Statistical geospatial modelling of arsenic concentration in Vaishali District of Bihar, India

Nihar Jangle¹ · Vaibhav Sharma¹ · David M. Dror¹

Received: 10 August 2015 / Accepted: 13 March 2016 / Published online: 13 April 2016
© Springer International Publishing Switzerland 2016

Abstract Arsenic contamination of the groundwater poses a major threat to millions of people living in the fertile plains of the Ganga river basin. In the Indian state of Bihar, which is located in the Ganga river basin and has a population of more than 100 million, millions of groundwater wells remain to be tested to estimate the scale of arsenic contamination. In this study, a geostatistical model is developed based on the measured arsenic concentration in the aquifers (samples collected from hand tube-wells) and hydro-geomorphological factors for one out of 38 districts in Bihar, namely Vaishali. Maps of arsenic concentration in the fresh water fetched from hand tube-well of different depths were prepared. The results reveal that high spatial variability of the arsenic concentration can be modeled well based on hydro-geomorphological parameters, most importantly proximity to river, and depth of hand tube-wells. The model might be adapted to estimate the arsenic risk in groundwater in entire Bihar.

Keywords Arsenic · Bihar · Geospatial modeling · Groundwater contamination · GIS

Introduction

Arsenic concentration above 0.01 mg/L in groundwater exceeds the limit set by the World Health Organization (WHO) as suitable for human consumption as drinking water (WHO 2011). The conditions under which arsenic can be released to the groundwater may expose the Indo-Gangetic plains to particular risks. The fluvial sediments from the Himalayas, which are composed of clay sand and silt, have been identified as the source of arsenic in the aquifers of the Indo-Gangetic plains (McArthur et al. 2004). The arsenic is released from the sediments into the groundwater through a microbiologically mediated reductive dissolution process (McArthur et al. 2001, 2004; Akai et al. 2004; Charlet and Polya 2006). This process is accelerated by high rates of vertical percolation of water from the surface to the aquifers through arsenic rich sediments, and facilitated by the presence of organic matter that influences the solubility and mobility of arsenic (Islam et al. 2004; Rowland et al. 2006).

Some knowledge is at hand about high exposure in Bangladesh, where nearly 50 % of the country's population is exposed to the risk of drinking arsenic contaminated water. In India, the most severe arsenic contamination has been reported from West Bengal (Khan 1997; Smith et al. 2000; Kinniburgh and Smedley 2001; Ravenscroft et al. 2005; Nickson et al. 2007; Shamsudduha et al. 2008). Other high arsenic concentrations in groundwater have been reported from Chandigarh (Datta and Kaul 1976), Jharkhand (Bhattacharjee et al. 2005), Uttar Pradesh (Ahamed et al. 2006), Manipur (Chakraborti et al. 2008), Assam (Hazarika and Bhuyan 2013), from highland upstream areas like Uttarakhand (Gaur et al. 2013) and Himachal Pradesh (Chakraborti et al. 2003; Shah 2010). Arsenic contamination in Bihar was first reported in 2002

✉ Vaibhav Sharma
vaibhav@mia.org.in

Nihar Jangle
nihar@mia.org.in

David M. Dror
david@mia.org.in

¹ Micro Insurance Academy, 86, First floor, Okhla Phase-III, New Delhi 110020, India

in Ojha Patti village, Shahpur block of Bhojpur district (Chakraborti et al. 2003). Nickson et al. (2007) have reported on arsenic concentration in 11 districts of Bihar; Singh and Ghosh (2012) have monitored arsenic contamination in Maner block in Patna district and found arsenic concentration within the range of 0.14–0.49 mg/L. The highest levels of arsenic concentration in the groundwater (2.18 mg/L) were reported from Buxar district of Bihar (Singh et al. 2014a, b). High arsenic concentrations have also been detected in Patna (0.498 mg/L) (Singh and Ghosh 2012), Samastipur (0.060 mg/L) (Gupta et al. 2014) and Bhagalpur districts (0.1 mg/L) (Kumar et al. 2014), Saran (0.15 mg/L), Begusari (0.0943 mg/L) (Agrawal et al. 2011), Khagaria, Munger and Katihar districts of Bihar (range of 0.05–0.1 mg/L) (Singh et al. 2014a, b; Singh 2015; Singh and Vedwan 2015). These publications suggest that in 16 out of 38 districts of Bihar, groundwater contains high arsenic concentrations (Ghosh et al. 2007; Saha et al. 2009; Singh and Ghosh 2012). The government of India has been testing arsenic contamination all over Bihar between 2005 and 2014, in some 114,737 deep tube-wells installed by the government under the national rural drinking water program (NRDWP) in 107,640 habitations. The results show that the concentration of arsenic significantly varies from one block to another even within a single district (Ghosh et al. 2007), and overall, in about 10 % of the habitations tested, arsenic levels were higher than the Indian permissible limit for drinking water of 0.05 mg/L (BIS 2012). This led to an estimate that more than 10 million people in Bihar state are exposed to arsenic concentrations of 0.01 mg/L or above (Singh et al. 2014a, b), and to acute arsenicosis. The screening of the wells is still ongoing, and new arsenic contaminated areas are being identified.

However, the testing under NRDWP is probably giving an underestimate of the risk exposure, considering that it covers only NRDWP installed deep hand tube-wells, whereas most rural households depend on privately installed hand tube-wells that are commonly installed at depths where the arsenic concentration in the aquifer would be much higher (Srikanth 2013). Moreover, the dependence on groundwater as a source of fresh water increased due to the combined effect of high population growth, overutilization of agricultural land and contamination of surface water (Central Ground Water Board 2007).

This situation raises the imperative need to conduct a systematic assessment of the vulnerable areas in this region, characterized by a dense river network and fluvial Holocene sediments (naturally rich in arsenic) that are ideal settings for arsenic dissolution and accumulation in groundwater in Bihar where arsenic concentrations in the aquifers are already known to be very high (Chakraborti et al. 2009; Singh et al. 2014a, b).

Estimation of the spatial manifestation and magnitude of the exposure to arsenic contamination throughout the state based on testing would take several years to complete. Consequently, this systematic assessment cannot rely on testing alone. In the past, researchers have proposed different models for the prediction of arsenic concentration in unsampled areas based on the geostatistical interpolation methods (Goovaerts et al. 2005; Hossain et al. 2007; Lee et al. 2007; Winkel et al. 2008). Applying similar methods, we present in this study a statistical regression model to predict the arsenic concentration of untested locations in the groundwater in all the 16 blocks of Vaishali district, Bihar. The model utilizes digitally available geomorphological and hydrogeological parameters, which are considered to impact physical processes governing the solubility of arsenic in the groundwater, as well as laboratory test results on arsenic concentration of water from sampled hand tube-wells across the district, carried out for calibration. The purpose of this analysis is to validate the model so that it could then be applied to obtain estimates in other untested locations in the Indo-Gangetic plain, where some one billion people live and where much of the food of South Asia is produced (Agrawal et al. 2011).

The article is structured as follows: the study area is described, followed by a section on methods. This is followed by results and a discussion of these results. The conclusions are then presented.

Methods

Study area

The study was conducted in Vaishali district in northern Bihar, India, (Fig. 1). Vaishali district covers an area of 2016 km² and consist of 16 blocks and 1531 villages inhabited by a total population of around 3.5 million (Census of India 2011). The study area is an interfluvial alluvial plain bounded by the Ganga River in the south and Gandak River on the western side with point of confluence at Hajipur. Hydrologically, the study area can be divided into two sub-basins; Gandak sub-basin and Burhi Gandak sub-basin. The majority of the area lies in the Burhi Gandak sub-basin. The study area is mostly flat with a maximum elevation of 78 m near the banks of the Gandak River in the north-west and an average gradient of −3.5 m/km towards the south-east. Vaishali's fertile plains are formed by regular deposition of arsenic rich alluvium brought by the Gandak and Ganga rivers over time from the Himalayas. The region encompasses three types of morphostratigraphic units, namely Hajipur (Late Pleistocene–Early Holocene), Vaishali (Middle–Late Holocene) and Diara (Late Holocene) formation and comprises a confined and

Fig. 1 Study area

prolific unconfined aquifer system formed by an alternate deposition of the sand, clay, silt and gravel of quaternary alluvial deposit (Central Ground Water Board 2007).

Data

Table 1 lists the data used along with the data sources comprising primary data on arsenic concentration from sampled hand tube-wells and secondary data used for the calibration of the prediction model.

Methods

Sampling and testing of arsenic content in water from hand tube-wells

For this study, 34 water samples from hand tube-wells were collected from all 16 blocks of Vaishali district in Bihar, 2–3 samples from each block (Fig. 2). The

villages and hand tube-wells within a village were randomly selected. The hand tube-wells served as a source of fresh water for humans, domestic and livestock needs. Each hand tube-well was flushed for 10 min prior to collection of the water samples to remove the stagnant water. After the water samples had been stored in sterilized opaque polyethylene bottles, concentrated nitric acid (HNO_3) had been added, and the samples were kept refrigerated ($2\text{--}4\text{ }^\circ\text{C}$) adhering to standard procedures, e.g., followed by UNICEF (2008). A private testing laboratory in India (Delhi Test House Pvt. Ltd, New Delhi) was mandated to test the arsenic concentrations of the collected water samples. The laboratory employed the inductively coupled plasma mass spectrometry (ICP-MS) technique to measure the arsenic content with minimum arsenic detection limit of 0.001 mg/L and uncertainty error of $\pm 2\text{ ppb}$ at 10 ppb . The latitude and longitude coordinates and the depths of the sampled hand tube-wells were captured while the water samples were collected.

Table 1 Primary and secondary data used

Data type	Data specifications	Source
Arsenic concentration	Laboratory test results on arsenic concentration in water from 34 sampled hand tube-wells	Primary data survey
Depth of hand tube-wells	Depth of each sampled hand tube-wells was noted while collection of water samples for arsenic testing	Primary data survey
Digital Elevation Model (DEM)—ASTER GDEM	Global coverage, 30 m pixel size, highly accurate, freely available	Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA) http://gdem.ersdac.jspacesystems.or.jp/
Soil map	Total coverage: Ganga basin inside the Indian boundaries, 1:250,000 scale, 90 m pixel size, freely available	National Bureau of Soil Survey Land Use Product NBSS-LUP, National Remote Sensing Centre NRSC http://www.gisserver.civil.iitd.ac.in/grbmp/
LULC map	Total coverage: entire Ganga basin, 1:250,000 scale, 90 m pixel size, freely available	NRSC http://www.gisserver.civil.iitd.ac.in/grbmp/
NDVI	Global coverage, 250 m pixel size, freely available	NASA http://www.LAADSWEB.com
Rainfall	Annual district average rainfall data for 5 years (2009–2013) and five districts (Vaishali, Saran, Muzzafarpur, Samastipur and Patna), freely available	Indian Metrological Department (IMD) http://www.imd.gov.in/

Digital spatial data processing

All digital spatial data sets were transformed into the geographic information system (GIS) framework in the Universal Transverse Mercator (UTM) coordinate system with a spatial resolution of 90 m, and modeled arsenic concentration maps were prepared using GIS software ArcGIS version 9.1 (ESRI 2005).

Digital elevation model (DEM)

Topographic information was extracted from the ‘ASTER GDEM’ digital elevation model (DEM), available at 30 m resolution. The DEM was projected into the UTM coordinate system to develop the raster products of elevation (m), slope (degrees) flow accumulation cell values and ‘distance to river’. Flow accumulation cell values represent accumulated water flowing into each corresponding downslope cell in a raster image. Pixels with high flow accumulation values denote areas with higher water flow along the gradient, whereas pixels with null values represent local peaks or ridges with no water inflow (Jenson and Domingue 1988). Flow accumulation has already been identified as a good predictor of arsenic concentration pathways (Weerasiri et al. 2013). Gao et al. (2007) studied the relationship of flow accumulation and arsenic speciation. They showed that cells with concentrated flow paths have high reducing conditions promoting reductive dissolution of arsenic. Furthermore, cells with high flow accumulation values also act as sinks for arsenic rich Holocene sediments.

‘Distance to river’ is another important factor believed to govern spatial variability of arsenic concentration

(Winkel et al. 2008). The distance from each point in the study areas to the river system was calculated with the ArcHydro tool in ArcGIS 9.1 using the DEM data. It has been observed that in Asian river basins in close proximity to main rivers and their tributaries, arsenic concentration in groundwater is higher (Erban et al. 2013). According to Central Ground Water Board (2007), the area comprising flood plains and riparian stripes in the Ganga basin are composed of deep Holocene sediments, which explain the higher concentration of arsenic in the groundwater.

NDVI (Normalized Difference vegetation Index) and spectral ratio

Twelve mid-month images of the ‘normalized difference vegetation index’ (NDVI) (Rouse et al. 1974) of the year 2011 were used to calculate a 12-months average NDVI value. The NDVI value was then algebraically converted into the spectral ratio. The spectral ratio is the quotient of near infrared and visible red spectral reflectance measurements. Vegetation indices are indicators for biomass, and hence availability of water. These indices can thus also serve as a proxy for organic matter in the soil, which is considered to enhance the release of arsenic from soils and sediments (Wang and Mulligan 2006; Rodríguez-Lado et al. 2013).

Land use land cover (LULC)

Different land use/cover classes were mapped to 1 and 0 depending on the vegetation. Agricultural land, forests and other vegetated areas were assigned a value of 1; and barren, waste land, and urban clusters were assigned a value of 0.

Fig. 2 Sampled locations and test results of arsenic in freshwater of 34 hand tube-wells across Vaishali district, Bihar. *Black and grey circles* indicate arsenic concentrations above 0.07 mg/L and between 0.05 and 0.07 mg/L, respectively

Agricultural land and other vegetated areas are regions with increased transport of organic material into the aquifer which instigates the reductive dissolution of arsenic to the groundwater (Wang and Mulligan 2006).

Soil map (estimation of hydraulic conductivity based on soil type)

The hydraulic conductivity for each soil type has been calculated based on the soil texture (combinations of sand, silt and clay) using a methodology developed by (Saxton et al. 1986).

Net groundwater recharge

The net groundwater recharge was calculated applying Eq. (1) following (Chaturvedi 1973) for the estimation of net recharge in the Ganga-Yamuna river basin:

$$R = 2.5 \sqrt{(P - 0.6)}, \quad (1)$$

where R is the net annual recharge in cm and P the total annual rainfall in cm. This simplified formula only takes account of rainfall, but not of local topography. The impacts of local topography on groundwater recharge are indirectly considered by incorporating the factors ‘flow accumulation cell values’ and LULC.

The total annual rainfall for Vaishali district was calculated by interpolating the past 5 years’ average annual rainfall (2009–2013) from Vaishali and four neighboring districts (Patna, Saran, Muzzafarpur, Samastipur) using the inverse distance weightage (IDW) method (Shepard 1968). Recent studies have reported that net recharge can play an important role in the mobilization of arsenic in the groundwater (Harvey et al. 2002; Ashfaq 2007; Meliker et al. 2009). Biodegradable organic carbon present on the

surface and in subsurface layers of the soil strata instigate the reductive dissolution processes of arsenic release in the groundwater and higher recharge would increase the transport of dissolved arsenic to the groundwater.

Depth of hand tube-wells

When we collected water samples, we also measured the depth of each hand tube-well sampled. The reason for measuring the depth is to include this parameter in modeling arsenic concentration, considering recent studies which indicate that higher concentrations have been observed for deep tube-wells in Bangladesh (Burgess et al. 2010), India (Chakraborti et al. 2009) and in North Vietnam (Winkel et al. 2008; Erban et al. 2013). In the studied area, (Saha et al. 2009, 2011) claimed that arsenic contaminated zone in the Ganga river basin is mostly located within the shallow aquifer range (<50 m) and deeper aquifers located below 190 m can be used to extract water for drinking purpose (Saha and Shukla 2013). However, our working assumption is that the depth of the arsenic rich Holocene sediments varies substantially in any aquifer.

Regression model with arsenic concentration as dependent variable

We ran a multiple linear regression model to predict the arsenic concentration (dependent variable) using statistical software “R” version 3.2.2. The following parameters were considered as potential independent variables: elevation, slope and flow accumulation value, NDVI, spectral ratio, LULC, hydraulic conductivity, distance to river, groundwater recharge and hand tube-well depth. The regression model was calibrated at the 34 sample sites where the arsenic concentration has been determined, by using the point information of all the selected parameters (independent variables) at the same locations for all hand tube-wells at different installation depths. These point values were extracted from the geospatial layers using the “spatial analyst tool” of the software ArcGIS version 9.1 (ESRI 2005).

The statistical significance between the arsenic concentration and each potential independent variable was initially evaluated through a simple linear regression analysis. Only parameters which showed significance above 95 % (p value <0.05) were included in the subsequent multiple linear regression. Then, the multiple linear regression were performed only with those significant parameters as independent variables.

Results

Arsenic concentration in the groundwater of entire Vaishali District based on in situ testing of hand tube-wells

In total, 34 groundwater samples were collected from hand tube-wells in all 16 blocks of the study area, whose depths ranged from 10 to 50 m. All tested samples displayed arsenic concentrations ranging from 0.050 to 0.088 mg/L, i.e., above the permissible limit of 0.05 mg/L set by government of India (BIS 2012) (Fig. 2). In passing we note that during the data collection, we observed that villagers in most of the locations displayed visible skin lesions, which are assumed to be arsenicosis as a result of chronic exposure to arsenic.

The highest arsenic concentration values (0.07 mg/L or higher) were measured in the low lying flood plain surfaces of the Gandak and Ganga river: (1) on the eastern flood plain of Gandak River in Vaishali block (0.079, 0.072 mg/L) and Lalganj block (0.076, 0.071 mg/L), (2) at the confluence of Ganga and Gandak at Hajipur block (0.070 mg/L), (3) on the northern bank of Ganga river in Bidupur block (0.088, 0.072 mg/L) and Desri block (0.072, 0.071 mg/L), and (4) on the island between two rivulets of the Ganga river in Raghapur block (0.071 mg/L).

Prediction of arsenic concentration across entire Vaishali District, Bihar

The p values of the simple linear regressions of measured arsenic concentration against each of the potential independent variables for the multiple linear regression are given in Table 2. We included only the spectral ratio, which can be derived through an algebraic transformation from the NDVI, which shows better results than the NDVI (higher R^2 value).

The characteristics of the multiple linear regression are presented in Table 3. The R^2 value is with 0.7157 quite high indicating that the identified parameters describe the arsenic concentration well.

The arsenic concentration in groundwater is positively correlated with flow accumulation value, spectral ratio, LULC, hydraulic conductivity, recharge, hand tube-well depth, slope, elevation, and negatively correlated with distance to river. Flow accumulation values, spectral ratio, LULC, distance to river, groundwater recharge and depth of hand tube-wells were significantly correlated with the arsenic concentration (p value <0.05), and were thus, included in the multiple linear regression.

Table 2 *p* values of simple linear regression with measured arsenic concentration as dependent variable

Variables regressed against measure as concentration	Unit	<i>p</i> values
Elevation	m	0.4116
Slope	degree	0.3164
Flow accumulation value	number of cells	0.0144
Vegetation index (spectral ratio)	–	0.0075
LULC	–	0.0163
Hydraulic conductivity	m/day	0.5649
Distance to river	km	0.0004
Groundwater recharge	cm/m	0.0156
Hand tube-well depth	m	0.0072

Table 3 Results of the multiple linear regression with arsenic concentration as the dependent variable

Independent variables	Coefficient	Std. error	<i>t</i> value	Pr (> <i>t</i>)
(Intercept)	−0.0568600	0.04368	−1.30200	0.20394
Flow accumulation value	0.0005547	0.00058	0.95200	0.34952
Vegetation index (spectral ratio)	0.0022470	0.00115	1.96200	0.06009
LULC	0.0050770	0.00222	2.29100	0.02997
Distance to river	−0.0004258	0.00010	−4.11400	0.00033
Groundwater recharge	0.0036940	0.00148	2.49000	0.01924
Hand tube-well depth	0.0001459	0.00009	1.54900	0.13307

Residual standard error: 0.005592 on 27 degrees of freedom

Multiple R^2 : 0.7157, Adjusted R^2 : 0.6525*F* statistic: 11.33 on 6 and 27 DF, *p* value: 2.573e−06

Observed and predicted arsenic concentration values are shown for all 34 sample sites in Fig. 3. The R^2 value of 0.7157 for the entire data set of 34 points corresponds to a correlation of 0.8460 between observed and predicted arsenic concentration values. To conduct a comparison between in-sample and out-of-sample accuracy, the set of 34 observations has been divided into sets of 27 and 7 data points. This division is based on the rule of thumb consideration that the sample size should be at least around 5

times the number of independent variables (Stoelting 2002). In this study we used a sample size of 27 for the in-sample training set to have a least 7 data points for the out-of-sample testing. For 100,000 random divisions, the correlations between observed and predicted arsenic values for in-sample and out-of-sample have been calculated and the averages determined. In-sample average correlation was 0.8536 and out-of-sample average correlation was 0.7239. The average out-of-sample root-mean-square error (RMSE), the standard deviation of the differences between predicted values and observed values for the 7 data points, was 0.006489. The analysis reveals that the predicted and actual arsenic concentrations even in untested areas can be expected to differ less than 0.01 mg/L in most of the cases.

Three arsenic maps were prepared illustrating the model predictions for arsenic concentrations at three different depths, namely 10, 30, and 50 m, see Fig. 4. The arsenic concentrations were grouped into five classes; 0.05–0.06, 0.06–0.07, 0.07–0.08, 0.08–0.09 and 0.09–0.1 mg/L. The arsenic concentration has a distinguished gradient with decreasing values from the rivers in the west and south to the north-east. The same gradient was observed in the 34 sample sites, and is introduced in the prediction model through the independent variable “distance to river”.

These results show that when a hand tube-well is 10 m deep, the majority of the district area (53 %) would be

Fig. 3 Comparative analysis of measured and predicted arsenic concentration at the 34 sample sites

Fig. 4 Predicted arsenic concentrations in Vaishali district, Bihar, at 10 (left), 30 (middle) and 50 m (right) hand tube-well depth

Fig. 5 Distribution of area predicted to fall under 0.05–0.06, 0.06–0.07, 0.07–0.08, 0.09–0.10 mg/L arsenic concentration for 10 (left), 30 (middle) and 50 m (right) hand tube-well depth in Vaishali district, Bihar

exposed to an arsenic concentration in the range of 0.061–0.07 mg/L (Fig. 5). At 30 to 50 ms depth, the majority of the area would be exposed to a higher concentration of 0.071–0.08 mg/L. And when hand tube-well depth exceeds 50 ms, virtually the entire area would be threatened by higher arsenic levels.

Discussion and conclusion

This study predicts the arsenic concentration in Vaishali district, Bihar, India based on a multiple linear regression model that has been calibrated with regional hydro-geomorphological characteristics. Water samples were collected from the sampled hand tube-wells at different depths, and tested for arsenic concentration levels. Results of the groundwater sample tests revealed that all blocks of the district are exposed to arsenic concentrations exceeding 0.05 mg/L (the permissible limit of arsenic in drinking

water set by the government of India; the limit for acceptable safe drinking water is 0.01 mg/L). Arsenic concentrations of 0.07 mg/L and higher were observed in sample sites located in low lying flood plain areas of Ganga and Gandak rivers. Moreover, arsenic concentrations higher than the safe limit have been found in previously considered “safe” blocks of North-Eastern Vaishali district, including Patepur, Sahdei Buzurg, Chehra Kala, Jandaha, Rajapakar, Mahnar, Mahua and Garauls. Generally speaking, arsenic concentrations gradually decrease as the distance from the rivers increases (and previous studies (Chakraborti et al. 2009; Shah 2010) claimed that the lower arsenic concentration in those locations is due to older alluvium from the Pleistocene era, covering the Holocene sediments).

The model predicted unsafe arsenic concentrations of above 0.05 mg/L across the entire study area and correctly identifies the blocks of Bidupur, Raghupur and Hajipur as most exposed—the same blocks where arsenic

contamination has already been detected (Central Ground Water Board 2007; Chakraborti et al. 2009). High arsenic concentrations have been predicted in the aquifers located in the vicinity of the rivers in the entire study area.

The spatial variability of the predicted arsenic concentration was relatively high, ranging from 0.05 to around 0.10 mg/L with proximity to river being an important predictor. Although distance to river is the major factor determining the arsenic concentrations (for a uniform tube-well depth), points of equal distance to the river system still exhibit differences in the arsenic levels mainly because of variations in flow accumulation values and land use and land cover, which are the conditions favorable for inducing reductive dissolution of arsenic.

One limitation of the present study is that the proposed model can only be applied for tube-well depths up to around 50 m. This is satisfactory for the majority of installation depths, and the depths of the hand tube-wells we sampled ranged from 10 to 50 m. However, some tube-wells are much deeper, for example, the government installed ones. The multiple linear regression models indicates that with increasing depth the arsenic concentration increases linearly, which would hold true only within the range of the sampled tube-well depths. Deeper tube-wells would have to be sampled to analyze the impact of larger ranges of depths, which cannot be assumed to be linear, considering various reports that beyond 50 m the arsenic concentration in the groundwater decreases due to the presence of old Holocene sediments (Yadav et al. 2012).

Another limitation is the relatively small sample size. To control the risk of overfitting, we conducted out-of-sample tests. Even in the out-of-sample sets the correlation between observed and predicted arsenic values was good and the average out-of-sample root-mean-square error (RMSE) low. This underlines that the prediction model is not over-fitted and that extending the training set used for model calibration would not lead to material changes in the prediction model. However, for other regions additional/different input parameters might have to be considered.

Those limitations notwithstanding, the model and results described in this article demonstrate a relatively easy and cheap way to identify the areas of high risk with high spatial resolution. This diagnosis is the essential prerequisite for targeted preventive and corrective interventions to protect the local population from excessive exposure to arsenic. As this model is based on easily and freely available input parameters, which can be obtained for the entire Bihar state, and even the entire Indian Indo-Gangetic Plain, we submit that the model could be applied to model and accurately predict the arsenic exposure in groundwater for larger regions, with only a relatively small sample size required to calibrate the model to other regions. Thus, the model adds to

the existing tools, at considerably reduced cost and time to obtain estimates, with comparable accuracy.

Acknowledgments The authors gratefully acknowledge funding support by the Climate Change and Development Division of the Embassy of Switzerland in India for this research, conducted under the “Climate Resilience through Risk Transfer (RES-RISK)” project led by the Micro Insurance Academy (MIA), New Delhi, and co-implemented by BASIX.

References

- Agrawal D, Kumar P, Avtar R, Ramanathan A (2011) Multivariate statistical approach to deduce hydrogeochemical processes in the groundwater environment of Begusarai District, Bihar. *Water Qual Expo Health* 3:119–126
- Ahamed S, Kumar Sengupta M, Mukherjee A et al (2006) Arsenic groundwater contamination and its health effects in the state of Uttar Pradesh (UP) in upper and middle Ganga plain, India: a severe danger. *Sci Total Environ* 370:310–322
- Akai J, Izumi K, Fukuhara H et al (2004) Mineralogical and geomicrobiological investigations on groundwater arsenic enrichment in Bangladesh. *Appl Geochem* 19:215–230
- Ashfaq K (2007) Effect of hydrological flow pattern on groundwater arsenic concentration in Bangladesh. Massachusetts Institute of Technology. Available at: <http://hdl.handle.net/1721.1/42218>. Accessed 6 Nov 2015
- Bhattacharjee S, Chakravarty S, Maity S et al (2005) Metal contents in the groundwater of Sahebgunj district, Jharkhand, India, with special reference to arsenic. *Chemosphere* 58:1203–1217
- BIS (2012) Indian standard drinking water—specification (second revision): IS 10500: 20102. New Delhi, India, Bureau of Indian Standards (BIS), Manak Bhavan, 9 Bahadur Shah Zafar Marg, New Delhi-110002, India
- Burgess W, Hoque M, Michael H et al (2010) Vulnerability of deep groundwater in the Bengal Aquifer System to contamination by arsenic. *Nat Geosci* 3:83–87
- Census of India (2011) Census of India. Government of India Publication. Available at: <http://censusindia.gov.in/2011-com-mon/censusdataonline.html>. Accessed 6 Nov 2015
- Central Ground Water Board (2007) Groundwater information booklet (Vaishali, District, Bihar, India). Government of India Publication. Available at: http://www.cgwb.gov.in/District_Profile/Bihar/Vaishali.pdf. Accessed 6 Nov 2015
- Chakraborti D, Mukherjee SC, Pati S et al (2003) Arsenic groundwater contamination in Middle Ganga Plain, Bihar, India: a future danger? *Environ Health Perspect* 111:1194
- Chakraborti D, Singh EJ, Das B et al (2008) Groundwater arsenic contamination in Manipur, one of the seven North-Eastern Hill states of India: a future danger. *Environ Geol* 56:381–390
- Chakraborti D, Das B, Rahman MM et al (2009) Status of groundwater arsenic contamination in the state of West Bengal, India: a 20-year study report. *Mol Nutr Food Res* 53:542–551
- Charlet L, Polya DA (2006) Arsenic in shallow, reducing groundwaters in southern Asia: an environmental health disaster. *Elements* 2:91–96
- Chaturvedi R (1973) A note on the investigation of ground water resources in western districts of Uttar Pradesh. *Annu Rep Irrig Res Inst* 1973:86–122
- Datta D, Kaul M (1976) Arsenic content of drinking water in villages in Northern India. A concept of arsenicosis. *J Assoc Physicians India* 24(9):599–604

- Erban LE, Gorelick SM, Zebker HA, Fendorf S (2013) Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping-induced land subsidence. *Proc Natl Acad Sci* 110:13751–13756
- ESRI (2005) An overview of spatial analyst. <http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?TopicName=An%20overview%20of%20Spatial%20Analyst>. Accessed 6 Oct 2015
- Gao S, Ryu J, Tanji K, Herbel M (2007) Arsenic speciation and accumulation in evapoconcentrating waters of agricultural evaporation basins. *Chemosphere* 67:862–871
- Gaur S, Joshi M, Jos H, Saxena S (2013) Analytical study of water safety parameters in ground water samples of Uttarakhand in India. *Indian J Public Health Res Dev* 4:185–189
- Ghosh A, Singh S, Bose N, Chowdhary S (2007) Arsenic contaminated aquifers: a study of the Ganga levee zone in Bihar, India. In: Symposium on arsenic: the geography of a global problem, Royal Geographical Society, London. <http://www.geog.cam.ac.uk/research/projects/arsenic>. Accessed 12 Nov 2015
- Goovaerts P, Avruskin G, Meliker J et al (2005) Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resour Res* 41(7):1–19
- Gupta V, Singh J, Kumari P et al (2014) Arsenic in hand pump water of different blocks of Samastipur district (Bihar), India. *Int J Basic Appl Res* 1(1):1–5
- Harvey CF, Swartz CH, Badruzzaman A et al (2002) Arsenic mobility and groundwater extraction in Bangladesh. *Science* 298:1602–1606
- Hazarika S, Bhuyan B (2013) Fluoride, arsenic and iron content of groundwater around six selected tea gardens of Lakhimpur District, Assam, India. *Arch Appl Sci Res* 5:57–61
- Hossain F, Hill J, Bagtzoglou AC (2007) Geostatistically based management of arsenic contaminated ground water in shallow wells of Bangladesh. *Water Resour Manag* 21:1245–1261
- Islam FS, Gault AG, Boothman C et al (2004) Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature* 430:68–71
- Jenson S, Domingue J (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogramm Eng Remote Sens* 54:1593–1600
- Khan A (1997) Arsenic contamination in ground water and its effect on human health with particular reference to Bangladesh. *J Prev Soc Med* 16:65–73
- Kinniburgh D, Smedley PL (2001) Arsenic contamination of groundwater in Bangladesh. BGS Technical Report. WC/00/19(2). Available at: <https://www.bgs.ac.uk/downloads/start.cfm?id=2223>. Accessed 6 Nov 2015
- Kumar P, Avtar R, Kumar A et al (2014) Geophysical approach to delineate arsenic hot spots in the alluvial aquifers of Bhagalpur district, Bihar (India) in the central Gangetic plains. *Appl Water Sci* 4:89–97
- Lee J-J, Jang C-S, Wang S-W, Liu C-W (2007) Evaluation of potential health risk of arsenic-affected groundwater using indicator kriging and dose response model. *Sci Total Environ* 384:151–162
- McArthur J, Ravenscroft P, Safiulla S, Thirlwall M (2001) Arsenic in groundwater: testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Water Resour Res* 37:109–117
- McArthur J, Banerjee D, Hudson-Edwards K 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. *Appl Geochem* 19:1255–1293
- Meliker JR, Slotnick MJ, Avruskin GA et al (2009) Influence of groundwater recharge and well characteristics on dissolved arsenic concentrations in southeastern Michigan groundwater. *Environ Geochem Health* 31:147–157
- Nickson R, Sengupta C, Mitra P et al (2007) Current knowledge on the distribution of arsenic in groundwater in five states of India. *J Environ Sci Health Part A* 42:1707–1718
- Ravenscroft P, Burgess WG, Ahmed KM et al (2005) Arsenic in groundwater of the Bengal Basin, Bangladesh: distribution, field relations, and hydrogeological setting. *Hydrogeol J* 13:727–751
- Rodríguez-Lado L, Sun G, Berg M et al (2013) Groundwater arsenic contamination throughout China. *Science* 341:866–868
- Rouse J Jr, Haas R, Schell J, Deering D (1974) Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec Publ* 351:309
- Rowland H, Polya D, Lloyd J, Pancost R (2006) Characterisation of organic matter in a shallow, reducing, arsenic-rich aquifer, West Bengal. *Org Geochem* 37:1101–1114
- Saha D, Shukla R (2013) Genesis of arsenic-rich groundwater and the search for alternative safe aquifers in the Gangetic Plain India. *Water Environ Res* 85(12):2254–2264
- Saha D, Dwivedi S, Sahu S (2009) Arsenic groundwater contamination in parts of middle Ganga plain, Bihar. *Curr Sci* 97:753–755
- Saha D, Sahu S, Chandra PC (2011) Arsenic-safe alternate aquifers and their hydraulic characteristics in contaminated areas of Middle Ganga Plain, Eastern India. *Environ Monit Assess* 175(1–4):331–348
- Saxton K, Rawls W, Romberger J, Papendick R (1986) Estimating generalized soil-water characteristics from texture. *Soil Sci Soc Am J* 50:1031–1036
- Shah BA (2010) Arsenic-contaminated groundwater in Holocene sediments from parts of middle Ganga plain, Uttar Pradesh, India. *Curr Sci* 98:1359–1365
- Shamsudduha M, Uddin A, Saunders J, Lee M-K (2008) Quaternary stratigraphy, sediment characteristics and geochemistry of arsenic-contaminated alluvial aquifers in the Ganges-Brahmaputra floodplain in central Bangladesh. *J Contam Hydrol* 99:112–136
- Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference. ACM, pp 517–524
- Singh SK (2015) Groundwater arsenic contamination in the Middle-Gangetic Plain, Bihar (India): the danger arrived. *Int J Environ Res* 4(2):70–76
- Singh SK, Ghosh AK (2012) Health risk assessment due to groundwater arsenic contamination: children are at high risk. *Hum Ecol Risk Assess Int J* 18:751–766
- Singh SK, Vedwan N (2015) Mapping composite vulnerability to groundwater arsenic contamination: an analytical framework and a case study in India. *Nat Hazards* 75(2):1883–1908
- Singh SK, Gean DS, Srikanta KP (2014a) Multiple groundwater contamination in the Mid-Gangetic Plain, Bihar (India): a potential threat. *Int J Adv Res Sci Tech* 3(3):175–179
- Singh S, Ghosh A, Kumar A et al (2014b) Groundwater arsenic contamination and associated health risks in Bihar, India. *Int J Environ Res* 8:49–60
- Smith AH, Lingas EO, Rahman M (2000) Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *Bull World Health Organ* 78:1093–1103
- Srikanth R (2013) Access, monitoring and intervention challenges in the provision of safe drinking water in rural Bihar, India. *J Water Sanit Hyg Dev* 3:61–69
- Stoelting (2002) Structural equation modeling/path analysis. Available at: <http://ocw.jhsph.edu/courses/structuralmodels/PDFs/Lecture3.pdf>. Accessed 6 Nov 2016
- UNICEF (2008) Arsenic primer: guidance for UNICEF country offices on the investigation and mitigation of arsenic contamination. Available at: http://www.unicef.org/eapro/Arsenic_primer_28_07_2008_final.pdf. Accessed 10 Nov 2015

- Wang S, Mulligan CN (2006) Effect of natural organic matter on arsenic release from soils and sediments into groundwater. *Environ Geochem Health* 28:197–214
- Weerasiri T, Wirojanagud W, Srisatit T (2013) Identifying arsenic pathway using electrical resistivity survey with GIS based flow modeling. *Int J Electrochem Sci* 8:12707–12718
- WHO (2011) Arsenic in drinking water-background document for development of WHO guidelines for drinking-water quality. http://www.who.int/water_sanitation_health/dwq/chemicals/arsenic.pdf. Accessed 6 Oct 2015
- Winkel L, Berg M, Amini M et al (2008) Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. *Nat Geosci* 1:536–542
- Yadav I, Singh S, Devi N et al (2012) Spatial distribution of arsenic in groundwater of Southern Nepal. In: Whitacre DM (ed) *Reviews of environmental contamination and toxicology*, vol 218. Springer, US, pp 125–140