

# Spatial Distribution of TDS in Drinking Water of Tehsil Jampur using Ordinary and Bayesian Kriging

Maqsood Ahmad  
College of Statistical and Actuarial Sciences  
University of the Punjab, Quaid-i-Azam Campus  
Lahore, Pakistan  
maqsoodzain5@gmail.com

Sohail Chand  
College of Statistical and Actuarial Sciences  
University of the Punjab, Quaid-i-Azam Campus  
Lahore Pakistan  
sohail.stat@gmail.com

## Abstract

In this research article, level of TDS in groundwater with spatial domain Tehsil Jampur, Pakistan is studied. Its enhanced level in drinking water produces harmful effects on human health and aquatic ecology. This study revealed that, among thirty observed locations, TDS level in 56% of the water samples exceeds the maximum permissible limit (1000g/ml) of TDS as described by WHO. Four estimation techniques were comparatively studied for fitting well known matern spatial covariance models. Model based Ordinary Kriging (OK) and Bayesian Kriging (BK) were used for spatial interpolation at unmonitored locations. The best interpolation technique was selected on the basis of least root mean squared prediction error (RMSPE). Prediction maps were generated for visual presentation of interpolated sites for both techniques.

**Keywords:** Spatial Interpolation, Box-Cox transformation, Bayesian Kriging, TDS

## 1. Introduction

Ground water is a significant source of drinking water in Pakistan. Fresh drinking water is essential for human life while polluted drinking water has negative effects on human health. In developing countries, a specific ratio of population is suffering from the water issues either due to its shortage or its impurification. Polluted drinking water is eventually linked to hundreds of thousands of deaths in whole world in every year and is regarded as a prevalent health risk (Hussain et al., 2014). In developing countries about 80% of all sicknesses are caused to poor water and cleanliness condition (Kahlowan, 2006). Total dissolved solids (TDS) represent the total concentration of dissolved substance in the water and one can measure the quality of drinking water based on level of TDS (Akhtar and Tang, 2013). In water Total dissolved solids are represented by the weight of deposit left when a water sample has been evaporated to dryness (WHO, 2006). The TDS in natural water concentration is usually less than 500 mg/l and water with TDS concentration with more than 500 mg/l is undesirable for drinking purposes. According to World Health organization (WHO) and US-Environmental Protection Agency (US-EPA) permissible limit of TDS is 1000 mg/l while Pakistan Council of Research in Water Resources (PCRWR), Pakistan Standard and Quality Control Authority (PSQCA), and Indian Water Quality Standards illustrate its permissible limits 500-1500 mg/l (WHO,

2006; PCRWR, 2007). Some other countries like Thailand, China and Indonesia showed maximum acceptable limit not more than 500 mg/l while in Japan it should be less than 1000 mg/l (WHO, 2006). The actual spatial distribution of infected groundwater display some heterogeneity and the pollutant concentration values are rarely readily available for every possible area of a region. The amount of pollutant concentration at each location is not really always achievable in view of time and the cost involved in information collection. For that reason prediction associated with values in other areas based upon selectively measured values could be among the alternatives. Within this context, in order to predict the actual concentration associated with pollutant in ungauged areas, the Geostatistical techniques might be applied. The fundamental assumption within using geostatistics is that the attributes in the earth have some spatial continuity up to certain lag distance (Webster and Oliver, 2007). The assumption of Gaussian random field is essential for interpolation. If normality assumption is violated, Box-Cox transformation (Box and Cox, 1964) could be used to cope with this issue (Pilz and Spöck, 2008). The actual Geostatistical ideas and its applications are documented by various researchers all over the world (Isaaks et al., 1989, Banerjee et al., 2008). A novel Geostatistical method known as ordinary kriging perform a great part for the best prediction (Adhikary et al., 2012; Jiang et al., 2013). It views the spatial correlation between sampled sites and is mainly used for mapping spatial variability. Many authors suggested the Bayesian Kriging which take into account the uncertainty of parameters. For Bayesian Kriging the selection of prior distribution of parameters is essential for efficient prediction. (Pilz and Spöck, 2008) addressed the need for Bayesian kriging and suggested it for unbiased interpolation of ungauged locations. Various packages associated with R language have an essential role with regards to analysis of point processed spatial and spatio-temporal data sets.

## **2. Material and methods**

### **2.1. Study area and Sampled observations**

Pakistan is located among 23° - 37° to the north latitude and 61° - 76° Eastern longitude. It has lot diversity within spatial as well as seasonal variance of the environment. Tehsil Jampur that is study domain of this research geographically situated at 29°38' North latitude and 70°35' Eastern longitude associated with south of Dera Ghazi Khan . The actual climate within the Jampur area is enjoyable, with very hot and dried out summers as well as mild and rainy winters (Rafique et al., 2011). To manage the biasness three samples were gathered from every location and their average value was utilized in research. Amount of TDS in every sample was measured during the field visit on the basis of conductivity meter (Hi 8033 Hana Hungry). The spatial distribution from the monitored locations is provided below in figure. 1.

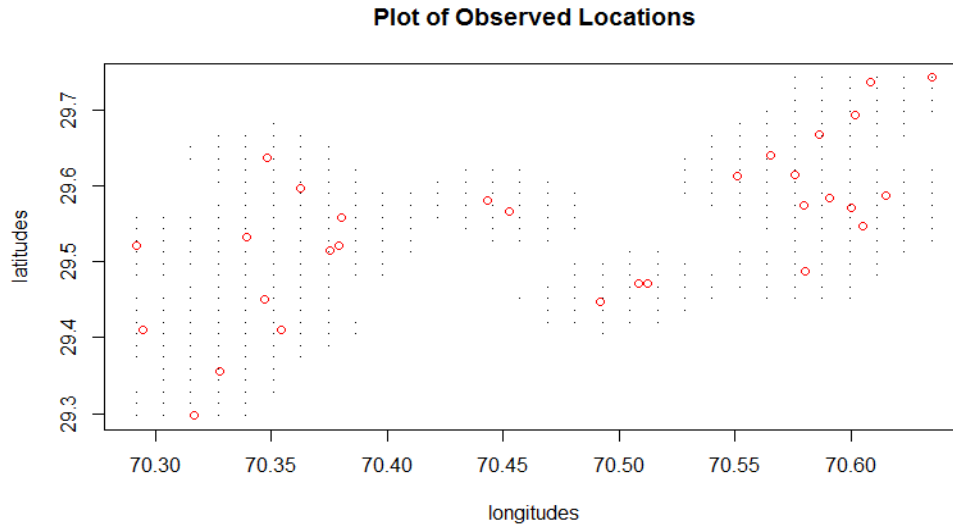


Figure 1: Plot of 30 Observed locations for water samples at Tehsil Jampur

**Table 1: Locations exceeding permissible limits of WHO**

<b>Names of Location</b>	<b>Latitude</b>	<b>Longitude</b>	<b>TDS(mg/l)</b>
Kot Janoon	29.7356	70.6083	1798
Kot Tahir	29.7422	70.6353	1009
Moza Mullan Wala	29.6936	70.6017	1020
Moza Pole Wala	29.6678	70.5867	1063
Moza Dohran Hajana	29.5742	70.58	2126
Muhammad Pur-1	29.4708	70.5122	1882
Bokhara	29.4472	70.4917	3648
Muhammad Pur-2	29.4706	70.5086	1998
Noorpur Mangwala	29.4108	70.3544	3020
Nawan Shaher	29.4506	70.3472	1702
Soon Wah	29.2972	70.3164	5661
Wah Lashari	29.5206	70.2917	1056
Miran Pur	29.4108	70.2944	1619
Moza Koloji Wala	29.6375	70.3478	7071
Basti Danwar	29.5142	70.3753	3859
Basti Nawaz Deena	29.5653	70.4528	3065
Rasool Pur	29.6403	70.5654	2009

## 2.2 Description of Spatial Variogram Model

Before interpolation of unsampled locations, it is necessary to evaluate the spatial correlation among neighboring sites (Le and Zidek, 2006). Variogram is actually used to capture the spatial dependence between different neighboring locations (Cressie, 1985). Spatial prediction requires estimations of the variogram  $\gamma(h)$  for those h's which are

unavailable in the underlying spatial data. If spatially stationary process has homogeneous covariance structure across the whole domain of the field than variogram is described as

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

$\gamma(h)$  is utilized to evaluate the spatial correlation. Wherever  $h$  is the spatial distance among locations and  $N(h)$  shows the number of observed pairs separated by range  $h$ . A well known Matern spatial covariance model (Ribeiro and Diggle, 2001) with fixed smoothness parameter utilized for fitting the actual spatial information and for modeling the spatial variations among sites which requires the form

$$\gamma(h) = \tau^2 + \sigma^2 \left[ 1 - (1 + \phi(h))e^{-\phi h} \right] \quad \text{If } h > 0 \tag{1}$$

Where  $\tau^2$  is the nugget effect, sill is  $\sigma^2$  and range of the Matern covariance model is described to be  $\phi$  at which the variogram is of 95% of the sill parameter. In this scenario  $h$  is the parting distance between any kinds of two spatial sites.

### 2.3 Model based Ordinary and Bayesian Kriging

Ordinary kriging is among the most strong method of kriging first developed by (Krige, 1951) and has already been used thoroughly in literature. This technique is founded on the supposition that the mean is continuous but not known. It is a basic most popular approach to spatial interpolation which generates both the forecasted values and also the prediction errors. In a model based ordinary Kriging, it is assumed that response variable  $Y$  follows the Gaussian distribution as described by (Diggle and Ribeiro, 2007) i.e. for observed values  $Y_{(S_i)}$  at different sites  $S_i$  for  $\{i = 1, 2, 3, \dots, n\}$

$$Y_{(S_i)} \sim N(\mu, \Sigma_{Y(S_i)})$$

Where  $\mu$  is the mean vector and  $\Sigma_{Y(S_i)}$  is the covariance matrix which may be decomposed as

$$\Sigma_{Y(S_i)} = \sigma^2 R(\alpha) + \tau^2 I$$

where  $R$  is a correlation matrix and it depends upon vector-valued of parameter  $\alpha$  where  $\alpha = c(\sigma^2 = sill, \phi = range, \tau^2 = nugget)$ . These three parameters could be estimated using four different estimation techniques like MLE, REML, OLS and WLS (Diggle and Ribeiro, 2007). A covariance model with least RMSPE is selected along with estimation technique for interpolation of unobserved locations. The prediction of the response variable at unobserved locations i.e.,  $\hat{Y}(S_0)$ , after estimation of the variogram model parameters can be made using a system of ordinary kriging as follows

$$\hat{Y}(S_0) = \sum_{i=1}^n w_i(s_0) Y(S_i) \tag{2}$$

Where  $w_i(s_0)$  are weights for  $\{i = 1, 2, 3, \dots, n\}$  having sum of weights equal to 1. Symbolically

$$\sum_{i=1}^n w_i(s_0) = 1$$

This is a necessary condition for unbiased prediction. These weights are obtained by Lagrange multiplier and results in reduction of Kriging variance  $\sigma^2(s_0)$  associated with each predicted estimate. Bayesian Kriging introduced by (Omre and Halvorsen, 1989) have advantage over ordinary Kriging (Diggle and Ribeiro, 2007) because it take into account the uncertainty of parameters for estimation purpose. Various prior distributions are utilized for the model parameters and prior distribution of these parameters is combined with the likelihood of the response variable using Bayes theorem to derive the posterior distribution (Banerjee et al., 2008). The mathematical form of posterior distribution is given as

$$\pi_1(\theta|Y_{S_i}) = \frac{L(\theta; Y_{S_i}).\pi_0(\theta)}{\int L(\theta; Y_{S_i}).\pi_0(\theta)d\theta}$$

Where  $\pi_0(\theta)$  is the joint prior distribution of parameters and  $L(\theta; Y_{S_i})$  is likelihood and  $\pi_1(\theta|Y_{S_i})$  is the posterior distribution. For detailed methodology of Bayesian Kriging see (Besag and Green, 1993; Diggle and Ribeiro, 2007).

### 3. Results and Discussion

Spatial data was analyzed by using geoR (Ribeiro and Diggle, 2001) package of R language and environment for statistical computing (R Core Team, 2014). Different functions of geoR package were used for graphical representations and numerical estimates of spatial parameters. Exploratory spatial data analysis (ESDA) used to detect some particular properties, pattern and distribution of spatial dataset.

**Table 2: Exploratory spatial data analysis**

Coordinates Summary				Latitude	Longitudes
Min				29.2972	70.2917
Max				29.7422	70.6353
Summary of Total Dissolved Solids (mg/l)					
Min	Q1	Median	Mean	Q3	Max
150	568.8	1038	1672	2006	7071

Normality of response variable is basic assumption of Geostatistical interpolation methods (Webster and Oliver 2007), Table 2 reports that observed spatial data of TDS is positively skewed so Box-Cox transformation which is a modified form of power transformation described by (Box and Cox, 1964) is used for analysis with transformation parameter  $\hat{\lambda} = 0.09485941$  estimated by MLE method using powerTransform function of CAR package (Fox and Weisberg, 2011) in R statistical computational software. This transformation is an attempt to get the behavior of variable as normally distributed and it also stabilizes the variability in measurements (Hussain et al., 2014)

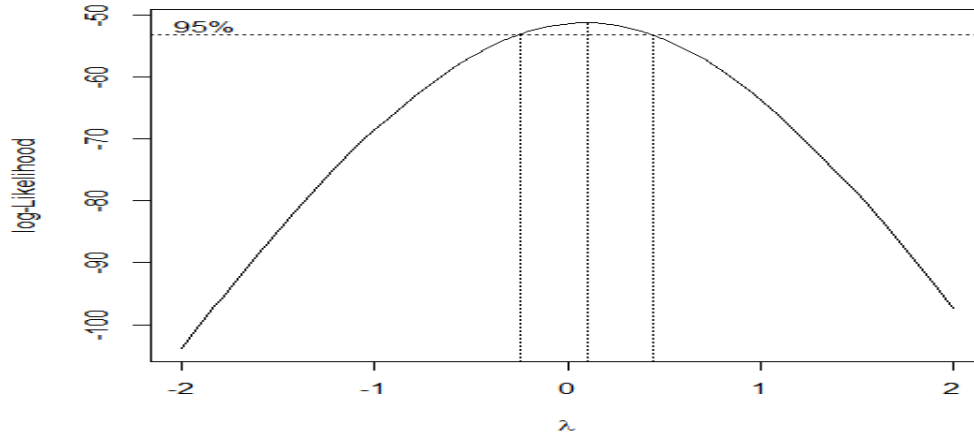


Figure 2: Plot of Normalized parameter lambda using Box Cox Transformation

Figure 2 Visual representation of normalized parameter for evaluating the value of lambda to transform the data into Gaussian settings.

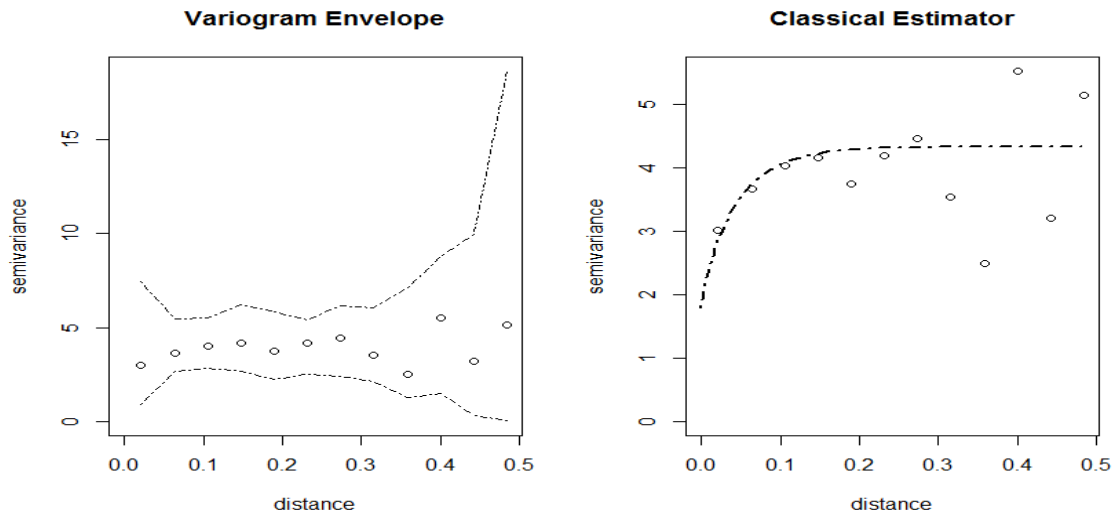


Figure 3 Left panel Variogram Envelop of Spatial dependence. Right panel matern spatial covariance model fitted on Empirical Variograms.

Variogram envelope plot for checking spatial correlation of the response variable TDS shown in above figure clearly indicates that there exists strong spatial correlation because all variogram points are inside the envelope limits and thus it confirming the presence of positive spatial correlation (Ribeiro and Diggle, 2001). This spatial dependence is helpful for further modeling and spatial interpolation of ungauged locations (Fuentes, 2007).

**Table 3: Estimation of Model Parameters using Matern Covariance model**

Estimation Method	Nugget ( $\tau^2$ )	Partial Sill ( $\sigma^2$ )	Range ( $\phi$ )	RMSPE
OLS	4.3355	3.2280	0.0396	695.926
WLS	4.0997	3.4122	0.0348	750.005
MLE	0.0894	0.0012	0.0013	664.529
REML	0.0937	0.0002	0.0375	674.676

In order to predict TDS of drinking water data associated with Jampur within the prediction grid we utilized ordinary as well as Bayesian kriging. REML and MLE are two basic estimation techniques for variogram parameters (Kitanidis, 1983). Results given table 3 are calculated on the basis of [Eq. 1] reports that among four estimation methods, parameters of MLE technique will be used for interpolation as it yields minimum root mean square prediction error (RMSPE) and least nugget effect. For Bayesian Kriging, prior distribution for the mean parameter  $\beta$  is taken as flat prior  $P(\beta) \propto 1$ , while the prior distribution for range parameter  $\phi$  is taken as uniform  $P(\phi) \propto 1$  and the prior distribution of partial sill parameter  $\sigma^2$  presumes the reciprocal  $P(\sigma^2) \propto 1/\sigma^2$ .

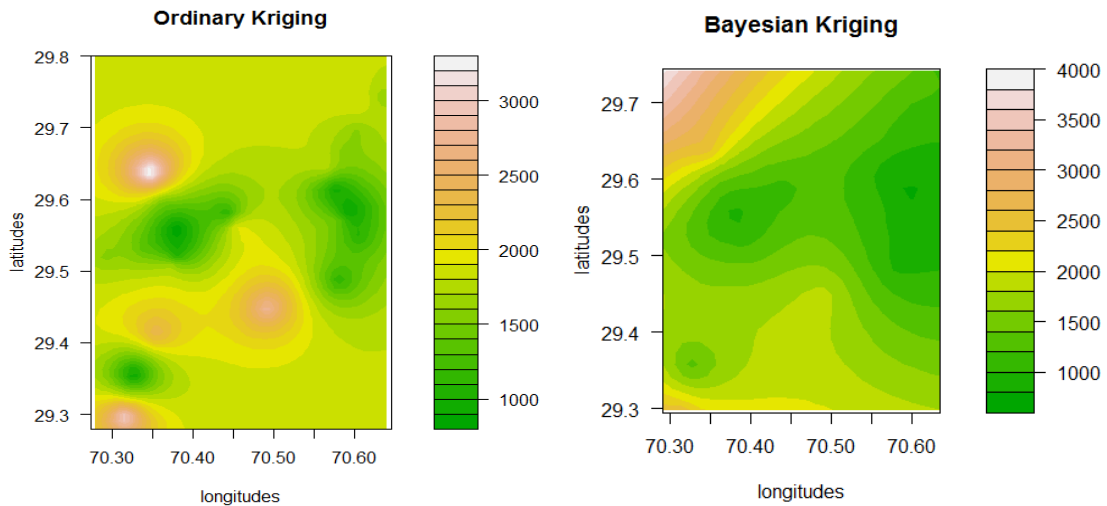


Figure 4 Left panel prediction map of ordinary Kriging (OK). Right panel prediction map of Bayesian Kriging (BK).

**3.1 Cross validation statistics: RMSPE**

For the purpose of selecting a suitable spatial covariance model, better estimation technique and effective interpolation technique between two (Model based Geostatistical) MBG methods, cross validation statistics root mean square prediction errors (RMSPE) is extensively used in literature (Hussain et al., 2014). If  $y(s_i)$  are actual observed values and  $\hat{y}(s_i)$  are predicted values at respective spatial locations  $S_1, S_2, \dots, S_{30}$  than the underlying cross validation statistics is described as

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

**Table 4: Comparison between Ordinary Kriging and Bayesian Kriging using cross-validated RMSPE**

Model Based Geo-statistical method	Ordinary Kriging (OK)	Bayesian Kriging (BK)
Root Mean Square Prediction Error	2325.694	1858.588

Table 4 reports the comparison between Ordinary Kriging and Bayesian Kriging on the basis of cross validated RMSPE. Here Bayesian kriging was comparatively selected as best geostatistical interpolation technique because of its least RMSPE value. K-fold cross validation statistics was used for diagnostic checking. 1000 samples of MCMC from posterior and predictive distribution under the above mentioned priors were obtained for the model parameters. Under Bayesian prediction setup, right panel of figure 4 illustrates that top right region having latitude 26.6°-26.8° and longitude 70.30° - 70.35° consists of excessive amount of Total dissolved solids (mg/l) in drinking water. This region has extremely contaminated and pollutant water which is harmful for human body and will cause several diseases until it would be properly cured.

#### 4. Conclusion and Recommendation

In this article we studied the nature of excessive amount of Total Dissolved Solids (TDS) in drinking water of Tehsil Jampur and narrated the harmful effect of TDS. Model based Geo-statistical analysis through spatial structure like Ordinary Kriging and Bayesian Kriging were applied for the purpose of interpolation of TDS at un-gauged locations. Best interpolation technique was determined on the criterion of least cross-validated RMSPE. To find out the valid estimation technique for fitting Metern spatial covariance model which could better predict the unmonitored site, we compare four estimation techniques (OLS, WLS, MLE, and REML) and finally suggested the Bayesian interpolation technique with reduced RMSPE. It was observed that at 17 (56%) sampled locations, amount of TDS was alarming that showed health risks and highly harmful to human body. So it is highly recommended that water purification plants should be installed in these areas and every household must check water quality before using the supplied water for drinking purposes.

#### Acknowledgement

The authors are appreciative to the Pakistan Council of Research in Water Resources, Ministry of Science and Technology for providing hydrological data of Jampur Region.

#### References

1. Adhikary, P. P., Dash, C. J., Chandrasekharan, H., Rajput, T., and Dubey, S. 2012. Evaluation of groundwater quality for irrigation and drinking using gis and geostatistics in a peri-urban area of Delhi, India. *Arabian Journal of Geosciences*. 5(6):1423-1434.
2. Akhtar M.M., Tang. Z. 2013. Identification of contamination sources and TDS concentration in groundwater of second biggest city of Pakistan. *International Journal of Environmental Science and Development*. 4(1):341-345.



3. Banerjee, S., Gelfand, A.E., Finely, A.O., and Sang, H. 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society; Series B (Statistical methodology)*. 70(4): 825-848.
4. Besag, J., and Green, P.J. 1993. Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society, Series B (Methodological)*. 55(1): 25-37.
5. Box, G., and Cox, D.R. 1964. An analysis of transformations (with discussion). *Journal of Royal Statistical Society B*. 26(2):211–252.
6. Cressie, N. 1985. Fitting variogram models by weighted least squares. *Mathematical geology*. 17(5): 563-586.
7. Diggle, P. J., and Ribeiro Jr, P. J. 2007. *Model-based geostatistics: Springer*, New York.
8. Fox, J. and Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/~jfox/Books/Companion>
9. Fuentes, M. 2007. Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*. 102(477): 321-331.
10. Isaaks, E. H., and Srivastava, R. M. 1989. An introduction to applied geostatistics. New York: Oxford University.
11. Hussain, I., Shakeel, M., Faisal, M., Soomro, Z. A., Hussain, M., and Hussain, T. 2014. Distribution of Total Dissolved Solids in Drinking Water by Means of Bayesian Kriging and Gaussian Spatial Predictive Process, *Water Quality Exposure Health*, DOI 10.1007/s12403-014-0123-9.
12. Jiang, Y., Ma, W., Luo, H., and Huang, W. 2013. Analysis of spatial distribution of groundwater quality in huaihe river basin. In *Geoinformatics, 2013 21st International Conference on*. IEEE.
13. Kahlowan, M., Ashraf, M., Hussain, M., Salam, H., and Bhatti, A. 2006. Impact assessment of sewerage and industrial effluents on water resources, soil, crops and human health in Faisalabad. Research Report, *Pakistan Council of Research in Water Resources*, Islamabad.
14. Kitanidis, P. K. 1983. Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*. 19(4):909-921.
15. Krige, D. G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society*. 52(1): 119–139.
16. Le, N. D. and Zidek, J. V. (2006). Statistical Analysis of Environmental Space-time Processes. *Springer Science*.
17. Omre, H. and Halvorsen, K. B. (1989). The bayesian bridge between simple and universal kriging. *Mathematical Geology*, 21(7):767-786.
18. Paulo J. Ribeiro Jr and Peter J. Diggle. 2001. geoR: a package for geostatistical analysis R-NEWS. 1(2):15-18.

19. PCRWR. 2007. National water quality monitoring programme water quality Islamabad. Report: 2005–2006, *Pakistan Council for Research in Water Resources*, Islamabad.
20. Pilz, J. and Spöck, G. 2008. Why do we need and how should we implement bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment*. 22(5):621-632.
21. R Core Team. 2014. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.URL <http://www.R-project.org/>.
22. Rafique, H.M., Abbas, I., Sohl, M.A., Shehzadi, R., Ramay, S.M., Imran, M., Zaghayer, Y., Mahmood, A., and Sohl, M.N. 2013. Appraisal of drinking water quality of tehsil Jampur, Pakistan, *Desalination and Water Treatment*, DOI:10.1080/ 19443994.2013.803936.
23. Webster, R. and Oliver, M.A. 2007. Geostatistics for Environmental Scientists/2nd Edition *John Wiley & Sons*, Ltd.
24. WHO. 2006. Guidelines of drinking water quality recommendation, 3rd edition. *World Health Organisation*, Geneva.