SPATIAL DISTRIBUTION OF ARSENIC CONCENTRATION IN DRINKING WATER USING KRIGING TECHNIQUES

Zahid Javed [1,*], Sana Saeed[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

[*]Corresponding author's e-mail: zzahidgee@yahoo.com

ABSTARACT: *Arsenic is an organic component of the actual earth's crusting and is broadly distributed through the entire environment within the 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 As 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 standards 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. There were used ordinary kriging and Bayesian Kriging for forecasting spatial structure of Arsenic in drinking water of district Faisalabad, Punjab, Pakistan.*

Keywords: Arsenic Concentration, drinking water, District Faisalabad, Spatial distribution.

INTRODUCTION

Geostatistics can be observed as a assortment of numerical techniques which deal with 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 created more environmental problems and difficulties regarding worldwide climate modifications, contaminants of groundwater resources, worldwide warming, nuclear and technical hazardous waste products and many more problems in the entire world[2]. In the developing countries, a massive part of population undergoes from health harms related with 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].

Various countries conventional their specific water quality requirements or recommendations for conference their nationwide priorities and necessities of the environment as well as cultural angles. A short listing of standards various countries and organizations with regard to maximum allowable limit with regard to As was actually presented within Table 1.

[3] Revealed that the As meditation in drinking water exceeding the standards. The relationship of high As standards with rice yield, known to stimulate reduced surroundings due to ponding, additional corroborates this assumption. [6] Evaluated that the Asia was 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 source of cancer in the skin, lungs, bladder and kidney. It can also be the reason in other skin fluctuations such as thickening and pigmentation. 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]. Different spatial statistical tools will be applied for the purpose of interpolation of the spatial distribution of this Physio-Chemical parameter Arsenic.

70 spatial samples were collected from different areas of Faisalabad Punjab, Pakistan and test in laboratory for Arsenic. The spatial techniques had been deliberated as the supreme advanced way of exclamation as well as yield best consequences in the logic which they give impartial assessments along with minimal worth 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 completed for normalization of reaction variable. Within the first stage, spatial exploratory exploration was built to detect the spatial performance. Within the

second stege, empirical variogram was attracted from the dataset and variables of theoretical variogram versions (Matern, Exponential, Spherical and Gaussian) are estimated by using ordinary least square estimation methods. The performance of the above models was checked by cross validation and it was noted that a model has least value of RMSPE was 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. This is a major developing city of Pakistan. The town Faisalabad is actually circumscribed within the northern through the area of Hafizabad as well as Chiniot , and in the east through district Nankana Sahib , on the South-East through Okara, within the South Sahiwal and Toba Tek Singh, and the to the west by Jhang. Faisalabad is situated at northeast side of the Punjab, amongst longitude 73°74 East, latitude 30°31.5 North, having an elevation associated with 184 metre distance (604 ft) above ocean level [9]. Due to its greatest evapotranspiration, Faisalabad has a very warm weather. The locations of each spatial sample are shown in the following map.

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 information variation each and every fine range and is the discontinuity in the origin. Usually variogram actually 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. Anywhere k may be the space between localities by and x+k and k may be the number of statement sets divided by range k [10]. Graphically the information gained by variogram is expressed as Variogram Models

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

:

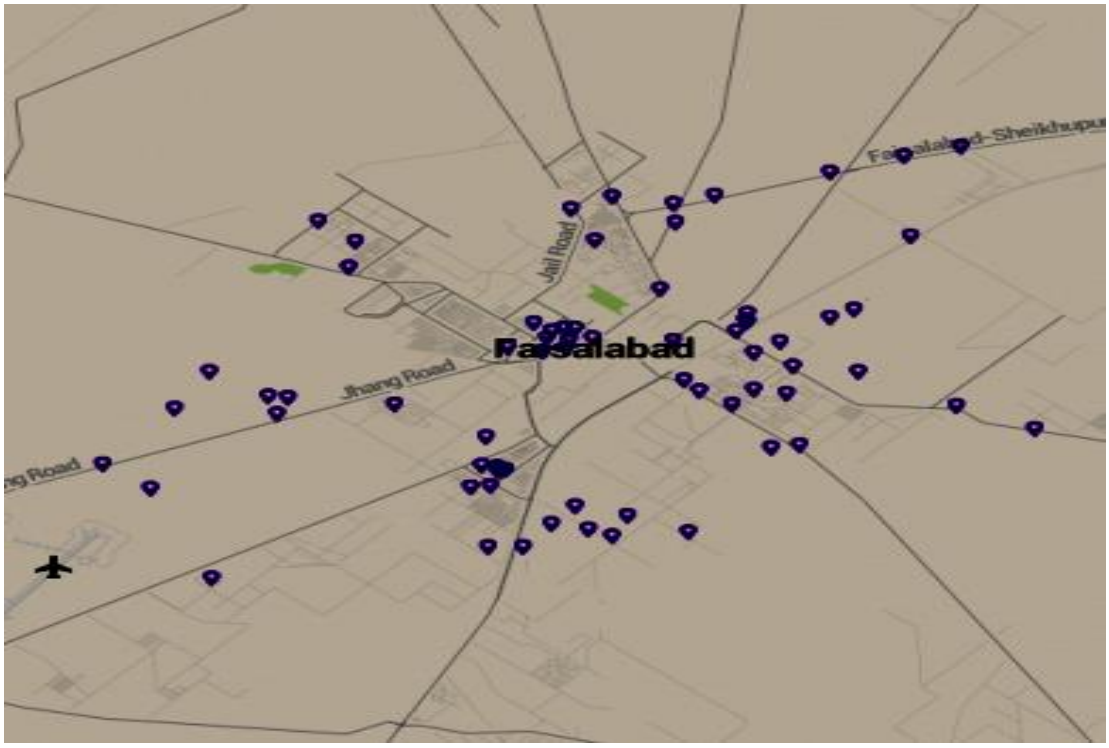

Figure 1 samples location

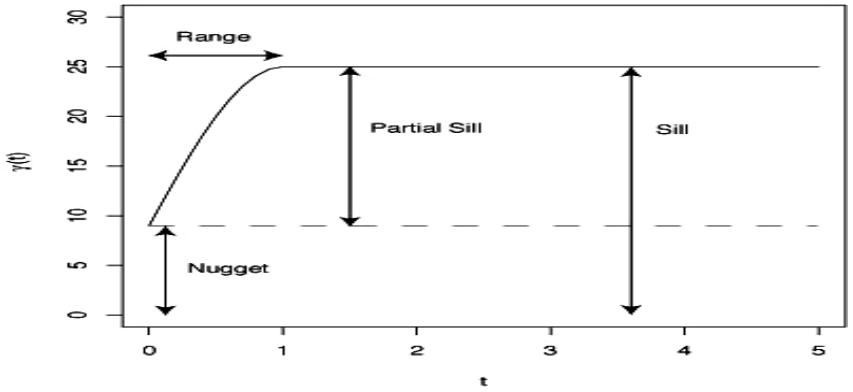

Figure 2 Sketch of Variogram

The actual explanation associated with popular spatial covariance versions with regard to spatially fixed process is actually provided beneath anywhere spatially fixed procedure offers homogeneous covariance conduct over the whole domain name from the area. A short introduction regarding these kinds of variograms types is actually described right here[11].

Matern Model

Additionally its entitled whittle Matern method. It had been submitted through Matern with regard to degree one

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

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 an rank v. This particular variogram product is an advanced method or even option among Gaussian product and exponential model. With regard to v=1/2 within Matern product, it is decreased into rapid model as well as for $\emptyset = \tilde{\emptyset}/(2\sqrt{k+1})$ and as $k \to \infty$ in

Matern model, this converts into square of the exponential distribution.

Exponential Model

The actual exponential method for spatial correlation is described below

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

With regard to $h > 0$ as well as τ^2, σ^2 *and* $\emptyset \geq 0$, where $\tau^2+\sigma^2$ would be the sill as well as τ^2 is known as the real nugget a direct result the model. The range associated with exponential method relies on the worth of factor $\emptyset$. The actual semi-variogram exponential model was shown below in the figure 3[11].

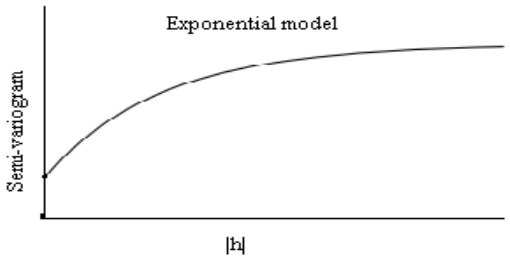

Figure 3 Sketch of Spatial Exponential Semi-Variogram Model

This increase tremendously as the range h among sites improves and it acts linearly close to the origin.

Gaussian Model

The actual mathematical product for the Gaussian models with regard to $h > 0$ is really as

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

Where τ^2, σ^2 *and* $\emptyset \geq 0$. The increasing act with embrace h is actually same as exponential model however it behaves parabolically near the source. Its common graphical draw is being demonstrated in the figure 4[12].

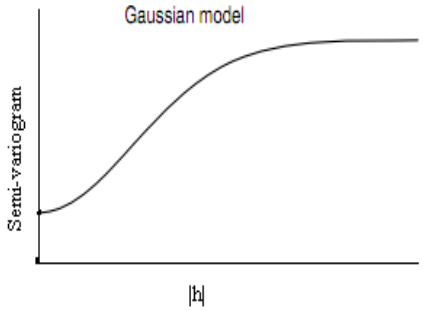

Figure 4 Sketch of Spatial Gaussian Semi-Variogram Model

This particular model features a practical selection of 3$\emptyset$.

Spherical Model

The equation of the model is described as follow

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

Wherever τ^2, σ^2 *and* $\emptyset \geq 0$. Spherical model slowly rises after the nugget result τ^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. These kinds of somewhat variogram classic are generally useful in withdrawal software. The model acts graphically shown below[11].

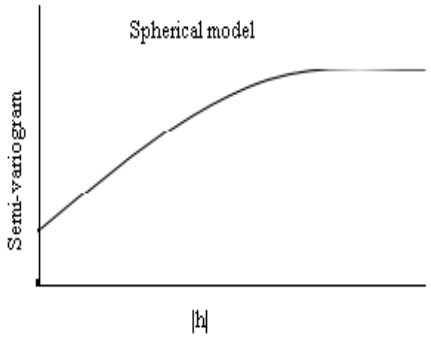

Figure 5 Sketch of Spatial Spherical Semi-Variogram Model

The actual Gaussian model, using its parabolic conduct in the source, signifies effortlessly varying characteristics. However, making use of the Gaussian model by yourself with no nugget impact can result in statistical volatilities within kriging procedure. The actual actual round as well as rapid model display linear conduct the foundation, right for symbolizing attributes having a upper level associated with small kind of variability [13]

Ordinary Least Squares Estimation Method

The famous competition fitting formula in spatial measurements is known as ordinary least square. It estimates the actual parameter θ by reducing the next qualifying standard.

Wherever in this qualifying criterion v_k are usually averaged

$$s_0(\theta) = \sum_{k=1}^{k} \{v_k - v(u_k;\theta)\}^2$$

empirical variogram values and $v(u_k;\theta)$ are the actual consistent theoretical variogram standards of v_k, m may be the whole number associated with cases within empirical variogram.

An additional advanced as well as upgraded type of ordinary least sq. is actually n-weighted minimum square by which identify the consequence associated with changing the actual numbers of n_k in various cases. The actual standard associated with n measured least sq. approximation is really as given below

$$s_n(\theta) = \sum_{k=1}^{k} n_k\{v_k - v(u_k;\theta)\}^2$$

n WLS is nearly corresponding to OLS anytime relate it to be able to empirical variogram, it will get precisely corresponding to OLS anytime the average of the variogram cases i-e u_k had accurate lags perhaps most obviously. The effectiveness of the two approaches typically in techniques are determined by value related to m along with u_k [14].

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 which interpolator 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,…,Y_n) were beliefs of reaction factor calculated through noticed sites. [16] offers planned program of OK in order to forecast worth of reaction factor in unnoticed area, where x= $\sum_i^n a\,(x_i)$, here the standards associated with a(x_i) are approximated exhausting the program of linear equality. [14] possess suggested product founded OK that can be printed 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 μ=0 , $\sigma = \tau^2$ and X_i actually 2nd degree fixed Gaussian unsystematic area. 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 versions [17]. This is apparent which reaction 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 worth of resposne 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 weights and approximated through OK program. Estimation Variance associated with indications in unnoticed area is

$Var\,((\hat{S}(x))|Y) = 1 - r'\,(\tau^2 I + R)^{-1} r$

This particular rely on prospective position x and information position x_i, beliefs of product factors however do not rely upon noticed quantities from reaction adjustable Y_i [14].

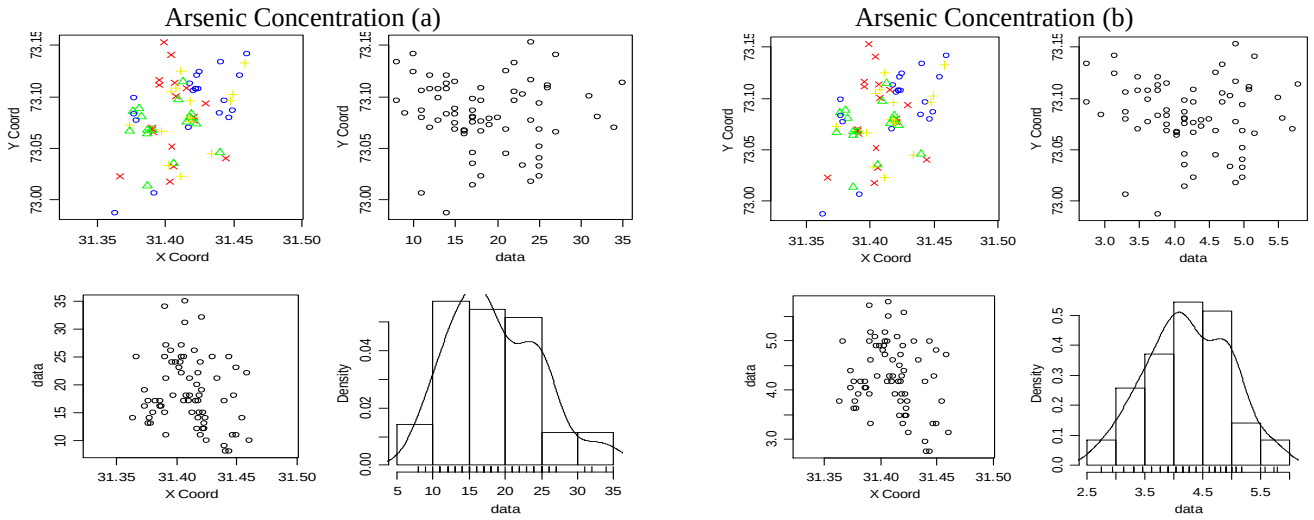

Figure 6. Exploratory data analysis of spatial data

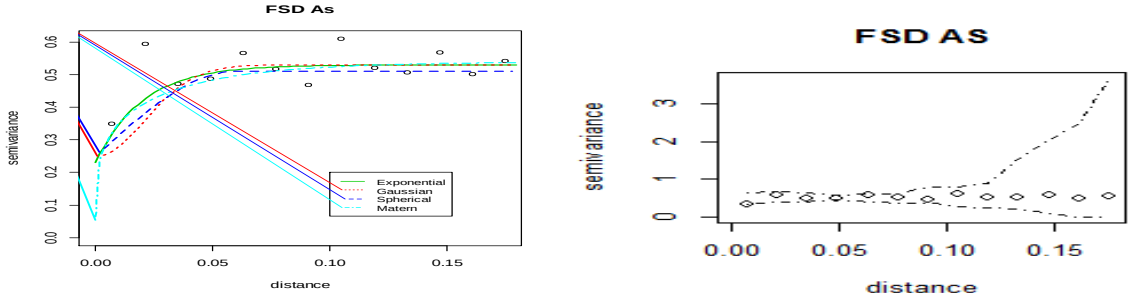

Figure 7. Empirical Variogram models and variogram envelope

Table 2: Parameter Estimation of Spatial Arsenic Concentration data

Parameter estimation Method	Variogram Model	Partial Sill ς2	Range φ	Nugget τ 2	Cross-validation (RMSPE)
Ordinary least square	**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	20.4004	**4.1633**

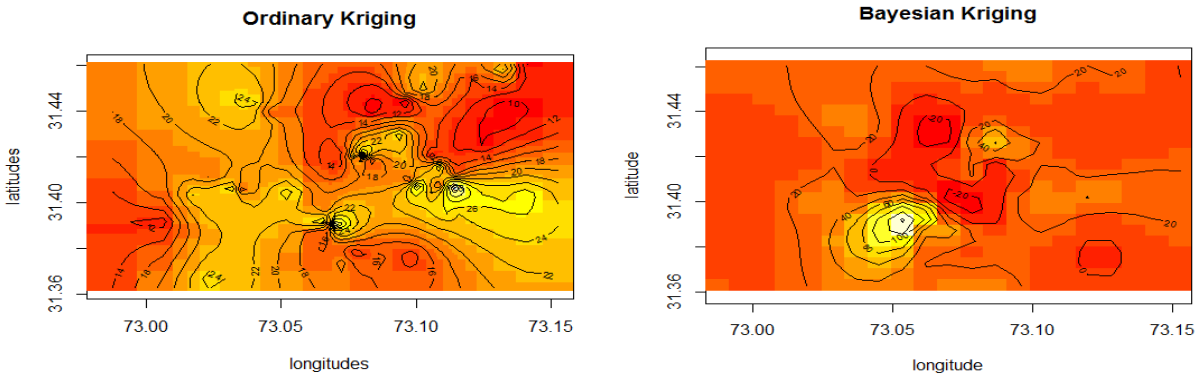

Figure 8. Spatial prediction map of As by OK and BK

Bayesian Kriging (BK)

BK estimation suggested through [18] had advantages on OK estimation that within BK entirely variables are handled as adjustable not set. In BK the actual factors tend to be assessed utilizing Bayes theorem standards that utilize the earlier distribution (θ) of particular parameters and also the

likelihood functionality (θ; y) wherever θ=(β, σ^2, ϕ) would be the parameters associated with covariance product. The producing posterior submission 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 π (β,σ^2,ϕ) =π (β, σ^2 |ϕ) π(ϕ) , here π (ϕ) was definitely an independent earlier distribution associated with π(ϕ) that stipulate this as a under the radar distribution comprising in practice. Here adopt the prior π (β, σ^2, ϕ) =π (β, σ^2 |ϕ) (ϕ), and π (ϕ) is definitely an independent earlier distribution associated with (ϕ) that identified like a discrete submission 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

[β, σ^2, |y] = [, σ^2 |y ,ϕ][ϕ| y]

Where $[\beta, \sigma^2, \phi\ |y] \sim N\chi^2_{sci}(\tilde{\beta}, v_{\tilde{\beta},} \mathrm{n}_\sigma + n, s^2)$ and the posterior density of ϕ is

$$p(\phi\ |y) \propto\ \pi(\phi)\left|v_{\tilde{\beta},}\right|^{\frac{1}{2}} (s^2)^{-(n+n_\sigma)/2}$$

Where v_β andS^2 are given below.

$$v_\beta = (v_b^{-1} + D'R^{-1}D)^{-1} \text{ and}$$

$$s^2 = \frac{n_\sigma s_\sigma^2 + m_b' v_b^{-1} m_b + y'R^{-1}y - \tilde{\beta}' V_\beta^{-1} \tilde{\beta}}{n_\sigma + n}$$

RESULTS AND DISCUUSIONS

The analysis of the underlying spatial data was done on the basis of geo-statistical techniques. Firstly descriptive analysis was done than the basic estimation techniques were used to estimate the parameters. After estimation of parameters different kriging techniques were used for the purposes of interpolation and checked that which technique was better by calculating RMSPE. R Language software was used for the whole analysis. The exploratory spatial data analysis of Arsenic (As) concentration was carried out using geoR package of R software is as follow.

It was shown in the figure 6 (a) that the data of Arsenic was non normal that's why Box Cox transformation was apply for normalization of the data as shown an the figure 6 (b).

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 and so on. There was installed only the matern, exponential, gaussian, spherical, versions on empirical variogram if take the value associated with parameters sill, range and nugget within the above stated models. The Graphical view of all the fitted models were given in figure 7 (left panel).

Variogram Envelope diagram for examining spatial relationship of reaction factor demonstrated under figure 7 (right 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 or visual display of all spatial covariance models are given in figure 4.7(left panel).

In table the combinations of Ordinary least square (OLS) estimation method with Gaussian spatial covariance model perform well because RMSPE has the minimum value. We fitted four different covariance structures like Matern, Exponential, spherical and Gaussian structures. Here Ordinary least square (OLS) estimation technique and Gaussian method were used for further prediction.

The comparison between ordinary kriging and Bayesian kriging for Arsenic was made in the following figure.

The maximum level of Arsenic is 35 with longitude 73.1139 and latitude 31.4067 as shown by yellow area in the figure 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

RMSPE	OK	BK
Arsenic Concentration	5.8732	0.1652

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 0.1652. Thus prediction of unmonitored location through Bayesian kriging will always be beneficiary for the underlying study domain.

CONCLUSIONS AND RECOMENDATIONS

In this article we studied the spatial data on Arsenic concentration at 70 locations of District Faisalabad. 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. It was highly recommended for the inhabitants of that locality to be aware of the harms of 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 GEOSTATISTICS And 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.