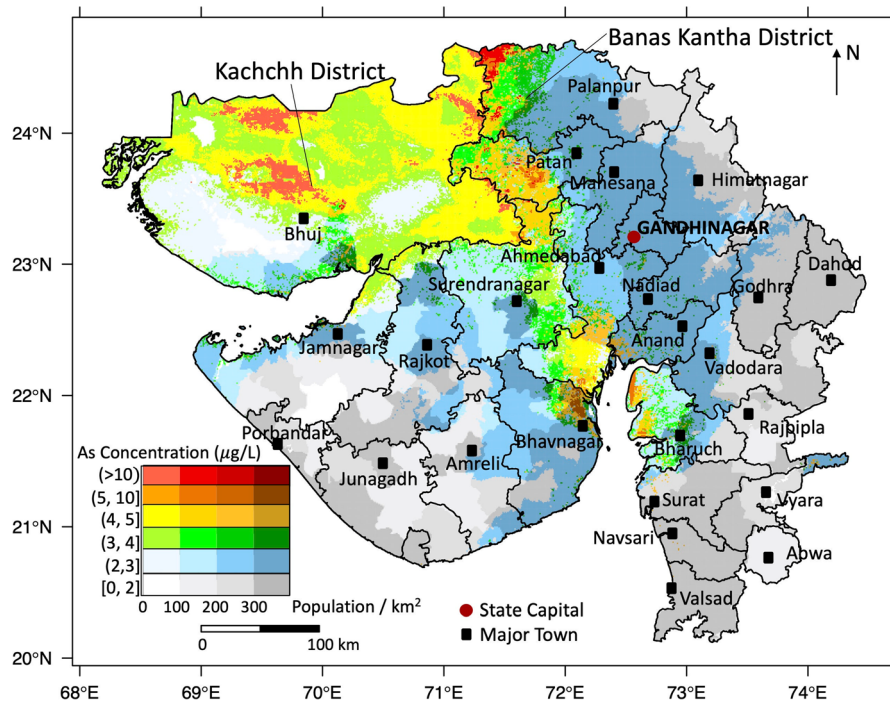
cut-off value for that concentration (being 0.5 for As = 2 µg/L; 0.57 for As = 3 µg/L; 0.61 for As = 4 µg/L; 0.66 for As = 5 µg/L; 0.69 for As = 10 µg/L)

However, many studies (Medrano et al. 2010; Moon et al. 2017; Polya et al. 2019b; Ahmad et al. 2020) pointed out that low concentrations of arsenic also pose health risks to humans, although the harm is not as serious as that arising from higher arsenic concentrations in drinking water.

Accounting for 48% of rural household water supplied being through hand pumps and tube wells (The World Bank 2006) and 29% of urban households using untreated taps, bore wells, hand pumps and wells as water supply infrastructure (IIHS 2014), we estimate that approximately 49,000 people in Gujarat are exposed to elevated arsenic contamination (> 10 µg/L) through domestic consumption of groundwater. The population exposed to other arsenic concentrations in groundwaters is summarized in Table 4.

Health effects of exposure to groundwater arsenic in Gujarat (Table 5) estimated using dose response functions of arsenic-induced cancers (Brown et al. 1989; NRC 1999, 2001; Yu et al. 2003) include a prevalence of 670 cases of skin cancer arising from

exposure to groundwater arsenic in Gujarat. However, in Gujarat, groundwater arsenic does not significantly contribute to internal cancers (lung cancer, bladder cancer, liver cancer) with a combined modelled incidence of only 12 cases—corresponding to just 0.001% of cancer-related fatalities in Gujarat. The low number of cancer cases modelled to be caused by groundwater arsenic reflects the relative low groundwater arsenic hazard in Gujarat State. These results are similar to those estimated by Yu et al. (2003) for low groundwater arsenic areas in Bangladesh (viz. Brahmaputra FP (Chandina regions), the Chittagong Coast (sandstone/shale regions), and the Terraces West/East (clays and alluvium regions) where mean groundwater arsenic concentrations are in the range of 1–6 µg/L. Notwithstanding this, there are method model and parameter uncertainties in the dose–response relations used and these warrant further investigation in order to obtain more accurate estimates of arsenic attributable health outcomes.



**Fig. 5** Population density (persons per 1 km<sup>2</sup>) co-plotted with modelled groundwater arsenic concentrations in Gujarat State, India. Cities are shown for illustrative purposes only. The proportion of people utilizing untreated groundwater for

drinking purpose differs substantially between urban and rural areas, so this map should not be utilized as an exposure map without appropriate correction for groundwater usage

**Table 4** Estimated population exposed to various groundwater arsenic concentration in Gujarat

Arsenic concentrations (µg/L)	Population living in areas with indicated range of groundwater arsenic concentrations	Population exposed to indicated range of groundwater arsenic concentrations
> 10	122,000	49,000
(5, 10)	206,000	82,000
(4, 5)	1,773,000	708,000
(3, 4)	5,011,000	2,000,000
(2, 3)	32,981,000	13,162,000
(0, 2)	30,351,000	12,113,000
Total	70,444,000	28,114,000

## Implications

The groundwater arsenic hazard and potential exposure maps for Gujarat produced in our study facilitate the calculation of the spatial distribution of groundwater arsenic attributable health outcomes. The predictive maps generated in this paper have high resolution and so provide a means of interpolating

existing data (CGWB 2016), thereby providing value added to such existing datasets. Our arsenic distribution map, potential exposure map and associated health risk estimation of population present an obvious improvement in the rendering of both the detailed distribution of different arsenic concentrations in groundwaters and in estimates of the number of people potentially affected in Gujarat State. Our

**Table 5** Modelled health effects in exposure to groundwater arsenic hazard in Gujarat: prevalence of skin cancer and incidence of internal cancers (lung cancer, bladder cancer, liver cancer) for males and females in rural and urban areas

	Rural areas		Urban areas		Total
	Males	Females	Males	Females	
Prevalence of skin cancer (number of persons)	380	80	180	30	670
Incidence of internal cancers (fatalities per year)					
Lung cancer	8	0	4	0	12
Bladder cancer	0	0	0	0	0
Liver cancer	0	0	0	0	0

models also provide a basis for applying to other parts of India and globally, particularly useful for estimating populations at risk of exposure to different levels of hazard. Notwithstanding the utility of these models, we note that they are not intended to be an authoritative indicator of the quality of individual groundwater sourced tube-well water. Accordingly, these findings should be used with caution. In particular, the well-known significant local scale spatial heterogeneity in arsenic in groundwater indicates that wells should be tested individually in order to obtain the most robust assessment of groundwater arsenic hazard.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Funding** This work was financially supported by the United Kingdom’s Engineering and Physical Sciences Research Council (IAA Impact Award) and the Swiss Agency for Development and Cooperation (project no. 7F-09963.01.01).

**References**

Ahmad, A., van der Wens, P., Baken, K., de Waal, L., Bhattacharya, P., & Stuyfzand, P. (2020). Arsenic reduction to < 1 µg/L in Dutch drinking water. *Environment*

*International*, 134, 105253–105262. <https://doi.org/10.1016/j.envint.2019.105253>.

Ahmed, K. M., Bhattacharya, P., Hasan, M. A., Akhter, S. H., Alam, S. M., Bhuyian, M. H., et al. (2004). Arsenic enrichment in groundwater of the alluvial aquifers in Bangladesh: An overview. *Applied Geochemistry*, 19(2), 181–200. <https://doi.org/10.1016/j.apgeochem.2003.09.006>.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control Mladenov*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.

Alarcón-Herrera, M. T., Bundschuh, J., Nath, B., Nicolli, H. B., Gutierrez, M., Reyes-Gomez, V. M., et al. (2013). Co-occurrence of arsenic and fluoride in groundwater of semi-arid regions in Latin America: Genesis, mobility and remediation. *Journal of Hazardous Materials*, 262, 960–969. <https://doi.org/10.1016/j.jhazmat.2012.08.005>.

Amini, M., Abbaspour, K. C., Berg, M., Winkel, L., Hug, S. J., Hoehn, E., et al. (2008). Statistical modeling of global geogenic arsenic contamination in groundwater. *Environmental Science and Technology*, 42(10), 3669–3675. <https://doi.org/10.1021/es702859e>.

Argos, M., Kalra, T., Rathouz, P. J., Chen, Y., Pierce, B., Parvez, F., et al. (2010). Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): A prospective cohort study. *The Lancet*, 376(9737), 252–258. [https://doi.org/10.1016/S0140-6736\(10\)60481-3](https://doi.org/10.1016/S0140-6736(10)60481-3).

Ayotte, J. D., Medalie, L., Qi, S. L., Backer, L. C., & Nolan, B. T. (2017). Estimating the high-arsenic domestic-well population in the conterminous United States. *Environmental Science and Technology*, 51(21), 12443–12454. <https://doi.org/10.1021/acs.est.7b02881>.

Berg, M., Stengel, C., Trang, P. T. K., Viet, P. H., Sampson, M. L., Leng, M., et al. (2007). Magnitude of arsenic pollution in the Mekong and Red River Deltas—Cambodia and Vietnam. *Science of the Total Environment*, 372(2–3), 413–425. <https://doi.org/10.1016/j.scitotenv.2006.09.010>.

Bhattacharya, P., Polya, D., & Jovanovic, D. (Eds.). (2017). *Best practice guide on the control of arsenic in drinking water*. London, UK: IWA Publishing. <https://doi.org/10.2166/9781780404929>.

Bretzler, A., & Johnson, C. (2015). Geogenic contamination handbook—Addressing arsenic and fluoride in drinking

- water. *Applied Geochemistry*, 63, 642–646. <https://doi.org/10.1016/j.apgeochem.2015.08.016>.
- Bretzler, A., Lalanne, F., Nikiema, J., Podgorski, J., Pfenninger, N., Berg, M., et al. (2017). Groundwater arsenic contamination in Burkina Faso, West Africa: Predicting and verifying regions at risk. *Science of the Total Environment*, 584–585, 958–970. <https://doi.org/10.1016/j.scitotenv.2017.01.147>.
- Brown, K. G., Boyle, K. E., Chen, C. W., & Gibb, H. J. (1989). A dose-response analysis of skin cancer from inorganic arsenic in drinking water. *Risk Analysis*, 9(4), 519–528. <https://doi.org/10.1111/j.1539-6924.1989.tb01263.x>.
- Buragohain, M., & Sarma, H. P. (2012). A study on spatial distribution of arsenic in ground water samples of Dhemaji district of Assam, India by using arc view GIS software. *Scientific Reviews and Chemical Communications*, 2(1), 7–11.
- CGWB. (2016). *Groundwater Year Book—2015–2016 Gujarat state and UT of Daman & Diu*. India Central Ground Water Board. Resource document. Retrieved May 2, 2019, from <http://cgwb.gov.in/Regions/GW-year-Books/GWYB-2015-16/GWYB%20WCR%202015-16.pdf>.
- Chakraborti, D., Mukherjee, S. C., Pati, S., Sengupta, M. K., Rahman, M. M., Chowdhury, U. K., et al. (2003). Arsenic groundwater contamination in Middle Ganga Plain, Bihar, India: A future danger? *Environmental Health Perspectives*, 111(9), 1194–1201. <https://doi.org/10.1289/ehp.5966>.
- Chakraborti, D., Rahman, M. M., Das, B., Nayak, B., Pal, A., Sengupta, M. K., et al. (2013). Groundwater arsenic contamination in Ganga–Meghna–Brahmaputra plain, its health effects and an approach for mitigation. *Environmental Earth Sciences*, 70(5), 1993–2008. <https://doi.org/10.1007/s12665-013-2699-y>.
- Chakraborti, D., Sengupta, M. K., Rahman, M. M., Ahamed, S., Chowdhury, U. K., Hossain, A., et al. (2004). Groundwater arsenic contamination and its health effects in the Ganga–Meghna–Brahmaputra plain. *Journal of Environmental Monitoring*, 6(6), 74–83. <http://search.proquest.com/docview/66681706>.
- Chandramouli, C. (2011). *Census of India 2011*. Office of Registrar General & Census Commissioner, India. Retrieved October 18, 2019, from <https://www.census2011.co.in/>.
- Charlet, L., & Polya, D. A. (2006). Arsenic in shallow, reducing groundwaters in southern Asia: An environmental health disaster. *Elements*, 2(2), 91–96. <https://doi.org/10.2113/gselements.2.2.91>.
- Chatterjee, A., Das, D., Mandal, B. K., Chowdhury, T. R., Samanta, G., & Chakraborti, D. (1995). Arsenic in ground water in six districts of West Bengal, India: The biggest arsenic calamity in the world. Part I. Arsenic species in drinking water and urine of the affected people. *Analyst*, 120, 643–650. <https://doi.org/10.1039/AN9952000643>.
- Chen, Y., & Ahsan, H. (2004). Cancer burden from arsenic in drinking water in Bangladesh. *American Journal of Public Health*, 94(5), 741–744. <https://doi.org/10.2105/AJPH.94.5.741>.
- Chowdhury, T. R., Basu, G. K., Mandal, B. K., Biswas, B. K., Samanta, G., Chowdhury, U. K., et al. (1999). Arsenic poisoning in the Ganges delta. *Nature*, 401(6753), 545–546. <https://doi-org.manchester.idm.oclc.org/10.1038/44056>.
- Chowdhury, U. K., Biswas, B. K., Chowdhury, T. R., Samanta, G., Mandal, B. K., Basu, G. C., et al. (2000). Groundwater arsenic contamination in Bangladesh and West Bengal, India. *Environmental Health Perspectives*, 108(5), 393–397. <https://doi.org/10.1289/ehp.00108393>.
- Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, 339(6122), 940–943. <https://doi.org/10.1126/science.1229881>.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Fendorf, S., Michael, H. A., & van Geen, A. (2010). Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science*, 328(5982), 1123–1127. <https://doi.org/10.1126/science.1172974>.
- Flanagan, S. V., Johnston, R. B., & Zheng, Y. (2012). Arsenic in tube well water in Bangladesh: Health and economic impacts and implications for arsenic mitigation. *Bulletin of the World Health Organization*, 90, 839–846. <https://doi.org/10.2471/BLT.11.101253>.
- Franke, G. R. (2010). Multicollinearity. *Wiley international encyclopedia of marketing* (pp. 197–198). Chichester, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781444316568.wiem02066>.
- García-Esquinas, E., Pollán, M., Umans, J. G., Francesconi, K. A., Goessler, W., Guallar, E., et al. (2013). Arsenic exposure and cancer mortality in a US-based prospective cohort: The strong heart study. *Cancer Epidemiology and Prevention Biomarkers*, 22(11), 1944–1953. <https://doi.org/10.1158/1055-9965.EPI-13-0234-T>.
- Ghosh, M., Pal, D. K., & Santra, S. C. (2019). Spatial mapping and modeling of arsenic contamination of groundwater and risk assessment through geospatial interpolation technique. *Environment, Development and Sustainability*, 22, 2861–2880. <https://doi.org/10.1007/s10668-019-00322-7>.
- Ghosh, A. K., Sarkar, D., Dutta, D., & Bhattacharyya, P. (2004). Spatial variability and concentration of arsenic in the groundwater of a region in Nadia district, West Bengal, India. *Archives of Agronomy and Soil Science*, 50(4–5), 521–527. <https://doi.org/10.1080/0365034042000220757>.
- Guo, H., Zhang, B., Li, Y., Berner, Z., Tang, X., Norra, S., et al. (2011). Hydrogeological and biogeochemical constrains of arsenic mobilization in shallow aquifers from the Hetao basin, Inner Mongolia. *Environmental Pollution*, 159(4), 876–883. <https://doi.org/10.1016/j.envpol.2010.12.029>.
- Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*. <https://doi.org/10.1029/2012GC004370>.
- Hengl, T. (2018). Global DEM derivatives at 250 m, 1 km and 2 km based on the MERIT DEM. *Zenodo*. <https://doi.org/10.5281/zenodo.1447210>.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>.

- Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, New Jersey: Wiley.
- IHS. (2014). *Sustaining policy momentum: Urban water supply & sanitation in India*. Bangalore: Indian Institute for Human Settlements. <https://doi.org/10.24943/ihsrffpps6.2014>.
- Islam, F. S., Gault, A. G., Boothman, C., Polya, D. A., Charnock, J. M., Chatterjee, D., et al. (2004). Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature*, 430(6995), 68–71. <https://doi.org/10.1038/nature02638>.
- ISRIC. (2017). *SoilGrids—Global gridded soil information*. SRIC—World Soil Information. Retrieved July 3, 2019, from <https://www.isric.org/explore/soilgrids>.
- IUSS. (2015). World Reference Base for Soil Resources 2014, Updated 2015. *World Soil Resources Reports 106*. Food and Agriculture Organization of the United Nations. Resource document. Retrieved September 10, 2019, from <http://www.fao.org/3/i3794en/I3794en.pdf>.
- Kinniburgh, D. G., & Smedley, P. L. (2001). *Arsenic contamination of groundwater in Bangladesh, vol. 2: final report*. British Geological Survey and Department of Public Health Engineering. Retrieved November 10, 2018, from <https://www.bgs.ac.uk/research/groundwater/health/arsenic/Bangladesh/reports.html>.
- Lado, L. R., Polya, D., Winkel, L., Berg, M., & Hegan, A. (2008). Modelling arsenic hazard in Cambodia: A geostatistical approach using ancillary data. *Applied Geochemistry*, 23(11), 3010–3018. <https://doi.org/10.1016/j.apgeochem.2008.06.028>.
- McArthur, J. M., Banerjee, D. M., Hudson-Edwards, K. A., Mishra, R., Purohit, R., Ravenscroft, P., et al. (2004). Natural organic matter in sedimentary basins and its relation to arsenic in anoxic ground water: The example of West Bengal and its worldwide implications. *Applied Geochemistry*, 19(8), 1255–1293. <https://doi.org/10.1016/j.apgeochem.2004.02.001>.
- McArthur, J. M., Nath, B., Banerjee, D. M., Purohit, R., & Grassineau, N. (2011). Palaeosol control on groundwater flow and pollutant distribution: The example of arsenic. *Environmental Science and Technology*, 45(4), 1376–1383. <https://doi.org/10.1021/es1032376>.
- McArthur, J. M., Ravenscroft, P., Safiulla, S., & Thirlwall, M. F. (2001). Arsenic in groundwater: Testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Water Resources Research*, 37(1), 109–117. <https://doi.org/10.1029/2000WR900270>.
- Medrano, M. J., Boix, R., Pastor-Barriuso, R., Palau, M., Damián, J., Ramis, R., et al. (2010). Arsenic in public water supplies and cardiovascular mortality in Spain. *Environmental Research*, 110(5), 448–454. <https://doi.org/10.1016/j.envres.2009.10.002>.
- Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, 13(3), 253–267. <https://doi.org/10.1016/j.envres.2009.10.002>.
- Mladenov, N., Zheng, Y., Simone, B., Bilinski, T. M., McKnight, D. M., Nemerget, D., et al. (2015). Dissolved organic matter quality in a shallow aquifer of Bangladesh: Implications for arsenic mobility. *Environmental Science and Technology*, 49(18), 10815–10824. <https://doi.org/10.1021/acs.est.5b01962>.
- Monrad, M., Ersbøll, A. K., Sørensen, M., Baastrup, R., Hansen, B., Gammelmark, A., et al. (2017). Low-level arsenic in drinking water and risk of incident myocardial infarction: A cohort study. *Environmental Research*, 154, 318–324. <https://doi.org/10.1016/j.envres.2017.01.028>.
- Moon, K. A., Oberoi, S., Barchowsky, A., Chen, Y., Guallar, E., Nachman, K. E., et al. (2017). A dose-response meta-analysis of chronic arsenic exposure and incident cardiovascular disease. *International Journal of Epidemiology*, 46(6), 1924–1939. <https://doi.org/10.1093/ije/dyx202>.
- Mukherjee, A., Gupta, S., Coomar, P., Fryar, A. E., Guillot, S., Verma, S., et al. (2019). Plate tectonics influence on geogenic arsenic cycling: From primary sources to global groundwater enrichment. *Science of the Total Environment*, 683, 793–807. <https://doi.org/10.1016/j.scitotenv.2019.04.255>.
- Nickson, R. T., McArthur, J. M., Shrestha, B., Kyaw-Myint, T. O., & Lowry, D. (2005). Arsenic and other drinking water quality issues, Muzaffargarh District, Pakistan. *Applied Geochemistry*, 20(1), 55–68. <https://doi.org/10.1016/j.apgeochem.2004.06.004>.
- NRC. (1999). *Arsenic in drinking water*. Washington, DC: National Academy Press. <https://doi.org/10.17226/6444>.
- NRC. (2001). *Arsenic in drinking water: 2001 update*. Washington, DC: National Academy Press. <https://doi.org/10.17226/10194>.
- Pages, W., Gallery, M., & Viewer, M. (2018) *Gridded population of the world (GPW), v4*. Socioeconomic Data and Applications Center. Retrieved October 24, 2019, from <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>
- Pearce, J., & Ferrier, S. (2000). An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 128(2–3), 127–147. [https://doi.org/10.1016/S0304-3800\(99\)00227-6](https://doi.org/10.1016/S0304-3800(99)00227-6).
- Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., et al. (2016). *Global 1-km gridded thickness of soil, regolith, and sedimentary deposit layers*. Distribution Active Archive Center for Biogeochemical Dynamics. Retrieved July 3, 2019, from [https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds\\_id=1304](https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds_id=1304).
- Podgorski, J. E., Eqani, S. A. M. A. S., Khanam, T., Ullah, R., Shen, H., & Berg, M. (2017). Extensive arsenic contamination in high-pH unconfined aquifers in the Indus Valley. *Science Advances*, 3(8), e1700935. <https://doi.org/10.1126/sciadv.1700935>.
- Podgorski, J. E., Labhasetwar, P., Saha, D., & Berg, M. (2018). Prediction modeling and mapping of groundwater fluoride contamination throughout India. *Environmental Science and Technology*, 52(17), 9889–9898. <https://doi.org/10.1021/acs.est.8b01679>.
- Polya, D. A., & Charlet, L. (2009). Environmental science: Rising arsenic risk? *Nature Geoscience*, 2(6), 383–384. <https://doi.org/10.1038/ngeo537>.
- Polya, D. A., & Middleton, D. R. (2017). Arsenic in drinking water: Sources & human exposure. *Best practice guide on the control of arsenic in drinking water* (pp. 1–23). London, UK: IWA Publishing. <https://doi.org/10.2166/9781780404929>.

- Polya, D. A., Sparrenbom, C., Datta, S., & Guo, H. (2019a). Groundwater arsenic biogeochemistry-key questions & use of tracers to understand arsenic-prone groundwater systems. *Geoscience Frontiers*, *10*(5), 1635–1641. <https://doi.org/10.1016/j.gsf.2019.05.004>.
- Polya, D. A., Xu, L., Launder, J., Goody, D. C., & Ascott, M. (2019b). Distribution of arsenic hazard in public water supplies in the United Kingdom—methods, implications for health risks and recommendations. *Environmental arsenic in a changing world* (p. 22). Boca Raton: CRC Press. <https://doi.org/10.1201/9781351046633-8>.
- Rahman, M. A., & Hasegawa, H. (2011). High levels of inorganic arsenic in rice in areas where arsenic-contaminated water is used for irrigation and cooking. *Science of the Total Environment*, *409*(22), 4645–4655. <https://doi.org/10.1016/j.scitotenv.2011.07.068>.
- Ravenscroft, P., Brammer, H., & Richards, K. (2009). *Arsenic pollution: A global synthesis*. Chichester, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781444308785>.
- Rodríguez-Lado, L., Sun, G., Berg, M., Zhang, Q., Xue, H., Zheng, Q., et al. (2013). Groundwater arsenic contamination throughout China. *Science*, *341*(6148), 866–868. <https://doi.org/10.1126/science.1237484>.
- Rowland, H. A., Omeregic, E. O., Millot, R., Jimenez, C., Mertens, J., Baciu, C., et al. (2011). Geochemistry and arsenic behaviour in groundwater resources of the Pannonian Basin (Hungary and Romania). *Applied Geochemistry*, *26*(1), 1–17. <https://doi.org/10.1016/j.apgeochem.2010.10.006>.
- Rowland, H. A. L., Pederick, R. L., Polya, D. A., Pancost, R. D., Van Dongen, B. E., Gault, A. G., et al. (2007). The control of organic matter on microbially mediated iron reduction and arsenic release in shallow alluvial aquifers, Cambodia. *Geobiology*, *5*(3), 281–292. <https://doi.org/10.1111/j.1472-4669.2007.00100.x>.
- Shamsudduha, M., Marzen, L. J., Uddin, A., Lee, M. K., & Saunders, J. A. (2009). Spatial relationship of groundwater arsenic distribution with regional topography and water-table fluctuations in the shallow aquifers in Bangladesh. *Environmental Geology*, *57*(7), 1521. <https://doi.org/10.1007/s00254-008-1429-3>.
- Shamsudduha, M., & Uddin, A. (2007). Quaternary shoreline shifting and hydrogeologic influence on the distribution of groundwater arsenic in aquifers of the Bengal Basin. *Journal of Asian Earth Sciences*, *31*(2), 177–194. <https://doi.org/10.1016/j.jseas.2007.07.001>.
- Sharma, K. D., & Kumar, S. (2008). Hydrogeological research in India in managing water resources. *Groundwater dynamics in hard rock aquifers: Sustainable management and optimal monitoring network design* (pp. 1–19). Dordrecht, Netherlands: Springer. [https://doi.org/10.1007/978-1-4020-6540-8\\_1](https://doi.org/10.1007/978-1-4020-6540-8_1).
- Smedley, P. L., & Kinniburgh, D. G. (2002). A review of the source, behaviour and distribution of arsenic in natural waters. *Applied Geochemistry*, *17*(5), 517–568. [https://doi.org/10.1016/S0883-2927\(02\)00018-5](https://doi.org/10.1016/S0883-2927(02)00018-5).
- Smith, A. H., Lingas, E. O., & Rahman, M. (2000). Contamination of drinking-water by arsenic in Bangladesh: A public health emergency. *Bulletin of the World Health Organization*, *78*, 1093–1103. <http://search.proquest.com/docview/57789950/>.
- Sovann, C., & Polya, D. A. (2014). Improved groundwater geogenic arsenic hazard map for Cambodia. *Environmental Chemistry*, *11*(5), 595–607. <http://search.proquest.com/docview/1620244406/>.
- The World Bank. (2006). *India: Water supply and sanitation—Bridging the gap between infrastructure and service*. Washington, DC: The World Bank. Retrieved October 27, 2019, from <http://documents.worldbank.org/curated/en/134931468770505493/India-water-supply-and-sanitation-bridging-the-gap-between-infrastructure-and-service>.
- The World Bank. (2017). *World slope model in degrees*. Washington, DC: The World Bank. Retrieved from July 10, 2019, from <https://datacatalog.worldbank.org/>.
- Thornton, I., & Farago, M. (1997). The geochemistry of arsenic. *Arsenic* (pp. 1–16). Dordrecht, Netherlands: Springer. [https://doi.org/10.1007/978-94-011-5864-0\\_1](https://doi.org/10.1007/978-94-011-5864-0_1).
- Trabucco, A., & Zomer, R. J. (2009). Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. CGIAR Consortium for Spatial Information. Retrieved July 3, 2019, from <http://www.cgiar-csi.org>.
- Trabucco, A., & Zomer, R. (2010). Global soil water balance geospatial database. CGIAR Consortium for Spatial Information. Retrieved July 3, 2019, from <http://www.cgiar-csi.org>.
- van Geen, A. (2008). Environmental science: Arsenic meets dense populations. *Nature Geoscience*, *1*(8), 494–496. <https://doi.org/10.1038/ngeo268>.
- WHO/UNICEF. (2018). *Arsenic Primer—Guidance on the investigation & mitigation of arsenic contamination*, 3rd ed. New York, USA: WHO/UNICEF. Resource document. Retrieved November 1, 2019, from [https://www.unicef.org/wash/files/UNICEF\\_WHO\\_Arsenic\\_Primer.pdf](https://www.unicef.org/wash/files/UNICEF_WHO_Arsenic_Primer.pdf).
- Winkel, L., Berg, M., Amini, M., Hug, S. J., & Johnson, C. A. (2008). Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. *Nature Geoscience*, *1*(8), 536–542. <https://doi.org/10.1038/ngeo254>.
- Yu, W. H., Harvey, C. M., & Harvey, C. F. (2003). Arsenic in groundwater in Bangladesh: A geostatistical and epidemiological framework for evaluating health effects and potential remedies. *Water Resources Research*, *39*(6), 1146–1162. <https://doi.org/10.1029/2002WR001327>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.