Journal of Himalayan Earth Sciences Volume 52, No. 1, 2019 pp. 65-85

Drinking water quality assessment using water quality index and geostatistical techniques, Mardan District, Khyber Pakhtunkhwa, Pakistan.

Wajid Ali[1], Muhammad Nafees[2], Syed Ali Turab[1], M. Younis Khan[1], Khaista Rehman[1]

[1]National Centre of Excellence in Geology, University of Peshawar

[2]Department of Environmental Sciences, University of Peshawar

**Corresponding authors email: wajeedaali@uop.edu.pk*

Submitted:15/1/2019 Accepted:12/3/2019 Published online:29/3/2019

Abstract

In this study, the Water Quality Index (WQI) was calculated for drinking water samples collected from an area of Mardan district. A total of 30 water samples were collected from water distribution systems at the end user end. The samples were analyzed for 21 parameters namely, C, pH, EC, Do, TDS, NTU, TH, Ca, Mg, Na, K, SO4, NO3, Mn, Cu, Fe, Zn, Pb, Ni, Cd and Cr. Standard methods were used for the analysis of the physicochemical parameters while heavy and trace metals analyzed using Atomic Absorption Spectrometer. For the calculation of Water Quality Index, all parameters were subjected to screening using principal component analysis, thus reducing the original number of parameters to a final list of ten. The final list of parameters was used to calculate and map WQI. Moreover, two methods of kriging Ordinary Kriging and Empirical Beysian Kriging were compared for their performance in modelling the spatial distribution of the selected parameters in the study area. It was found that EBK performance was more appropriate as compared to other spatial variability models for the majority of the variables.

Keywords: Water quality, Water quality index, Effective weights, Empirical bayesian kriging, Ordinary kriging, Principal component analysis.

1. Introduction

Water is a vital natural resource, besides its impending scarcity owing to unplanned exploitation in third world countries contamination and pollution of water sources remain a primary concern (Azizullah, 2011). A large majority of the world population remain at risk of water-borne diseases (WHO, 2010). Contamination of drinking water and its potential health risks in different areas of Pakistan are extensively reported in scientific literature (Khan et al., 2013; Muhammad, 2010; Shah & Danishwar, 2003, Memon, 2011, Khan et al., 2013). The main causes of water contamination are heavy metals, agrochemicals and biological agents (Azizullah et al., 2011; Nabeela et al., 2014, Ayaz et al., 2011) Water quality assessments are necessary to identify water contamination related issues and, therefore, help in the prevention of water-borne diseases and planning remedial measures. One way of presenting water quality data is the Water Quality Index (WQI), which takes into account the combined effect of all the water quality parameters and categorizes water quality samples into categories of suitability that are understandable to the general public and policymakers. WQI is commonly calculated for drinking water sources, both groundwater and surface water (Abdel-Satar et.al. 2017; Bhuiyan et al., 2016; Chaurasia et al., 2018; Ewaid & Abed, 2017; Kumar et. al. 2015; Ponsadailakshmi et. al. 2018; Yousefi et al., 2017). It can be calculated keeping in view the different uses of the water such as drinking and irrigation purposes as well (Mora-Orozco et.al. 2017; Misaghi et. al. 2017). The method of WQI calculations may vary and a number of different methods are available in the literature (Romanto & Wardiatno, 2015, Córdoba, 2010). A wide variety of studies have demonstrated the integration of WQI with geographic information systems to produce easy to understand, legible maps of water suitability for different purposes (Khosravi et al., 2017; Şener et. al. 2017, Shahid et al. 2017, Li et al., 2019).

Spatial interpolation methods such as kriging may come in handy for mapping the distribution of environmental variables (Renard et. al. 2005; Webster & Oliver, 2007). The capacity of the method has been demonstrated by other workers using different environmental variables (Robinson & Metternicht, 2006; Wu, 2006, Khan et al., 2019). The effectiveness of the various Kriging methods in mapping the spatial distribution of water quality parameters in comparison with other methods of interpolation has been discussed earlier in several studies (Arslan, 2012; Johnson et. al. 2018; Liang et. al. 2018; Mir et. al. 2017; Murphy et. al. 2010; Narany et. al. 2014; Tiwari et. al. 2017).

Multivariate statistical techniques such as principal components analysis (PCA), Factor Analysis (FA) and Cluster Analysis (CA) remain the mainstay of studies conducted on drinking water contaminations. Authors commonly use the methods to find patterns in water quality data in an attempt to conduct source apportionment of water contaminants (Ahmed et al., 2019; Mohammad et. al. 2011; Liu et. al. 2003; Masoud, 2014; Paz-Ferreiro et. al. 2010; Singh et al. 2004; Viswanath et. al. 2015, Gul et. al. 2015; Muhammad et al., 2010). Recently, keeping in view the ability of PCA to reduce the dimensionality of a dataset, it has been used to reduce the burden of the parameters for WQI calculations (Tripathi & Singal, 2019). PCA and other multivariate methods in integration with geostatistical techniques and water quality indices have been used successfully for mapping and characterization of water quality (Bodrud-Doza et al., 2016; McLeod et.al. 2017; Tao, et al. 2016, Sahoo et. al. 2015, Liu et al. 2011) .

The study area lies to the south of Mardan City, mainly consisting of rural areas and towns (Fig.1.). Mardan district is the second largest district of Khyber-Pakhtunkhwa province. According to the 2017 census conducted by the Government of Pakistan, the total population of the district was 2.373 million. A large fraction of the population (1,933,736) in the district resides in rural areas (Anonymous, 2017). The water contamination problem in the district has been highlighted previously by researchers, reporting high levels of Nitrates contamination in the drinking water sources (Ali & Nafees, 2015). Ali, et. al. (2009), investigated the concentration of heavy metals in tap and hand pump water of Mardan city, they found that the concentrations of Pb, Cr, Cd and Ni exceed the permissible limits. Furthermore, they suggested that the high concentrations of the metals may be due to high corrosivity and aggressive nature of the water.

The main aim of this study is to ascertain the drinking water quality status of the study area based on Water Quality Index (WQI) and find the major contributing factor to WQI calculations. The study also aims to compare the prediction performance of Ordinary Kriging and Empirical Bayesian Kriging for water quality parameters.

2. Materials and methods

2.1. Sampling and laboratory analysis

A total of thirty drinking water samples were collected randomly from water distribution systems of the study area, mainly from taps and hand pumps. The sampling locations along with sample codes are shown in Figure. 1. The samples were collected without flushing the water, in order to collect the representative samples of water in the distribution system. Samples for physical and chemical parameters were collected in 1-litre polyethene bottles while for heavy metals analysis samples were preserved using HNO3 in separate 500 ml bottles. A total of 21 parameters namely, temperature, pH, EC, TDS, DO, Turbidity, Ca, Mg, Na, K, NO3, SO4, Total Hardness (TH), Fe, Cu, Zn, Mn, Cr, Pb, Cd and Ni were ascertained for this study. Concentrations of Temperature pH, EC, Turbidity, DO and TDS were analyzed on the spot using portable meters. Standard methods were followed for the analysis of major ions and heavy metals in the drinking water samples in the laboratory by means of Perkin Elmer Atomic Absorption Spectrophotometer (Perkin-Elmer AAS).

2.2. Water quality index

Water Quality Index reflects the combined influence of the water quality parameters. Water Quality index calculation was carried out in the following steps (Sahu & Sikdar, 2008; Yidana & Yidana, 2010). First, the Relative Weights of each selected parameter was calculated using the following equation:

$$Wi = \frac{wi}{\sum_{i=1}^{n} wi} \qquad (1)$$

Where Wi is the relative weight of each parameter, wi is the assigned weight of each parameter and n is the number of parameters. Each parameter was assigned a weight (wi) ranging from 1 to 5 based on its importance in affecting the water quality for domestic and health purposes, whereby 1 is assigned to parameters with least potential effect 5 to parameters whose concentrations beyond a certain range are considered critical for health i.e. Cr and Cd. Then a quality rating for each parameter was assigned by dividing its actual concentration in the water sample by its guidelines values recommended by the World Health Organization (WHO, 2004) and multiplying it by 100.

$$qi = \frac{Ci}{Si} \times 100 \qquad (2)$$

where qi is the quality rating, Ci is the actual concentration of the parameter in each water sample, Si is the guideline value of the parameter in mg/l (WHO, 2004). Calculation of the sub-index value (SI) is the next step in WQI calculation.

$$SI_{\mathrm{i}} = Wi \times qi \qquad (3)$$

Where SIi is the sub-index for the ith parameter, qi is the quality rating of the parameter (Eq.2) and Wi is the relative weight (Eq.1). The water quality index is given by Eq.4.

$$WQI = \sum_{i=1}^{n} SI \qquad (4)$$

Based on WQI, water samples can be divided in to four categories namely Excellent Water (<50), Good Water (50-100), Poor Water (100-200), Very Poor Water (200-300) and Unsuitable for Drinking (>300) (Sahu & Sikdar, 2008; Yidana & Yidana, 2010).

Further, the effective weight of each parameter was calculated by dividing its sub-index value SI, by the overall water quality index value (WQI) and the result multiplied by 100.

$$Ewi = \frac{SI_{\mathrm{i}}}{WQI} \times 100 \qquad (5)$$

where Ewi, is the effective weight for the ith parameter, SI is its sub-index value and WQI is the overall WQI value. The effective weights for each parameter were defined in order to identify the parameter with the most influence on the water quality index calculation (Şener, et. al., 2017).

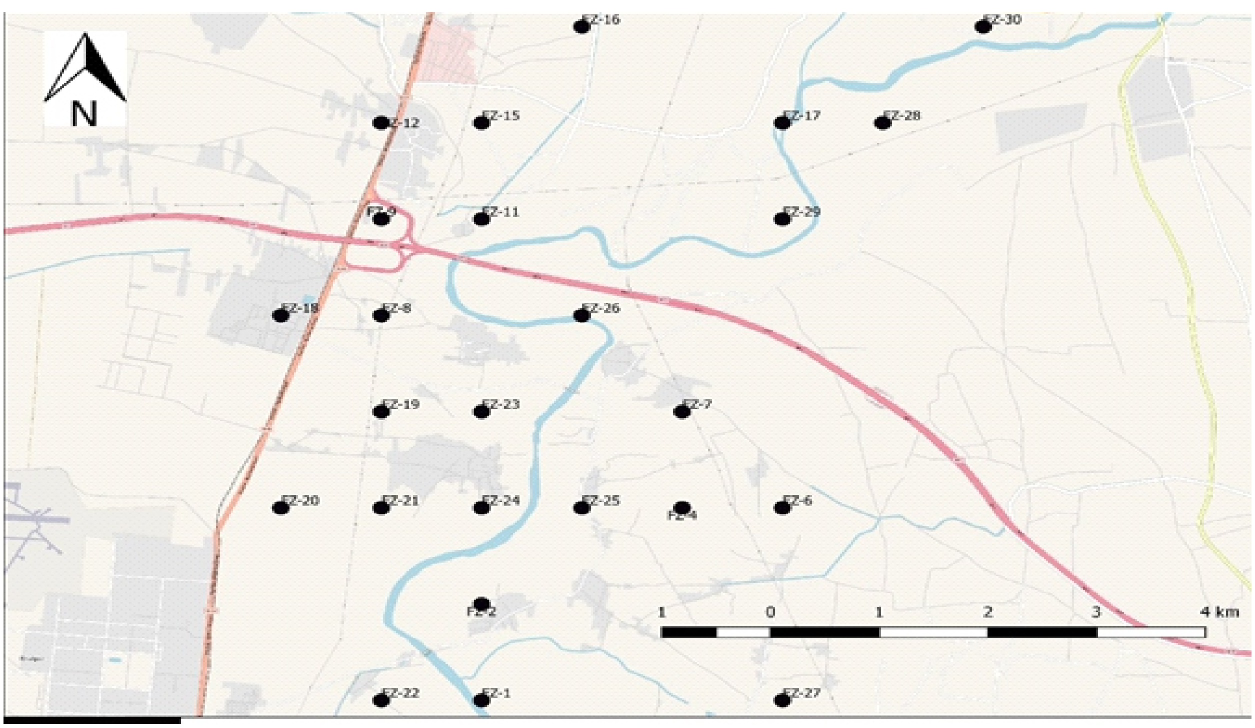


Fig. 1. Map of the Study area showing sample locations.

2.2. Selection of parameters

Tripathi and Singal (2019), used a method of parameter selection based on Principal Component Analysis for Water Quality Index calculation adopted for this study. Their method can be summarized as:

1) Principal Component Analysis.
2) The selection of drinking water quality parameters based on the contribution (>0.35, positive or negative) of the parameters in the selected components.
3) Further screening of the parameters based on a correlation matrix. Parameters with the least correlation are retained for further analysis.

The Kaiser-Meyer-Olkin (KMO) test was conducted to ascertain the suitability of the dataset for conducting PCA analysis. The KMO test is an indicator of sampling adequacy for each variable utilized in the analysis and for the overall model. It is preferable to have a Mean Sampling Adequacy (MSA) closer to 1, whereas an MSA value less than 0.5 is generally considered unsuitable for carrying out PCA (Kaiser, 1958). The KMO test returned an MSA value of 0.513 for the dataset utilized in this study. The Bartlett test of sphericity returned a significance value of 2.2e-16 thus indicating that the correlation matrix is not an identity matrix and therefore, the data is suitable for carrying out PCA (Bartlett, 1950).

The main aim of the Principal Component Analysis (PCA) is to reduce the dimensionality of a multivariate data set. The method reduces the data into fewer components that explain most of the information contained in the data. It is recommended that the data shall be normalized before the application of the PCA technique (Reimann, et. al., 2008). Hence, the original dataset in this study was subjected to log transformation before the application of the PCA technique. Components were selected based on their eigenvalues and cumulative variances. For the purpose of this study, the first six components with eigenvalue (λ) > 1.0 (Fig.3) were selected for further analysis as suggested by Kaiser (1958). The selected components explain about 79.48% of the cumulative variance as shown in the scree plot (Fig.4). The applications of PCA depend on a-prioiri knowledge and its usage implies both heuristics and statistical techniques (Jackson, 1993). Given the importance of pH and DO in water quality index calculations the method was applied to the screening of all parameters except these two. Drinking water quality parameters exhibiting high contributions (> 0.35, negative or positive) among the selected six components (Table 1), were shortlisted for further screening (Tripathi and Singal, 2019). This resulted in the reduction of the total number of parameters from 21 to 11. The 11 shortlisted parameters were EC, TDS, NTU, TH, Mg, SO_4, Cu, Mn, Cr, Zn and Cd. These parameters were then subjected to correlation analysis using Pearson correlation (Table. 2). Correlation analysis estimates the relationship between two variables based on their covariance. The correlation coefficients in a correlation matrix vary between -1 and +1 (Liu, et. al., 2003, Reimann, et.al., 2008). The least correlated parameters were shortlisted while parameters showing significantly higher correlations (> 0.7) were excluded from the analysis. The above-mentioned criteria resulted in the further exclusion of parameters such as EC, TDS and Cu, thus leaving a final set of ten parameters behind i.e. pH, DO, NTU, TH, Mg, SO_4, Mn, Cr, Zn and Cd.

2.3. Geostatistical analysis

Kriging is a popular statistical interpolation method which estimates values of a random field at an unobserved location using samples. Kriging results are based on the spatial autocorrelation of variables, instead of using distance as the sole determinant. The method is based on George Matheson's "The Theory of Regionalized Variables" and was named after Diane Krige (Hengl, 2008). The spatial variation in this method is quantified by a semi-variogram model. The empirical semivariogram is then fitted using spatial variability models such as spherical, exponential, circular and, gaussian etc. Ordinary Kriging (OK), Simple Kriging (SK), Universal Kriging (UK), Indicator Kriging, Probability Kriging, Disjunctive Kriging and Empirical Bayesian Kriging (EBK) are different types of Kriging methods in use.

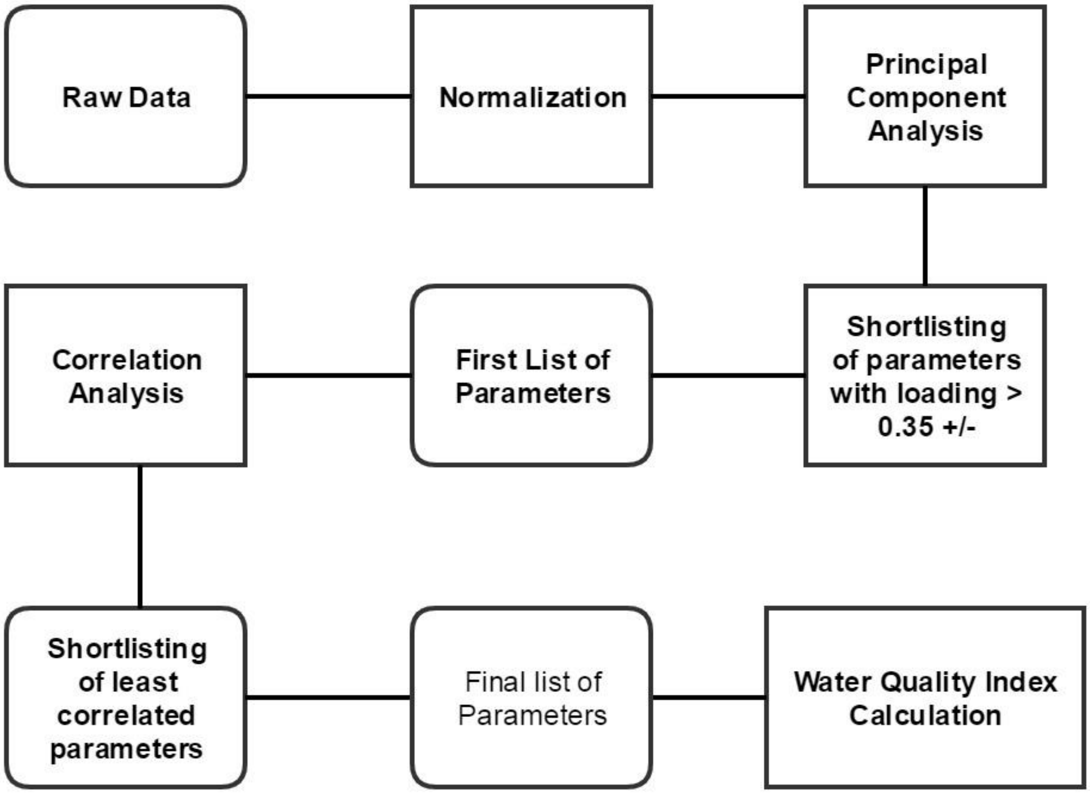


Fig. 2. Flowchart summarizing the methodology used for parameter selection in this study.

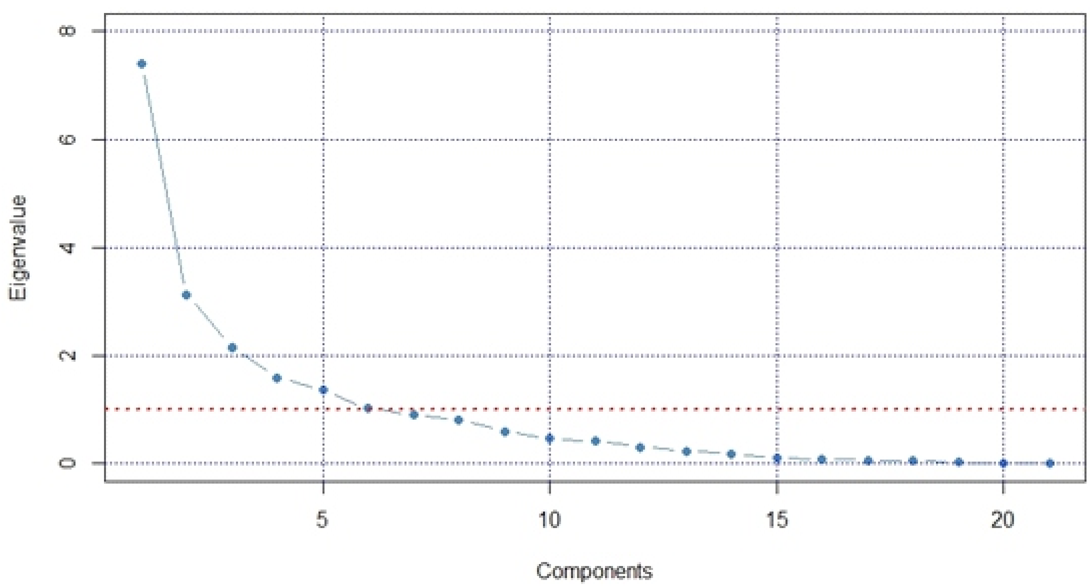


Fig. 3. Eigenvalues (λ) of the Components derived in this study. The first six components with λ > 1 (red dotted line in the graph) were selected for further analysis.

Table 1. Factor Loadings for Water Quality Parameters

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
C	0.006322	-0.0202	0.011879	0.024112	-0.01358	-0.0282
pH	0.000568	-0.00136	0.003243	-0.00102	0.003434	0.001846
EC	-0.22336	-0.01094	0.041821	0.039213	-0.42738	-0.067
TDS	-0.22326	0.001342	0.039598	-0.01588	-0.39099	-0.07057
NTU	-0.33229	-0.57421	-0.14504	-0.39918	0.359886	0.104158
DO	0.011805	0.005546	-0.02792	0.040974	-0.00703	-0.04918
TH	-0.30378	0.086481	0.554131	-0.04674	0.067542	-0.09659
Ca	-0.23389	0.02101	-0.10154	0.11619	-0.11604	0.259665
Mg	-0.27577	0.101877	0.484503	-0.0194	0.122831	0.022091
Na	-0.02542	-0.1745	-0.00826	-0.04609	-0.21194	-0.0058
K	-0.08161	-0.09305	0.049	-0.07821	0.128955	0.14539
SO4	-0.0953	-0.26677	-0.22095	0.068854	-0.53855	0.190985
NO3	-0.26658	-0.11973	-0.19996	-0.17005	0.029953	-0.20854
Fe	-0.1107	0.071949	-0.0079	-0.13996	-0.03399	-0.20727
Cu	-0.43933	0.257119	0.074198	-0.19585	-0.17424	0.060791
Mn	-0.35355	-0.30487	-0.09762	0.61634	0.168763	-0.44159
Pb	-0.12012	0.184425	0.002091	0.156749	-0.03548	0.113555
Cr	-0.26506	0.539953	-0.54555	-0.19125	0.160046	-0.20458
Zn	-0.2429	0.045682	-0.11825	0.234898	0.182171	0.691363
Ni	-0.04296	0.098826	0.033484	-0.05079	0.080095	0.149478
Cd	-0.05812	0.158637	-0.05917	0.471083	0.15048	0.036828

In this study, Ordinary Kriging (OK) and Empirical Bayesian Kriging (EBK) methods were evaluated for performance in interpolating water quality parameters. In OK the predictions are based on:

$$Z(s) = \mu + \varepsilon'(s) \quad (6)$$

In Eq.6. μ, is the constant stationary function and ε' (s) is the spatially correlated part of the variation (Barnett, 2004, Hengl, 2008).

The spatial variability models evaluated in this study include Circular, Spherical, Exponential, Gaussian, K Bessel, J Bessel and Stable.

The Empirical Bayesian Kriging (EBK) differs from other kriging methods in that, it simulates new values for input locations based on a semivariogram estimated from the data. Then a new semivariogram is developed from the simulated data, the weight for the semivariogram is calculated using Bayes rules (Krivoruchko, 2012).

Geostatistical analysis were carried out in ArcGIS software version 10.2.

2.4. Exploratory data analysis

Application of geostatistical techniques to environmental variables requires a normal distribution of the data (Webster & Oliver, 2007). Exploratory data analysis is the precursor to any geostatistical analysis. It is carried out in order to familiarize with one's data and detect patterns of regularity. In other words, the exploratory analysis gives the experimental or empirical distribution and behaviour of the data regardless of its location (Kitanidis, 1997). Histograms, Normal QQ plots, Ogives and Boxplots are the visual aids for exploring the experimental or empirical distribution of data. In this study Histograms were used for exploratory analysis of the data. Histograms show the frequency distribution of the given data. It helps in the assessment of the location, spread and shape of the data (Peck, et.al., 2008). Another important characteristic of the data is its symmetry. Skewness coefficient provides a measure of the symmetry of the data. The skewness coefficient of symmetric distribution is zero. If the data contains many values slightly smaller than the mean and few values much larger than the mean the data is considered positively skewed, on the contrary, if the data contains many values slightly larger than the mean and few values much smaller than the mean then the skewness coefficient is negative (Kitandis, 1997). For strong positive skewness values (> 1) log transformation is suggested while for data sets exhibiting skewness coefficients values < 1 square root transformation is considered appropriate (Webster & Oliver, 2007, Cressie, 1992).

2.5. Model performance evaluation

The evaluation of the models was based on Mean Error (ME), Root Mean Square Error (RMSE), Root Mean Square Error Standardized (RMSES), Mean Squared Error (MSE) and Average Standard Error (ASE) (Cressie, 1992). These metrics are expressed in Equations 7-11, given below:

$$ME = \frac{1}{N} \sum_{i=1}^{n} [Z(si) - z(si)] \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} [Z(si) - z(si)]^2}{n}} \quad (8)$$

$$MSE = \frac{\sum_{i=1}^{n} [Z(si) - z(si)]/\sigma(si)}{n} \quad (9)$$

$$\text{RMSES} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [Z(si) - z(si)/\sigma(si)\ ]^2} \quad (10)$$

$$ASE = \sqrt{\frac{\sum_{i=1}^{n} \sigma^2(si)}{n}} \quad (11)$$

The most commonly used statistic for the comparison of the performance of the interpolation methods is RMSE. RMSE is a measure of the difference between locations that are known and locations that have been interpolated or digitized, it is useful while comparing different models. The smaller its values the better the model results. ME is the average difference between observed values and values predicted by the model, it is an indicator of the bias in model predictions and shall ideally be 0, the lower its value the lower is the bias in model predictions. It is suggested that the value of MSE shall be closer to 0 and that of RMSSE shall be closer to 1, a value greater than 1 represents underestimation in the model prediction while a value less than 1 may represent that the model is overestimating the variability in predictions (Johnston, 2001, Cressie, 1992).

R Studio version 3.4.3 was used for statistical analysis and data visualization in the study (R Studio Team, 2015).

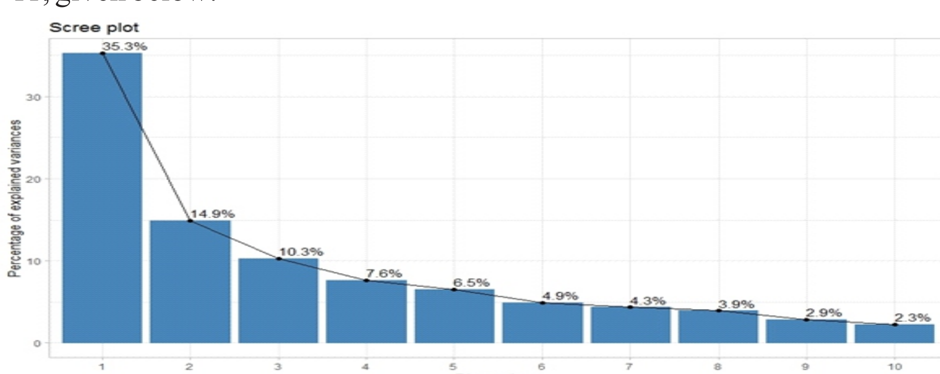


Fig. 4. Scree plot of the first ten components showing the variance explained by the individual components. The first six components explain about 79.48 % of the variance for the original dataset.

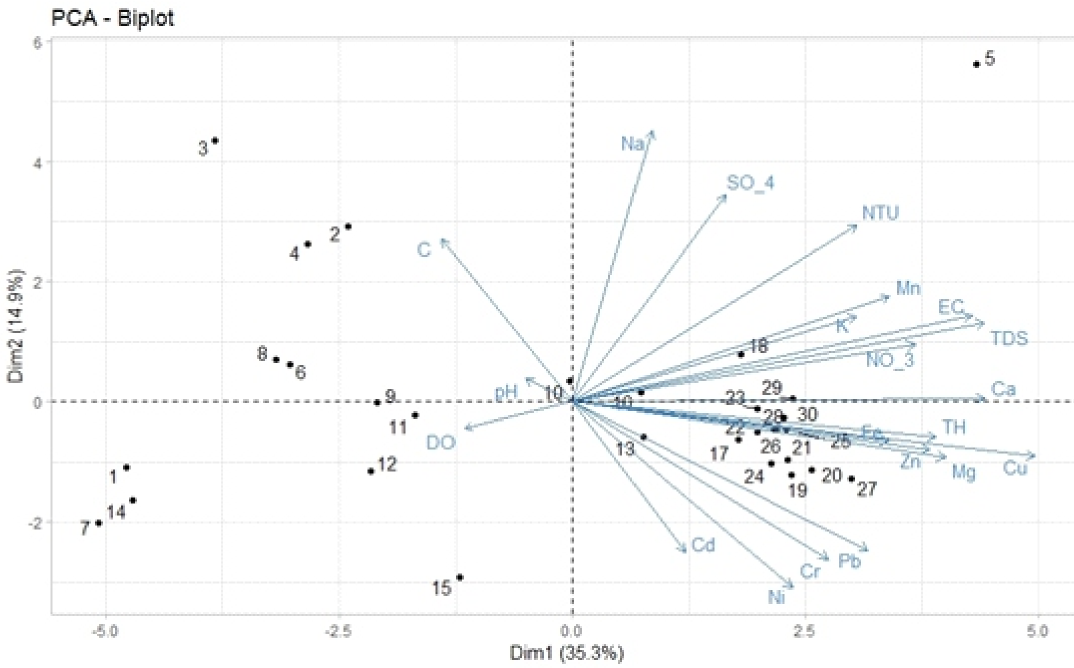


Fig. 5. Biplot showing individuals and variables.

Table 2. Pearson Correlation Matrix of the shortlisted drinking water quality parameters.

Parameters	EC	TDS	NTU	TH	Mg	SO4	Cu	Mn	Cr	Zn	Cd
EC	**1**										
TDS	**0.9526**	**1**									
NTU	0.1837	0.2101	**1**								
TH	**0.5835**	**0.6367**	0.3657	**1**							
Mg	**0.5611**	**0.6154**	0.3701	**0.9984**	**1**						
SO4	0.1809	0.1466	0.3165	-0.1802	-0.2051	**1**					
Cu	**0.6572**	**0.7117**	0.0294	**0.6583**	**0.6459**	-0.1519	**1**				
Mn	0.3500	0.3264	**0.6832**	0.3470	0.3415	0.2527	0.0897	**1**			
Cr	0.1435	0.2536	-0.1095	0.1730	0.1693	-0.1824	**0.5549**	-0.0109	**1**		
Zn	0.4023	0.3955	**0.5233**	0.4728	0.4701	0.1260	0.3437	0.4897	0.0858	**1**	
Cd	0.1853	0.0976	-0.1671	0.0870	0.0836	-0.1119	-0.0038	0.4245	-0.0743	0.2918	**1**

3. Results and discussion

The total number of parameters included in this study was 21 for 30 drinking water samples, which were effectively reduced to 10 using PCA for inclusion in water quality index and effective weights calculations. The descriptive statistics of the selected water quality parameters are shown in table 4, the discussion of the whole range of parameters is out of the scope of the current manuscript. The spatial distribution of the selected parameters was mapped employing suitable interpolation methods using ArcGIS 10.2 Geostatistical Analyst. The interpolated surfaces were then utilized as an input for final WQI map, using Raster Calculator.

Two methods of spatial interpolation namely, Ordinary Kriging and Empirical Bayesian Kriging were both compared to assess their performance in modelling prediction surfaces for pH, DO, NTU, TH, Mg, SO4, Mn, Cr, Zn and Cd. During exploratory data analysis the skewness, kurtosis and outliers were also ascertained for all the parameters (Table 4). Outliers were detected in the case of NTU, Cr, and Zn. All parameters show different degrees of asymmetry, therefore attempts were made to achieve approximate normality using appropriate transformation methods. Data with skewness coefficients between -0.5 and 0.5 are considered approximately normal, whereas, data with skewness coefficients > 1 and < -1 are considered highly skewed. Generally, it is believed that for parameters with strong positive skewness (> 1) coefficient log-normal transformation shall be applied while parameters with skewness values between 0.5 and 1, square root transformation method shall be used to achieve the objective of approximate normality . As shown in Table 4. pH, DO and Sulphates (skewness < 1) were subjected to square root transformation. NTU, Mn, Cr and Cd show strong positive skewness (> 1). Though in such a case log-normal transformation is a preferred method, it was not helpful in all the cases, except Cd. In the case of NTU the removal of outlier value (127.5) and subsequent transformation using the square root method, helped in normalizing the data (skewness 0.0306). The removal of outlier value (0.856) for Cr helped in the reduction of non-normality of the data and no transformation was needed thereafter. For Mn, the approximate normality was achieved only with the help of Boxcox transformation. Zn has a skewness coefficient of 0.351352 and satisfies the condition of approximate normality and therefore no transformation was required.

On the other hand, Mg and TH express moderate negative skewness -0.6506 and -0.6873 and high kurtosis values 2.5862 and 2.6506 respectively. In both cases, data was reflected first and then subjected to log-normal transformation. The non-normal distribution of the variables is depicted using normal histograms in Fig 6. A comparison of skewness coefficients before and after the application of transformation methods is shown in Table 3.

Table 3. Comparison of Skewness before and after transformation

	Skewness	
Parameter	Before Transformation	After Transformation
pH	0.789477	0.312618
NTU	5.156437	0.030622
DO	0.719904	0.194931
TH	-0.687353	0.410255
Mg	-0.650678	0.379307
SO4	0.753246	0.060971
Mn	3.031993	0.156512
Cr	3.945009	-0.339106
Zn	0.351352	NA
Cd	1.789642	0.497529

Table 4. Descriptive Statistics of the selected Water Quality Parameters

Parameter	pH	NTU	DO	TH	Mg	SO4	Mn	Cr	Zn	Cd
Min	6.8	1.1	2.96	19.8	9.55	9	0.015	0.016	0.034	0.084
Max	7.2	127.5	4.67	535	128	94	0.458	0.856	0.32	0.642
Mean	6.96	7.526667	3.523	301.99	72.07867	42.73	0.0695	0.139	0.141967	0.2253
S.D.	0.0788	22.71822	0.4101	125.2577	29.57542	19.729	0.107101	0.147	0.0652	0.115766
Q1	6.905	1.9	3.2025	208.25	50.1325	30.865	0.0255	0.070	0.0837	0.15025
Q2	7.017	4.7175	3.905	389.1	92.165	48.962	0.052	0.155	0.1777	0.2525
IQR	0.1125	2.8175	0.7025	180.85	42.0325	18.097	0.0265	0.084	0.094	0.10225
L Bound	6.736	2.32625	2.1487	63.025	12.9163	3.7187	0.01425	0.056	0.0572	0.00313
U Bound	7.186	8.94375	4.9587	660.37	155.2138	76.108	0.09175	0.281	0.3187	0.40587
Skewness	0.831	5.431894	0.7583	-0.7240	-0.68544	0.7934	3.20800	4.136	0.4103	1.90161
Kurtosis	1.855	29.65695	0.2006	0.1853	0.26183	0.7025	9.76160	20.49	0.5273	4.67460

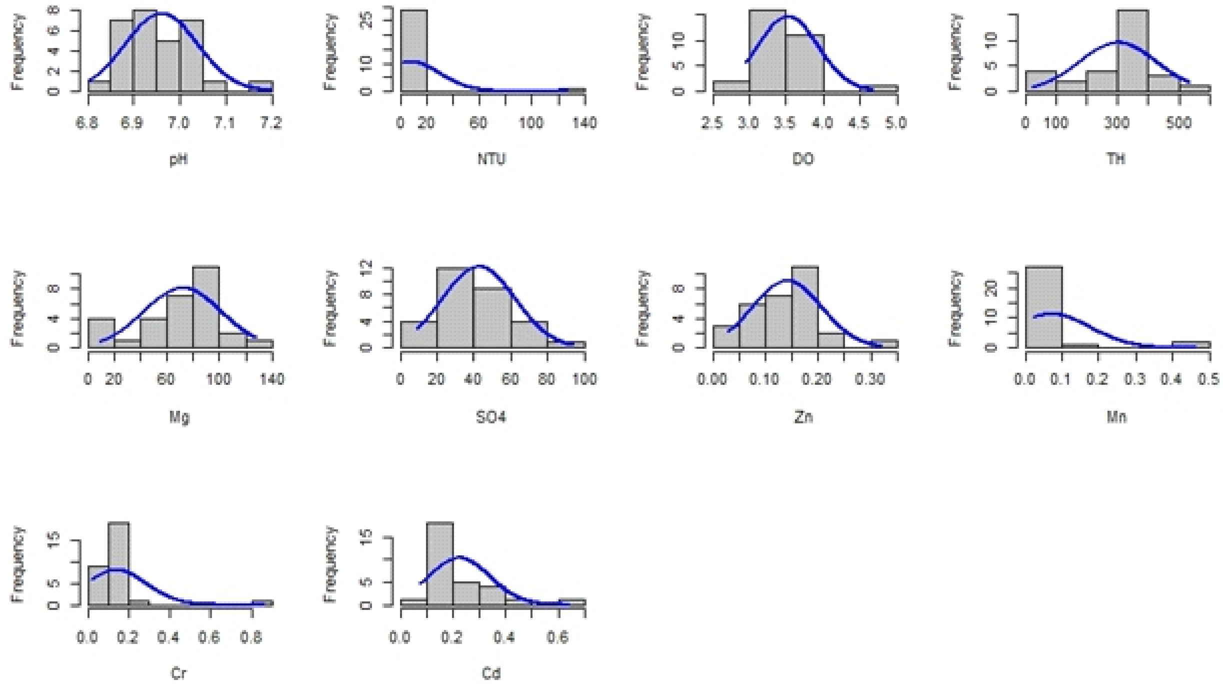


Fig. 6. Showing the non-normal distribution of the water quality parameters included in this study.

The model prediction performance statistics for the water quality parameters are summarized in table 5. A review of the literature shows that the usage of these error metrics in model selection has differed among researchers. Though RMSE is the most widely used criterion for the selection of optimal model (Shahid et al., 2017, Gundogdu & Guney, 2007; Sakizadeh, 2019) other authors have used ME as the sole criterion for the purpose (Audu & Usman, 2015). Though it is recommended that other error metrics such as RMSES which is an indicator of overestimation or underestimation in model prediction results shall also be taken into account. For a model to be considered appropriate the RMSE and ASE values shall be closer enough (ESRI, 2019). In this study, a combination of the above criterion was utilized for the selection of an appropriate model for spatial variability of the selected parameters. For pH, the Stable model shows the lowest ME and RMSE values in comparison to all other models. However, the RMSES value for EBK is closer to 1, while the rest of the models have RMSES values significantly higher than 1 indicating a possible underestimation in the prediction of variability. Further, the ASE is closer in value to the RMSE, this and the above-mentioned criteria suggests that the EBK model predictions for pH are appropriate. Models performance in terms of RMSE did not show much variation in this study for Dissolved Oxygen (DO) and were closely related. The EBK model performance was preferred over other models based on the fact that it had the least ME and RMSES, the RMSES was closer to 1, also RMSE and ASE were closer as compared to other models. Among the OK models, the performance of J Bessel showed promising results for DO, in terms of ME and RMSE. Error metrics for modelling results of Turbidity using EBK show an overall better performance as compared to other models, The RMSES values for all OK models for NTU show an overestimation in spatial variability predictions except J Bessel. With an RMSES value slightly higher than 1 the results may be underestimated by a fraction as compared to the EBK, yet the overall performance of J bessel for Turbidity is better when it comes to the comparison with other OK models. In the case of TH, EBK show lowest ME and RMSE values in comparison to the OK models. OK models have recorded RMSES values significantly lower than 1, suggesting overestimation in results. On the contrary, RMSES value for EBK model is slightly higher and may suggest underestimation in the model results to a certain degree. Further, the closer values of RMSE and ASE suggest that EBK has performed better

better and model results are more appropriate as compared to other models used in this study. Mg presents a case similar to that observed in the evaluation of model performance for TH, EBK in both cases outperformed all OK models. Similarly, error metrics for SO4 also indicate EBK as the optimal model for spatial distribution modelling. In case of Mn, EBK has the highest RMSE as compared to other models, yet OK models show a high degree of underestimation in prediction results in terms of RMSES and least bias in terms of ME as compared to it. Validation results for Cr are of special interest, as the Exponential model has the lowest ME and J Bessel model has the lowest RMSE. EBK has the highest ME and RMSE in comparison to all other models, in contrast, it shows the lowest RMSES values. The RMSE and ASE values for all the models are far apart from each other for all the models evaluated. The exponential model was selected based on its lowest ME and the fact that its RMSE and ASE are relatively closer to each other as compared to the other models under evaluation. For Zn, the overall performance of the EBK model was considered satisfactory. Once again in the case of Cd, RMSE for all the models are closely associated, with EBK showing slightly higher RMSE than other models. However, other models were not considered optimal owing to the high RMSES values. Also, RMSE and ASE for EBK are closer in value thus representing good performance for prediction of the spatial variability of the parameter. The prediction surfaces for the water quality parameters are shown in figure 7 and 8.

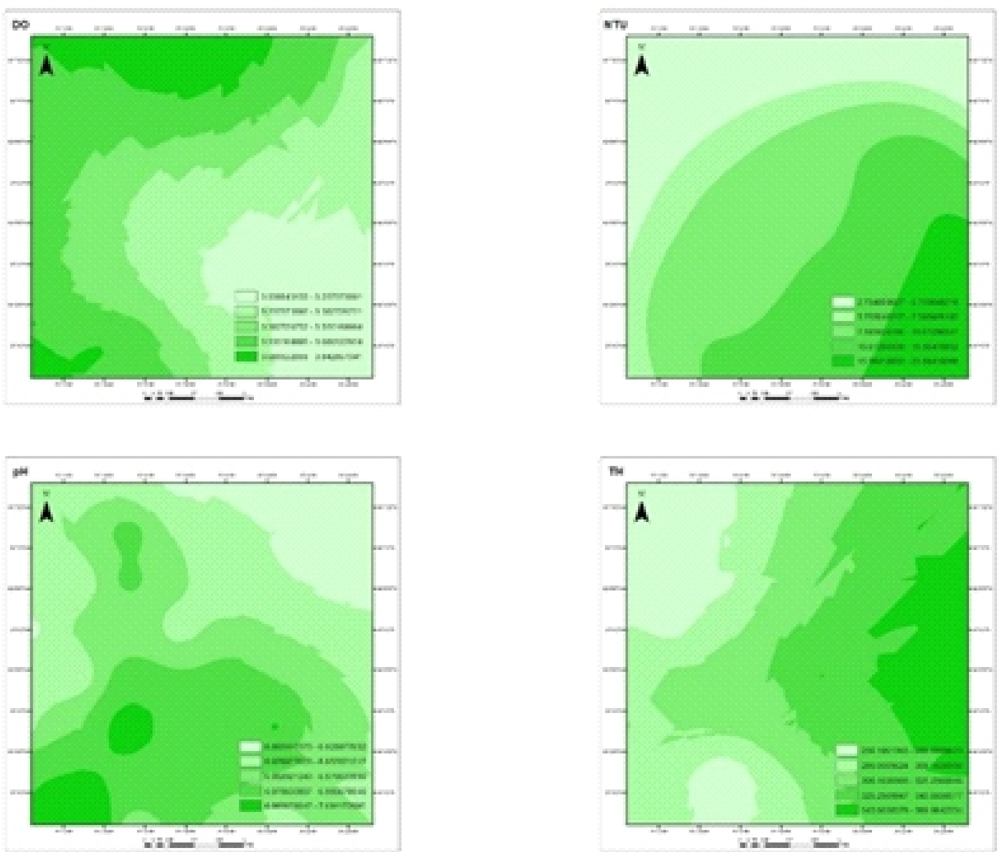


Fig. 7. Prediction maps of Do, NTU, pH and TH.

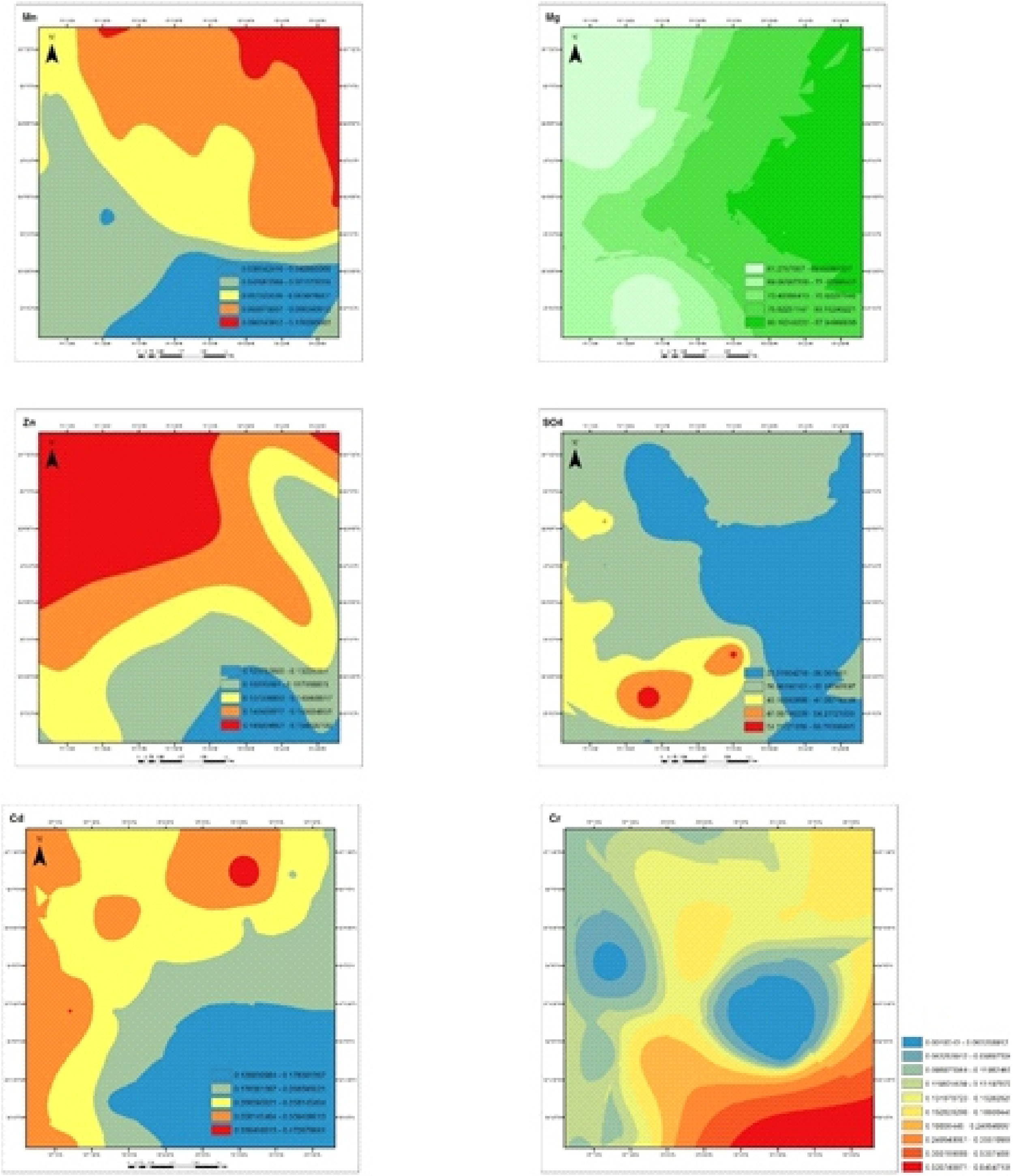


Fig. 8. Prediction maps for Mn. Mg, Zn, SO4, Cd and Cr.

Table 5. Cross-Validation results for the models evaluated in this study

Parameter	Kriging Method	Model	Mean	RMSE	MSE	RMSES	ASE
pH	**OK**	Circular	0.006588	0.102958	0.070510	1.172763	0.087115
		Spherical	0.006749	0.102915	0.072353	1.170480	0.087259
		Exponential	0.007146	0.102852	0.076958	1.165386	0.087617
		Gaussian	0.006574	0.102887	0.070326	1.173146	0.087034
		K Bessel	0.006643	0.102871	0.071108	1.172028	0.087104
		J Bessel	0.006468	0.104190	0.069052	1.170785	0.088380
		Stable	0.002892	0.086273	0.038605	1.091364	0.076406
	EBK		0.001535	0.093926	0.012969	0.990092	0.094568
DO	**OK**	Circular	0.020911	0.417232	0.061278	1.092151	0.376354
		Spherical	0.026297	0.419326	0.076448	1.098313	0.375343
		Exponential	0.033262	0.421496	0.093156	1.107047	0.374915
		Gaussian	0.023165	0.418950	0.067858	1.098434	0.375775
		K Bessel	0.025072	0.419086	0.073172	1.097881	0.375465
		J Bessel	0.041075	0.410783	0.109679	1.086339	0.371879
		Stable	0.023165	0.418950	0.067858	1.098434	0.375775
	EBK		0.003966	0.415649	0.004980	0.981799	0.421902
Turbidity	**OK**	Circular	0.072086	1.507014	0.022675	0.732670	2.066522
		Spherical	0.062827	1.498308	0.030354	0.747515	2.042163
		Exponential	0.063433	1.514421	0.032462	0.756677	2.028477
		Gaussian	0.069878	1.511076	0.026694	0.743328	2.053745
		K Bessel	0.062226	1.512821	0.033425	0.757791	2.027116
		J Bessel	0.034399	1.581929	0.149315	1.045685	1.901890
		Stable	0.064172	1.515658	0.032077	0.756412	2.029188
	EBK		0.002187	1.365918	0.002525	0.993654	1.377531
TH	**OK**	Circular	10.173654	113.20	0.017970	0.765677	158.653
		Spherical	11.113683	112.18	0.002377	0.732979	159.365
		Exponential	11.541855	113.73	0.002887	0.711841	161.171
		Gaussian	9.987394	113.97	0.018045	0.755322	159.277
		K Bessel	11.301087	113.57	0.003294	0.735657	159.794
		J Bessel	6.400817	109.57	0.033013	0.743474	157.029
		Stable	9.987394	113.97	0.018045	0.755322	159.277
	EBK		0.321677	98.49	0.003680	1.019285	97.137

Continued..

Parameter	Kriging Method	Model	Mean	RMSE	MSE	RMSES	ASE
Mg	**OK**	Circular	2.555267	27.375	0.018538	0.760017	38.549
		Spherical	2.831271	27.097	0.001065	0.726561	38.699
		Exponential	2.945851	27.434	0.004955	0.704100	39.134
		Gaussian	2.513665	27.546	0.018428	0.749762	38.694
		K Bessel	2.870068	27.440	0.002443	0.729413	38.822
		J Bessel	1.685110	26.621	0.033419	0.742236	38.146
		Stable	2.513665	27.546	0.018428	0.749762	38.694
	EBK		0.249505	23.515	0.011032	1.011738	23.337
SO4	**OK**	Circular	-0.361770	13.262	0.158120	1.183916	14.2167
		Spherical	-0.487875	13.583	0.179399	1.222681	14.0413
		Exponential	-0.035465	13.212	0.092404	0.958667	15.6209
		Gaussian	-0.800069	14.232	0.217799	1.273017	13.8983
		K Bessel	-0.474142	14.080	0.172282	1.188830	14.1032
		J Bessel	-0.542834	14.676	0.321776	1.545956	13.4082
		Stable	-0.450053	14.004	0.169212	1.179991	14.2494
	EBK		-0.026991	12.090	0.001979	0.979570	12.3766
Mn	**OK**	Circular	-0.010416	0.093448	0.425786	2.248174	0.050659
		Spherical	-0.010498	0.093345	0.421502	2.227878	0.050672
		Exponential	-0.010288	0.093499	0.423425	2.256023	0.049978
		Gaussian	-0.010525	0.093216	0.425220	2.252402	0.049737
		K Bessel	-0.010577	0.093219	0.425122	2.257222	0.049234
		J Bessel	-0.010867	0.093107	0.421311	2.220084	0.048956
		Stable	-0.010525	0.093216	0.425220	2.252402	0.049737
	EBK		-0.008208	0.096988	0.083648	0.997920	0.096973
Cr	**OK**	Circular	-0.027389	0.182435	0.344772	1.746761	0.108471
		Spherical	-0.026154	0.183275	0.354195	1.772450	0.113085
		Exponential	-0.012631	0.183472	0.245176	1.688710	0.148121
		Gaussian	-0.031681	0.180487	0.428674	1.734887	0.100113
		K Bessel	-0.029577	0.181062	0.418336	1.698064	0.104535
		J Bessel	-0.028636	0.179558	0.482756	1.913579	0.104043
		Stable	-0.030165	0.180737	0.408541	1.690220	0.103774
	EBK		-0.019444	0.192684	0.053452	1.135615	0.121993

Continued..

Parameter	Kriging Method	Model	Mean	RMSE	MSE	RMSES	ASE
Zn	**OK**	Circular	0.002127	0.059558	0.201382	1.334334	0.057534
		Spherical	0.002482	0.059527	0.184291	1.282972	0.058683
		Exponential	0.003838	0.055225	0.077615	1.000810	0.065088
		Gaussian	0.001973	0.059914	0.193687	1.293811	0.057607
		K Bessel	0.002099	0.059894	0.192067	1.290899	0.058058
		J Bessel	0.002905	0.058959	0.162233	1.250638	0.057536
		Stable	0.001973	0.059914	0.193687	1.293811	0.057607
	EBK		-0.000403	0.050358	0.007438	1.017631	0.049739
Cd	**OK**	Circular	-0.009656	0.124235	0.202558	1.728979	0.092467
		Spherical	-0.009473	0.124382	0.199429	1.731555	0.092621
		Exponential	-0.009473	0.124382	0.199429	1.731555	0.092621
		Gaussian	-0.008897	0.124328	0.191608	1.727403	0.092339
		K Bessel	-0.008872	0.124405	0.195012	1.751263	0.092409
		J Bessel	-0.007989	0.124433	0.184015	1.745166	0.091962
		Stable	-0.008897	0.124328	0.191608	1.727403	0.092339
	EBK		-0.004890	0.128018	0.037676	0.968993	0.131914

In this study, the water samples from water distribution system were evaluated for drinking water quality using the Water Quality Index. The Water Quality Index was calculated using a total of ten parameters shortlisted from an array of 21 parameters as described earlier using PCA. pH, NTU, DO, TH, Mg, SO4, Mn, Cr, Zn and Cd were taken into account for the calculation of Water Quality Index. In order to calculate the WQi for each sampling site the parameters were assigned a weight value based on its relative importance in impacting water quality and it's bearing on human health. Weight values vary between 1 and 5, suggesting the least impact and highest impact on water quality respectively (Yidana & Yidana, 2010, Varol & Davraz, 2015). Cd, Cr and Mn were assigned the highest weight value of 5, considering their importance in drinking water quality assessments. Trace metals in water may come from both geogenic and anthropogenic sources, their accumulation in water beyond a certain limit may cause various health problems. Cr concentration above a limit of 0.05 mg/l in water is considered hazardous to public health. Similarly, Cd and Mn may cause several health-related issues if present in concentrations above the recommended limits proposed by WHO (2008). The main sources of Cd in drinking water is the zinc impurities in galvanized pipes and solders commonly used for water distribution. Mn is abundantly found in nature and primarily contributes to the drinking water sources from geogenic sources. Higher concentrations of Mn in drinking water is linked with neurological disorders. WHO maintains a guideline value of 0.4 mg/l for Mn. pH and Do were assigned wi value of 4, SO4 was assigned a value of 3 and the lowest value of 2 was assigned to TH, Mg and Zn. TH and Mg do not have any health-based guidelines values, and its concentration in water is not considered as a health concern. The primary impact of Zn in water quality is its impact on water potability and consumer acceptability (Cotruvo, 2017; WHO, 2010). Relative Weights (Wi) were calculated for each parameter using Eq (1), the results for Wi calculations along with assigned weights and guideline values for each parameter are given in table 6. The WQI was then calculated using Eq (2) and Eq.(3) and (4) and the results are shown in Figure 9. Figure 10 shows the spatial distribution of WQI in the study area. The WQI values for the study areas range from 438.23 to 3151.24 with a mean value of 547.51, which shows that all the samples analyzed in this study have waters classified as “unsuitable for drinking purposes” i.e. WQI > 300.

The effective weight (Ewi) for each parameter was calculated using Eq (5) to ascertain the distinct influence of water quality parameters on the overall WQI. The results of effective weight calculations and its comparison to the relative weights of each parameter are shown in table 7. Cd has the highest mean effective weight (91.045 %) followed by Cr (3.8611 %), thus suggesting a major influence of these parameters on overall WQI calculations. Cd and Cr along with Mn also have the highest relative weights of 14.28 % each, assigned by WQi calculations. Conversely, the effective weight calculated for Mn (0.2640 %) is far lower than Cd and Cr, which may be due to the lower concentration of Mn in water samples (Şener et al., 2017b). Other parameters also have similar relative weights whereas, their effective weights differ from each other showing different degrees of influence on overall WQI, such as pH and DO with a relative weight of 11.48 % show different mean effective weights of 1.1466 % and 0.2866% respectively. Likewise, NTU and SO4 have the same relative weights assigned to them by WQI but NTU has a higher effective weight of 1.3763% as compared to a value of 0.1558% calculated for SO4. The lowest relative weight of 5.7142% was calculated for TH, Mg and Zn, all these parameters show lower values of effective weights as well. Overall Zn has the lowest effective weight as compared to other parameters. In summary, Cd contributed the most to WQI as compared to other parameters.

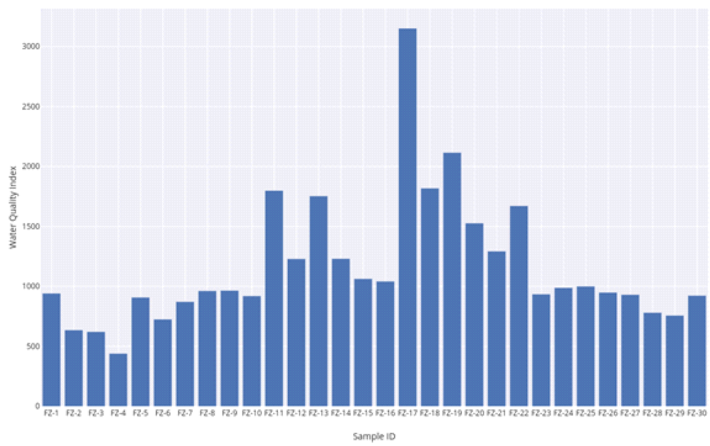
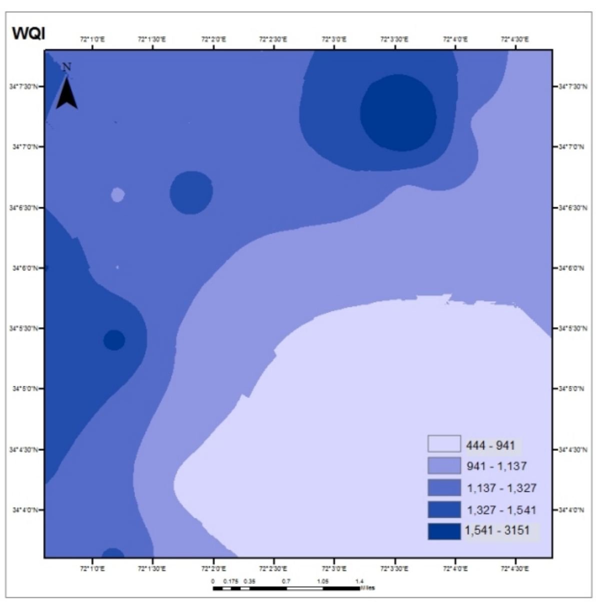


Fig. 9. Water Quality Index values calculated for sample locations in this study.



438

Fig. 10. Water Quality Index (WQI) map of the study area.

Table 6. Assigned Weights and Calculated Relative Weights for drinking water quality parameters.

No	Parameter	Standard limits	Weight (wi)	Relative Weight (Wi)
1	pH	7	4	0.114285714
2	Turbidity	5	3	0.085714286
3	DO	14	4	0.114285714
4	Hardness	300	2	0.057142857
5	Mg	30	2	0.057142857
6	SO4	250	3	0.085714286
7	Mn	0.4	5	0.142857143
8	Cr	0.05	5	0.142857143
9	Zn	3	2	0.057142857
10	Cd	0.003	5	0.142857143
			∑wi=35	∑Wi=1

Table 7. Descriptive Statistics of Effective Weights

No	Parameters	Effective Weight (%)				Relative Weight
		Min	Max	Mean	S.D.	
1	pH	0.3595	2.6041	1.1466	0.4477	11.4285
2	NTU	0.1283	24.1084	1.3763	4.3110	8.5714
3	DO	0.0994	0.5551	0.2866	0.1081	11.4285
4	TH	0.0306	1.1240	0.5592	0.2903	5.7142
5	Mg	0.1934	2.5932	1.4603	0.5992	5.7142
6	SO4	0.0328	0.3428	0.1558	0.0719	8.5714
7	Mn	0.0569	1.7398	0.2640	0.4068	14.2857
8	Cr	0.6896	26.3145	3.8611	4.6800	14.2857
9	Zn	0.0085	0.0672	0.0251	0.0126	5.7142
10	Cd	67.2303	97.0139	91.045	6.9522	14.2857

4. Conclusions

It has been observed that PCA is an effective method in minimizing the burden of parameters for Water Quality Index calculations. WQI show that all samples fall in the unsuitable for drinking category. Consequently, remedial measures shall be implemented before any use. Based on Effective weight calculations Cd is the major influencer of WQI calculations as compared to other parameters. The rest of the parameters in combination contribute only 9 % of the total to WQI. Empirical Bayesian Kriging was compared to Ordinary Kriging in this study. It was concluded that EBK prediction of spatial variability for the majority of the variables except Cr were fitting than spatial variability models tested for the OK method. For Cr Exponential model using OK method produced appropiate results.

Acknowledgements

The Director, National Centre of Excellence in Geology, University of Peshawar, is dually acknowledged for this work. We are also indebted to the comments and suggestions of the anonymous reviewers which helped in improving this manuscript.

Authors Contributions

Wajid Ali and Muhammad Nafees conceived the idea. Wajid Ali carried out the sampling, analysis in laboratory and geostatistical analysis. Syed Ali Turab, Younis Khan and Khaista Rehman contributed to the statistical analysis. All authors contributed equally to the preparation of the manuscript.

References

Abdel-Satar, A. M., Ali, M. H., Goher, M. E., 2017. Indices of water quality and metal pollution of Nile River, Egypt. The Egyptian Journal of Aquatic Research, 43(1), 21–29.

Ahmed, N., Bodrud-Doza, M., Towfiqul Islam, A. R. M., Hossain, S., Moniruzzaman, M., Deb, N., Bhuiyan, M. A. Q., 2019. Appraising spatial variations of As, Fe, Mn and NO3 contaminations associated health risks of drinking water from Surma basin, Bangladesh. Chemosphere, 218,

726–740.
Ali, W., Khan, H. Q., Latif, K., 2009. Pb, Cr, Ni and Cd contamination in Hand pump and Tap water of Mardan District. The Environ Monitor, 9(11–12), 13–26.
Ali, W., Nafees, M., 2015. Nitrates and other Major Ions Concentration in the Groundwater of Mardan City and its vicinity. Journal of Science and Technology, University of Peshawar, 39(2), 13–20.
Anonymous. 2017. District and Tehsil level population. Summary with region break up population no of HH District / Tehsil region. Retrieved from:
http://www.pbscensus.gov.pk/sites/default/files/bwpsr/kp/MARDAN_SUMMARY.pdf.
Arslan, H., 2012. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: The case of Bafra Plain, Turkey. Agricultural Water Management, 113, 57–63.
Audu, I., Usman, A., 2015. An Application of Geostatistics to Analysis of Water Quality Parameters in Rivers and Streams in Niger State, Nigeria. American Journal of Theoretical and Applied Statistics, 4(5), 373.
Ayaz, S., Khan, S., Khan, S. N., Bibi, F., Shamas, S., Akhtar, M., 2011. Prevalence of Zoonotic Parasites in Drinking Water of Three Districts of Khyber Pakhtunkhwa Province, Pakistan. Pakistan Journal of Life and Social Sciences, 9(1), 67–69.
Azizullah, A., Khattak, M. N. K., Richter, P., Häder, D. P., 2011. Water pollution in Pakistan and its impact on public health - A review. Environment International, 37(2), 479–497.
Barnett, V., 2004. Environmental Statistics : Methods and Applications. J. Wiley.
Bartlett, M. S., 1950. Tests of significance in factor analysis. British Journal of Statistical Psychology, 3(2), 77–85.
Beamonte Córdoba, E., Casino Martínez, A., Veres Ferrer, E., 2010. Water quality indicators: Comparison of a probabilistic index and a general quality index. The case of the Confederación Hidrográfica del Júcar (Spain). Ecological Indicators, 10(5), 1049–1054.
Bhuiyan, M. A. H., Bodrud-Doza, M., Islam, A. R. M. T., Rakib, M. A., Rahman, M. S., Ramanathan, A. L., 2016. Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis. Environmental Earth Sciences, 75(12), 1020.
Bhuiyan, M. A. H., Rakib, M. A., Dampare, S. B., Ganyaglo, S., Suzuki, S., 2011. Surface water quality assessment in the central part of Bangladesh using multivariate analysis. KSCE Journal of Civil Engineering, 15(6), 995–1003.
Chaurasia, A. K., Pandey, H. K., Tiwari, S. K., Prakash, R., Pandey, P., Ram, A., 2018. Groundwater Quality assessment using Water Quality Index (WQI) in parts of Varanasi District, Uttar Pradesh, India. Journal of the Geological Society of India, 92(1), 76–82.
Cotruvo, J. A., 2017. WHO Guidelines for Drinking Water Quality: First Addendum to the Fourth Edition. Journal American Water Works Association, 109, 44–51.
Cressie, N., 1992. Statistics for spatial data. Terra Nova, 4(5), 613–617.
De La Mora-Orozco, C., Flores-Lopez, H., Rubio-Arias, H., Chavez-Duran, A., Ochoa-Rivero, J., 2017. Developing a Water Quality Index (WQI) for an Irrigation Dam. International Journal of Environmental Research and Public Health, 14(5).
ESRI. 2019. Comparing models—Help | ArcGIS Desktop. Retrieved February 17, 2019, from http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/comparing-models.htm
Ewaid, S. H., Abed, S. A., 2017. Water quality index for Al-Gharraf River, southern Iraq. The Egyptian Journal of Aquatic Research, 43(2), 117–122.
Gundogdu, K. S., Guney, I., 2007. Spatial analyses of groundwater levels using universal kriging. Journal of Earth System Science, 116(1), 49–55.
Hengl, T., 2008. A Practical guide to Geostatistical Mapping. EUR - Scientific and Technical Research Reports.
Jackson, D. A., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology, 74(8), 2204–2214.

Johnson, C. D., Nandi, A., Joyner, T. A., Luffman, I., 2018. Iron and Manganese in Groundwater: Using Kriging and GIS to Locate High Concentrations in Buncombe County, North Carolina. Groundwater, 56(1), 87–95.

Kaiser, H. F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23(3), 187–200.

Khan, A. R., Rafique, M., Rahman, S. U., Basharat, M., Shahzadi, C., Ahmed, I., 2019. Geo-spatial analysis of radon in spring and well water using kriging interpolation method. Water Science and Technology: Water Supply, 19(1), 222–235.

Khan, K., Lu, Y., Khan, H., Zakir, S., Ihsanullah, Khan, S., Wang, T., 2013. Health risks associated with heavy metals in the drinking water of Swat, northern Pakistan. Journal of Environmental Sciences, 25(10), 2003–2013.

Khan, S., Shahnaz, M., Jehan, N., Rehman, S., Shah, M. T., Din, I., 2013. Drinking water quality and human health risk in Charsadda district, Pakistan. Journal of Cleaner Production, 60, 93–101.

Khosravi, R., Eslami, H., Peirovi, R., Heidari, M., Fallahzadeh, R. A., Almodaresi, S. A., Taghavi, M., 2017. Use of geographic information system and water quality index to assess groundwater quality for drinking purpose in Birjand City, Iran. Desalinization and Water Treatment, 67, 74–83.

Kitanidis, P. K. 1997. Introduction to geostatistics : applications to hydrogeology. Cambridge University Press. Retrieved from https://www.cambridge.org/gb/academic/subjects/earth-and-environmental-science/hydrology-hydrogeology-and-water-resources/introduction-geostatistics-applications-hydrogeology?format=PB&isbn=9780521587471

kumar, K. S., Logeshkumaran, A., Magesh, N. S., Godson, P. S., Chandrasekar, N., 2015. Hydro-geochemistry and application of water quality index (WQI) for groundwater quality assessment, Anna Nagar, part of Chennai City, Tamil Nadu, India. Applied Water Science, 5(4), 335–343.

Krivoruchko, K. 2012. Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst. Retrieved from https://www.esri.com/NEWS/ARCUSER/1012/files/ebk.pdf

Li, H., Smith, C., Wang, L., Li, Z., Xiong, C., Zhang, R., Zhang, R., 2019. Combining Spatial Analysis and a Drinking Water Quality Index to Evaluate Monitoring Data. International Journal of Environmental Research and Public Health, 16(3), 357. https://doi.org/10.3390/ijerph16030357

Liang, C. P., Chen, J. S., Chien, Y.C., Chen, C. F. 2018. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: A kriging approach. Science of The Total Environment, 627, 1048–1057.

Liu, C.W., Lin, K. H., Kuo, Y. M. 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. The Science of the Total Environment, 313, 77–89.

Liu, W. C., Yu, H. L., Chung, C. E., 2011. Assessment of water quality in a subtropical alpine lake using multivariate statistical techniques and geostatistical mapping: a case study. International Journal of Environmental Research and Public Health, 8(4), 1126–1140.

Masoud, A. A., 2014. Groundwater quality assessment of the shallow aquifers west of the Nile Delta (Egypt) using multivariate statistical and geostatistical techniques. Journal of African Earth Sciences, 95, 123–137.

McLeod, L., Bharadwaj, L., Epp, T., Waldner, C. L. 2017. Use of principal components analysis and kriging to predict groundwater-sourced rural drinkingwater quality in saskatchewan. International Journal of Environmental Research and Public Health.

Memon, M., Soomro, M. S., Akhtar, M. S., Memon, K. S., 2011. Drinking water quality assessment in Southern Sindh (Pakistan). Environmental Monitoring and Assessment, 177(1–4), 39–50.

Mir, A., Piri, J., Kisi, O., 2017. Spatial monitoring and zoning water quality of Sistan River in the wet and dry years using GIS and geostatistics. Computers and

Electronics in Agriculture, 135, 38–50.
Misaghi, F., Delgosha, F., Razzaghmanesh, M., Myers, B., 2017. Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2017.02.226
Muhammad, S., Tahir Shah, M., Khan, S., 2010. Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques in Kohistan region, northern Pakistan. Food and Chemical Toxicology, 48(10), 2855–2864.
Murphy, R. R., Curriero, F. C., Ball, W. P., Asce, M., 2010. Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay. Journal of Environmental Engineering, 136(2), 160–171.
Nabeela, F., Azizullah, A., Bibi, R., Uzma, S., Murad, W., Shakir, S. K., Häder, D. P., 2014. Microbial contamination of drinking water in Pakistan—a review. Environmental Science and Pollution Research, 21(24), 13929–13942.
Narany, T. S., Ramli, M. F., Aris, A. Z., Sulaiman, W. N. A., Fakharian, K., 2014. Spatial assessment of groundwater quality monitoring wells using indicator kriging and risk mapping, Amol-Babol Plain, Iran. Water (Switzerland), 6(1), 68–85.
Paz-Ferreiro, J., Vázquez, E. V., Vieira, S. R., 2010. Geostatistical analysis of a geochemical dataset. Bragantia, 69 (suppl), 121–129.
Ponsadailakshmi, S., Sankari, S. G., Prasanna, S. M., Madhurambal, G., 2018. Evaluation of water quality suitability for drinking using drinking water quality index in Nagapattinam district, Tamil Nadu in Southern India. Groundwater for Sustainable Development, 6, 43–49.
Reimann, C., Filzmoser, P., Garrett, R. G., Dutter, R. 2008. Statistical Data Analysis Explained: Applied Environmental Statistics with R. Chichester: John Wiley & Sons, Ltd.
Renard, phillipe, Demougeot-Renard, H., Froidevaux, R., 2005. Geostatistics for Environmental Applications. Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications. Berlin: Springer Verlag.
Robinson, T. P., Metternicht, G. 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture, 50(2), 97–108.
Romanto, Wardiatno, Y. 2015. Water Quality Status of Ciambulawung River, Banten Province, Based on Pollution Index and NSF-WQI. Procedia Environmental Sciences, 24, 228–237.
Roxy Peck, Chris Olsen, Jay L. Devore. 2008. Introduction to Statistics and Data Analysis (Third Edition). Blemont, Thomson Brooks.
Saha, N., Das, S., Islam, A. R. M. T., Bodrud-Doza, M., Ahmed, F., Rahman, M. S. 2016. Characterization of groundwater quality using water evaluation indices, multivariate statistics and geostatistics in central Bangladesh. Water Science, 30(1), 19–40.
Sahoo, M. M., Patra, K. C., Khatua, K. K. 2015. Inference of Water Quality Index Using ANFIA and PCA. Aquatic Procedia, 4, 1099–1106.
Sahu, P., Sikdar, P. K. 2008. Hydrochemical framework of the aquifer in and around East Kolkata Wetlands, West Bengal, India. Environmental Geology, 55(4), 823–835.
Sakizadeh, M. 2019. Spatial analysis of total dissolved solids in Dezful Aquifer: Comparison between universal and fixed rank kriging. Journal of Contaminant Hydrology, 221, 26–34.
Şener, Ş., Şener, E., Davraz, A. 2017. Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). Science of The Total Environment, 584–585, 131–144.
Shahid, S. U., Iqbal, J., Khan, S. J. 2017. A comprehensive assessment of spatial interpolation methods for the groundwater quality evaluation of Lahore, Punjab, Pakistan. NUST Journal of Engineering Sciences, 10(1), 1–13.
Singh, K. P., Malik, A., Mohan, D., Sinha, S. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River

(India)—a case study. Water Research, 38(18), 3980–3992.

Tahir Shah, M., Danishwar, S. 2003. Potential Fluoride Contamination in the Drinking Water of Naranji Area, Northwest Frontier Province, Pakistan. Environmental Geochemistry and Health, 25(4), 475–481.

Tao, X. F., Huang, T., Li, X. F., Peng, D. P., 2016. Application of a PCA based water quality classification method in water quality assessment in the Tongjiyan Irrigation Area, China. International Conference on Energy and Environmental Protection (ICEEP 2016) Application,, Atlantis Press.

Tiwari, K., Goyal, R., Sarkar, A. 2017. GIS-Based Spatial Distribution of Groundwater Quality and Regional Suitability Evaluation for Drinking Water. Environmental Processes, 4(3), 645–662.

Tripathi, M., Singal, S. K., 2019. Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India. Ecological Indicators, 96, 430–436.

Varol, S., Davraz, A. 2015. Evaluation of the groundwater quality with WQI (Water Quality Index) and multivariate analysis: a case study of the Tefenni plain (Burdur/Turkey). Environmental Earth Sciences, 73(4), 1725–1744.

Viswanath, N. C., Kumar, P. G. D., Ammad, K. K., Kumari, E. R. U. 2015. Ground Water Quality and Multivariate Statistical Methods. Environmental Processes, 2(2), 347–360.

Webster, R., Oliver, M. 2007. Geostatistics for Environmental Scientists (2nd ed.). John Wiley and Sons, Ltd.

World Health Organization. 2010. Water for Health WHO Guidelines for Drinking-water. Retrievedfrom:http://www.who.int/water_sanitation_health/publications/guideline_policy_procedure/en/.

World health Organization. 2004. Guidelines for Drinking-water Quality, Volume 1. Geneva.

Wu, J., Norvell, W. A., Welch, R. M. 2006. Kriging on highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon. Geoderma, 134(1–2), 187–199.

Yidana, S. M., Yidana, A. 2010. Assessing water quality using water quality index and multivariate analysis. Environmental Earth Sciences, 59(7), 1461–1473.

Yousefi, M., Saleh, H. N., Mohammadi, A. A., Hossein Mahvi, A., Ghadrpoori, M., Suleimani, H. 2017. Data on water quality index for the groundwater in rural area Neyshabur County, Razavi province, Iran. https://doi.org/10.1016/j.dib.2017.10.052