Journal of Spatial Hydrology

Volume 7 | Number 2 Article 5

2007

Spatial Variability and Prediction Modeling of Groundwater Arsenic Distributions in the Shallowest Alluvial Aquifers in Bangladesh

Journal of Spatial Hydrology
Vol.7, No.2, Fall 2007

Spatial Variability and Prediction Modeling of Groundwater Arsenic Distributions in the Shallowest Alluvial Aquifers in Bangladesh

M. Shamsudduha

Abstract

Elevated arsenic in groundwater is the greatest environmental problem in Bangladesh. Spatial variability of arsenic in groundwater has been examined by semivariogram analysis that revealed high degree of small-scale spatial variability in alluvial aquifers. Small-scale variability of arsenic concentrations, indicated by high "nugget" values in semivariograms, is associated with heterogeneity in local-scale geology and geochemical processes. In unsampled locations, arsenic concentrations have been predicted using both deterministic and stochastic prediction methods. Natural neighbor (NN) method predicted better than inverse distance to power (IDP) method, and small-scale variations of arsenic concentrations are preserved. Ordinary kriging (OK) method on the untransformed arsenic data and their residual values performed considerably in predicting spatial arsenic distributions on regional-scale. Predicted results are evaluated by cross-validation, mean prediction error, and root mean square methods. Results show that approximately 25% area of Bangladesh, excluding Chittagong Hill Tracts and southern coastal parts, is below the concentration of 10 µg L^{-1} of arsenic. Approximately, 43% area in Bangladesh has arsenic concentrations of 10-50 µg L^{-1} at shallow depth (< 25 m). More than 17% area has arsenic concentrations between 50 µg L^{-1} and 100 µg L^{-1}. High density dataset and small-scale modeling would perform better in prediction of spatial distributions of groundwater arsenic. Sequential simulation and co-kriging methods can be applied to evaluate the spatial distributions of arsenic in groundwater in Bangladesh.

Keywords: Arsenic, Bangladesh, distribution, spatial variability, semivariogram, prediction models.

Introduction

Bangladesh is affected by one of the worst cases of groundwater contamination by elevated arsenic (As) in the world. For more than a decade, many studies have explored the complex biogeoscience of the elevated and wide-spread arsenic occurrences in groundwater of Bangladesh and West Bengal, India (Bhattacharya et al. 1997, Nickson et al. 2000, BGS and DPHE 2001, McArthur et al. 2001). A systematic survey throughout the country by Department of Public Health Engineering (DPHE) of Bangladesh and British Geological Survey (BGS) in 1998 and 1999 revealed that nearly 35 million people were drinking groundwater containing arsenic of a concentration of more than 50 µg L^{-1} (Bangladesh Standard) and about 57 million people were exposed to a concentration exceeding 10 µg L^{-1} (World Health Organization Standard). Arsenic contaminated groundwaters are mostly extracted from the alluvial aquifers, located within the depth range of 10-50 m below the ground surface (BGS and DPHE 2001).

Department of Geology and Geography, Auburn University, Auburn, AL 36849, USA
E-mail: shamsmo@auburn.edu

Spatial variability of arsenic in groundwater of Bangladesh is extremely high. Inadequate studies are performed on spatial patterns of groundwater arsenic distributions and prediction in unsampled areas of the country (e.g., Karthik 2001, Ravenscroft et al. 2001, Yu et al. 2003, Gaus et al. 2003, van Geen et al. 2003, Shamsudduha 2004, Hossain et al. 2005, Hossain and Sivakumar 2005 and ref. therein). Arsenic concentrations in the very shallow alluvial aquifers (i.e., depth within 25 m from the ground surface) vary greatly within a small distance even at similar depths. Spatial variability of arsenic in Bangladesh is observed fairly high within the neighboring wells, especially at local-scale of 10's to 100's of meters (van Geen et al. 2003, Yu et al. 2003). However, large-scale spatial dependencies in arsenic concentrations are observed that are consistent with geologic-geomorphic divisions of Bangladesh (BGS and DPHE 2001, Ravenscroft et al. 2005). Variations in arsenic concentrations are observed as small as < 5 μg L^{-1} to about 900 μg L^{-1} that occur at relatively shallow depths between 6 m and about 25 m within the alluvial aquifers (Ahmed et al. 2004). One of the most challenging tasks is to predict arsenic concentrations at unknown locations as the concentrations in adjacent tubewells bear very little similarities even if these wells are screened within the same aquifers (BGS and DPHE 2001, Gaus et al. 2003, Yu et al. 2003). Geostatistical analyses and prediction modeling on groundwater arsenic has been performed by others using ordinary and indicator kriging methods (e.g., Ravenscroft et al. 2001, Gaus et al. 2003, Hossain et al. 2005). However, some further scopes are present in analyzing the spatial distributions of groundwater arsenic and also compare and contrast different deterministic and stochastic prediction methods to learn more about the nature of geospatial distributions of arsenic in aquifers.

In this study, spatial patterns of arsenic concentrations in groundwater of Bangladesh are examined using geostatistical methods. Distributions of arsenic in shallow alluvial aquifers of Bangladesh are examined and modeled using deterministic and stochastic prediction methods on 1,145 data points. Results from these predictions models are evaluated and validated using cross-validation, root mean square error, and mean prediction error methods. Further guidelines for improving the spatial predictions of groundwater arsenic and other dissolved ions are also recommended in this study.

Materials and Methods

Dataset

Dataset from the National Hydrochemical Survey (NHS) in Bangladesh is used for spatial variability and prediction analyses in this study. Data are available in the British Geological Survey (BGS) official website (http://www.bgs.ac.uk/arsenic/bangladesh/reports.htm). During the summers of 1998 and 1999 the Department of Public Health Engineering (DPHE) of the Government of Bangladesh and British Geological Survey (BGS) and Mott MacDonald Ltd. surveyed groundwater from private and government tubewells and collected 3,534 samples across the country. Collected samples were analyzed for arsenic by hydride generation-atomic fluorescence spectrometry (HG-AFS) and hydride generation-ICP-AES under the supervision of British Geological Survey in the United Kingdom. This database is the largest hydrochemical database of Bangladesh and provides the most complete spatial coverage of groundwater arsenic distributions, and other geochemical parameters in the country. A total of 1,145 data points are found ranging in depth from a few meters to 25 m that are separated from the NHS dataset (Fig. 1). Most shallow aquifers in Bangladesh occur within this depth range. Approximately 27% of 1,145 water samples having a concentration of arsenic below the practical limits of instrumental measurement are categorized as non-detect data. In environmental studies, non-detect or censored data are very common and in many cases they cover a major portion of a dataset. Moreover, the non-detect arsenic values are critical for the prediction modeling since these low concentrations are close to the limit of acceptable arsenic concentrations in dirking water standards (e.g., WHO standard 10 μg L^{-1} and Bangladesh standard 50 μg

L^{-1}). For the simplicity of analysis, these non-detect values were substituted with their half detection limits following a standard statistical norm. There are two minimum detection limits (0.5 µg L^{-1} and 6.0 µg L^{-1}) in the BGS and DPHE (2001) arsenic dataset and these non-detect values were substituted with 0.25 µg $L^{-1}$and 3.0 µg L^{-1} respectively.

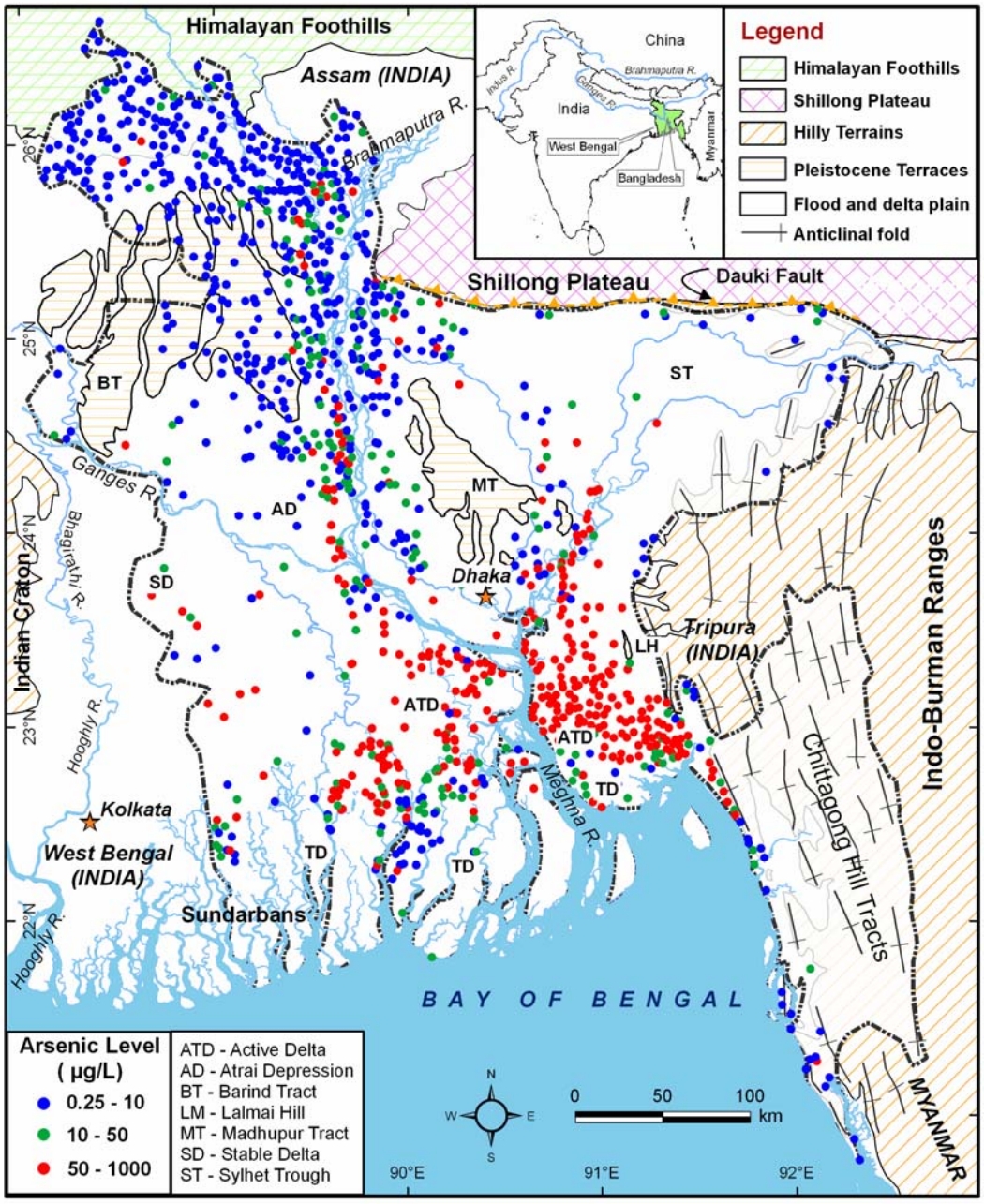

Figure 1. Simplified geological map of Bangladesh with shallow (depth < 25 m) tubewell locations and their arsenic concentrations. Different geomorphic units in the country are shown on this map. High arsenic concentrations are found in south-central parts of the country within the stable and active delta parts. A distinct increasing trend in arsenic concentrations in groundwater tubewells from northwest to south-central part of Bangladesh is observed.

Data Transformation

Distributions of arsenic concentrations in shallow wells are highly skewed (Gaus et al. 2003). There is a regional trend in the arsenic dataset where the general concentrations increase toward the south-southeast from northwest of the country. Temporal variations in arsenic concentrations in Bangladesh have been studied by several authors (e.g., Cheng et al. 2005, Zheng et al. 2005). Little variation in arsenic levels particularly at shallow wells (< 50 m) has been noticed over a period of three years in the Araihazar Thana in north-central Bangladesh (Cheng et al. 2005). These temporal variations are appeared to be related to seasonal precipitation and monsoonal recharge in Araihazar area and most likely in other areas in Bangladesh. However, due to lack in long-term arsenic concentrations data in many places across the country no definite temporal trend can be certainly predicted (Ravenscroft et al. 2006, Sengupta et al. 2006).

The minimum arsenic concentration is 0.25 µg L^{-1}, whereas the maximum concentration is 668 µg L^{-1} in this selected dataset. However, the maximum arsenic concentration in the NHS dataset is about 1,670 µg L^{-1}, and only three samples have arsenic concentrations exceeding 1,000 µg L^{-1} (BGS and DPHE 2001). These extreme values are considered as "outliers" and are not considered for geostatistical analyses in this study. The mean arsenic value is about 76 µg L^{-1} with a standard deviation of about 146 µg L^{-1} in the selected samples. The variance of the dataset is large compared to the standard deviation and mean, which indicates that the skewness greatly exceeds the mean. It is a pre-requisite to transform skewed data into a normal distribution before applying to any geostatistical analysis and kriging interpolation. Log-transformation is one of the widely used methods for data normalization. But log-transformation of the arsenic concentrations in this case does not transform the skewed NHS data into an acceptable normal distribution and therefore has not been applied in this study (Gaus et al. 2003). Rather, highly skewed arsenic data are transformed into Gaussian or normal-score distribution, which is a unique transformation system in geostatistics. This method is widely used because of its extreme analytical simplicity and for being the limit distribution of many analytical theorems (Anderson 1958, Johnson and Wichern 1982). Normal-score transformation is simply done by ranking the data in a standardized normal distribution *N* (0,1) (Olea 1999). Normal-score values for arsenic concentrations are back-transformed after the stochastic interpolation methods (Deutsch and Journel 1998). Histograms of the original data and normal-score transformed data for the shallow aquifers illustrate how skewed data become normally distributed after transformation (Fig. 2).

Prediction Methods

Natural neighbor

Natural neighbor (NN) is a common deterministic interpolation method widely used in mining industries and also for environmental studies. This is a local weighted average method and it uses geometrical constructions to generate weights for each datum in the natural neighbor subset (Watson 1999). In NN method, the weights are taken as the natural neighbor coordinates of any point. One of the greatest advantages of using natural neighbor interpolation methods is that the original values of the data points are recovered at their observation points, and the interpolation parameters, such as the search radius is not required like other interpolation methods. Extrapolation, beyond the existing data set is highly speculative, and undependable. A full set of natural neighbors is not available in the area where there is no data point (Watson 1999).

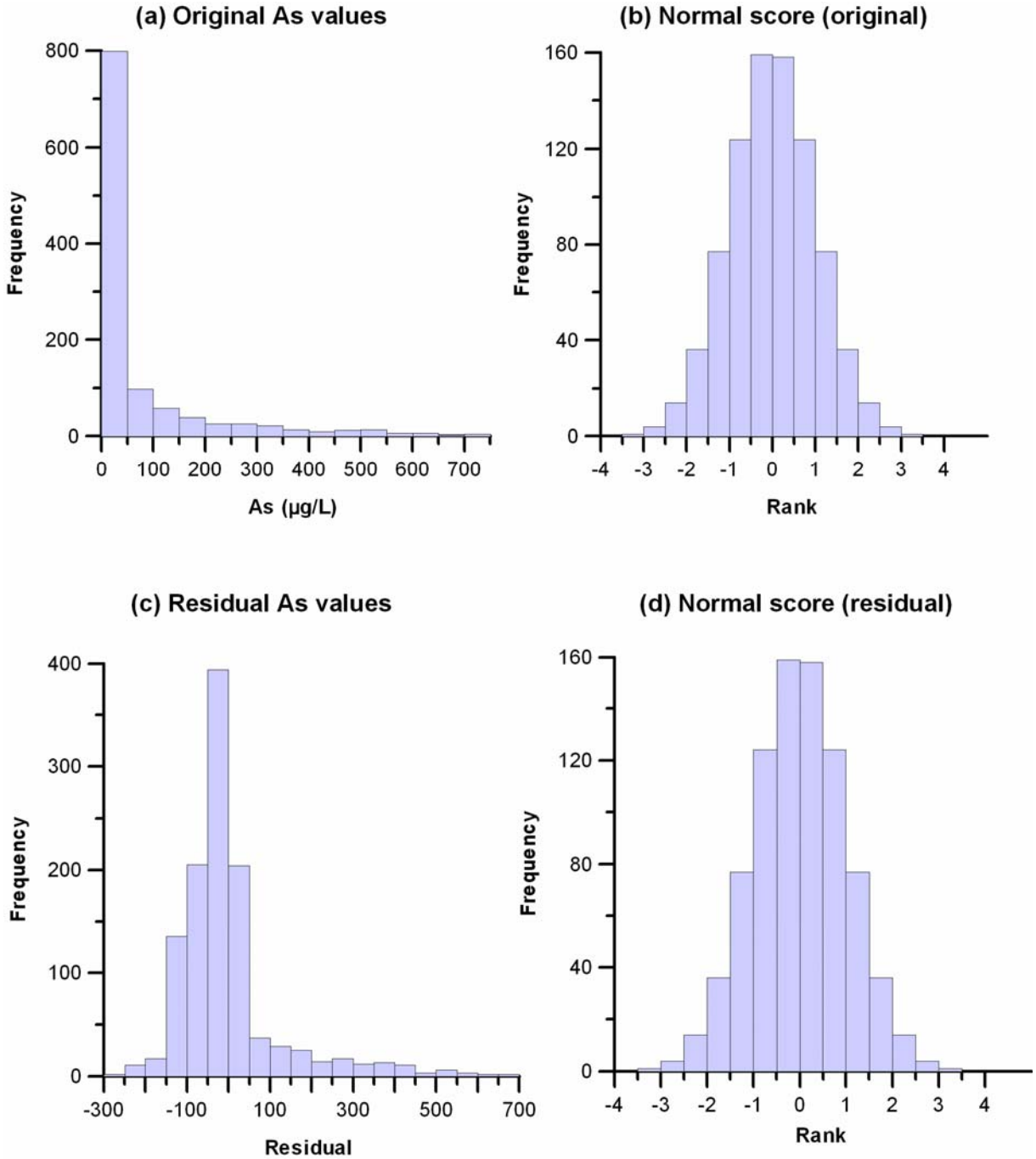

Figure 2. Histograms of original untransformed arsenic concentrations ("a" and "c") and their corresponding normal score transformed values ("b" and "d") in shallow aquifers of Bangladesh.

Inverse distance to power

One of the most commonly used deterministic interpolation methods is the inverse distance to power (IDP) method (Shepard 1968), which was developed on the concept that zones of influence decay with distance. This is a weighted average interpolator where data are weighted during interpolation such that

the influence of point relative to another declines with distance from the grid nodes. Weighting is assigned to data through applying a weighting power that controls the weighting factors, which drop off as the distance from grid-node increases. The greater the weighting powers, the less effect points far from the grid node have during interpolation. As the power increases, the grid node value approaches the value of the nearest point. This method also behaves as an exact interpolator when a particular smoothing parameter is applied for the interpolation.

Ordinary kriging

Kriging is a generalized least-square regression technique that allows one to account for the spatial dependence between observations, as revealed by the semivariogram into spatial predictor (Goovaerts 2000). In the kriging family, ordinary kriging (OK) is one of the most widely used interpolation methods (Deutsch and Journel 1998). The algorithm of this method is the moderate form of simple kriging (SK), where the kriging weights sum to one. The formula of ordinary kriging estimator is (Deutsch and Journel 1998):

$$Z^{*}_{OK}(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{(OK)}(u) Z(u_{\alpha}) \quad \textit{with} \quad \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{(OK)}(u) = 1 \qquad \text{Equation (1)}$$

The stationary OK system is expressed as follows:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_{\beta}^{(OK)}(u) C(u_{\alpha} - u_{\beta}) + \mu(u) = C(u_{\alpha} - u_{\beta}), \alpha = 1,...,n \\ \sum_{\beta=1}^{n} \lambda_{\beta}^{(OK)}(u) = 1 \end{cases} \qquad \text{Equation (2)}$$

Here, $\lambda_{\alpha}^{(OK)}(u)$'s are the OK weight and $\mu(u)$ is a parameter associated with the constraint $\sum_{\beta=1}^{n} \lambda_{\beta}^{(OK)}(u) = 1$. The information required by the kriging system are the semivariograms values for different lag distances, which are derived from variogram modeling (Goovaerts 2000).

In this study, the ordinary kriging (OK) on original data with the sample variogram model, and on normal-score (NS) data with normal score variogram, and on normal-score residual data with normal-score residual variogram are performed in order to observe the differences in arsenic prediction models.

Spatial Variability Analysis

Semivariogram analysis

Spatial variability of groundwater arsenic in Bangladesh is high at shallow depths (van Geen et al. 2003, Yu et al. 2003). High concentrations of groundwater arsenic, and the highest probability of exceeding Bangladesh standard of 50 µg L^{-1}, most often occur in tube wells screened between 20 m and 60 m (Ravenscroft et al. 2005). Most of the low arsenic concentrations are located in the northwestern portion of the country. In contrast, the higher values are concentrated in the south-central southwest deltaic parts and Sylhet trough (Fig.1).

Spatial pattern of variables are well described by 'semivariogram' which measures the dissimilarity or similarity between observation points as a function of the separation distance (Goovaerts 1997). The spatial dependency between separation vectors is measured by the experimental semivariogram, $\gamma(h)$, which is half of the average squared distance between two attribute values (u) in every data pair.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_{\alpha}) - z(u_{\alpha} + h)]^2 \qquad \text{Equation (3)}$$

The semivariance at short lag distances is the most important part of the semivariogram as the relationship reduces as the distance increases. A discontinuity of the semivariogram at the origin at zero lag distance, (h=0), which is called the "nugget". The variances arise from measurement errors or spatial sources of variation at distances smaller than the sampling interval or both (Journel and Huijbregts 1978). Generally, from the nugget, the value of semivariance increases with distance to a constant value, which is known as the "sill" of the variogram. The corresponding distance below the sill point is the "range" and beyond this range no correlation or covariance exists between two points in space. The structure of the semivariogram largely depends on the number of data and their locations in space and distance between sampling points. This study used 1,145 data points for spatial analysis and prediction modeling of arsenic concentrations in unsampled areas of Bangladesh. Omnidirectional semivariograms are computed from both original arsenic data and their normal-score transformed values of the wells of depth within 25 m from the surface (Fig. 3). Arsenic concentrations appear to vary more continuously up to a distance ranging from 30 km to 80 km as indicated in the semivariograms (Fig. 3). The semivariograms having a maximum lag distance of 50 km have shown continuous spatial correlation in arsenic variations. However, the spatial dependency of arsenic variations becomes very little with shorter lag distances and shows significant variability in concentrations as indicated in the semivariograms. The nugget variances representing the variability of wells located within a short distance, devoid of some measurement errors, are fairly high in the shallow wells (< 25 m), indicating the presence of high spatial variability at small spatial scales. It has been found that the measurement error of arsenic concentrations of the country is much smaller than 0.36 $(\log\text{-}\mu g/L)^2$ and thus most of these high nugget variances represents the small-scale variability of arsenic (Yu et al. 2003). In another study, it is found that about 88% wells having arsenic concentrations above the median are located within a distance of 100 m from the wells having arsenic concentrations below the median value (van Geen et al. 2003). This small- scale variability of arsenic concentrations also points toward the heterogeneity of local near-surface geology and variations in biogeochemical environments. The overall variation of arsenic concentrations increases with distance and this spatial structure is well shown in the semivariograms with larger lag distance. Beyond a distance of 80-100 km, the spatial dependency of arsenic concentrations in groundwater becomes random and more independent.

Semivariogram modeling

The semivariance is a measure of degree of spatial dependency between samples in space. A variogram model needs to be fitted through the experimental variograms to get the semivariances at all points. Modeling the spatial distribution of attribute values is the pre-requisite of spatial prediction of variables. A Gaussian model is fitted with the omnidirectional semivariograms of the original data, their residuals, and normal-score transformed residual values (Fig. 3). Both residual semivariograms and normal score semivariograms on detrended values are modeled that are used for spatial prediction modeling of arsenic concentrations.

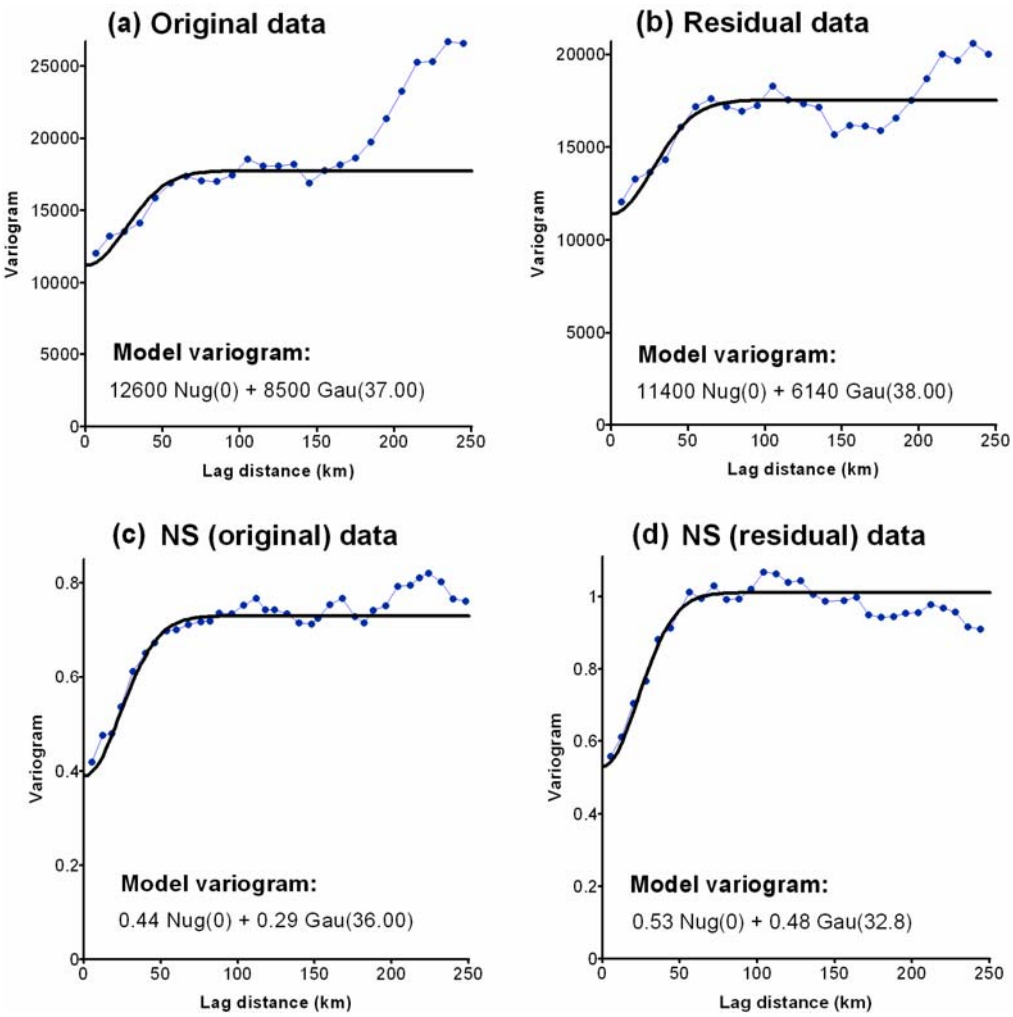

Figure 3. Semivariogram models on original (untransformed) arsenic concentrations, residual and their normal score transformed values. Gaussian models are fitted with all experimental semivariograms that are used in kriging interpolation. Spatial dependency significantly varies in the semivariograms with increasing lag distances. High "nugget" values on variograms indicate high spatial variability of arsenic within smaller distance in Bangladesh.

Results and Interpretations

Spatial prediction results

All predicted grids are shown as classified post maps (Fig. 4) that represent the predicted arsenic distributions in the shallowest alluvial aquifers (< 25 m) in Bangladesh. Conventional prediction methods (Natural neighbor and Inverse distance to power) on untransformed arsenic concentrations are applied. Interpolations by ordinary kriging method both on untransformed values (original arsenic concentrations) and their normal-score transformed values and their residual values are used (Fig. 4). Ordinary kriging method is also used on residual arsenic concentrations and their normal-score transformed values. At the end of the interpolation, the resulting grid values were back-transformed to create interpolated arsenic distribution maps.

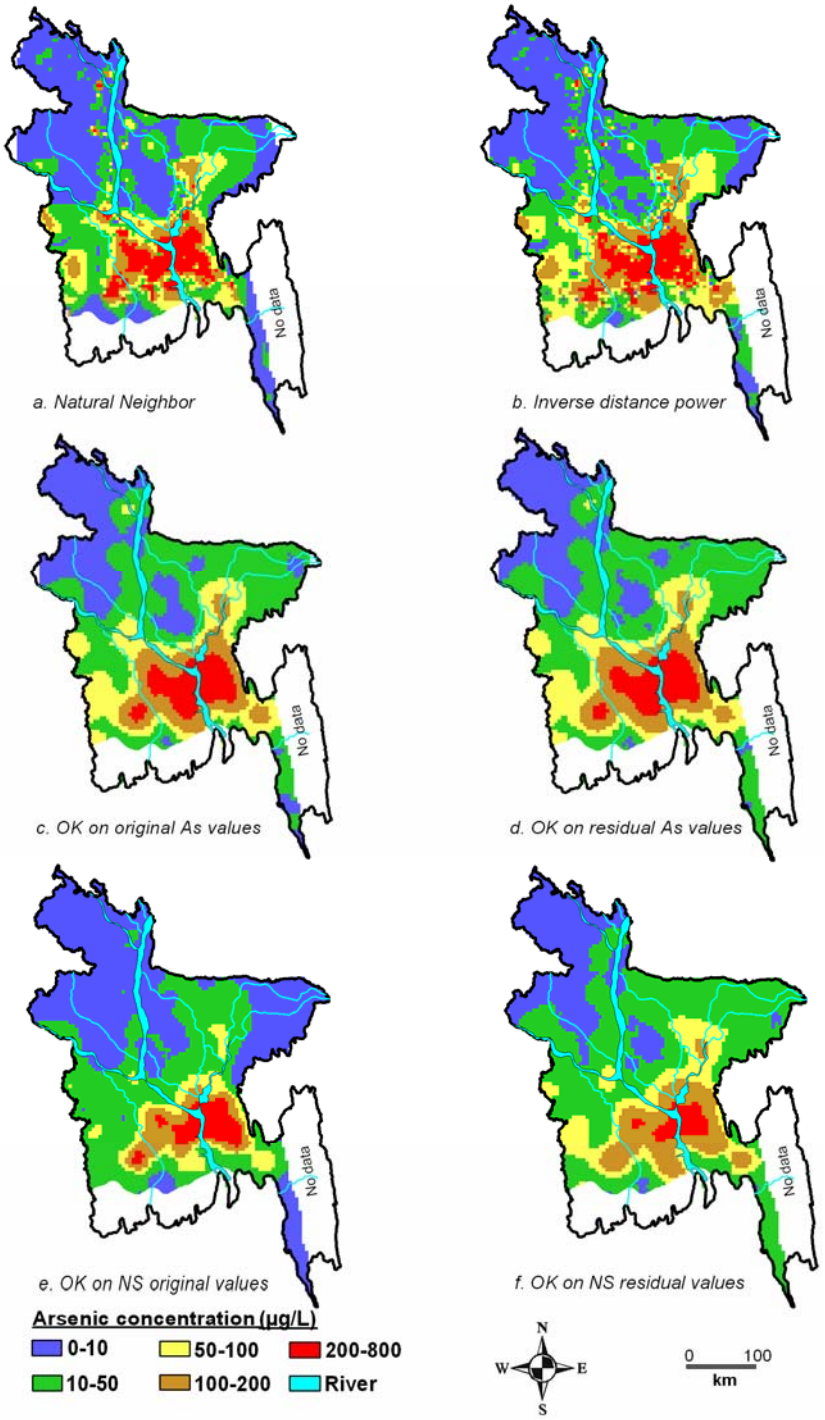

Figure 4. Predicted groundwater arsenic distribution maps in shallow alluvial aquifers obtained through interpolation of 1,145 data points using natural neighbor (a), inverse distance to power (b), ordinary kriging (OK) on original arsenic concentrations (c), residual arsenic values (d), normal-score transformed original (e), and normal-score transformed residual (f). No interpolation is performed in Chittagong Hill Tracts and coastal Bangladesh due to lack of substantial data.

Natural neighbor and inverse distance to power methods are the non-geostatistical deterministic interpolation methods, of which the natural neighbor has produced more realistic and detailed distribution patterns of groundwater arsenic concentration in space. Local arsenic variations are predicted fairly well by this method. Among the geostatistical methods, ordinary kriging on normal-score transformed values has predicted wells the regional arsenic distributions in the country. Due to smoothing effects, small-scale arsenic variations are underestimated by ordinary kriging method.

From interpolated result of the OK grids, a statistical analysis was performed on arsenic distributions and spatial coverage of groundwater arsenic concentrations in Bangladesh. Prediction modeling was performed based on 5,221 square grid nodes of a length of 5 km on each side of the whole model area. However, the total area of interpolation was estimated to be approximately 120, 500 sq. km, whereas the total area of Bangladesh is approximately 144,000 sq. km. Chittagong Hill Tracts and southern coastal parts of the country (Fig. 4) are not included in arsenic modeling since no sufficient well data are available from these areas. Estimation from the ordinary kriging on original arsenic data shows about 23% of the model area is found with arsenic concentrations below 10 µg L^{-1}. Approximately, 65% area of Bangladesh has arsenic concentrations within the country arsenic standard of 50 µg L^{-1}. More than 15% area has arsenic concentrations exceeding 50 µg L^{-1} up to a concentration of 100 µg L^{-1}. Approximately 17% model area has arsenic levels higher than 100 µg L^{-1}, which are located in southern deltaic areas in Bangladesh.

Model validation

One of the most commonly used validation methods is cross-validation, which has been used in this study. The interpolation error is computed from the observed and estimated values of arsenic for the prediction method prior to gridding on all the prediction locations. Cross-validation method is used to find the estimation errors or residuals associated with spatial prediction methods.

The interpolated results are analyzed by comparing the estimated values (z_i) with the observed values (z^*_i) for a number of the validation points (n) locations in order to assess systematic error, calculated as the mean prediction error (MPE):

$$\text{MPE} = \frac{1}{n} \cdot \sum_{i=1}^{n} [z_i^* - z_i] \qquad \text{Equation (4)}$$

The prediction errors are calculated as root mean square error (RMSE) that indicates the accuracy of the prediction. RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}} \cdot \sum_{i=1}^{n} (z_i^* - z_i)^2 \qquad \text{Equation (5)}$$

MPE method is quite useful in evaluating the performance of interpolation methods as this method is less sensitive to extreme values (Willmott 1984). The value of MPE should be close to zero if the interpolation method is to consider as a suitable one and the value of RMSE should be small. The larger the value of the RMSE, the greater is the discrepancy between the observation and estimated values.

The range of NN grided arsenic values are found quite identical with the input original data set. However, other gridding methods, particularly OK on the original data, and IDP method ignored the higher arsenic values and smoothing effects on arsenic distributions are observed in these prediction models. OK on normal-score transformed data and normal-score residual data have lower sample variances. OK methods show lower standard deviations indicating lower prediction errors associated with the arsenic modeling. Histograms of the corresponding errors are constructed to compare the prediction performance

and the accuracy of these methods (Fig. 5). Mean values for prediction residuals for NN and OK on original arsenic concentrations and their residual are small compared with the other prediction models. Error variances for OK on original values and their residual are small, which suggest higher degree of prediction accuracy of the distribution modeling of groundwater arsenic in Bangladesh.

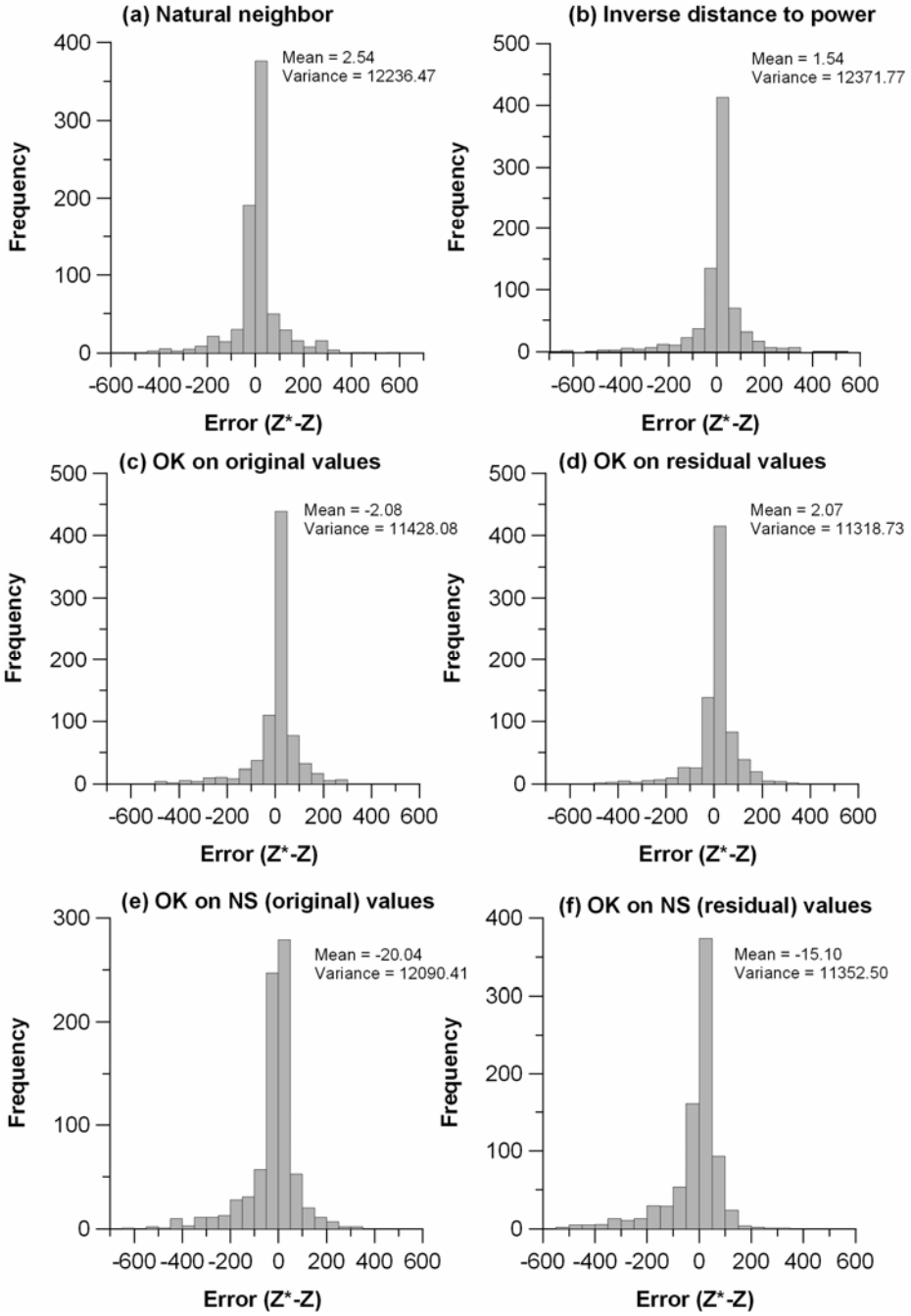

Figure 5. Histograms of estimation error or residual of each prediction method applied in this study. Figures “a” through “f” illustrate the prediction errors of different prediction models. Mean and variances of prediction errors indicate ordinary kriging on untransformed arsenic concentration and their residual values performed reasonably in predicting groundwater arsenic in shallow alluvial aquifers in Bangladesh.

Modeling the spatial distributions of groundwater arsenic concentrations has been a challenging task over the years. Many authors tried to explore the spatial dependency of arsenic levels and estimate in unsampled areas using available datasets (e.g., Karthik 2001, Ravenscroft et al. 2001, Yu et al. 2003, Gaus et al. 2003, van Geen et al. 2003, Shamsudduha 2004, Hossain et al. 2005). It has been understood that low density data and extremely high small-scale spatial variability coupled with vertical variations in arsenic concentrations, especially within the Holocene alluvial aquifers and presumably high variations in the biogeochemical processes make prediction of arsenic levels difficult and uncertain. Results from many previous researches and this study suggest that the application of other geostatistical methods such as probability kriging, co-kriging and sequential Gaussian simulation methods might be interesting and worthwhile to generate more accurate arsenic distribution maps and thus manage this disastrous environmental problem in Bangladesh and other arsenic-affected areas around the world.

Conclusions

Spatial variability of groundwater arsenic distribution at very shallow depths in Bangladesh is extreme with variability increasing at smaller distances in alluvial aquifers. Variogram analysis suggests that the high degree of spatial variability of arsenic levels are resulted not from analytical errors or due to poor locationing of samples, rather closely associated with high degree of spatial heterogeneity of aquifer materials and biogeochemical processes. Spatial patterns of arsenic concentrations in shallow wells also indicate that these broad-scale patterns in arsenic distributions are possibly controlled by regional geologic-geomorphic divisions in Bangladesh.

Different interpolation procedures are applied and results are compared to evaluate the most appropriate prediction method for the estimation of arsenic concentrations in the shallow aquifers of Bangladesh. From cross-validation, mean prediction error and biasness analyses it is found that ordinary kriging on original arsenic concentrations and their residual values produced better prediction models for spatial arsenic distribution in shallow aquifers. However, natural neighbor method shows the local and small-scale variability of arsenic in shallow aquifers. The inverse distance method gives realistic estimation results, but prediction variance is found higher than the other methods. Ordinary kriging on normal-score residual data under estimated high arsenic concentrations and smoothed out the regional distribution patterns in the country.

Statistical analysis on grided results suggests that approximately 65% model area in Bangladesh has arsenic concentrations within the country arsenic standard (50 µg L^{-1}). More than 15% area has arsenic concentrations exceeding 50 µg L^{-1} up to a concentration of 100 µg L^{-1}. About 17% model area has arsenic concentrations higher than 100 µg L^{-1}, which are mainly located in the south-central deltaic areas, known as “hotspots” of arsenic contamination in Bangladesh.

Detailed geostatistical analysis on arsenic distributions will help to gather more information on spatial correlation and variability of arsenic concentrations in alluvial aquifers. More careful statistical analyses should be performed to find out other continuous and categorical variables well correlate with arsenic that can be incorporated with geostatistical arsenic prediction models both on local- and regional-scales. More local-scale studies on high density datasets should be performed to evaluate prediction models of arsenic distributions in Bangladesh.

Acknowledgments

The author thanks Australian Agency for International Development for supporting this research at University of Technology Sydney (UTS), Australia. Thanks to Dr. Bryce Kelly of UTS for his training. The author acknowledges British Geological Survey (BGS) and Department of Public Health Engineering (DPHE) Bangladesh for creating a complete public-domain database used in this study. Comments and suggestions from two anonymous reviewers that improved this manuscript are greatly acknowledged.

References

Ahmed, K.M., Bhattacharya, P., Hasan, M.A., Akhter, S.H., Alam, S.M.M., Bhuyian, M.A.H., Imam, M.B., Khan, A.A., Sracek, O. (2004) Arsenic enrichment in groundwater of the alluvial aquifers in Bangladesh: an overview, Applied Geochemistry 19, 181-200.

Anderson, T. (1958) An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons.

BGS and DPHE (2001) Arsenic contamination of groundwater in Bangladesh, British Geological Survey Report WC/00/19, British Geological Survey, Keyworth, UK.

Bhattacharya, P., Chatterjee, D., Jacks, G. (1997) Occurrence of arsenic-contaminated groundwater in alluvial aquifers from delta plains, eastern India: Options for safe drinking water supply, Water Resources Development 13(1), 79-92.

Deutsch, C.V., Journel, A.G. (1998) GSLIB: Geostatistical Software Library and User's Guide, 2nd ed., New York: Oxford University Press.

Gaus, I., Kinniburgh, D.G., Talbot, J.C., Webster, R. (2003) Geostatistical analysis of arsenic concentration in groundwater in Bangladesh using disjunctive kriging. Environmental Geology, 44, 939-948.

Goovaerts, P. (1997) Geostatistics for Natural Resources Evaluation. New York: Oxford University Press.

Goovaerts, P. (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of Hydrology, 228, 113-129.

Hossain, F., Sivakumar, B. (2005). Spatial pattern of arsenic contamination in shallow wells of Bangladesh: regional geology and nonlinear dynamics. Stoch. Environ. Res. Risk Assess., 20, 66–76. DOI 10.1007/s00477-005-0012-7.

Hossain, F., Bagtzoglou, A.C., Nahar, N, Hossain, M.D. (2005) Statistical characterization of arsenic contamination in shallow tube wells of western Bangladesh. Hydrol. Proc., 19 DOI 10.1002/hyp.5946.

Johnson, R., Wichern, D. (1982) Applied Multivariate Statistical Analysis, New Jersey: Englewood Cliffs, Prentice Hall.

Journel, A.G., Huijbregts, C.J. (1978) Mining Geostatistics. San Diego: Academic Press, New York.

Karthik, B. (2001) Spatial variability of groundwater arsenic in Bangladesh: An evaluation of geologic and physical controls. M.Sc. thesis (unpublished), Department of Civil and Environmental Engineering, University of Cincinnati, USA.

McArthur, J.M., Ravenscroft, P., Safiullah, S., Thirlwall, M.F. (2001) Arsenic in groundwater: testing pollution mechanisms for sedimentary aquifer in Bangladesh, Water Resources Research, 37, 109-117.

Nickson, R.T., McArthur, J.M., Ravenscroft, P., Burgess, W.G., Ahmed, K.M. (2000) Mechanisms of arsenic release to groundwater, Bangladesh and West-Bengal, Applied Geochemistry, 15, 403-413.

Olea, R.A. (1999) Geostatistics for engineers and earth scientists. Boston: Kluwer Academic Publishers.

Ravenscroft, P, McArthur J.M., Hoque B.A. (2001) Geochemical and palaeohydrological controls on pollution of groundwater by arsenic. In: Arsenic Exposure and Health Effects IV (Eds. W.R. Chappell, C.O. Abernathy and R.L. Calderon), pp. 53-77, Elsevier, Oxford.

Ravenscroft, P., Burgess, W.G., Ahmed, K.M., Burren, M., Perrin, J. (2005) Arsenic in groundwater of the Bengal basin, Bangladesh: distribution, field relations, and hydrogeological setting, Hydrogeology Journal, 13, 727-751.

Ravenscroft, P., Howarth, R.J., McArthur, J.M. (2006). Comment on "Limited temporal variability of arsenic concentrations in 20 wells monitored for 3 years in Araihazar, Bangladesh". Environ. Sci. Technol., 40(5), 1716-1717.

Sengupta, M.K., Mukherjee, A., Ahamed, Hossain, S.M.A., Das, B., Nayak, B., Chakraborti, D. (2006). Comment on "Limited temporal variability of arsenic concentrations in 20 wells monitored for 3 years in Araihazar, Bangladesh ". Environ. Sci. Technol., 40(5), 1714-1715.

Shamsudduha, M. (2004) Geostatistical and multivariate statistical analyses on the widespread arsenic problem in the groundwater of Bangladesh, M.Sc. thesis (unpublished), University of Technology Sydney, NSW, Australia.

Shepard, D. (1968) A two-dimensional interpolation function for irregularly-spaced data. Proceedings of 23rd National Conference ACM, pp. 517-524.

van Geen, A., Zheng, Y., Versteeg, R., Stute, M., Horneman, A., Dhar, R., Steckler, M., Gelman, A., Small, A., Ahsan, H., Graziano, J., Hussein, I., Ahmed, K.M. (2003) Spatial Variability of Arsenic in 6000 Contiguous Tube Wells of Araihazar, Bangladesh, Water Resources Research, 39 (6), 1140.

Watson, D. (1999) The natural neighbor series manuals and source codes, Computers & Geosciences, 25, 463-466.

Willmott, C.J. (1984) On the evaluation of model performance in physical geography. In: Spatial Statistics and Models, edited by G.L. Gaile, and C.J. Willmott, pp. 443-460. Holland: Dordrecht.

Yu, W.H., Harvey, C.M., Harvey, H.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.

Zheng, Y., van Geen, A., Stute, M., Dhar, R., Mo, Z., Cheng, Z., Horneman, A., Gavrieli, I., Simpson, H.J., Versteeg, R., Steckler, M., Grazioli-Venier, A., Goodbred, S., Shahnewaz, M., Shamsudduha, M., Hoque, M.A., Ahmed, K.M. (2005). Geochemical and hydrogeological contrasts between shallow and deeper aquifers in two villages of Araihazar, Bangladesh: Implications for deeper aquifers as drinking water sources. Geochim. Cosmochim. Acta., 69 (22), 5203-5218.