SPATIAL DISTRIBUTION OF ARSENIC CONCENTRATION IN DRINKING WATER USING KRIGING TECHNIQUES

Sana Saeed [1], Zahid Javed[1,*], Sohail Chand[1], Naimatullah Hashmi[2], Maqsood Ahmad[1]
[1]College of Statistical and Actuarial Sciences, University of Punjab, Lahore, Pakistan
[2]Department of Statistics, Govt. P.G. College Satellite Town, Rawalpindi, Pakistan
[*]Corresponding author's e-mail: zzahidgee@yahoo.com

ABSTRACT: *Arsenic is an organic component of the actual earth's crusting and is broadly distributed through the entire environment within air, drinking water and land. People are subjected to elevated amounts of inorganic arsenic through consuming contaminated drinking water, using infected water within food preparation along with irrigation associated with food vegetation, eating infected food and cigarette smoking. Pores and skin lesions with skin malignancy are the majority of characteristic results. The WHO, PCRWR and US Environment Protection Agency suggests Arsenic standards levels at 10 ppb for drinking water. In this research only 6% of the samples were below WHO standard and other 94% were above standard limits. Box-Cox transformation was executed for making the response variable normal. Empirical variogram was graphed on the dataset and factors of theoretic variogram techniques (Matern, Exponential, Spherical and Gaussian) were estimated by using ordinary least square estimation method. The performance of the models was checked by cross validation and it was noted that Gaussian model has least value of RMSPE so it was the best model. Ordinary Kriging and Bayesian Kriging were used for forecasting spatial structure of Arsenic in drinking water of district Faisalabad, Punjab, Pakistan.*

Keywords: Arsenic Concentration, Drinking Water, District Faisalabad, Spatial Distribution, Variogram.

INTRODUCTION

Geostatistics can be observed as aassortment of numerical techniques which deal with the description of spatial features, using mostly random models in a way alike to the manner in which the time series exploration illustrates temporal data[1]. Recently, human population surge, urbanization sprawl and quick industrialization possess have created many environmental problems and difficulties regarding worldwide climate modifications, contaminants of groundwater resources, worldwide warming, nuclear and technical hazardous waste models and many more problems in the entire world[2].In the developing countries, a massive part of the population undergoes health harms related to either scarcity of ground water or because of the existence of microbiological adulteration in water. Underprivileged water eminence is blamable for the decease of probable five million youngsters in developing countries[3]. Pakistan is also dealing with drinking water high quality crisis and its ranking place is at quantity 80 amongst 122 countries regarding drinking water quality [4]. Every year, about 250,000 kids die because of waterborne illnesses only within Pakistan[5]. It was observed that 40% of all deaths and 30% of diseases were because of poor drinking water quality in Pakistan [7]. Diarrhea, water paid for disease was actually reported since the leading reason for death within infants as well as children in the country while each and every fifth resident suffers from sickness and illness caused by the actual polluted drinking water[8].

Various countries straight their specific water quality requirements and recommendations for meeting their nationwide priorities and necessities of the environment as well as cultural position [3]. A short listing of the standard maximum allowable limits of Arsenic of different countries and organizationswas presented in Table 1 which revealed that the Arsenic meditation in drinking waternis exceeding the standards. The relationship of high Arsenic standards with rice yield, known to stimulate reduced surroundings due to ponding, additional corroborates this assumption. [6]Evaluated the environmental condition of Asia as it is the most pointedly affected zone for arsenic adulteration around the world. 100 million peoples in Asia were at danger of arsenic polluted water sources and further more than 700,000 cases had been described for arsenic associated diseases. The long term exposure to arsenic in drinking water can be the source of cancer in the skin, lungs, bladder and kidney. It can also be the reason of other skin fluctuations such as thickening and pigmentation. Different geostatistical tools help for the purpose of interpolation of the spatial distribution of physio-chemical parameters of water.

The spatial techniques had been deliberated as the supreme advanced way of exclamation as well as yield best consequences as they give impartial assessments along with minimal value of spatial prediction error. There was used model based Ordinary Kriging and Bayesian Kriging with regard to predicting the actual spatial framework of Arsenic in pure consuming groundwater associated with district Faisalabad, Punjab Pakistan. Box-Cox transformation was used for normalization of observed variable. Within the first stage, spatial exploratory analysis was built to detect the spatial performance. In the second stage, empirical variogram was attracted from the dataset and variables of theoretical variogram models (Matern, Exponential, Spherical and Gaussian) are estimated by using ordinary least square estimation methods. The performance of the variogram models and interpolation techniqueswere checked by cross validation and it was noted that a model with least value of RMSPE is considered the best model.

MATERIAL AND METHODS

Description of Study Area

Faisalabad is the 3rd biggest city of Pakistan and 2nd largest city in Punjab province. Itis a major developing city of Pakistan. It is situated at northeast side of the Punjab, amongst longitude 73°74 East, latitude 30°31.5 North, having an elevation associated with 184 meters distance (604 ft) above sea level [9].Due to its greatest evapotranspiration, Faisalabad has a very warm weather. We collected 70 spatial samples from different areas of Faisalabad, and Arsenic concentration (ppb) was tested in every sample through laboratory.The locations of each spatial sample are shown in the Figure 1.

Table 1. Different Organization Maximum Acceptable Limit Standard for As

Sr.#	**Country or organization Name**	**Arsenic (ppb)**
1	World Health Organization (WHO)	10
2	US Environmental Protection Agency (US EPA)	10
3	Pakistan Council of Research in Water Resources (PCRWR)	10
4	Pakistan Standard and Quality Control Authority (PSQCA)	10
5	Indian Water Quality Standards	10

Variogram

The actual variogram tackles three primary parameters which are Sill, Range and Nugget. The incomplete sill may be the quantity of variance in the procedure which is expected to create data. Range is the distance beyond that data which don't have significant record dependence. Nugget is the information variation because of measurement mistake and is the discontinuity in the origin. Usually variogram designated as follow.

$$\gamma(k) = \frac{1}{2N(k)} \sum_{i=1}^{N(k)} [Z(x_i + k) - Z(x_i)]^2$$

$\gamma(k)$ is actually used to measure the spatial relationship. Here k is the space between localities and x+k be the number of statement sets divided by range k [10]. Graphically the information gained by variogram is expressed in figure 2.

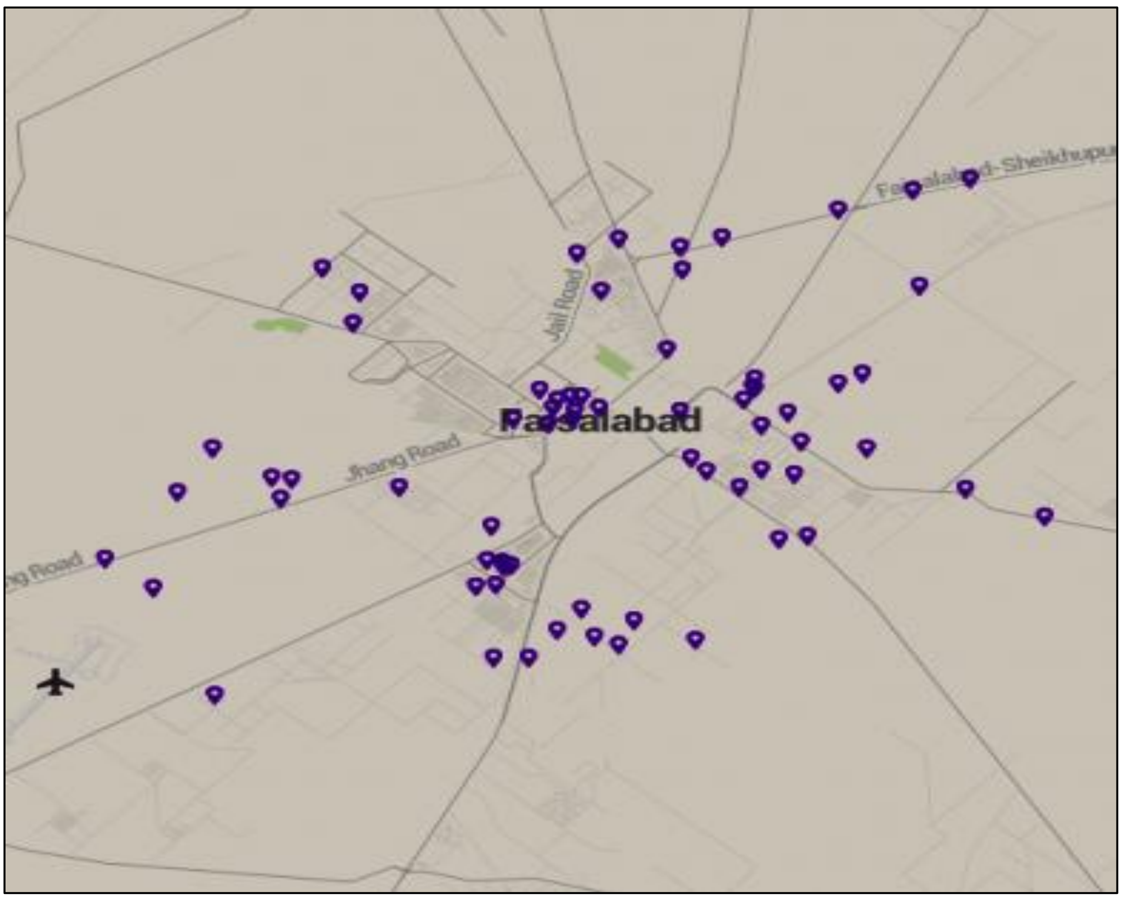

Figure 1 Sampled locations of Arsenic Concentration in district Faisalabad

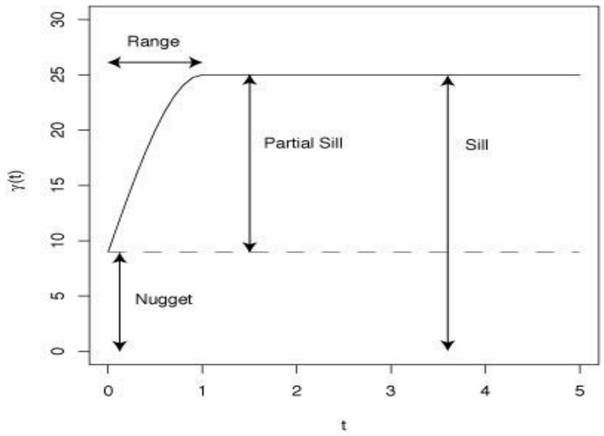

Figure 2 Sketch of Variogram along with its parameters: Sill, Range and Nugget

Description of Spatial Covariance Functions

A suitable spatial covariance functions capture the autocorrelation structure between observations of study domain [11] . Well known spatial covariance functions are Matern, Exponential, Gaussian and Spherical.TheMatern covariance model takes the form

$$\gamma(h)=\tau^2+\sigma^2(1-(|h|/\phi)^v)k_v(|h|/\phi)$$

With regard to $h>0$ as well as all τ^2, σ^2, $v and \emptyset \geq 0$ where K_v is altered bassel functionality of rank v. This particular variogrammodel is an advanced method among Gaussian model and Exponential model. The Exponential model for spatial correlation is described by

$$\gamma(h)=\tau^2+\sigma^2(1-\exp(-|h|/\phi))$$

With regard to $h>0$ as well as τ^2, $\sigma^2 and \emptyset \geq 0$.The range associated with Exponential method relies on the value of factor $\emptyset$[11]. The Gaussianmathematical model with regard to $h>0$ is given as

$$\gamma(h)=\tau^2+\sigma^2(1-\exp(-|h|^2/\phi))$$

Where τ^2, $\sigma^2 and \emptyset \geq 0$.The increasing act with embrace h is actually same as exponential model however it behaves parabolically near the source[12]. This particular model features a practical selection of $3\emptyset$.The equation of the spherical model is described as follow:

$$\begin{cases} \tau^2+\sigma^2\left(\frac{3|h|}{2\emptyset}+\frac{|h|^3}{2\emptyset^3}\right) & 0<|h| \leq \alpha \\ \tau^2+\sigma^2|h|>0 & \end{cases}$$

Whereτ^2, $\sigma^2 and \emptyset \geq 0$. Spherical model slowly rises after the nugget effectτ^2 to sill quantity τ^2, σ^2 is spatial lag quantity $h \geq \emptyset$ and continuously reduce through its maximum quantity σ^2 to absolutely zero[11,13].

Cross Validation Statistics

Root Mean Square Prediction Error (RMSPE) was used to predict the overall performance of spatial interpolation and Kriging techniques. RMSPE is thought as precision extent for spatial interpolation techniques which how accurate the forecasts. Little beliefs of this statistics reveal whichinterpolator gives the dependable approximation for unidentified areas whereas Mean Prediction Error is steps of interpolator bias. With regard to balanced interpolator the beliefs of Mean Prediction Error must be near to zero. Actual equations of most the cross validation statistics with regard to spatial interpolator techniques get under[15].

$$RMSPE=\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}(s_i)-y(s_i))^2}{n}}$$

Ordinary Kriging (OK)

Suppose Y is a variable in which ($Y_1,\ldots,Y_n$) arevalues of observedvariable calculated through monitored sites. [16]Offers planned program of OK in order to forecast value of observedvariable in unnoticed area, where x= $\sum_i^n a(x_i)$, here the standards associated with a(x_i) are approximated exhausting the system of linear equality. [14]suggestedmodelbased OK that can be written in the shape of Gaussian technique like $Y_i=X_i+Z_i$: i =1,…,n. in which the Z_i tends to be mutually independent as well as normally dispersed having $\mu=0$, $\sigma=\tau^2$ and S(X_i) actually 2nd degree fixed Gaussian unsystematic field. Suppose S(x) = (S(x_1),…,S(x_n)) remain the actual unnoticed indicators at noticed sites as well as $S(x)$ is actually Multivariate Gaussian Distributed along with mean direction μI wherever I means a vector in which every components is actually 1 as well as Variance Matrix $\sigma^2 R$, wherever R may be $n*n$ relationship matrix along components $r_{ij}=\rho\,|x_i-y_i|$. The fundamentals r_{ij} associated with matrix R is approximated through consuming legitimate covariance models[17]. This is apparent which observed factor might be Gaussian along with mean direction μI as well as Variance Marix$\sigma^2 v=\sigma^2(R+vI)=\sigma^2 R+\tau^2 I$.

Right it would like to predict the value of response factor in ungauged areas so to exclaim actual (x) in undetected area focus to minimization of MSE extrapolation associated with (x). Beliefs of indicators of undetected position could assess;

$$\hat{S}(x)=\mu+\sum_{i=1}^{n}a_i(x)(y_i-\mu)$$

$$\hat{S}(x)=\left[1-\sum_{i=1}^{n}a_i(x)\right]\mu+\sum_{i=1}^{n}a_i(x)\,y_i$$

Wherever $a_i(x)$ is conjecture weight and approximated through Ordinary Kriging[14].

Bayesian Kriging (BK)

BK estimation suggested through [18] had advantages on OK estimation that within BK entirely variables are handled on the basis of uncertainty. In BK the actual models tend to be assessed utilizing Bayes theorem that utilize the prior distribution (θ) of particular parameters and also the likelihood functionality $(\theta;\, y)$ wherever $\theta=(\beta, \sigma^2, \phi)$ would be the parameters associated with covariance model. The producing posterior distribution of parameter θ might be conveyed below;

$$p(\theta|y)=\frac{l(\theta;y).\pi(\theta)}{\int l(\theta;y).\pi(\theta)d\theta}$$

Additional the Bayesian kriging extrapolative distribution within model dependent frame function is recommended by[19].

Adopt a prior $\pi\ (\beta,\sigma^2,\phi) = \pi\ (\beta,\ \sigma^2\ |\phi)\ \pi(\phi)$, here $\pi\ (\phi)$ was definitely an independent prior distribution associated with $\pi(\phi)$ that stipulate this as under the radar distribution comprising in practice. Here adopting the prior $\pi\ (\beta,\ \sigma^2,\ \phi) = \pi\ (\beta,\ \sigma^2\ |\phi)\ (\phi)$, and $\pi\ (\phi)$ is definitely an independent prior distribution associated with (ϕ) that identified like a discrete distribution covering used[20]. The discrete distribution containing the additional sensible variety because it is makes calculation simpler. The posterior distribution for factors after that becomes

$[\beta,\ \sigma^2\ |y] = [\beta,\ \sigma^2|\phi][\phi|\ y]$

The selection of posterior and prior distribution has fundamental importance in Bayesian interpolation of spatially latentvariable [19].

RESULTS AND DISCUSSIONS

The analysis of the underlying spatial data was carried out on the basis of geo-statisticalkriging techniques using geoR package [14] of R statistical computational language. Firstly descriptive exploratory spatial data analysis of Arsenic (As) concentration was carried out and it was noticed that observed variable was positively skewed while the fundamental assumption about spatially varied latent variable is normality. Box Cox transformation was used to transform the variable with lambda= 0.2551355 using the expression

Transformed Variable = [{Original Variable} $^{0.2551355}$]-1 / (0.2551355)

The transformed value of lambda= 0.2551355 could be inserted in every function of underlying software to analyze the underlying spatial data.

There are plenty of theoretical models accessible to fit upon empirical variogram, the common theoretical models tend to be Matern, Exponential, Gaussian, Spherical, Cubic as well as Powered etc. There was installed only the Matern, Exponential, Gaussian and Sphericalmodels on empirical variogram it take the value associated with parameters sill, range and nugget within the above stated models. Figure 4 reported the level of observed Arsenic concentration (ppb) level in the studied area.

Variogram Envelope diagram for examining spatial relationship of observed factor demonstrated under figure 5 (left panel) which shows that most of the points were inside the variogram envelope so there was a durable spatial correlation. Variogram model's parameters are predicted by using ordinary least square (OLS) method. The presentation of these estimation methods was equated with the help of cross validation. The root mean square prediction errors were also estimated to choose the best estimation technique. There was a fitted different combination of spatial covariance models on the empirical variogram in Arsenic concentration level. RMSPE is also calculated for these methods for selection of best model fit. The graphical display of all spatial covariance models are given in figure 5 (right panel).

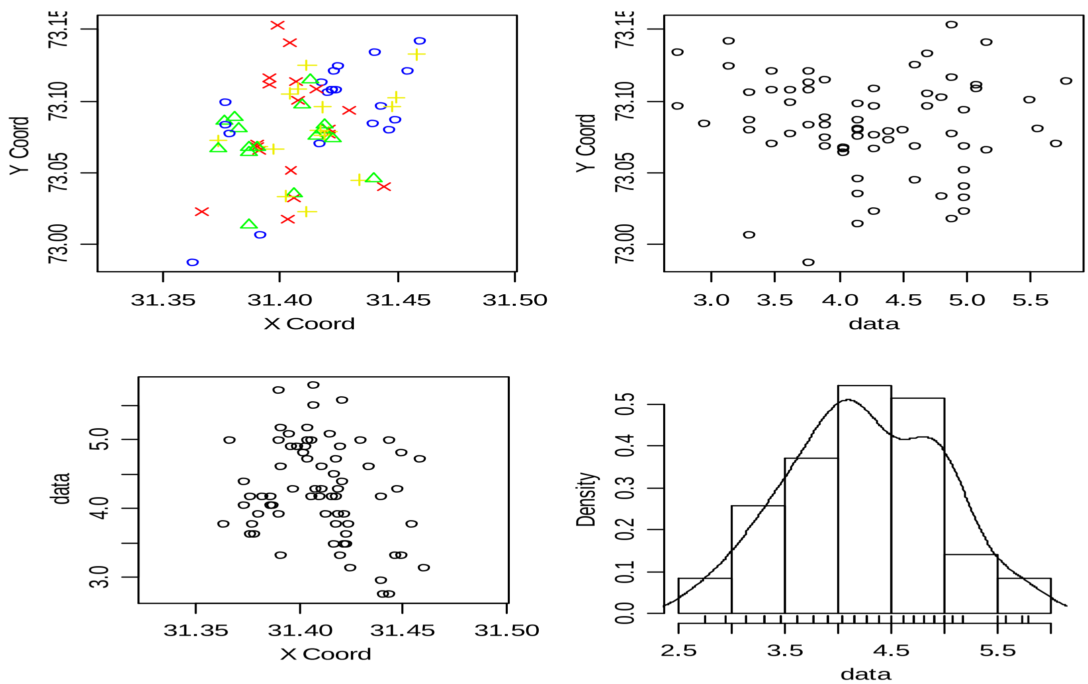

Figure 3. Exploratory data analysis of underlying spatial data of Arsenic concentration after Box-Cox transformation

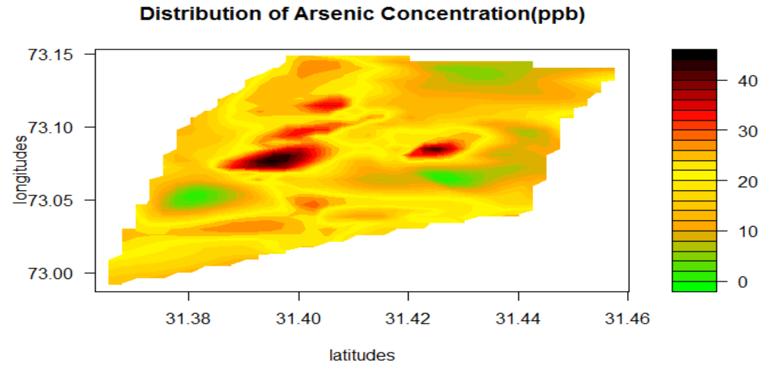

Figure 4. Distribution of observed Arsenic Concentration (ppb)

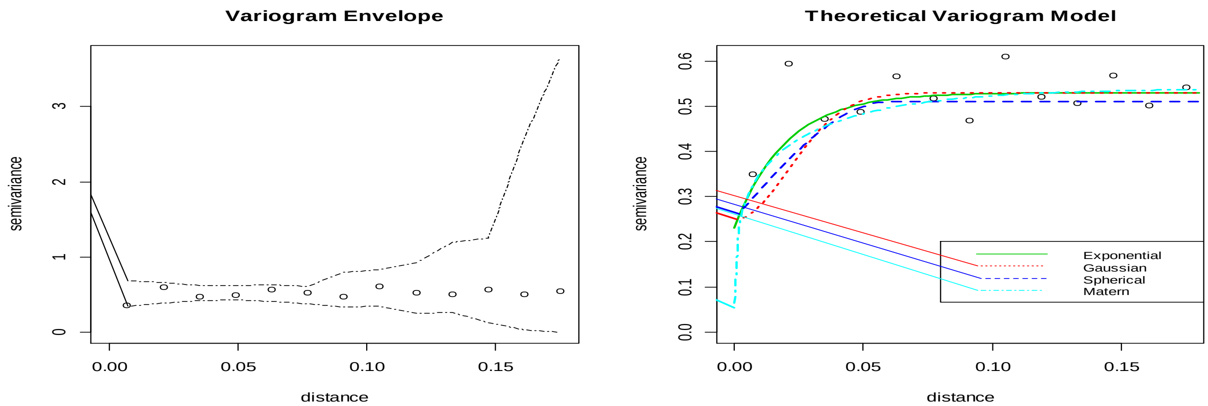

Figure 5.Variogram envelope for detecting spatial serial correlation is given in left panel while Empirical variogram models fitted over theoretical model is given in right panel.

Table 2: Parameter Estimation of Spatial Arsenic Concentration data

Estimation Method	Variogram Model	Sill σ^2	Range φ	Nugget τ^2	Validation (RMSPE)
Ordinary Least Square (OLS)	**Matern**	17.6089	0.005613	20.5280	4.2277
	Exponential	17.3603	0.005212	20.8038	4.2302
	Spherical	18.3329	0.00005	19.4412	4.1633
	Gaussian	17.3737	0.000043	18.51214	4.1412

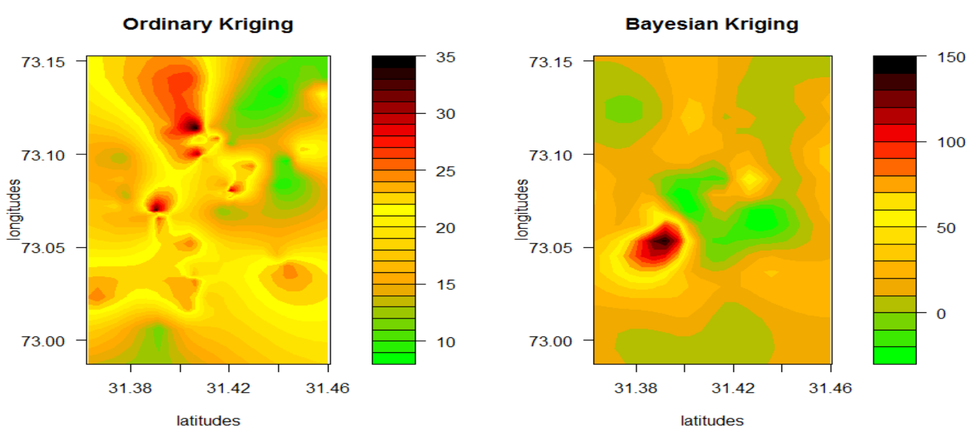

Figure 6. Spatial prediction map of Arsenic Concentration (ppb)using Ordinary Kriging and Bayesian Kriging

In table 2, combinations of ordinary least square (OLS) estimation method with Gaussian spatial covariance model perform well because RMSPE has the minimum value.Four differentcovariance structures were fitted like Matern, Exponential, Spherical and Gaussian structures. Here ordinary least square (OLS) estimation technique along with Gaussian variogram model is used for further prediction because it yields minimum prediction error. The comparison between Ordinary kriging and Bayesian kriging for Arsenic was made in the figure 6.

The maximum level of Arsenic is 35 with longitude 73.1139 and latitude 31.4067 as shown by yellow area in the figure 6 and the minimum level is 8 having longitude 73.0964 and latitude 31.4435 by red area. Top right area of study domain is showing least level of Arsenic concentration while bottom right area showed its average value. Several contours of varying levels could easily be judged through these interpolations maps.

Table 3 Comparison of different kriging techniques using RMSPE

Kriging Method	Ordinary	Bayesian
Validation RMSPE	5.8732	2.1652

In the table 3 the comparison of Ordinary and Bayesian Kriging was made by calculating the RMSPE. In the underlying data, the Bayesian Kriging technique is chosen for prediction because it yields low value of RMSPE which is 2.1652. Thus prediction of unmonitored location through Bayesian Kriging will always be beneficiary for the underlying study domain.

CONCLUSIONS AND RECOMENDATIONS

In this article spatial data on Arsenic concentration at 70 locations of District Faisalabad was studied. The maximum value of Arsenic (As) was observed to be 35 ppb and minimum value was 8 ppb whereas the permissible limit according to WHO and PCRWR was 10 ppb. As the level of Arsenic increasing 10, thus water was not suitable for drinking purposes because Arsenic is a poison and it is injurious to health. It causes many severe diseases. The high value of Arsenic can cause of cancer. We predict the Arsenic concentration level using model based Ordinary Kriging (OK) and Bayesian Kriging (BK) approach at some unmonitored locations. On the basis of interpolation maps, varying level of Arsenic concentration at gauged and ungauged sites were observed. It was highly recommended for the inhabitants of that locality to be aware of the harms of contaminated drinking water. Government should take steps to fulfill sufficient health based necessities for peoples of district Faisalabad. Water purification plants should be installed to avoid the harmful effects of drinking water.

BIBLOGRAPHY

[1] R.A. Olea, "Geostatistics for Engineers and Earth Scientists," *Kluwer Acad. Publ.*, p. 303, 1999.

[2] P. C. Mishra, "Some Aspects of the Quality of Water in and around Rourkela," 2005.

[3] A. I. A. S. S. Andrade and T. Y. Stigter, "The distribution of arsenic in shallow alluvial groundwater under agricultural land in central Portugal : Insights from multivariate geostatistical modeling," *Sci. Total Environ.*, vol. 449, pp. 37–51, 2013.

[4] D. H. Azizullah, Peter Richeter, "Ecotoxicological Evaluation of Wastewater Samples from Gadoon Amazai Industrial Estate (GAIE), Swabi , Pakistan," *Int. J. Environ. Sci.*, vol. 1, no. 5, pp. 959–976, 2011.

[5] "World Wide Foundation," *http: //www. worldwildlife. org/initiatives/fresh-water*, 2007.

[6] A. A. Uqaili, A. H. Mughal, and B. K. Maheshwari, "Arsenic Contamination in Ground Water Sources of District Matiari , Sindh," *Int. J. Chem. Environ. Eng.*, vol. 3, no. 4, pp. 259–266, 2012.

[7] "Country Report – Pakistan," *www. unesco. org/education/wef/countryreprts/pakistan/rapport-1.html*, 2000.

[8] P. B. Kahlown, M. A., Tahir, M. A., Hifza, R., & Khalid, "Water Quality Status of Pakistan: Forth Technical Report National Water Quality Monitoring Programe," *Pakitan Counc. Reserach Water Resour. Islam. Pakistan*, 2005.

[9] "Population Statistics and Maps," 2014.

[10] G. Bohling, "Introduction to Geostatisticsand Variogram analysis," *Kansas Geol. Surv.*, 2005.

[11] J. V. Le, N. D., & Zidek, "Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging," *J. Multivar. Anal.*, vol. 43, no. 2, pp. 351–374, 2006.

[12] A. Hengl, T., Heuvelink, G. M. B., Stein, "A generic frame work for spatial prediction of soil variables based on regression kriging.," *Geoderma*, vol. 122, no. 1–2, pp. 75–93, 2004.

[13] T. Hengl, *A Practical Guide to Geostatistical Mapping*. 2009.

[14] P. J. Diggle, P. J., Ribeiro Jr, "Model-based Geostatistics. Springer Series in Statistics," *Springer*, p. 288, 2007.

[15] N. Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, "Using ArcGIS geostatistical analyst," *Esri Redlands, CA.*, vol. 300, 2001.

[16] G. Matheron, "Traité de géostatistique appliquée," *Mémoires du Bur. Rech. Géologiques Minières.*, vol. 14, 1962.

[17] B. D. Venables, W. N., Ripley, "Modern applied statistics with 99 S, 4th Edition," *Springer-Verlag, New York*, p. 481, 2002.

[18] K. . Omre, H. and Halvorsen, "The Bayesian Bridge between Simple and Universal Kriging," *Math. Geol.*, vol. 21, no. 07, pp. 767–786, 1989.

[19] P. J. Diggle, P. J., & Ribeiro Jr, "Model-based geostatistics," *Springer*, 2006.

[20] D. J. Pebesma, E. J., de Jong, K., Briggs, "Visualising uncer- 80 tain spatial and spatio-temporal data under different scenarios," *Int. J. Geogr. In- 82 Form. Sci.*, vol. 21, no. 5, pp. 515–527, 2007.