Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.26087

198 (2020) 61–70
September

Geospatial analysis of groundwater quality in the major cities of Pakistan

Zahid Javed[a], Sohail Chand[a,*], Reda Amer[b]

[a]*College of Statistical and Actuarial Sciences, University of the Punjab, Lahore 54590, Pakistan, emails: sohail.stat@gmail.com (S. Chand), zahid.stat86@gmail.com (Z. Javed)*
[b]*Department of Earth and Space Sciences, Lamar University, Beaumont, TX 77710, USA, email: ramer1@tulane.edu (R. Amer)*

Received 29 July 2019; Accepted 10 May 2020

ABSTRACT

The aim of this study is to assess the spatial variability of groundwater quality parameters. Geographic information systems multivariate statistical techniques and nonparametric kriging methods were used to analyze the correlational structure and spatial pattern of groundwater quality parameters in the major cities of Pakistan. The hydro-chemical results of 366 water samples were taken from the Pakistan Council of Research in Water Resources (PCRWR) report 2015–2016 of 25 major cities of Pakistan. The correlation matrix was used to identify the highly correlated groundwater quality parameters. The principal component analysis and cluster analysis categorized the quality parameters according to the variation. The results indicated that seven water quality parameters including electric conductivity, calcium (Ca), magnesium (Mg), hardness, sodium (Na), sulfate (SO_4) and total dissolved solids were found exceeding the permissible limits of World Health Organization (WHO). Due to highly skewed data, nonparametric kriging methods were used to estimate the probability of concentration of groundwater quality parameters and to produce prediction maps. Cross-validation statistics demonstrated that the indicator kriging method showed better performance than ordinary kriging and co-kriging methods for mapping groundwater quality parameters. Overall water quality results showed that only 113 (31%) out of 366 water samples were suitable for drinking, whereas 253 (69%) were not safe drinking water. This study highlights the use of nonparametric kriging methods for non-normal data and concludes that solid waste and sewage systems should be developed to reduce the contamination of groundwater resources.

Keywords: Groundwater quality; Health effects; Geographic information systems, Geostatistical analysis; Pakistan

1. Introduction

Water is necessary for sustaining life and development on earth. Groundwater is an important source of freshwater and plays a vital role in sustaining industrial, agricultural and human activities [1]. The rapid increase in population, urbanization, and industrialization have increased groundwater contamination [2]. It has been observed that urban growth increased anthropogenic activities like wastewater discharge and nutrient excess which lead to impaired quality of water [3]. In developing countries, a huge part of the population has undergone health problems due to a shortage of drinking water or contaminated drinking water [4]. Pakistan has been observed as a water-stressed region and the potential of water-scarce in near future [5]. It was detected that 40% of all deaths and 30% of all diseases were due to unsafe drinking water in Pakistan [6]. The World Bank report revealed that Pakistan is ranked at 140 out of 180 countries on the environmental performance index and 64% of peoples in Pakistan have no access to safe

* Corresponding author.

1944-3994/1944-3986 © 2020 Desalination Publications. All rights reserved.

and clean drinking water [7]. Contaminants include pesticides and many other metals have been found in the drinking water of major cities of Pakistan [8–13].

The drinking water quality is assessed through the existence of eminent levels of physiochemical toxins and their health effects [14]. Environment Protection Agency (EPA) has examined that more than 200 organic compounds are present in drinking water [15]. Some of the coastal groundwater aquifers were found contaminated by seawater intrusion [16]. The World Health Organization (WHO) notifies that the hazardous power of hydrogen (PH) levels may cause many skin and eye infections. The arsenic-contaminated water causes cancer of lungs, bladder, and membrane. It can also be the cause of skin thickness and pigmentation [17]. Drinking water with a higher total dissolved solids (TDS) level causes gastric problems and bone diseases [18].

Geographic information systems (GIS) have been used for evaluating the spatial variability in water quality parameters [19]. Various nonparametric kriging techniques like ordinary kriging, Co-kriging, and Indicator kriging are being used widely for assessment and prediction of spatial variability of groundwater quality parameters [20,21]. Talaee [22] used ordinary kriging to estimate the spatial pattern of groundwater depth and quality parameters in the Ardabil plain in the northwest of Iran. Their results showed that the water quality parameters (chloride, pH sulfate, sodium, bicarbonate, calcium, magnesium, and hardness) were exceeding the permissible limits. In another study, Adhikary [23] used two nonparametric indicator kriging and probability kriging methods for assessment of Cu, Fe, and Mn concentration in drinking water in Delhi, India. They found that the study area is under risk for a higher concentrations of Cu, Fe, and Mn at 26.34%, 65.36%, and 99.55%, respectively. Using the indicator kriging and ordinary kriging, Delbari [24] reported that in Iran the groundwater quality parameters (electric conductivity, sodium, chlorine, and sodium absorption ratio) were crossing the threshold limits. The indicator kriging and ordinary kriging were used to evaluate the spatial distribution of groundwater quality parameters in Western and Southern Algeria. Their results indicated that alkalinity was slightly higher than the permissible limits [25]. Multivariate statistical techniques such as cluster analysis (CA) and principal component analysis (PCA) have been used for monitoring and predicting groundwater quality parameters in eastern Tunisia [26]. Their results showed that sulfate, sodium, chloride, and TDS were higher in the collected groundwater samples from saltwater intrusion [27]. Djemai [28] examined that the PCA allows to confirm the principal chemical facies and also favors to differentiate the aters of upper and middle Seaou River.

The main objective of this study was to identify the correlational structure of the considered groundwater quality parameters and to assess the spatial variability of the significant parameters in the major cities of Pakistan. GIS multivariate statistical techniques, CA and PCA were applied to identify the group of correlated parameters. Moreover, the geospatial nonparametric ordinary kriging, co-kriging, and indicator kriging techniques were used to develop the prediction maps for the significant water quality parameters.

2. Material and methods

2.1. Study area

Pakistan is located in southern Asia, having neighbors India in the east, Afghanistan in the northwest, China in northeast and Iran in the west (Fig. 1). Pakistan covers an area of 881,913 km^2 and the total population is 220 million according to the census of 2017.

The primary water resources in Pakistan are rainwater, groundwater, and rivers. The leading rivers are the Indus, Sutlej, Chenab, Beas, Jhelum, Sindh, and Kabul. The climate in the major parts of the region is dry; however humid condition prevails over a small area in the north. The average yearly rainfall in the most of area of Pakistan is below 250 mm and the rainiest area in Murree (Rawalpindi) with an average annual rain of around 1,484 mm. Pakistan has four seasons; a cool and dry winter from December to February; dry spring from March to May; summer/rainy season from June to September and autumn from October to November. Monsoon precipitation is the lifeline of Pakistan's water resources which falls in summer from July to September. Topographically, Pakistan can be divided into six main areas; the northern mountains, western mountains, Baluchistan plateau, Pothohar plateau and salt range, Indus River plain, and desert areas [29].

2.2. Groundwater sampling

A water quality monitoring survey was conducted by the Pakistan Council of Research in Water Resources (PCRWR) in 2015–2016 that included 25 major cities including Karachi, Lahore, Faisalabad, Rawalpindi, Gujranwala, Peshawar, Multan, Hyderabad, Islamabad, Quetta, Bahawalpur, Sargodha, Sialkot, Sukkur, Sheikhupura, Gujrat, Mardan, Kasur, Mingora, Muzaffarabad, Abbottabad, Badin, Loralai, Ziarat and Gilgit [30]. In this study, groundwater quality data of 366 locations (Fig. 1) from the 25 major cities of Pakistan were taken from the PCRWR survey (2015–2016). Samples were collected from selected wells, taps, hand pumps, streams, and water supply schemes. A minimum distance of 1 km and a maximum of 16 km was maintained between the two monitoring sites. Preferences were given to the permanent public places and the water samples were collected in clean sterile plastic bottles of 0.5 and 1.5 L capacities. Before collecting the samples bottles were washed properly. Boric acid and nitric acid were added as a preservative in the sampling bottles for trace elements. The samples were kept cool and in the dark while transporting to the laboratory. Further detail of the chemical analysis of the samples can be accessed by following the link provided in the supplementary material section. Groundwater samples were examined for eleven physiochemical parameters including electric conductivity, power of hydrogen, bicarbonate, calcium, magnesium, hardness, sodium, potassium, sulfate, TDS, and arsenic.

2.3. Multivariate statistics

Multivariate statistical tools are widely used in the analysis of groundwater quality parameters. The correlation coefficients r_{ij} between the *i*th and *j*th variables are as follows:

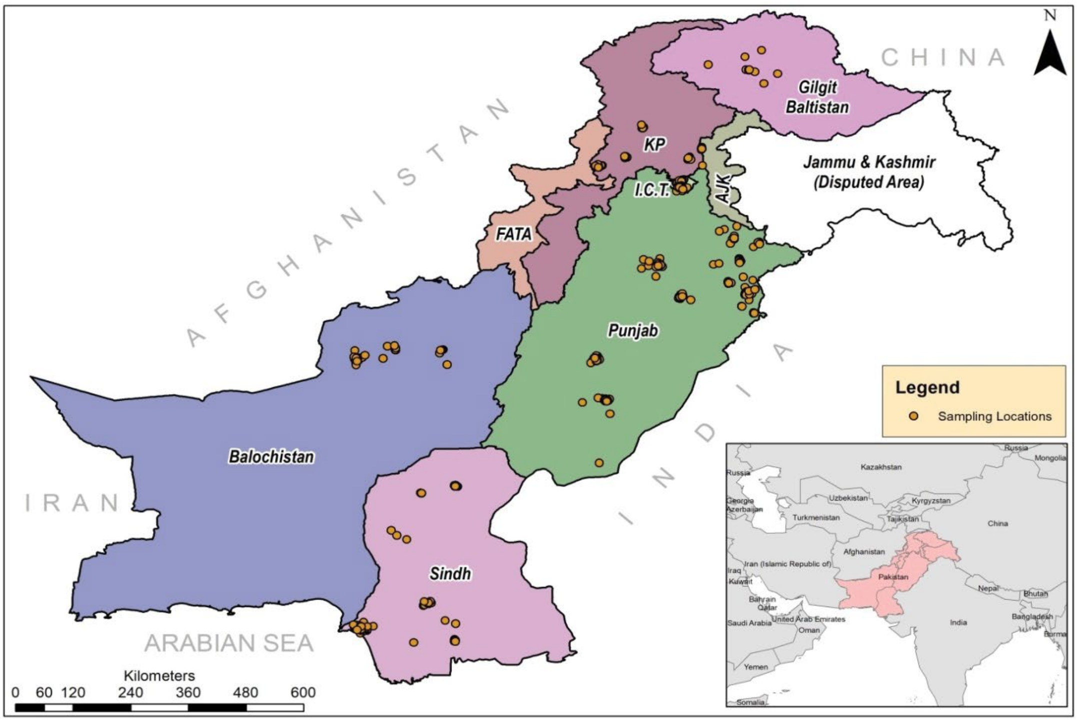

Fig. 1. The study site and sampling locations of major cities of Pakistan.

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \tag{1}$$

where S_{ij} is the covariance between ith and jth variables and S_{ii} and S_{jj} are variances.

The correlation matrix R is widely used for the identification of most correlated parameters. The correlation matrix for the sample data is described as:

$$R = \begin{bmatrix} 1 & r_{12}\cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & r_{p2}\cdots & 1 \end{bmatrix} \tag{2}$$

PCA is extensively used for dimensionality reduction. PCA transforms the correlated variables into uncorrelated orthogonal factors by using the linear transformation. The first component carries the most information regarding the original data [31]. PCA facilitates the identification of the most significant groundwater quality parameters and provides information about the acquired chemical properties [32]. Cluster analysis (CA) is mostly used for grouping a set of objects in such a way that the objects in the same group (cluster) are more similar while dissimilar to the objects in other groups or clusters [31].

2.4. Variogram models

Variogram models are fitted because the spatial prediction (kriging) requires the estimates of the variogram. The matern model is also called Whittle–Matern model after the name of Whittle. It is used to define the spatial covariance between two points.

$$\gamma(h) = \tau^2 + \sigma^2 \left(1 - \frac{(|h|)}{\varnothing}\right)^{v} k_v\left(\frac{|h|}{\varnothing}\right) \tag{3}$$

where $|h| > 0$ and τ^2, σ^2, v and $\varnothing \geq 0$ where K_v is Bassel functionality of order v. This particular variogram model is an intermediate option among Gaussian and exponential model [33].

The exponential method for spatial correlation is:

$$\gamma(h) = \tau^2 + \sigma^2 \left\{1 - \exp\left(-\frac{|h|}{\varnothing}\right)\right\} \tag{4}$$

for $|h| > 0$ and τ^2, σ^2, and $\varnothing \geq 0$ where, where $\tau^2 + \sigma^2$ is the sill and τ^2 are called the real nugget effect of the model [34].

The mathematical model of the spherical family is described as:

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 \left(\frac{3|h|}{2\varnothing} + \frac{|h|^3}{2\varnothing^3}\right) 0 < |h| \leq \alpha \\ \tau^2 + \sigma^2 |h| > 0 \end{cases} \tag{5}$$

for τ^2, σ^2 and $\varnothing \geq 0$. Spherical model gradually increases from the nugget effect τ^2 to sill quantity $\tau^2 + \sigma^2$ when the spatial lag quantity $h \geq \varnothing$ [34].

2.5. Kriging

Kriging techniques are frequently used for the prediction and estimation of spatial data. Ordinary kriging (OK) method assumes constant unknown mean over the search neighborhood. OK estimator is defined as following [35]:

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i), \quad \sum_{i=1}^{n} \lambda_i = 1 \tag{6}$$

where $Z(x_0)$ is the estimated value at ungauged locations x_0, n is the numbers of neighboring points and λ_i is the weight allocated to the measured value $Z(x_i)$.

Co-kriging is a multivariate extension of the kriging system when auxiliary variables can be used to improve the accuracy of the kriging estimate [36]. Co-kriging is a very flexible spatial interpolation technique that allows the user to examine the graphs of autocorrelation and cross-correlation. The general equation of co-kriging estimator is:

$$\gamma_{uv}(x_j, x) = \sum_{l=1}^{V} \sum_{i=1}^{n_j} \bar{\lambda}_{il} \gamma_{lv}(x_i, x_j) + u_v \tag{7}$$

for all $v = 1, 2, \ldots, V$ and $j = 1, 2, \ldots, n_j$. The quantity $\bar{\lambda}_{il}$ is the weight function and $\gamma_{lv}(x_i, x_j)$ is the cross-semivariance between variables l and v at sites x_i and x_j. Where u_v is the langrage multiplier for vth variable [35].

Indicator kriging (IK) uses a binary variable (indicator variable) to generate probabilities that a critical value was exceeded or not at each location in the study area. The indicator variable divides the original data into binary data according to threshold values. The indicator variable $I(x_i;z_k)$ for the given continuous variable $Z(x_i)$ is given as described in [36]:

$$I(x_i; z_k) = \begin{cases} 1 & \text{if } Z(x_i) \le z_k \\ 0 & \text{otherwise} \end{cases} \quad k = 1, 2, \ldots, K \tag{8}$$

where K is the number of thresholds and z_k is the desired threshold. The experimental semivariogram is defined for each set of indicators.

$$\gamma(h) \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[I(x_i; z_k) - I(x_i + h; z_k) \right]^2 \tag{9}$$

where h is the distance between x_i and $x_i + h$ and $N(h)$ is the number of pairs after binary transformation. $I(x_i;z_k)$ and $I(x_i + h;z_k)$ are the indicator variables separated by the vector h. The indicator kriging estimator [36] can be calculated as:

$$I^*(x_0; z_k) = \sum_{i=1}^{n} \bar{\bar{\lambda}}_i \cdot I(x_i; z_k) \tag{10}$$

where $\bar{\bar{\lambda}}$ is the weight coefficient.

2.6. Evaluation of kriging methods

Cross-validation was used to evaluate the predictive performance. Three non-parametric kriging methods were considered in this study and were compared for mean error (ME), mean square error (MSE), and root mean square error (RMSE). Cross-validation can be calculated as following [37]:

$$\text{ME} = \frac{1}{n} \sum_{i=1}^{n} \left[z(x_i) - z^*(x_i) \right] \tag{11}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left[z(x_i) - z^*(x_i) \right]^2 \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[z(x_i) - z^*(x_i) \right]^2} \tag{13}$$

where $z(x_i)$ is the observed value at location x_i, $z^*(x_i)$ is the predicted measure at the same location x_i and n is the number of pairs of measured and predicted values. The ME value near to zero is an indicator of a better model. A model with a minimum value of MSE and RMSE is considered the best-fit model among the others.

3. Results and discussion

3.1. Descriptive analysis

The hydrochemical results of the water quality parameters of electric conductivity (EC), bicarbonate (HCO_3), calcium (Ca), magnesium (Mg), hardness (HARD), sodium (Na), potassium (K), sulfate (SO_4), power of hydrogen (pH), TDS, and arsenic (As) were used for assessment of the groundwater quality in the major cities of Pakistan. R-Language and ArcGIS 10.7 software was used for the descriptive and geostatistical analysis of the measured water quality parameters. The descriptive statistics of water quality parameters along with their corresponding WHO permissible limits are listed in Table 1. The results show that the measured water quality parameters are higher than the WHO permissible limits. The EC ranges from 74.1 to 30,000 µS/cm whereas its permissible limit is ≤300 µS/cm. The measured water hardness at some sites reaches 3,400 mg/L; the WHO limit is ≤500 mg/L. The Na ranges from 0.70 to 3,820 mg/L and the mean value is 133.72 mg/L. The TDS ranges widely from 37 to 13,997.4 mg/L and the WHO permissible limit is ≤1,000 mg/L. As concentration reaches up to 106 µg/L at some locations. The WHO standard limit of As is (≤50 µg/L). The standard deviations of the water quality parameters show a large variation in the measured parameters. The values of skewness and kurtosis for the measure water quality parameters, except pH, show a non-normal distribution of the considered water quality parameters.

3.2. Principal component analysis

The PCA was used to identify the factors of the water quality parameters EC, pH, HCO_3, Ca, Mg, HARD, Na, K, SO_4, TDS, and As. The analysis generated three axes that collectively account for 83.96% of total variation (Table 2). The first and the most important axis accounts for 70.96% of total variation and associated with higher loading of seven

Table 1
Descriptive statistics of physicochemical parameters in groundwater of major cities of Pakistan

Parameters	Units	Permissible WHO limits	Minimum	Maximum	Mean	Standard deviation	Skewness	Kurtosis
EC	µS/cm	≤300	74.10	30,000.00	1,156.64	24,98.57	8.40	80.90
pH	mg/L	6.5–8.5	6.70	8.70	7.60	0.30	0.01	0.36
HCO_3	mg/L	Not specified	20.00	1,190.00	239.09	152.98	2.20	7.94
Ca	mg/L	≤250	7.50	480.00	60.59	46.05	4.43	30.11
Mg	mg/L	≤150	0.97	583.20	34.64	53.65	6.21	47.95
HARD	mg/L	≤500	0.00	3,400.00	276.87	321.45	5.86	45.10
Na	mg/L	≤200	0.70	3,820.00	133.72	365.03	7.52	65.43
K	mg/L	≤12	0.00	355.00	17.20	51.03	4.63	22.76
SO_4	mg/L	≤500	0.00	2,640.00	136.48	293.69	5.48	34.03
TDS	mg/L	≤1000	37.00	13,997.40	671.07	1,264.82	7.41	64.29
As	µg/L	50	0.00	106.40	10.05	20.45	2.72	7.18

Table 2
The pattern of rotated factors of eleven water quality parameters using varimax rotation along with variance

Variable	Factor 1	Factor 2	Factor 3
TDS	0.972	0.027	–0.154
EC	0.960	–0.069	–0.162
HARD	0.958	–0.034	–0.117
Mg	0.947	0.060	–0.044
Na	0.946	–0.104	–0.132
SO_4	0.934	0.569	–0.095
Ca	0.838	0.188	–0.200
HCO_3	0.683	–0.112	0.506
K	0.044	–0.083	0.199
pH	–0.351	–0.544	–0.266
As	–0.030	–0.056	–0.440
% Variance	70.960	12.010	10.990
Cumulative % variance	70.960	82.970	83.960

water quality parameters EC, Ca, Mg, HARD, Na, SO_4, and TDS. The higher loadings (>0.80) of the seven water quality parameters indicated that they are the major parameters controlling the water quality dataset. The second axis accounts for only 12.01% of total variation and has positive loadings for water quality parameters Ca, Mg, SO_4, and TDS. The third axis accounts for 10.99% of total variation and has a positive loading for HCO_3. Moreover, the bivariate correlations show that EC is highly correlated with Mg, HARD, Na, TDS ($r \geq 0.90$), and also has a higher correlation with Ca and SO_4 ($r \geq 0.80$). The Mg is correlated with HARD, EC, Na, SO_4, and TDS ($r \geq 0.85$). The Na is also correlated with TDS and SO_4. The HARD is correlated with Ca, SO_4, Na, and TDS (Fig. 2).

3.3. Cluster analysis

Cluster analysis was performed to evaluate the most significant parameters of the water quality dataset (Fig. 3). The eleven water quality parameters were clustered into four classes (clusters). The dendrogram demonstrated that the first cluster consists of EC, TDS, Na, Mg, HARD, HCO_3, SO_4, and Ca. The pH has a negative correlation with all other water quality parameters and was clustered in a separate class. The dendrogram showed that the K and As have lower correlation with the other parameters, and therefore they were clustered in a separate class. It is also noted that the results of Ca confirm the classification of PCA.

3.4. Geostatistical analysis

The CA indicated that the EC, Ca, Mg, HARD, Na, SO_4, and TDS are the most significant parameters which affect the groundwater quality in the study area. The best-fitted theoretical models for the semivariogram of the seven water quality parameters are listed in Table 3. The nugget, sill, and range values of best-fitted models were estimated by the restricted maximum likelihood (REML) and weighted least square (WLS) methods (Table 3). The results showed that the REML estimation technique and Matern model were best fitted for EC concentration ($R^2 = 0.92$). The Matern semivariogram model was best fitted for Mg, Na, and TDS concentration, whereas the spherical model was appropriate for Ca and SO_4. The best estimation for the water quality parameters EC, Ca, HARD, and SO_4 was given by the REML method. The best estimate for Mg, Na, and TDS was given by WLS. The exponential semivariogram model was the best for the hardness water quality parameter.

Cross-variogram was used to assess the cross-correlation structure of the water quality parameters. The semivariograms indicated that the spatial structure has the largest component of the nugget effect for most of the parameters (Fig. 4 and Table 3). It shows that the variogram models Na–SO_4; TDS-SO_4; EC-TDS; EC-Mg; Mg-HARD; TDS-HARD; EC-HARD; Na–Mg have correlation 0.89, 0.92, 0.990.91, 0.95, 0.93, 0.92, and 0.88, respectively. The lowest spatial cross-correlation was found between the pairs of Ca–SO_4, Ca–Na, Ca–Mg.

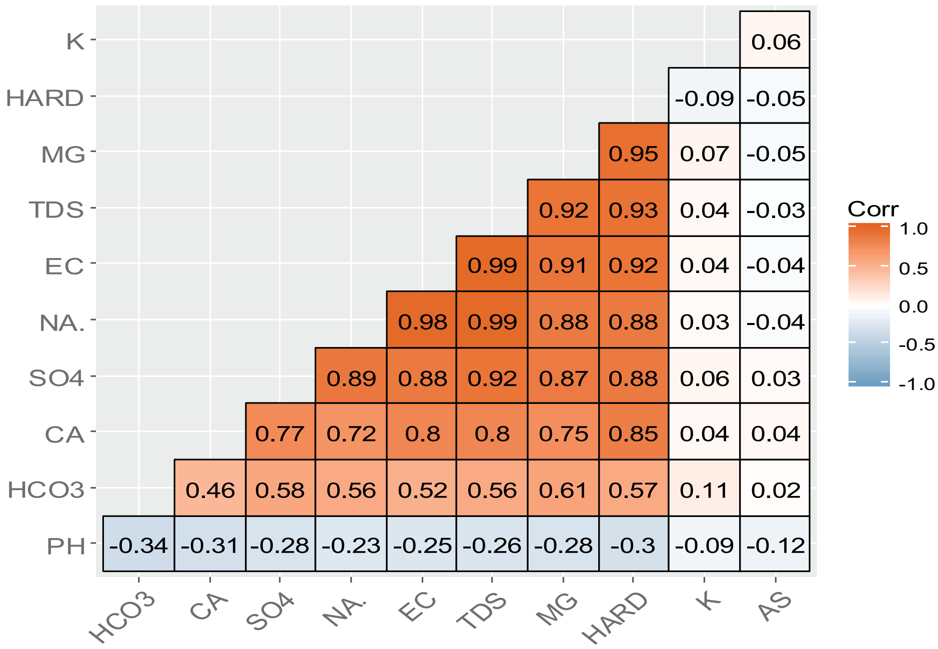

Fig. 2. Correlation matrix for water quality parameters.

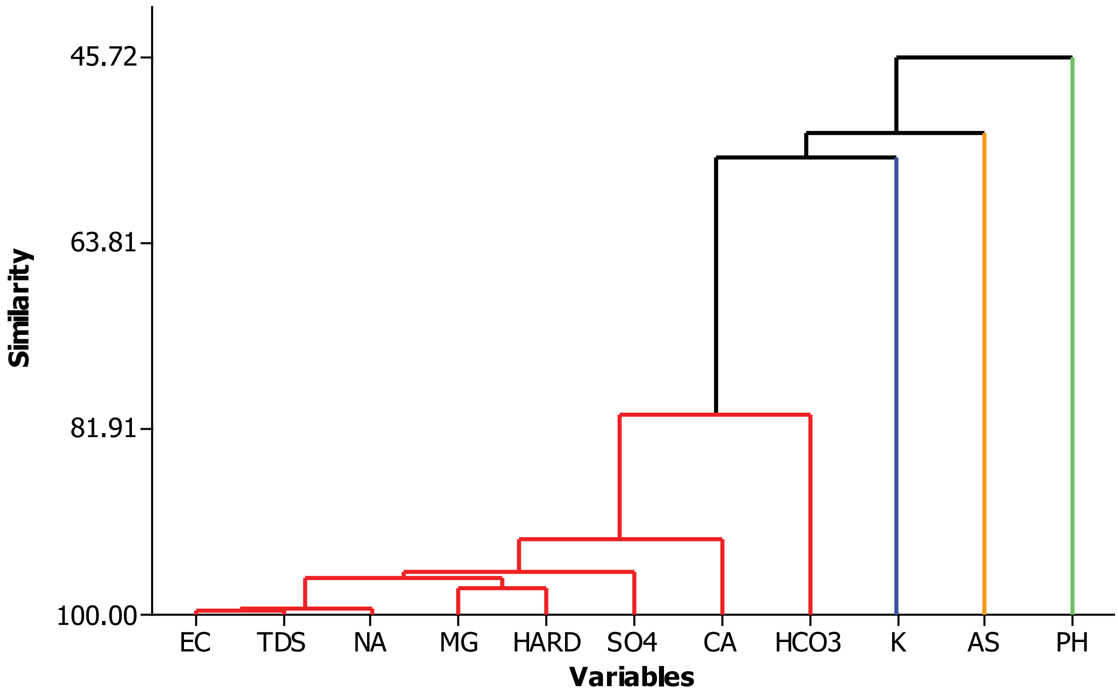

Fig. 3. Dendrogram for grouping the most similar and correlated factors.

3.5. Comparison of kriging methods

Cross-validation statistics were calculated for the water quality parameters to determine the predictive performance of theoretical models. The ME, MSE, and RMSE were calculated for the significant water quality parameters EC, Ca, Mg, HARD, Na, SO_4, and TDS (Table 4). It was observed that the ME of Indicator kriging method is smaller than that for ordinary kriging and co-kriging. Ideally, the ME should converge to zero. The MSE and RMSE values of the seven water quality parameters are very high due to higher variation in the water quality dataset. The statistical results show that the Indicator kriging method has the lowest values of ME, MSE, and RMSE for the considered water quality parameters except for the Ca. This indicates that the accuracy of indicator kriging is higher than the accuracy of

Table 3
Semivariogram parameters of the best fitted theoretical models of water quality parameters

Groundwater parameters	Best-fitted model	Estimation method	Sill (σ^2)	Range (ϕ) (meters)	Nugget (τ^2)	R^2
EC	Matern	REML	2.02×10^8	2.50×10^2	4.82×10^6	0.92
Ca	Spherical	REML	4.37×10^2	0.123	1.55×10^3	0.83
Mg	Matern	WLS	3.59×10^3	1.40×10^4	2.87×10^3	0.90
HARD	Exponential	REML	2.95×10^5	19.26	7.79×10^4	0.76
Na	Matern	WLS	3.80×10^5	3.86×10^4	1.33×10^5	0.87
SO_4	Spherical	REML	2.01×10^4	0.48	5.86×10^4	0.83
TDS	Matern	WLS	4.25×10^6	4.16	9.25×10^4	0.91

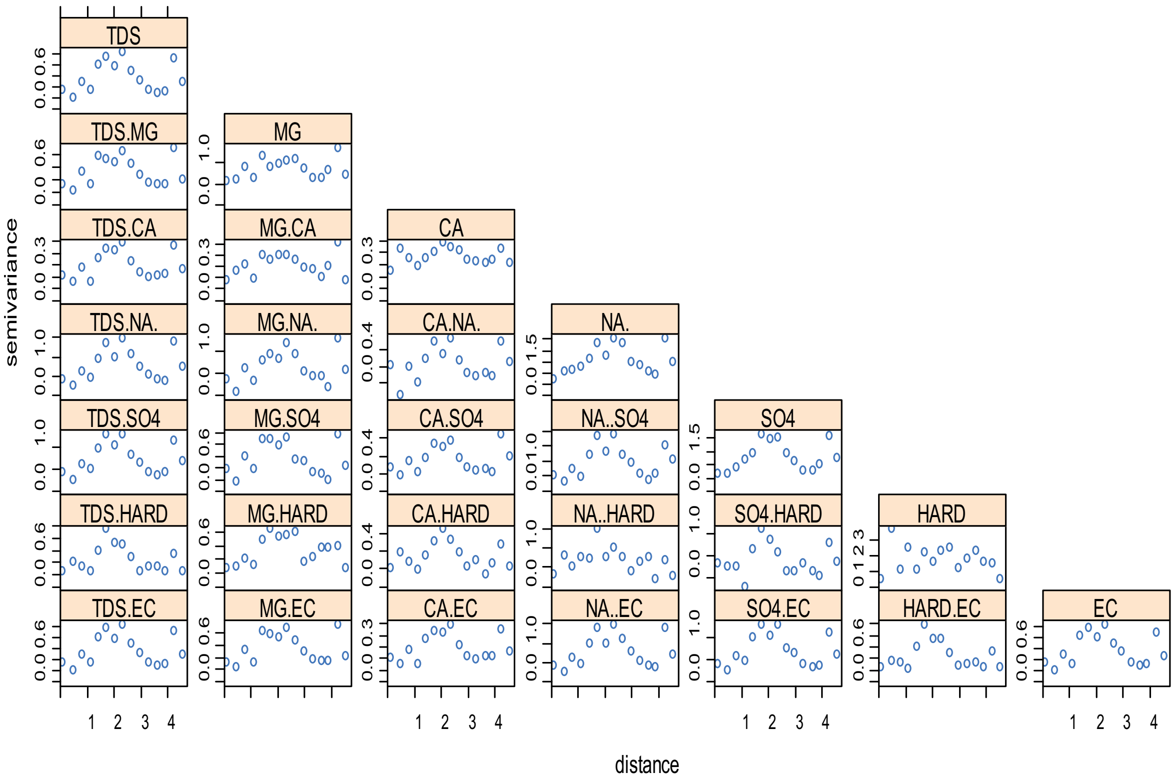

Fig. 4. Cross variography of water quality parameters to assess cross-correlation structure.

Table 4
Cross-validation statistics for kriging methods

Parameters	Ordinary kriging			Co-kriging			Indicator kriging		
	ME	MSE	RMSE	ME	MSE	RMSE	ME	MSE	RMSE
EC	5.34	233.57	15.28	3.76	215.78	14.69	1.81	128.98	11.13
Ca	3.24	115.23	10.73	2.33	75.09	8.66	3.31	91.78	9.58
Mg	0.98	18.57	4.31	0.45	6.35	2.52	0.30	3.56	1.25
HARD	2.09	49.76	7.05	0.78	23.25	4.82	0.02	8.45	3.61
Na	4.90	52.45	7.24	1.45	28.91	5.38	0.44	15.38	3.92
SO_4	8.35	56.37	7.51	4.09	23.84	4.88	1.06	5.72	2.39
TDS	0.78	112.05	10.58	0.05	40.46	6.36	0.01	13.23	3.08

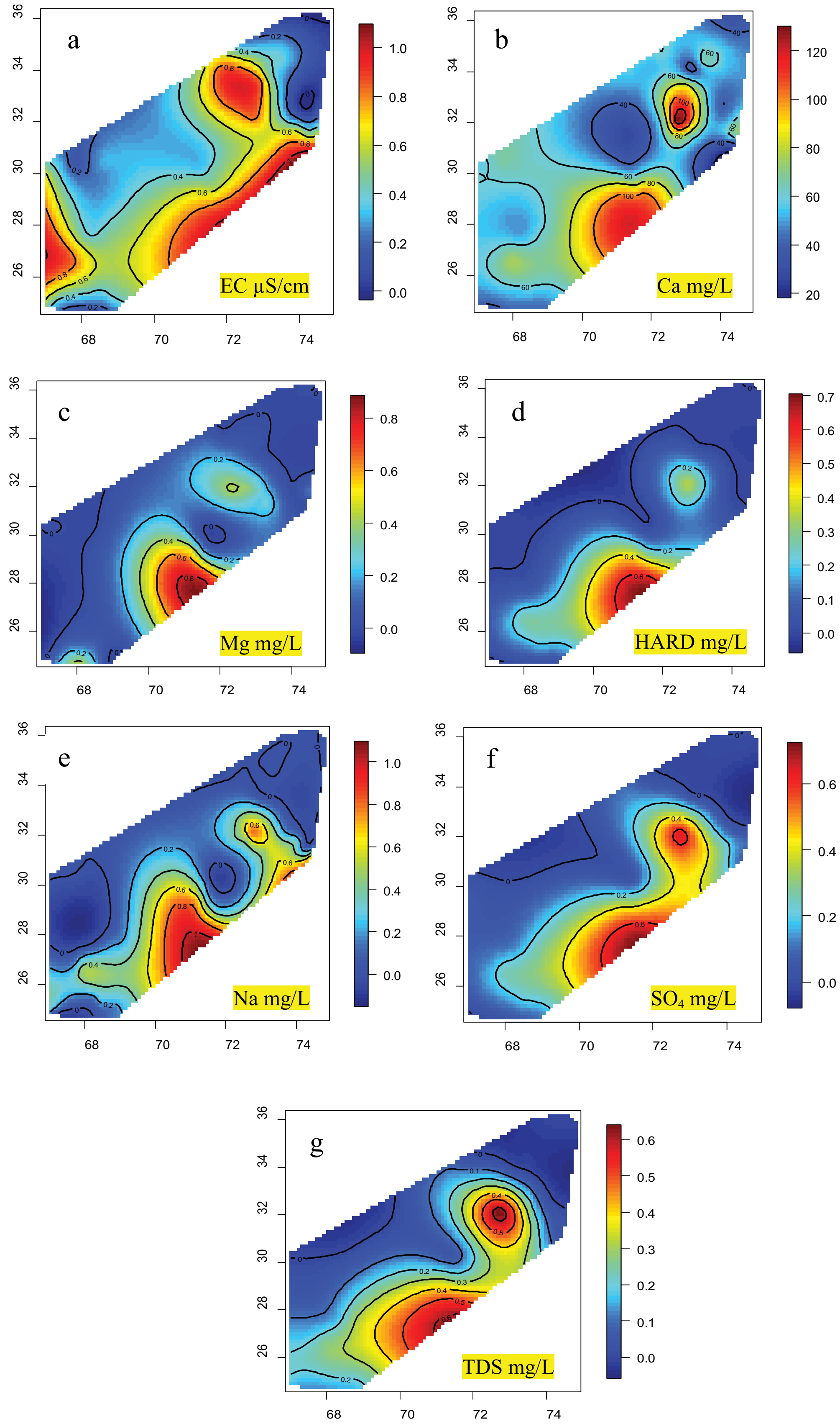

Fig. 5. Prediction maps of groundwater quality parameters in Pakistan: (a) electric conductivity EC, (b) calcium Ca, (c) magnesium Mg, (d) hardness, (e) sodium Na, (f) sulfate SO_4, and total dissolved solids TDS (g).

ordinary kriging and co-kriging. The co-kriging method has the lowest values of ME, MSE and RMSE, therefore it works well for the Ca parameter.

3.6. Spatial variability of water quality

The indicator kriging was used to produce prediction maps of water quality parameters (Fig. 5). The EC prediction map (Fig. 5a) shows that the values of EC concentration are higher towards the southern and southwestern parts of the study area. The higher values of EC indicate that the water salinity is higher than the permissible limits which make it unsuitable for drinking purposes. The prediction map of Ca shows that it exceeds the permissible limit in the southwestern part of the study area (Fig. 5b). The increased concentration of Ca in drinking water may cause cancer [38]. The prediction map of Mg and hardness show higher concentration that exceeds the WHO permissible limits in the southern part of the study area (Figs. 5c and d). The higher concentration of hardness, Ca, and Mg in drinking water impacts human health and may result in laxative effects and cancer [39]. Na concentration is higher in the southeast region of the study area (Fig. 5e). The SO_4 concentration map highlights two regions of higher values that exceed the WHO permissible limits (Fig. 5f). Dehydration is a common symptom of consuming water with a higher concentration of SO_4. The TDS concentration is higher in the eastern and southern regions of the study area as shown in Fig. 5g.

The general direction of groundwater flow in Pakistan is from east to west. The concentration maps of EC, Ca, Mg, HARD, Na, SO_4, and TDS groundwater quality parameters demonstrated that the quality of groundwater decreases from east to west which coincides with the general groundwater flow direction. Statistical analyses of the measured water quality parameters revealed that only 113 (31%) out of 366 water wells are suitable for drinking. 253 (69%) out of 366 water wells are unsuitable for drinking purposes. It is also indicated that 65% of water wells in Punjab, 47% in KPK, and 81% in Sindh and Baluchistan were unsafe for drinking purposes. Groundwater contamination most probably related to a rapid increase in population and industrialization. More research is needed to determine the point and non-point contamination sources.

4. Conclusion

This study investigated the spatial distribution of groundwater quality parameters in the major cities of Pakistan. The PCA and CA indicated that seven parameters including EC, Ca, Mg, HARD, Na, SO_4, and TDS of the measured water quality parameters were higher signaling to an alarming situation. The concentration of the seven water quality parameters was found exceeding the WHO permissible limits of safe drinking water. The non-parametric kriging techniques ordinary kriging, co-kriging, and Indicator kriging were used for mapping the spatial variability of groundwater quality parameters. Cross-validation statistics ME, MSE and RMSE demonstrated that the Indicator kriging is more suitable for mapping groundwater quality parameters than both ordinary kriging and co-kriging methods. Concentration maps of EC, Ca, Mg, HARD, Na, SO_4, and TDS indicated that the quality of drinking water decreases from east to west which coincides with the general direction of groundwater flow in Pakistan. Statistical analyses revealed that only 113 water wells (about 31%) out of 366 water wells are suitable for drinking and human consumption use. Yet, 253 (about 69%) of the tested water wells are unsafe for drinking. Deterioration of groundwater quality is most probably related to the rapid increase in population and industrialization. More research is required to determine the point and non-point contamination sources and to minimize the discharge of contaminants into the water resources. Simultaneously, more purification and infiltration processes should be applied to keep the water quality parameters within the WHO permissible limits of drinking water.

Supplementary information

The data used in this research, available at the Pakistan Council of Research in Water Resources (PCRWR) website http://www.pcrwr.gov.pk/publication.php?view_quality.

Acknowledgment

Authors are thankful to Ms. Saiqa Imran and Dr. Muhammad Ashraf, Chairman, Pakistan Council of Research in Water Resources (PCRWR) Islamabad, Pakistan for their help in accessing and understanding the data on water quality parameters.

We would like to thank the reviewers for thoughtful comments and constructive suggestions, which help to improve the quality of this manuscript.

References

[1] D. Long, X. Chen, B.R. Scanlon, Y. Wada, Y. Hong, V.P. Singh, Y.N. Chen, C.G. Wang, Z.Y. Han, W.T. Yang, Have GRACE satellites overestimated groundwater depletion in the Northwest India Aquifer?, Sci. Rep., 6 (2016) 243–258.

[2] P.C. Mishra, Some Aspects of the Quality of Water in and Around Rourkela, Thesis Doctor of Philosophy in Chemistry, National Institute of Technology, Rourkela, 2005.

[3] D. Carstens, R. Amer, Spatio-temporal analysis of urban changes and surface water quality, J. Hydrol., 569 (2019) 720–734.

[4] A.I.A.S.S. Andrade, T.Y. Stigter, The distribution of arsenic in shallow alluvial groundwater under agricultural land in central Portugal: insights from multivariate geostatistical modeling, Sci. Total Environ., 449 (2013) 37–51.

[5] R. Ullah, R.N. Malik, A. Qadir, Assessment of groundwater contamination in an industrial city, Sialkot, Pakistan, Afr. J. Environ. Sci. Technol., 12 (2009) 429–446.

[6] UNESCO, Country Report, Education Transforms Lives, United Nations Educational, Scientific and Cultural Organization, 2000. Available at: www.unesco.org/education/wef/countryreprts/pakistan/rapport-1.html.

[7] WB Report, 2019. Available at: https://www.pakistantoday.com.pk/2019/03/22/64-pakistanis-deprived-of-safe-drinking-water-says-wb-report.

[8] M. Ahmad, S. Chand, H.M. Rafique, Geostatistical cokriging and multivariate statistical methods to evaluate groundwater salinization in Faisalabad, Pakistan, Desal. Water Treat., 84 (2017) 93–101.

[9] A. Ali, S. Javed, S. Ullah, S.H. Fatima, F. Zaidi, M.S. Khan, Bayesian spatial analysis and prediction of groundwater contamination in Jhelum city (Pakistan), Environ. Earth Sci., 1 (2018) 77–87.

[10] A. Farooqi, H. Masuda, M. Kusakabe, M. Naseem, N. Firdous, Distribution of highly arsenic and fluoride contaminated groundwater from east Punjab, Pakistan, and the controlling role of anthropogenic pollutants in the natural hydrological cycle, Geochem. J., 41 (2007) 213–234.
[11] S. Javed, A. Ali, S. Ullah, Spatial assessment of water quality parameters in Jhelum city (Pakistan), Environ. Monit. Assess., 189 (2017) 119–135.
[12] S. Saeed, Z. Javed, S. Chand, N. Hashmi, M. Ahmad, Spatial distribution of arsenic concentration in drinking water using kriging techniques, Sci. Int., 27 (2015) 949–954.
[13] M.I. Tariq, S. Afzal, I. Hussain, Pesticides in shallow groundwater of bahawalnagar, Muzafargarh, D.G. Khan and Rajan Pur districts of Punjab, Pakistan, Environ. Int., 30 (2004) 471–479.
[14] R. Reza, G. Singh, Assessment of ground water quality status by using water quality index method in Orissa, India, World Appl. Sci. J., 9 (2010) 1392–1397.
[15] EPA, Wellhead Protection: A Guide for Small Communities, Office of Research and Development Office of Water, Washington, 1993.
[16] R. Ayadi, K. Zouari, H. Saibi, R. Trabelsi, H. Khanfir, Determination of the origins and recharge rates of the Sfax aquifer system (southeastern Tunisia) using isotope tracers, Environ. Earth Sci., 75 (2016) 1–21.
[17] I.A. Toor, S.N.A. Tahir, Study of arsenic concentration levels in Pakistani drinking water, Polish J. Environ. Stud., 18 (2009) 907–912.
[18] C.K. Jain, A. Bandyopadhyay, A. Bhadra, Assessment of ground water quality for drinking purpose, District Nainital, Uttarakhand, India, Environ. Monit. Assess., 166 (2010) 1–4.
[19] C.M. Liu, J.J. Yu, E. Kendy, Groundwater exploitation and its impact on the environment in the North China Plain, Water Int., 26 (2001) 265–272.
[20] K.-W. Juang, D.Y. Lee, Comparison of three nonparametric kriging methods for delineating heavy-metal contaminated soils, J. Environ. Qual., 29 (2000) 197–205.
[21] A.G. Journel, Nonparametric estimation of spatial distributions, J. Int. Assoc. Math. Geol., 15 (1983) 445–468.
[22] P. Hosseinzadeh Talaee, Analysis of groundwater quality in the northwest of Iran, Desal. Water Treat., 56 (2015) 2323–2334.
[23] P.P. Adhikary, C.J. Dash, R. Bej, H. Chandrasekharan, Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India, Environ. Monit. Assess., 176 (2011) 663–676.
[24] M. Delbari, M. Amiri, M.B. Motlagh, Assessing groundwater quality for irrigation using indicator kriging method, Appl. Water Sci., 6 (2016) 371–381.
[25] F. Bahri, H. Saibi, Characterization, classification, and determination of drinkability of some Algerian thermal waters, Arabian J. Geosci., 4 (2011) 207–219.
[26] I. Triki, N. Trabelsi, M. Zairi, H. Ben Dhia, Multivariate statistical and geostatistical techniques for assessing groundwater salinization in Sfax, a coastal region of eastern Tunisia, Desal. Water Treat., 52 (2014) 1980–1989.
[27] F. Bahri, H. Saibi, Evaluation of Groundwater from 24 Wells in Six Departments of Algeria, February, 2012.
[28] M. Djemai, H. Saibi, M. Mesbah, A. Robertson, Spatio-temporal evolution of the physico-chemical water characteristics of the Sebaou river valley (Great Kabylia, Algeria), J. Hydrol., 12 (2017) 33–49.
[29] https://www.climatestotravel.com/climate/pakistan
[30] M. Soomro, M. Khokhar, W. Hussain, M. Hussain, Drinking Water Quality Challenges in Pakistan, Pakistan Council of Research in Water Resources, Lahore, 2011, pp. 17–28.
[31] C. Chatfield, A.J. Collins, Principal Component Analysis, An Introduction to Multivariate Analysis, Springer, 1980, pp. 57–81.
[32] H.S.M. Mesbah, A.S.M.A.H. Guendouz, Principal component, chemical, bacteriological, and isotopic analyses of Oued-Souf groundwaters (revised), Environ. Earth Sci., 75 (2016) 1–17.
[33] A.E. Gelfand, P.J. Diggle, M. Fuentes, P. Guttorp, Handbook of Spatial Statistics, Chapman & Hall/CRC, Boca Raton, 2010.
[34] N.D. Le, J.V. Zidek, Interpolation with uncertain spatial covariances: a Bayesian alternative to Kriging, J. Multivar. Anal., 43 (2006) 351–374.
[35] P.J. Diggle, P.J. Ribeiro, Geostatistics, Springer Series in Statistics, Springer, 2007.
[36] P. Goovaerts, Geostatistics for Natural Resources Evaluation, Oxford University Press on Demand, 1997.
[37] R. Srivastava, E. Isaaks, An Introduction to Applied Geostatistics, Oxford University Press, New York, 1989.
[38] K.D. Cashman, Calcium intake, calcium bioavailability and bone health, Br. J. Nutr., 87 (2002) 169–177.
[39] C. Coudray, J. Bellanger, C. Castiglia-Delavaud, C. Rémésy, M. Vermorel, Y. Rayssignuier, Effect of soluble or partly soluble dietary fibres supplementation on absorption and balance of calcium, magnesium, iron and zinc in healthy young men, Eur. J. Clin. Nutr., 51 (1997) 375–380.