A comprehensive assessment of spatial interpolation methods for the groundwater quality evaluation of Lahore, Punjab, Pakistan

Syed Umair Shahid[1]*, Javed Iqbal[1], Sher Jamal Khan[2]
1Institute of Geographical Information Systems (IGIS), National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan
2Institute of Environmental Science and Engineering (IESE), National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan
* Corresponding author E-mail: umairphd13@igis.nust.edu.pk, rollno1@hotmail.com
Received: 29 September 2016 Accepted: 24 January 2017

Abstract

Spatial interpolation is commonly used to generate water quality surfaces, but different spatial interpolation methods yield different surfaces from the same data. The water quality map produced using one model of spatial interpolation method may be significantly different from the map produced using another model of the same spatial interpolation method. The purpose of this study was to evaluate the performance of different spatial interpolation methods to depict the water quality of Lahore correctly. The water samples (n = 73) were collected from tube wells and tested for physicochemical parameters (pH, turbidity, hardness, total dissolved solids, alkalinity, calcium, and chlorides). The data exploration was performed using SPSS software. The inter-comparison of different powers of inverse distance weighting (IDW) and different functions of radial basis functions (RBF) was completed using geostatistical analyst extension in ArcGIS 10.3. Moreover, these deterministic interpolation methods (IDW and RBF) were compared with geostatistical interpolation methods (ordinary kriging and ordinary co-kriging) based on cross-validation statistics; root means square error (RMSE). The analysis showed that ordinary co-kriging performed much better than ordinary kriging, RBF, and IDW, for water quality assessment of Lahore. Hence, ordinary co-kriging with appropriate auxiliary variable and the best-fitted semi-variogram model was used to generate the spatial distribution map for each water quality parameter. The water quality index (WQI) was computed using the tested physicochemical parameters, and the results showed that 98% of the tube wells were providing 'excellent' to 'good' water quality in Lahore city. However, there were few areas of City and Anarkali subdivisions where it indicated poor to very poor water quality. The procedure used in this study is valuable for the water management authorities to better understand and monitor the groundwater quality.

Keywords: Water Quality Index; Spatial Interpolation; Inverse Distance Weighting (IDW); Radial Basis Functions (RBF); kriging; co-kriging

Introduction

About one-third of the world's population relies on groundwater for drinking purposes. The scenario is not much different in Pakistan as groundwater is the major source of drinking water for most Pakistanis. Lack of safe drinking water is a major problem in rural as well as urban parts of Pakistan [1]. The organic substances and minerals present in drinking water can disturb human health, so water should be treated before drinking. The safe and sustainable use of groundwater requires a regular evaluation of its quality. The Water Quality Index (WQI) is considered as an effective tool to convey the information about overall water quality in a comprehensible and useful manner [2]. An important advantage of WQI is that it combines the data related to all the tested physicochemical parameters for a specific location to produce a single value that makes it very easy to understand the overall quality of water at that location [3].

As water sampling cannot be done at every location, the use of procedures that reflect trustworthy estimates of groundwater quality have become indispensable for monitoring this valuable resource [4]. Nowadays the usage of geospatial technologies has smartly reduced the complexities involved in the evaluation of natural resources and their related environmental concerns. Geographic information systems can support in providing a better solution to a wide range of problems associated with water resources, water availability and water quality assessment at a regional or local level.

The use of spatial interpolation methods to generate water quality surfaces for a region, based on data collected from sampling, is a common practice worldwide. The spatial interpolation methods mostly used in GIS software include Radial Basis Functions (RBF), Inverse Distance Weighting (IDW), kriging and co-kriging. The RBF is a simple but computationally fast spatial interpolation method that does not consider any external variable; rather it uses mathematical functions that represent the variable behavior with a continuous surface [5]. The RBF surface passes through the measured points and can predict beyond the maximum and minimum values of the variable. The method yields better results for gradually fluctuating surfaces. It's unsuitable when large differences exist in the surface values within a short horizontal distance or when there is a suspicion that the sample data is prone to error or uncertainty. Being an exact interpolator, it can be locally sensitive to outliers [6]. Scientists have applied these functions to generate raster data for the estimation of groundwater quality. Giang, et al. [7] investigated the correlation between arsenic concentrations and tube well depth in Thanh Tri, a district located in the southern part of Hanoi City, Vietnam. They applied spline with tension and completely regularized spline functions of RBF to examine their efficiency. The study resulted that both the SWT and CRS functions produced reasonable predictions in terms of arsenic concentration estimation in groundwater. The IDW makes predictions using a linear weighted combination based on the inverse of the distance between the points [8]. It is computationally fast and has the ability to accommodate barriers that reflect the linear discontinuity in the surface. The IDW surface does not pass through the samples. It is recommended when the set of points are evenly distributed throughout the area to capture the extent of local surface variation needed for analysis [9]. The quality of the prediction surface is compromised if the sampling points are clustered. Aminu, et al. [10] measured total suspended solids (TSS), dissolved oxygen, ammonia nitrate (NH3–N), biochemical and chemical oxygen demands (BOD and COD) and pH from seven sampling points to examine the water quality of Bertam River, a main stream in the rapidly growing tourist destination of Cameron Highlands, Malaysia. They preferred IDW method for the generation of water quality surface data as it is more intuitive and efficient. IDW method has also been used in water quality index zonation and in the production of spatial distribution maps of water quality parameters [11]. Kamińska and Grzywna [12] quantitatively compared RBF and IDW with groundwater level data sets of Sosnowica, West Polesie, Poland, to determine the accuracy of both interpolation methods. They used cross-validation statistics based on two criteria: mean error closer to zero, and lowest root mean square error (RMSE) to compare both interpolation surfaces. The results revealed that RBF created the best representation of reality. Kriging method uses spatial autocorrelation values among the sampled locations to estimate values at unsampled locations. It assumes that the data comes from a stationary stochastic process and some techniques require that the data be normally distributed [13]. The kriging surface does not pass through the measured points, and the pixel values can go even beyond the value range of samples. In addition to prediction maps, prediction standard errors, probability, and quantile maps can also be drawn using kriging. In terms of computing time, kriging interpolation method is moderately fast. Its disadvantage is the requirement of many parameter decisions on transformations, trends, models, parameters, and neighborhoods for its computation. It suffers limitations when there is outlier and nonstationarity in the data. Kriging has also been widely used to identify groundwater facies, water vulnerability zones [14] and spatial variability of water quality parameters. Cokriging can be considered as an extension of traditional kriging interpolation to predict the less good intensively sampled primary variable of interest using intensively sampled auxiliary variables [15]. It is a multivariate interpolation method in which one or more auxiliary variables that are correlated with the target variable can be used for prediction. The literature shows that co-kriging has been used for the prediction and estimation of groundwater quality parameters [16]. Hooshmand, et al. [17] applied kriging and co-kriging methods to evaluate the chloride content and sodium adsorption ratio in the groundwater of Boukan area, Iran. The estimated and observed values of both the parameters were compared based on RMSE and coefficient of determination (R^2). They found co-kriging method more accurate than kriging method.

The use of different spatial interpolation methods yields different surfaces from the same data. Each of these interpolation methods includes different models with slight variations to predict the surfaces, but their accuracy also differs greatly. It means that the water quality map produced using one model of spatial interpolation method may be significantly different from the map produced using another model of the same spatial interpolation method. Therefore, it is important to have the knowledge of the most suitable interpolation method and the model of that interpolation method for production of a map that correctly depicts the water quality of the study area.

The comparison of different models of geostatistical methods should be based on mean absolute error closer to zero and root RMSE as small as possible [6]. The values of mean absolute error should be used to determine the best method only when the RMSE of two methods are equal [18]. As the deterministic interpolation methods IDW and RBF provide information about the RMSE, it is appropriate to compare deterministic techniques with geostatistical techniques based on least RMSE [19]. The cross-validation statistics RMSE is calculated using the formula:

$$RMSE = \sqrt{\frac{\left[\sum_{i=1}^{n}\{Z_{(xi)} - z_{(xi)}\}^2\right]}{n}} \tag{1}$$

Where:

$Z_{(xi)}$ is the predicted value, and $z_{(xi)}$ is the observed value at respective spatial locations $x_1, x_2, \ldots, x_n$.

The RMSE is a widely used statistic to measure the error of the prediction surface. It's the least value specifies the most accurate predictions [20]. The literature shows that researchers have kept smallest RMSE a criterion to choose the most suitable interpolation method among different kriging types and variogram models [21], besides using it for the comparison of different deterministic and geostatistical methods [22]. Hence, each spatial distribution map should be produced using the model that shows least RMSE among all the models of all the spatial interpolation methods for that particular water quality parameter.

In the recent years, a number of studies have been published that involve the comparison of spatial interpolation methods, but they usually either compare few spatial interpolation methods [23-24] for water quality evaluation or compare different components of a particular spatial interpolation method [25-26]. This paper does not only involve the evaluation of deterministic and geostatistical spatial interpolation methods in detail, but it also compares their associated powers, functions and models. In order to evaluate the most suitable spatial interpolation method for the groundwater quality assessment of Lahore city, a comprehensive geostatistical analysis was required. Furthermore, analysis of groundwater quality of Lahore city using WQI was also an important issue.

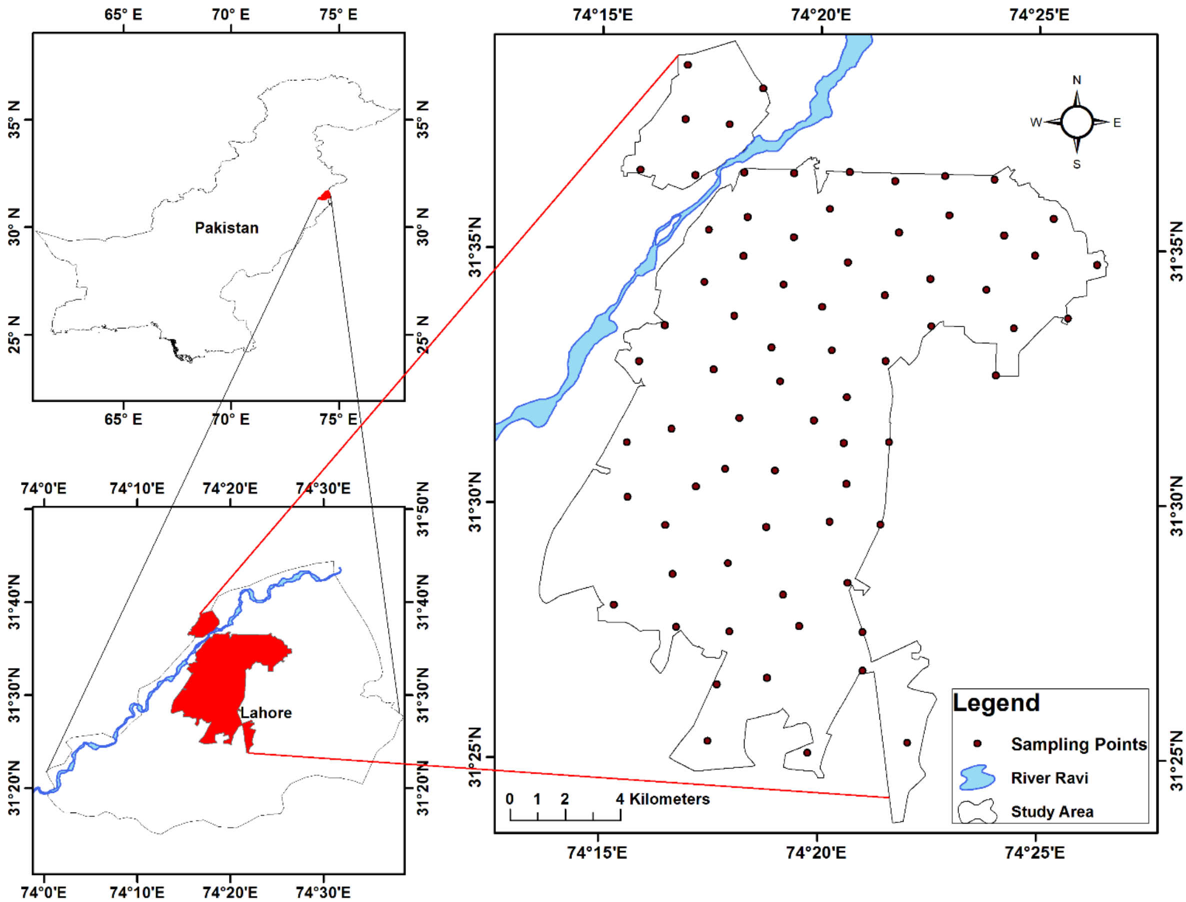

Fig. 1: Study area and the sampling locations in Lahore City

Materials and Methods

Study area

Lahore is the second largest metropolitan of Pakistan. It lies on the eastern border of Pakistan with India. It is surrounded by Sheikhupura District in the northwest and Kasur District in the south. The climate here is semi-arid. It is the responsibility of the Water and Sanitation Agency (WASA) to provide water to the residents of Lahore. It manages the water supply from groundwater using more than 480 tube wells. The WASA has divided its jurisdiction into 27 subdivisions covering an area of 245 km^2. The study area and the sampling locations are shown in figure-1.

Data collection and data preparation

A field survey was conducted to collect the water samples from the study area. The water samples were collected in such a way that they cover the entire area without any clustering. The study involved samples from 73 tube wells. They were tested for pH, turbidity and total dissolved solids using digital meters, whereas, titration method was adopted to test chlorides, alkalinity, hardness, and calcium. The geographic coordinates of the tube wells and the boundary of WASA's administrative units (sub-divisions) were acquired from WASA Lahore. The descriptive statistics of the data collected from water testing was analyzed in SPSS version 20 software. It was very useful in terms of outlier identification. The attribute data containing information about the physicochemical parameters was joined with the geographic coordinates of the respective sampling points. A geodatabase was created in ArcCatalog to keep the data integrated.

Geostatistical analysis

The first step in the geostatistical analysis is the exploratory spatial data analysis (ESDA). The purpose of ESDA is to understand the data quantitatively and notice the spatial patterns that eventually help in better decision making for the construction of interpolation models. There are several interpolation methods available in ArcGIS software. In this study, the deterministic interpolation methods (IDW and RBF) and geostatistical interpolation methods (ordinary kriging and ordinary co-kriging) were performed on 73 sampling points with the help of geostatistical analyst extension in ArcGIS 10.3. The IDW method is simple and requires very few inputs for the interpolation. IDW interpolation was executed on the data set using its powers 1-4 and the optimal power as well. The power at which the prediction surface has smallest RMSE is termed as the optimal power. The RBF interpolation makes predictions using kernel functions. The kernel functions used for performing RBF involved Completely Regularized Spline, Spline with Tension and Thin Plate Spline. Although, every RBF kernel is computed using its own equation for interpolation yet there exists very little differences among them [27]. In order to perform kriging and co-kriging interpolations, the data was first analyzed with ESDA tools including histograms, normal QQ plots, trend analysis tool and semivariogram clouds. Other than exposing the outliers in the data, the normal QQ plot and histogram tool help in identifying whether data is normally distributed or not. The trend analysis tool shows the trends in the data with respect to different directions. The semi variogram cloud shows the autocorrelation in the dataset. The models are fitted to the semivariogram based on functions. The model functions available to fit the empirical semi variogram include Rational Quadratic, Circular, Gaussian, Hole Effect, Spherical, Tetraspherical, Pentaspherical, J-Bessel, Exponential, K-Bessel and Stable. Each of the spatial interpolation methods was performed using its different powers, functions, and models to analyze their accuracy in terms of RMSE. The best model for a particular parameter showing least RMSE was used to make the spatial distribution map of that water quality parameter.

Water quality index

The model builder utility and spatial analyst extension in ArcGIS 10.3 software were used to computing the WQI. The WQI was based on seven parameters (pH, turbidity, chlorides, total dissolved solids, alkalinity, hardness, and calcium). These physicochemical parameters were used to calculate the relative weights for each parameter. Then the WQI was computed at all the seventy-three sampling points using the following formula:

$$WQI = Antilog\ [\textstyle\sum_{n=i}^{n} W\ log\ 10\ q_{\mathrm{ni}}] \tag{2}$$

Where:

Weightage factor (W) was calculated by the following equation,

$$W_n = \frac{K}{S_n} \tag{3}$$

and K, Proportionality constant was derived from,

$$K = \frac{1}{(\sum_{n=i}^{n} \frac{1}{S_i})} \tag{4}$$

Where:

S_n and S_i are the WHO standard values of the water quality parameter.

Quality rating (q) is calculated using the formula,

$$q_{\mathrm{ni}} = \frac{(V_{\mathrm{actual}} - V_{\mathrm{ideal}})}{(V_{\mathrm{standard}} - V_{\mathrm{ideal}})} * 100 \tag{5}$$

Where:

q_{ni} = Quality rating of i^{th} parameter for a total of n water quality parameters.

V_{actual} = Value of the water quality parameter obtained from laboratory analysis.

V_{ideal} = Value of that water quality parameter can be obtained from the standard tables.

V_{ideal} for pH = 7 and for other parameters it is equal to zero.

$V_{standard}$ = WHO standard of the water quality parameter.

The point values obtained as a result of computed WQI at each sampling point were interpolated using ordinary kriging to get the scenario for the whole study area. The surface thus generated was comprised of the derived value of WQI on each pixel. Hence, it was further classified as ‘excellent’ for values 0-25, ‘good’ for values 26-50, ‘poor’ for values 51-75, ‘very poor’ for values 76-100 and ‘unfit for drinking’ for values greater than 100, based on the criteria adopted by Shahid, et al. [19] and Asadi, et al. [28]. Instead of manual calculations, the major benefits of computing WQI in ArcGIS model builder are the reduction of errors in large computations and time efficiency. After validating the model, the WQI was calculated at a vigorous speed of about two minutes with computer specifications as follows:

- Processor: Intel (R) Core(TM) i7-4790 CPU @ 3.60 GHz 3.60 GHz
- Installed memory (RAM): 8GB,
- System Type: 64-bit Operating System

If the data scales up, then the computing time may increase depending upon the processing capabilities of the computer. The data requirements for WQI model are the values of water quality parameters at the sampling locations. These values can be from any field of the input feature dataset's attribute table having data types such as short, integer, float or double. The model immediately converts these point values to grid rasters, and the calculations are performed in raster format. The data types of inputs need to be corrected in order to validate and run the WQI model in the model builder. The limitations of this approach are that the same water quality parameters should be used again to calculate the WQI because the subtraction or addition of any water quality parameter will require subsequent changes (e.g. weightage factor, proportionality constant, constant rasters, spatial relationships, etc.) in the model. If different water quality parameters are to be evaluated, then the weightage factor and the proportionality constant would be manually re-calculated, and their constant rasters would be generated and adjusted in the model. Secondly, the field names of water quality parameters in the attribute table of geodatabase feature classes must be same as the field names provided in the WQI model.

The computation of WQI in ArcGIS model builder has robust spatial modelling capabilities. The spatial analysis flow diagram for WQI was designed in model builder. The complex WQI analysis comprising different spatial functions was performed, and a resulting WQI map was drawn that explained the outcomes of the investigation. The datasets used in the model and their weightage can be easily modified in future. The results obtained using WQI model can also serve as an input for another model. Once a model is designed, it can be run with a variety of parameters to identify data sensitivity or can be used to evaluate geographically different but structurally similar data sets.

Results and Discussion

Table 1: Descriptive statistics for physicochemical parameters

Parameter	Samples	Minimum	Maximum	Mean	Std. Deviation	Desirable Limit
pH	73	6.41	8.06	7.35	0.33	6.5 -8.5
Turbidity (NTU)	73	0.10	9.00	0.60	1.13	< 5 NTU
TDS (mg/L)	73	134.00	884.00	311.26	148.72	< 1000 (mg/L)
Hardness (mg/L)	73	33.33	523.33	150.23	82.70	< 500 (mg/L)
Calcium (mg/L)	73	12.00	112.00	40.49	17.82	< 200 (mg/L)
Chlorides (mg/L)	73	1.00	148.22	22.72	26.75	< 250 (mg/L)
Alkalinity (mg/L)	73	128.10	558.60	260.34	96.73	< 120 (mg/L)

The descriptive statistics (Table-1) for physicochemical parameters showed that pH, TDS, calcium, and chlorides values were well within the permissible limits. There was only one sample that showed turbidity beyond the threshold value of 5 Nephelometric turbidity units (NTU) so it was not acceptable. Similarly, the hardness value of one sample exceeded the 500 (mg/L) limit. The alkalinity values for all the samples were above 120 (mg/L)

Inverse Distance Weighting

The IDW uses power function to predict the surfaces. It assumes that the local variations have an important role in the phenomenon being modelled. Therefore, the number of closest neighboring samples affect the precision of IDW surface [29]. The greater the power used for IDW prediction, lesser the weightage of the farther points in prediction. The results showed that the optimal power using IDW for turbidity was 3.356, reflecting the fact that the farther points had lesser weightage in the interpolation process. The range for turbidity values was large, i.e. 0.1 to 9 NTU (Table-1) having a mean value of turbidity 0.6 NTU, yet the resulting surfaces (Figure-2) using IDW interpolation resulted in such a way that the area indicating turbidity values more than 5 NTU expanded as the power used for IDW surface increased. This clearly showed that the maximum value 9 NTU has a significant effect in the nearest areas due to the limited influence of farther points, using greater powers of IDW. On the other hand, if simply the IDW power 1 would have been selected in making predictions about turbidity than the area influenced by the maximum value of turbidity would have been smaller due to relatively more weightage of lower values of turbidity, even being farther. On the contrary, the variation between the pH values was very low, i.e. 6.41 to 8.06. Thus its optimal power also lied between 1 and 2, i.e., 1.232. Similarly, the optimal powers of other water quality parameters could be seen in Table-2 to understand the influence of values in predicting the estimates of their surroundings.

Radial Basis Functions

The RBFs are like a rubber sheet fitted to the sampled points. Figure-2b shows that the predicted area having turbidity levels more than 5 NTU varied with the RBF kernel used. The said area had an expanding trend with spline with tension, completely regularized spline and thin plate spline, respectively. The results were obtained using the optimal kernel parameter for each kernel. The thin plate spline is like fitting a rubber sheet to the sampled points with the formation of nice curves whereas the spline with tension is like pulling the fitted rubber sheet on the edges, hence lessening the curves. In the case of turbidity surfaces, the area showing values more than 5 NTU was almost equal for completely regularized spline kernel and spline with tension kernel and their RMSE, as described in Table-3, were also smaller than the RMSE of thin plate spline kernel. It might be inferences from the results as the sampling points had small distances in between and they belong to the same aquifer. Hence, there were very few fluctuations in the data. So, the spline with tension mostly produced smaller RMSE instead of curvy thin plate spline that showed highest RMSE for all the water quality parameters among RBF kernels.

Kriging

Instead of making predictions based on the inverse of the distance between the points as performed in the deterministic methods, geostatistical methods make predictions based on spatial autocorrelation among the data values. They assume that the data must be from a normal distribution. As the data of turbidity and pH was close to a normal distribution, it did not require the transformation, whereas the data of other parameters were not normally distributed, so the logarithmic transformation was applied to the data before making predictions. The semivariogram varies along different angles; the directional influences were also incorporated considering the anisotropy. It can be inferences from the results in Table-4 that no semi variogram model alone most accurately capture the spatial dependence of all the water quality parameters because of the fact that semi variogram models are merely mathematical models that are fitted to read the spatial autocorrelation for a particular parameter in the area of interest. Due to the substantial spatial variability of different water quality parameters in Lahore city, a single semi variogram model did not fit all water quality parameters equally good. The models showing lowest RMSE among all the kriging models for each water quality parameter are given in Table 4

Table 2: Inverse Distance Weighting powers and their root mean square error (RMSE)

Parameter	IDW (1)	IDW (2)	IDW (3)	IDW (4)	IDW (optimal)	
Turbidity (NTU)	1.1472	1.1309	1.1237	1.1246	(3.356)	1.1232
pH	0.3339	0.3353	0.3422	0.3528	(1.228)	0.3338
Alkalinity (mg/L)	75.5822	73.8884	73.1605	73.3204	(3.27)	73.1289
Calcium (mg/L)	16.4144	16.1575	16.1029	16.2500	(2.73)	16.0950
Chlorides (mg/L)	24.1155	23.9875	24.4682	25.3233	(1.70)	23.9582
Hardness (mg/L)	79.5177	78.1324	78.3067	79.6523	(2.37)	78.0225
TDS (mg/L)	132.9962	131.1209	131.1964	132.8053	(2.45)	130.918

Table 3: Radial Basis Function kernels and their RMSE

Parameter	Completely Regularized Spline	Spline with Tension	Thin Plate Spline
Turbidity (NTU)	1.110	1.110	1.227
pH	0.335	0.333	0.375
Alkalinity (mg/L)	74.136	73.841	90.679
Calcium (mg/L)	15.736	15.757	16.995
Chlorides (mg/L)	23.776	23.601	28.348
Hardness (mg/L)	76.935	76.935	89.765
TDS (mg/L)	129.752	129.552	153.311

Table 4: Details of kriging method with lowest RMSE

Parameter	Transformation applied	Anisotropy	Model	RMSE
Turbidity (NTU)	No	True	J-Bessel	0.9727
pH	No	True	Rational Quadratic	0.3220
Alkalinity (mg/L)	Log	True	J-Bessel	67.8567
Calcium (mg/L)	Log	True	Hole Effect	15.8498
Chlorides (mg/L)	Log	True	Rational Quadratic	22.2581
Hardness (mg/L)	Log	True	Exponential	75.5510
TDS (mg/L)	Log	True	Exponential	124.961

Table 5: Showing lowest RMSE obtained from best-fitted semi variogram model using co-kriging method for the estimation of each water quality parameter

		Auxiliary variable						
		Turbidity	**pH**	**Alkalinity**	**Calcium**	**Chlorides**	**Hardness**	**TDS**
Water quality parameter	**Turbidity** (NTU)		0.968 JB	0.831 JB	1.0374 PS	0.9541 JB	0.9872 PS	0.8136 JB
	pH	0.3365 SP		0.3386 SP	0.3072 EX	0.3364 GA & ST	0.3229 CR	0.3346 CR
	Alkalinity (mg/L)	64.1056 PS	70.8117 CR		71.6795 JB	55.8167 RQ	56.0207 JB	55.8905 EX
	Calcium (mg/L)	14.0301 HE	15.7428 RQ	16.8432 GA & ST		16.3083 RQ	12.865 JB	13.4551 HE
	Chlorides (mg/L)	20.5508 GA	22.3661 RQ	18.0544 RQ	22.2841 RQ		17.1778 RQ	10.2958 RQ
	Hardness (mg/L)	61.2895 CR	72.2095 CR	48.9595 PS	60.9037 ST	53.1465 RQ		39.8010 RQ
	TDS (mg/L)	105.9405 PS	116.071 JB	63.4487 KB	118.6024 CR	84.0342 RQ	86.4202 SP	

JB J-Bessel; *PS* Penta Spherical; *SP* Spherical; *EX* Exponential; *GA* Gaussian; *ST* Stable; *CR* Circular; *RQ* Rational Quadratic; *HE* Hole Effect; *KB* K-Bessel

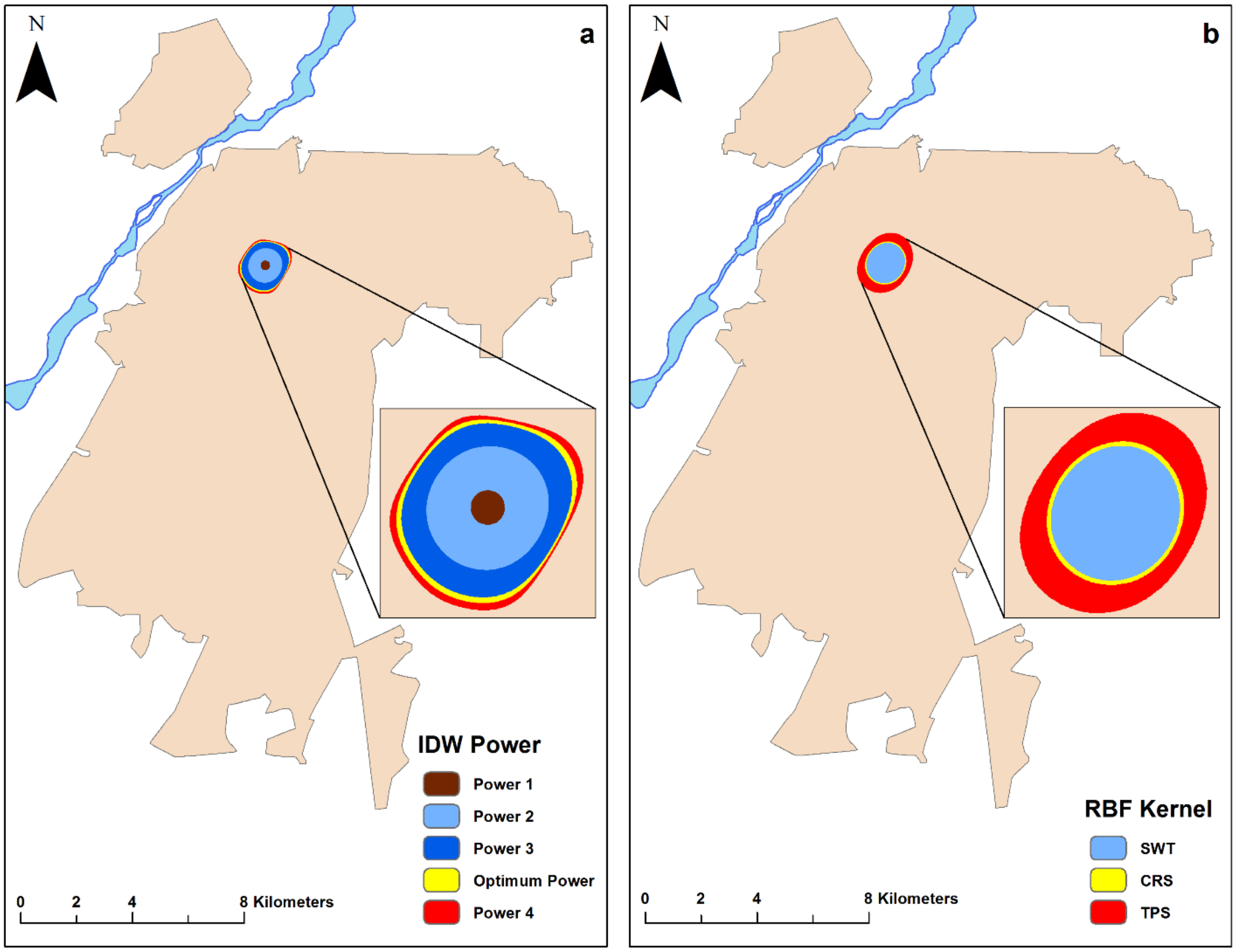

Fig. 2: Showing variation in the results (area having Turbidity >5NTU) obtained using different (a) power of IDW and (b) kernels of RBF

SWT Spline with Tension; *CRS* Completely Regularized Spline; *TPS* Thin Plate Spline

Co-kriging

The co-kriging method is like kriging model that has an additional characteristic of involving an auxiliary variable based on which the values of the target variable are predicted. Usually, the variable showing highest correlation with the target variable is selected as an auxiliary variable. Table-5 revealed that the auxiliary variables that showed lowest RMSE for the prediction of pH, turbidity, chlorides, total dissolved solids, alkalinity, hardness, and calcium were calcium, TDS, TDS, Alkalinity, chlorides, TDS and hardness, respectively. Similar to the kriging results, no semi variogram model alone presented best results using co-kriging interpolation for all the water quality parameters. The smallest RMSE for the prediction of pH, turbidity, chlorides, total dissolved solids, alkalinity, hardness, and calcium were 0.3072, 0.8136, 10.2958, 63.4487, 55.8167, 39.8010 and 12.865 using co-kriging models Exponential, J-Bessel, Rational Quadratic, K-Bessel, Rational Quadratic, Rational Quadratic, and J-Bessel, respectively. After examining the results described in Tables 2-5, it clearly indicated that the RMSE using co-kriging method were quite lower than the other three spatial interpolation methods used in this study. The reason for such a lower RMSE was the use of highly appropriate auxiliary variables. For instance, the RMSE for the prediction of chlorides using TDS as an auxiliary variable was much lower than using turbidity. It could be justified as chlorides were also a component of TDS concentrations in water. Similarly, the lowest RMSE for the prediction of calcium was obtained using hardness as an auxiliary variable. Shahid, et al. [19] and Khosravi, et al. [22] also compared different deterministic and geostatistical techniques and found co-kriging is the best method for modeling spatial distribution of groundwater quality.

Spatial distribution maps

The spatial distribution map of pH (Figure-3) indicated that the water is provided in the city is neither severely acidic nor extremely basic in nature. According to World Health Organization (WHO) guidelines, the pH of the water should be between 6.5 -8.5. If the water has a very low value of pH, it may be toxic, and if its value is very high, then it may have a bitter taste. Turbidity is mainly a result of suspended particles in water. Usually a variety of smaller particles e.g. decaying plants, clay, silt, etc. can be found in water which contributes to turbidity. The WHO standard for turbidity in drinking water is 5 NTU. The turbidity map indicated that only in the upper northern parts of the study area the turbidity values had crossed WHO standard for turbidity in drinking water, whereas, in rest of the areas it is within the desirable limits. The biological problems may arise in these areas as water turbidity is directly associated with the growth of pathogens. The chloride concentrations should be below 250 (mg/L) in drinking water. It is inferred from the chlorides map that there was no issue in the study area in terms of chlorides concentrations as it remained under 160 (mg/L) in the entire study area. The alkalinity map showed that most of the areas have alkalinity above 150 (mg/L). The southeastern parts of the study area had even higher values of alkalinity, but its concentration mostly below 500 (mg/L) was not a serious threat to the population, rather the aesthetic issues might arise due to higher alkalinity in those areas. Calcium is not only a significant component of human bones and teeth, but it also assists as a signal in important physiological processes. The calcium intake through drinking water can be important for people who are deficient in it [30]. The calcium intake is inversely correlated with blood pressure [31]. The calcium concentration map in Figure-4 revealed that there was no tube well in the study area having values even higher than 100 (mg/L). There is absolutely no issue regarding excessive calcium concentrations in the study area. As calcium is an important component of hardness in water, the hardness map showed that the areas are having higher values of hardness, e.g., in the central northern parts of the study area, also had relatively higher values in the calcium map. The reason for calcium and water hardness might be the presence of limestone in the alluvial deposits underlain the study area. People from different communities can have varying water hardness acceptability. Depending on the interactions, a hardness greater than 200 (mg/L) together with alkalinity and pH may be a cause of scale deposition in water tanks, distribution systems, treatment plants, etc. The weight of residue left after a water sample is evaporated to dryness is denoted by the TDS in water. According to WHO guidelines, water with TDS value less than 600 (mg/L) is generally acceptable to the people in terms of its taste. The TDS map showed that the TDS concentrations were highly variable in the study area. It might be due to the presence of different solubility materials in the aquifer. The lesser concentrations were near river Ravi, and they increased towards the east. There were a patch showing TDS concentrations higher than 500 (mg/L) in the central upper half of the study area, i.e., Anarkali subdivision. In order to calculate the WQI, the relative weights for seven physicochemical parameters were calculated using Eqs. (3) and (4). In these equations, the standardized values of S_n and S_i for pH, turbidity, chlorides, total dissolved solids, alkalinity, hardness and calcium were 8.5, 5 NTU, 250 (mg/L), 1000 (mg/L), 120 (mg/L), 500(mg/L) and 200 (mg/L) , respectively. Hence, the relative weights for pH, turbidity, chlorides, total dissolved solids, alkalinity, hardness, and calcium were 0.34808, 0.59175, 0.01183, 0.00296, 0.02466, 0.00592, and 0.01479, respectively. Equation (5) was used to compute the quality ratings for each parameter, and the final results were obtained by using equation (2). Although the range of WQI varies from 1.83 to 91.93 most of the samples, i.e., 66 out of 73, had shown WQI value less than 25 so they fall into the category of 'excellent' water quality. Similarly, 6 out of 73 samples were regarded as 'good' with WQI values ranging between 25 to 50 and only one sample having 91.93 WQI value fall into 'very poor' category. The main reason for this high value of WQI was a high value of turbidity, i.e., 9 NTU. The WASA installs a tube well only after clearance of water quality examination. As the water from surrounding tube wells does not have such a high turbidity level, it could be inferenced as this area is densely populated and the water extraction has increased significantly, the resulting water-table drawdown exerts pressure on the surrounding areas for more water intrusion. As a result, a solid material/stone with immense water pressure may have caused a rupture in 'fiber glass' screen of the tube well, which eventually increased the water turbidity. Overall the water quality index map (Figure-5) showed that the physicochemical water quality in Lahore city was acceptable. Some areas like Farrukhabad, Gulberg, City and Johar Town had good water quality. However, there were some patches in Anarkali area where the physicochemical quality of water was determined as poor to very poor. Chattergee et al. [32] applied the same WQI on surface water and shallow wells in coal mining area of Jharkhand, India. He also found the majority of the area showing physicochemical WQI excellent to good, but some areas were identified having poor to unfit for drinking water quality. In our study, all these tube wells are in the deep aquifer, so they are safe from the contamination caused by anthropogenic activities. Hence, the WQI for most of the areas is satisfactory. However, there might be issues regarding bacteriological water quality.

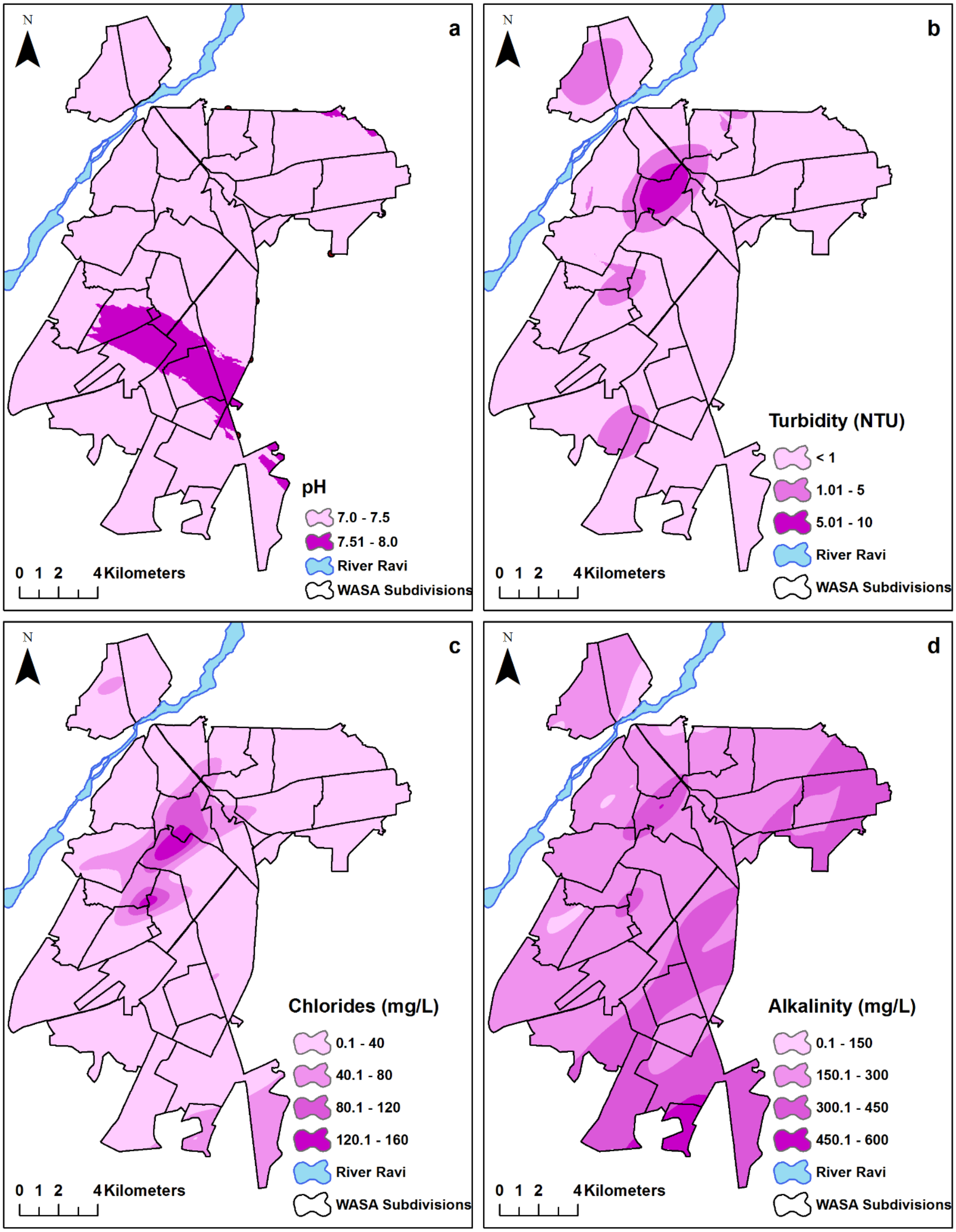

Fig. 3: Spatial distribution maps of (a) pH, (b) turbidity, (c) chlorides and (d) alkalinity in Lahore City.

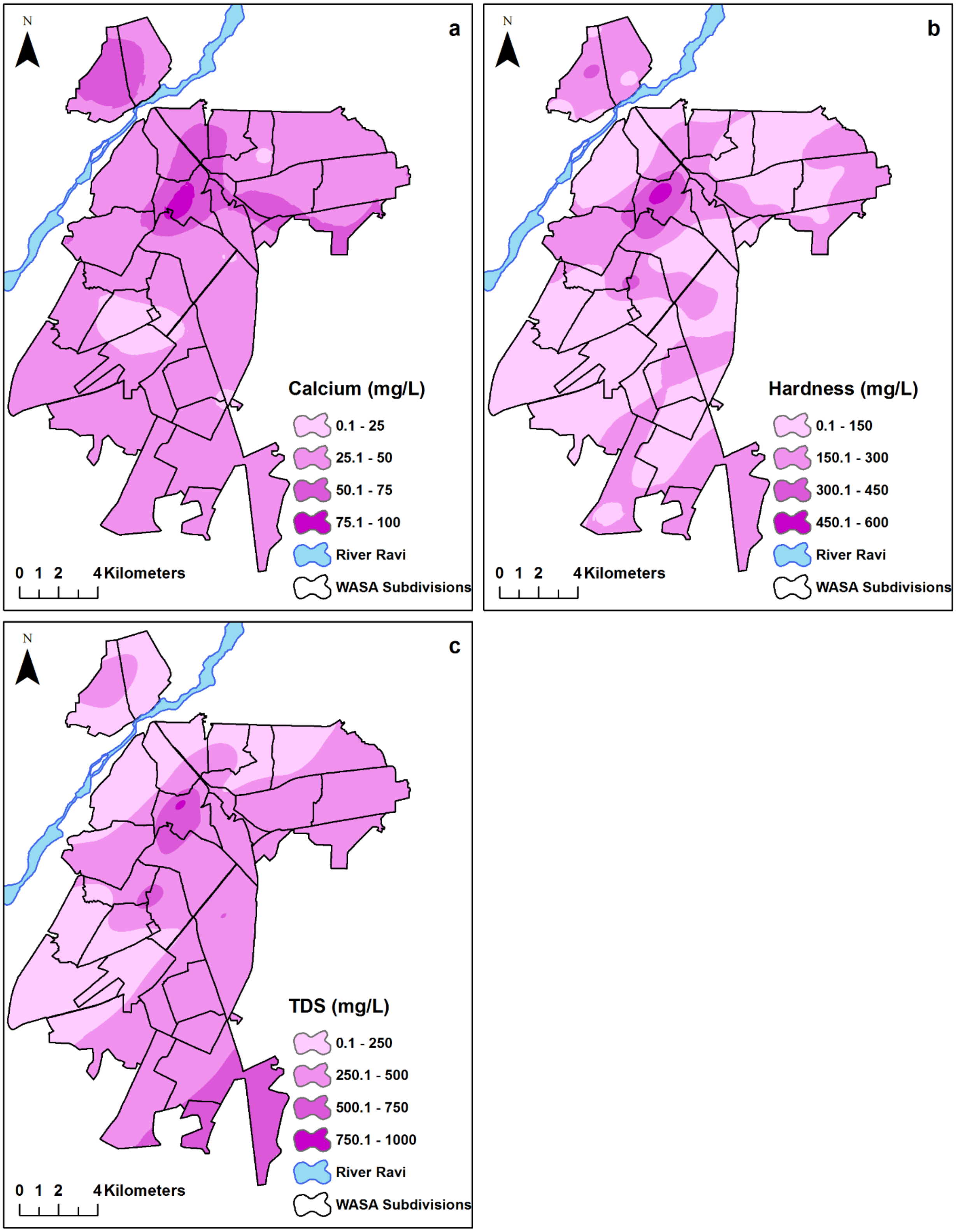

Fig. 4: Spatial distribution maps of (a) calcium, (b) hardness and (c) TDS in Lahore City

Water quality index

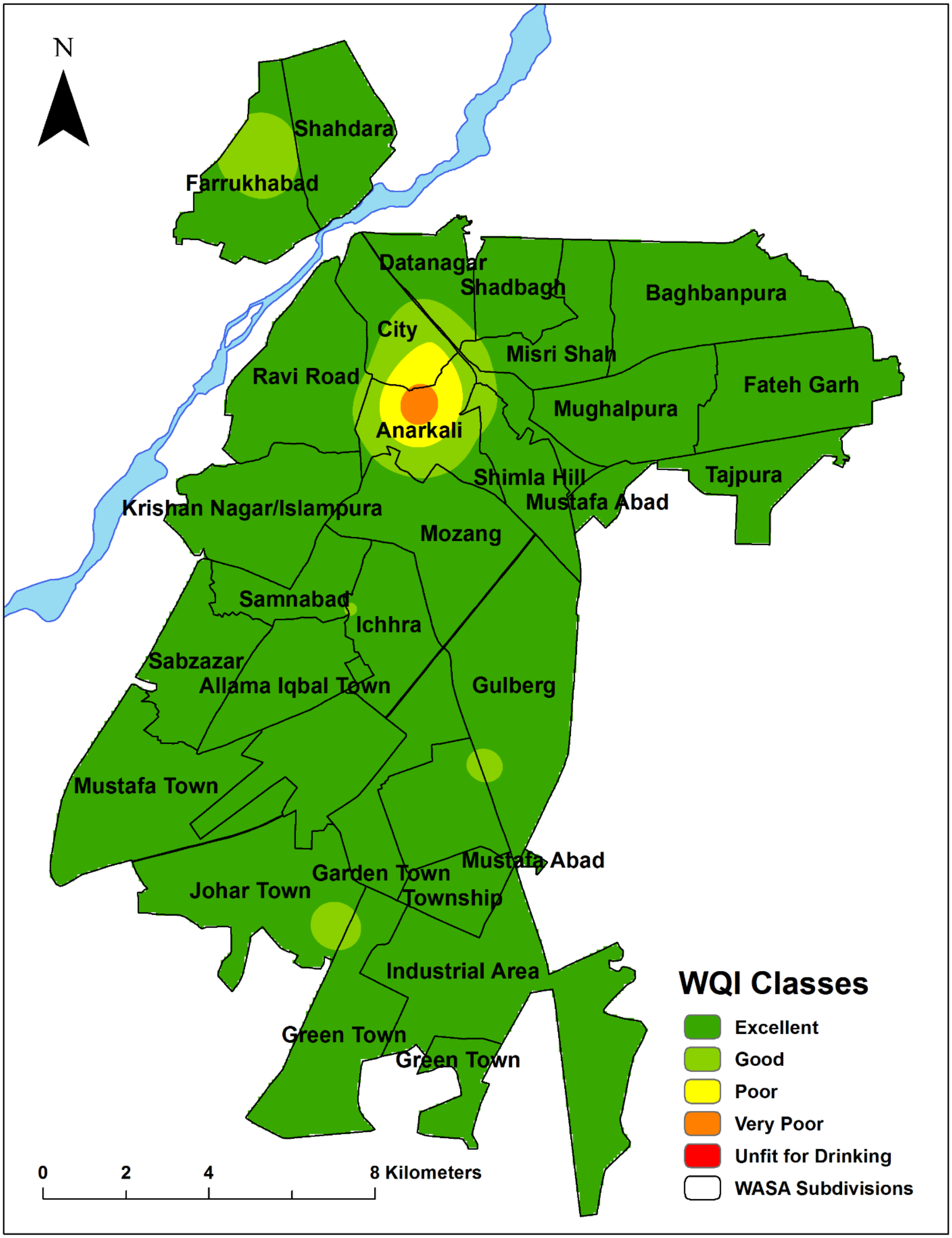

Fig. 5: Water quality index map of Lahore City

Conclusions

The spatial distribution maps and water quality index maps which nowadays play a key role in the water quality management are a product of GIS tools and spatial interpolation methods. The convenience of using readily available spatial interpolation methods pave a way to investigate more and more techniques to find the most suitable one for each water quality parameter so as to represent the true picture of existing water quality. The intercomparison of the IDW powers showed that the optimal power for variable increases as the spatial variation in the data increases. The less curvy spline with tension produced better results in the intercomparison of RBF kernels. As the data of water quality parameters did not have too many fluctuations, the RMSE values using RBF were generally lower than using IDW method. Hence, it indicates that the interpolation based on RBF is better among deterministic methods when we have minor variations in the data because it results in the smoother surfaces. However, the use of statistically strong geostatistical methods for spatial interpolation outperformed the deterministic methods in this study. The spatial distribution maps of each parameter were generated using different models of a co-kriging method that showed lowest RMSE so as to get more reliable predictions.

The WQI is an appropriate tool for analyzing the water quality of a large area at ease. The results of WQI indicated that the physicochemical water quality was mostly within the desired limits in Lahore. As this study analyzed the water samples from tube wells, it is highly recommended that the people instead of taking drinking water from house taps should get it directly from point-of-use water treatment systems or taps nearest to tube wells so as to avoid presence of harmful pathogens normally observed in the water distribution system due to leakage from sewage lines and old pipelines. As some of the water quality parameters had relatively higher concentrations in the Anarkali subdivision and nearby areas, the WASA authorities should take this issue seriously and set up filtration plants in the area. It is recommended that a further study with increased number of water samples in that area should be conducted to get detailed information about the spatial variability of physio-chemical parameters in that region. Moreover, the procedure adopted in this study to determine a reliable prevailing scenario about water quality is valuable for the water management authorities to better understand and monitor the groundwater quality and implement a revised water quality strategy in future.

Acknowledgements

The study was conducted using the student research fund from National University of Sciences and Technology (NUST), Islamabad, Pakistan. The authors are thankful to Water and Sanitation Agency, Lahore, for providing the help necessary in the field survey. The authors also acknowledge the management of Water and Wastewater Laboratory, Institute of Environmental Science and Engineering, School of Civil and Environmental Engineering, NUST, for facilitation in the water quality analysis.

References

1. A. Alamgir, et al., "Assessment of groundwater quality in the coastal area of Sindh Province, Pakistan", *Environmental Monitoring and Assessment,* vol. 188:78, 2016, pp.1-13.
2. R. Shabbir and S.S. Ahmad, "Use of geographic information system and water quality index to assess groundwater quality in Rawalpindi and Islamabad", *Arabian Journal for Science and Engineering*, vol. 40, 2015, pp. 2033-2047.
3. C. R. Ramakrishnaiah, C. Sadashivaiah and G. Ranganna, "Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India", *E-Journal of Chemistry,* vol. 6, no. 2, 2009, pp. 523-530.
4. A. Mahmood, et al., "Application of multivariate statistical techniques for the characterization of groundwater quality of Lahore, Gujranwala and Sialkot (Pakistan)", *Pakistan Journal of Analytical and Environmental Chemistry,* vol. 12, no. 1, 2011, pp.102-112.
5. A. D. Piazza, et al., "Comparative analysis of spatial interpolation methods in the Mediterranean Area: Application to temperature in Sicily", *Water*, vol. 7, 2015, pp. 1866-1888.
6. K. Johnston, et al., "Using ArcGIS Geostatistical Analyst", ESRI, Redlands, California, 2003.
7. P. Q. Giang, et al., "Investigating and mapping spatial patterns of arsenic contamination in groundwater using regression analysis and spline interpolation technique", Journal of *Water Supply: Research and Technology-AQUA*, vol. 62, no. 6, 2013, pp.385-394.
8. S.U. Shahid and J. Iqbal "Groundwater quality assessment using averaged water quality index: a case study of Lahore City, Punjab, Pakistan", *Proceedings of World Multidisciplinary Earth Sciences Symposium*, Prague, Czech Republic, September 5-9, 2016.
9. C. Childs, "Interpolating surfaces in ArcGIS spatial analyst", https://www.esri.com/news/arcuser/0704/files/interpolating.pdf , ArcUser July-September, 2004.
10. M. Aminu, et al., "A GIS-based water quality model for sustainable tourism planning of Bertam River in Cameron Highlands, Malaysia", *Environmental Earth Sciences*, vol. 73, 2015, pp. 6525-6537.
11. K. Rina, et al., "Characterization and evaluation of processes governing the groundwater quality in parts of the Sabarmati Basin, Gujarat using hydrochemistry integrated with GIS",

Hydrological Processes, vol. 26, 2012, pp. 1538-1551.

12. A. Kamińska and A. Grzywna, "Comparison of deterministic interpolation methods for the estimation of groundwater level", *Journal of Ecological Engineering*, vol. 15, no. 4, 2014, pp. 55–60.
13. I. Audu and A. Usman, "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*, vol.4, no. 5, 2015, pp. 373-388.
14. C. Yan, et al., "Assessment of water quality and identification of polluted risky regions based on field observations & GIS in the Honghe River watershed, China", *PLoS ONE*, vol. 10, no. 3, 2015, DOI:10.1371/journal.pone.0119130.
15. Z. Han, R. Zimmermann and S. Görtz, "A new cokriging method for variable-fidelity surrogate modeling of aerodynamic data", *Proceedings of 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, Orlando, Florida, January4-7, 2010.
16. M. Delbari, et al., "Investigating spatio-temporal variability of groundwater quality parameters using geostatistics and GIS", *International Research Journal of Applied and Basic Sciences*, vol. 4, no. 12, 2013, pp. 3623-3632.
17. A. Hooshmand, et al., "Application of kriging and cokriging in spatial estimation of groundwater quality parameters", *African Journal of Agricultural Research*, vol. 6, no. 14, 2011, pp. 3402-3408.
18. H. Arslan, and N. A. Turan, "Estimation of spatial distribution of heavy metals in groundwater using interpolation methods and multivariate statistical techniques; its suitability for drinking and irrigation purposes in the Middle Black Sea Region of Turkey", *Environmental Monitoring and Assessment*, vol. 187:516, 2015, DOI 10.1007/s10661-015-4725-x.
19. S.U. Shahid, J. Iqbal and G. Hasnain, "Groundwater quality assessment and its correlation with gastroenteritis using GIS: a case study of Rawal Town, Rawalpindi, Pakistan", *Environmental Monitoring and Assessment*, vol. 186, no. 11, 2014, pp. 7525-7537.
20. M. Samin, et al., "Spatial estimation of groundwater quality parameters based on water salinity data using kriging and cokriging methods", *Proceedings of International Conference on Transport, Environment and Civil Engineering*, Kuala Lumpur, Malaysia, August 25-26, 2012.
21. N. Mubarak, et al., "Spatial distribution of sulfate concentration in groundwater of South-Punjab, Pakistan", *Water Quality, Exposure and Health*, vol. 7, 2015, pp. 503-513.
22. H. Khosravi, et al., "Investigation of spatial structure of groundwater quality using geostatistical approach in Mehran Plain, Iran", *Pollution*, vol. 2, no. 1, 2016, pp. 57-65.
23. R. Taghizadeh-Mehrjardi, M. Zareiyan-Jahromi, and F. Asadzadeh, "Mapping the spatial variability of groundwater quality in Urmia (Iran): comparison of different interpolation methods", *Journal of International Environmental Application & Science*, vol. 8, no. 3, 2013, pp. 359-368.
24. A. Khashei-Siuki, and M. Sarbazi, "Evaluation of ANFIS, ANN, and geostatistical models to spatial distribution of groundwater quality (Case study: Mashhad plain in Iran)", *Arabian Journal of Geosciences*, vol. 8, 2015, pp. 903-912.
25. Z. Javed, et al., "Spatial distribution of arsenic concentration in drinking water using kriging techniques", *Science International (Lahore)*, vol. 27, no. 2, 2015, pp. 1089-1095.
26. M. Jalali, S. Karami and A. F. Marj, "Geostatistical evaluation of spatial variation related to groundwater quality database: Case study for Arak Plain Aquifer, Iran", *Environmental Modeling & Assessment*, 2016, DOI 10.1007/s10666-016-9506-6.
27. M. Yang, "Benchmarking rainfall interpolation over the Netherlands", M.S. Thesis, University of Twente. Enschede, Netherlands, 2015.
28. S. S. Asadi, P. Vuppala and M. A. Reddy, "Remote sensing and GIS techniques for evaluation of groundwater quality in Municipal Corporation of Hyderabad (zone-V), India", *International Journal of Environmental Research and Public Health*, vol. 4, no. 1, 2007, pp. 45-52.
29. X. Yao, et al., "Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment", *PLoS ONE*, vol. 8, no.1, 2013,
30. World Health Organization. "Guidelines for drinking-water quality", 4th ed. Geneva: WHO, 2011.
31. N. Munir, et al., "Physio-Chemical differences between Aab-E-ZamZam and mineral water effect of water intake on blood pressure", *American-Eurasian Journal of Toxicological Sciences*, vol. 7, no. 2, 2015, pp. 83-87.
32. R. Chatterjee, G. Tarafder and S. Paul, "Groundwater quality assessment of Dhanbad District, Jharkhand, India", *Bulletin of Engineering Geology and the Environment*, vol. 69, 2010, pp. 137-141.