GIS-based spatio-temporal and geostatistical analysis of groundwater parameters of Lahore region Pakistan and their source characterization

Sadia Ismail1 · M. Farooq Ahmed1

Received: 16 April 2021 / Accepted: 6 October 2021 / Published online: 21 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
Assessment of groundwater quality is critical, especially in the areas where it is continuously deteriorating due to unplanned industrial growth. This study utilizes GIS-based spatio-temporal and geostatistical tools to characterize the groundwater quality parameters of Lahore region. For this purpose, a large data set of the groundwater quality parameters (for a period of 2005–2016) was obtained from the deep unconfined aquifers. GIS-based water quality index (WQI) and entropy water quality index (EWQI) models were prepared using 15 water quality parameters pH (power of hydrogen), TDS (Total dissolve solids), EC (Electrical conductivity), TH (Total hardness), Ca2+ (Calcium), Mg2+ (Magnesium), Na+ (Sodium), K+ (Potassium), Cl− (Chloride), As (Arsenic), F (Fluoride), Fe (Iron), HCO3− (Bicarbonate), NO3− (Nitrate), and SO42− (Sulfate). The data analysis exhibits that 12% of the groundwater samples fell within the category of poor quality that helped to identify the permanent epicenters of deteriorating water quality index in the study area. As per the entropy theory, Fe, NO3−, K, F, SO42− and As, are the major physicochemical parameters that influence groundwater quality. The spatio-temporal analysis of the large data set revealed an extreme behavior in pH values along the Hudiara drain, and overall high arsenic concentration levels in most of the study area. The geochemical analysis shows that the groundwater chemistry is strongly influence by subsurface soil water interaction. The research highlights the significance of using GIS-based spatio-temporal and geostatistical tools to analyze the large data sets of physicochemical parameters at regional level for the detailed source characterization studies.

Keywords Groundwater · EWQI · GIS · Spatio-temporal · Geostatistics

Introduction
Water is an essential component of the hydrologic cycle, which travels in many different forms like evaporation, evapotranspiration, precipitation, and finally, as runoff on the earth’s surface. This valuable natural resource is essential for life and the environment (Fetter 2013). Groundwater is the primary resource everywhere globally for irrigation, domestic and drinking purposes, and its quality preservation is the need of the day. Overexploitation and excessive groundwater withdrawal result in a gradual decrease in water table levels. The erratic supply of water has created water shortages all over the world.

Groundwater contamination is a critical problem almost everywhere in the world (Hudak 2000; Sadat-Noori et al. 2014; Tiwari et al. 2017). Its quality is adversely affected by the number of anthropogenic activities like demographic migration, urbanization, industrialization, over pumping and excessive use of fertilizers and in the same way by geogenic activities, amount and quality of atmospheric precipitation, underground hydrogeochemical processes, and origin and quality of recharge water (Vasanthavigar et al. 2010; Tiwari et al. 2018). Appropriate methodologies should be adopted to eliminate water scarcity and assess the new water potential zones (Ketata et al. 2012; Pazand 2016). At present, aquifer water is the primary source of potable water in Pakistan’s major cities. Most of the groundwater reserves got contaminated because of the disposing of untreated municipal, commercial, residential, and industrial waste directly
into the nearby streams, channels, drains, ponds, rivers, and open fields or agricultural lands (Azizullah et al. 2011; Qureshi and Sayed 2014). This untreated industrial effluent waste contains many metals and metal ions contaminating the groundwater and soil, hazardous for humans and other living things (Qureshi and Sayed 2014). Just 20% of Pakistan’s population uses safe drinking water, and the rest is compelled to utilize untreated contaminated water. Sewage contamination in freshwater sources must be considered a critical ecological and medical concern (Azizullah et al. 2011).

At present, several techniques have been applied to groundwater characterization and quality assessment. Fuzzy mathematical method (Kamrani et al. 2016), Cluster analysis (Hosseininmarandi et al. 2014), set of pair analysis techniques (Pei-Yue et al. 2011), and blind number approach (Yan and Zou 2014) are several of them. However, these techniques cannot be easily applied at a regional scale by incorporating many factors for groundwater characterization. Modern tools and more innovative ways are essential to study spatial and temporal studies on a regional level to understand and identify the natural and anthropogenic contamination sources. Geographic Information Systems (GIS) have gained significant importance in different earth sciences, including geological sciences and engineering, geomorphology, environment, groundwater and surface water contamination, waste disposal and landfill site selection studies, climate changes, and environmental sciences. Most researchers concluded that GIS is a useful and effective tool for organizing physico-chemical databases. It has a broad application in detection, monitoring, assessment, and modeling the spatio-temporal variation for the surface and subsurface water resource evaluation and characterization (Strassberg 2005; Faunt and Ed Douglas 2015; Pazand 2016; Islam et al. 2017). A number of studies are successfully performed to model the water quality in a GIS environment by considering the spatial distribution of physicochemical parameters, using GIS-based inverse distance weighted (IDW) method (Tiwari et al. 2017), water quality index (Charizopoulos et al. 2018), entropy water quality index (Gorgij et al. 2017), and advanced statistical tools (Sadat-Noori et al. 2014; Pazand 2016; Chaurasia et al. 2018).

The regional-level groundwater characterization studies (i.e., with a large amount of data set) must be conducted by applying more robust, quick, and reliable spatio-temporal analysis tools. The GIS-based water quality index method is widely utilized to assess the quality parameters of surface and groundwater resources (Al-hadithi 2012; Ketata et al. 2012; Sadat-Noori et al. 2014; Shabbir and Ahmad 2015). The water quality index (WQI) is a rating approach that assists in studying the overall quality of water by considering individual physicochemical parameters. Initially, Horton introduced the Water quality index (WQI) concept that consists of water quality parameters. Later, it was updated by Brown and Deininger for the Scottish Development Department (Sadat-Noori et al. 2014). The most crucial components of healthy water parameters that should be incorporated in preparing water quality index maps include: pH, TDS, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, and SO$_4^{2-}$ (Saeedi et al. 2010). Numerous researchers found this index useful to evaluate the groundwater quality based on different physico-chemical parameters (Ketata et al. 2012; Sadat-Noori et al. 2014; Islam et al. 2017; Tiwari et al. 2018).

The newly developed entropy water quality index (EWQI) is a more efficient, innovative, and practical tool to assess the suitability of groundwater quality parameters and source characterization (Pei-Yue et al. 2010; Amiri et al. 2014; Islam et al. 2017). First, in this method, the weights are assigned to quality parameters based on expert knowledge and secondly, the entropy weights are calculated considering the quality of groundwater by employing the entropy-based water quality index (EWQI). This method further helps to reduce any bias related to personal judgment about the assignment of initial weights. In addition, EWQI is more efficient, as other tools have some limitations in considering the environmental impacts of the study area which are also important for the precise evaluation of groundwater quality (Amiri et al. 2014).

Multivariate statistical analysis is a widely used approach among the several other methods to identify the potential contamination sources and possible factors affecting water quality (Wishrat 1928; Anderson 1958; El Alfy 2010; Okiongbo and Douglas 2015; Pazand 2016; Islam et al. 2017). It is a useful tool to evaluate the complex water quality parameters and identify the main anthropogenic activities and potential contamination sources affecting aquifer resources. Physicochemical parameters having a significant correlation coefficient, indicating a similar source of origin, which could be geogenic (weathering and disintegration of soil-forming minerals), or anthropogenic (Okiongbo and Douglas 2015; Pazand 2016). The chemical composition of the groundwater can be assessed by the Piper plot and Gibbs diagram, which are used to understand the degree of disintegration of rocks and the mechanism of processes in the groundwater aquifer systems (Al-Ahmadi 2013; Pazand 2016; Selvam et al. 2016; Tiwari et al. 2017).

For the reliability of larger data sets, the impact of sample density on spatial variability of physicochemical parameters can also be an important aspect to consider especially when it changes over time. Many researchers studied the effect of sample density on the accuracy of interpolation surfaces (Zhang et al. 2015, 2018; Long et al. 2018). The acquisition and accuracy of the spatio-temporal pattern are mainly influenced by sample size and its prediction error which can be reduced by improving the sample density (Zhang et al. 2018). The spatial autocorrelation index method is used to
examine the relationship between the sample density and the accuracy of prediction surfaces. In this research, the sample density issue is also considered to check its likely impact on the resultant variation in the spatio-temporal patterns of physicochemical parameters generated by the IDW method (Bhunia et al. 2018; Long et al. 2018).

The current study aims to provide a holistic view of the groundwater source characterization using different modern tools, including GIS-based geostatistical techniques, WQI, EWQI, and multivariate statistics. This study will help to identify the major contributing factors (either anthropogenic or geogenic) to groundwater contamination at the regional scale. Previously for this region, a few scattered studies were performed by incorporating small data sets to identify and evaluate the surface contamination zones using conventional statistical methods and simple GIS techniques, including WQI buffering, and Kriging on a limited scale. This study deals with the considerable amount of physicochemical parameters data acquired in several years (2005–2016) of hundreds of water wells with different depths (mostly deep wells). This research evaluates and identifies potential groundwater contamination sources and prepares a groundwater quality map based on the water quality index and entropy water quality index. The study outcome could help in understanding the spatio-temporal deviations in groundwater quality and understand the local authorities' primary groundwater contamination sources and their characterization.

**Study area**

Lahore city is the capital of Punjab Province of Pakistan, covering a total area of 1772 km². It has a flat surface with a variation of altitude from 208 to 213 m above MSL. Lahore’s population has recently crossed the mark of 10 million and is considered one of the world’s most populated areas (Qureshi and Sayed 2014). The study area’s climate can be classified as semi-arid with long and intensely hot summers. The mean annual rainfall of Lahore is 680 mm, with heavy downpouring that usually occurs during the monsoon season, providing 40% of the total groundwater recharge (Mahmood et al. 2016). The Lahore’s unconfined aquifer is composed of unconsolidated alluvial deposits up to 400 m thickness with a hydraulic transmission rate of about 2100 m²/day and alternate layers of sand, silt, and clay formations (Farooqui et al. 2007; Qureshi and Sayed 2014).

Geologically the Lahore region is resting over Bari Doab, which is the portion of Indo-Gangetic Quaternary alluvial plains, deposited by the Indus River and its tributaries. These alluvium deposits overlie the unconsolidated Tertiary soil and igneous rocks of the Precambrian age (Kadwai and Siraj 1964). From the late Pleistocene to the recent era, the deposition of sand, silt, and clay particles was taken place which indicates the last glacial cycle and Chung Fun formation. During the Pleistocene–Recent age, the sediments complex has a thickness of more than 1300 feet (400 m). The alluvial sedimentary deposits formed a large aquifer, which on a regional scale behaves as an unconfined homogeneous aquifer where the groundwater occurs under water table conditions (see Fig. 1) (Greenman et al. 1967; Ahmad et al. 2002).

The groundwater is the significant source for drinking, domestic and industrial usage in the area under study. 82% of the groundwater aquifer of Lahore is recharged by the River Ravi (the primary source), 12% from monsoon rainfall, and 6% from the return flows from irrigation (Qureshi and Sayed 2014).

Water is supplied by the Water and Sanitation Agency (WASA) in more than 150 union councils of Lahore. Until the year 2009, WASA installed approximately 380 pumping wells to meet the local population requirements of water. Afterward, 200 more wells were added to the system by considering the increasing water demand due to the rapid growth in industrial and population (Qureshi and Sayed 2014). Lahore is a larger metropolitan city, accordingly for this study, a map with the City Districts boundaries, Towns/ Tehsil, and Union Councils levels were used to show the trends of spatio-temporal variations in the groundwater quality for convenience (The Local Government Ordinance in 2001, Administrative bodies for Districts) (see Table 1 and Fig. 1).

**Data and methods**

Data of 1305 valid sampling points were collected from different departments of Lahore including Pakistan Council of Research in Water Resource (PCRWR), Public Health Engineering Department and NESPAK, Lahore from 2005 to 2016 (see Fig. 2). Over the years, these departments have established reliable groundwater data repositories by employing the professional expertise in field water sampling and subsequent testing in their respective state of the art laboratories. During this period, acid washed high density polyethylene (HDPE) bottles were used to collect water samples as per the standard procedure. Global Positioning System (GPS) was used to record the geographic coordinates of each sampling location. Based on the data records, the water sampling was done in the second quarter of the year. The data were analyzed to measure the concentration levels of hydrochemical parameters as indicated by the American Public Health Association (APHA 1995). These yearly examined groundwater quality data include pH, TDS, EC, TH, Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, As, F, Fe, HCO₃⁻, NO₃⁻, and SO₄²⁻ for the period 2005–2016.
For the verification of the data's precision, the analyses were carried out for cation and anion ionic charge balances (< 10%), results more significant than 10% were not accepted (Islam et al. 2017). The expression used for the calculation of ionic charge balance is written as follows:

\[
\text{ICBE(\%)} = \left( \frac{\sum \text{Cations} - \sum \text{Anions}}{\sum \text{Cations} + \sum \text{Anions}} \right) \times 100. \quad (1)
\]

ArcGIS 10.7 software was used for this study, a valuable tool for organizing a physicochemical database (Sadat-Noori et al. 2014; Shabbir and Ahmad 2015; Thapa et al. 2017). GIS is a preferred option to present the sampling data geographically as a shape file and prepare spatial–temporal distribution maps to assess groundwater quality change. GIS-based Interpolation techniques have been utilized to deal with the data with a continuous distribution of variables that estimate the parameter value at unsampled locations (Muhammad and Zhonghua 2014; Gong et al. 2014; Charizopoulos et al. 2018).

Different types of Kriging interpolation models were used for different mechanisms of prediction, such as indicator, ordinary, universal, Inverse distance weighted (IDW), and simple Kriging. All the aforementioned interpolation methods were carried out for each physicochemical parameter. With the help of cross-validation criteria, the best-optimized approach of Inverse distance weighted (IDW) was finally used to prepare the spatio-temporal maps of each quality parameter.

IDW is generally suitable for the studies where spatial continuity is required to determine through interpolation (Balakrishnan 2011; Ketata et al. 2012; Selvam et al. 2016; Charizopoulos et al. 2018; Bairu et al. 2020). Interpolated data show variation spatially to estimated and
measured data points and creates different thematic layers for each ion (Selvam et al. 2016). These IDW assessed values are the weighted average values of the surrounding sample locations (Magesh et al. 2011; Selvam et al. 2016) and can be improved using a power function, mostly with its default value 2, whereas some other optimized values are also used for better results. A time series analysis was performed after developing precise interpolated (IDW) surfaces on the same spatial scale for the quality parameters of pH, TDS, EC, TH, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, As, F, Fe, HCO$_3^-$, NO$_3^-$, and SO$_4^{2-}$. A multivariate statistical technique (principal component analysis) was run using the statistical package for social sciences (IBM® SPSS® version 22.0) software for the computation of experimental groundwater data. The relationship between the pairs of different physicochemical can be identified with the help of a correlation matrix. Identification of hydrogeochemical processes has been computed using the Grapher® software (version 16.1) for groundwater classification, resulting in variations in groundwater composition.

**Water quality index**

The water quality index is the best way to summarize drinking water quality. It is used to identify the areas with suitable water for drinking purposes. WQI method was applied in the GIS environment on the available groundwater quality data of different years (i.e., 2005–2016). The WQI was calculated through three steps. In the first step a weight ($w_i$) has been allocated to 15 quality parameter pH, TDS, EC, TH, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, As, F, Fe, HCO$_3^-$, NO$_3^-$, and SO$_4^{2-}$ as per their perceived significance in the drinking water quality. The parameters such as EC, TDS, Cl$^-$, NO$_3^-$, Na$^+$, F, As, and SO$_4^{2-}$ have been allocated a maximum weight of 5 considering their significant importance in drinking water quality (Srinivasamoorthy et al. 2008; Vasanthavigar et al. 2010; Thapa et al. 2017; Tiwari et al. 2018). The minimum weight of 1 has been given to HCO$_3^-$ in water quality assessment due to its insignificant role. The remaining quality parameters (pH, TH, Ca$^{2+}$, Mg$^{2+}$, K$^+$, and Fe) have been assigned a weight between 1 and 5 (Ketata et al. 2012; Sadat-Noori
et al. 2014). Each physicochemical parameter’s relative weight ($W_i$) is calculated in the second step with the help of the following Eq. (2):

$$W_i = \frac{w_i}{\sum_{i=1}^{n} w_i}.$$  \hspace{1cm} (2)

Each physicochemical parameter’s input weight is represented by ($w_i$), total number of quality parameters is represented by ($n$), and $W_i$ is the ith parameter’s relative weight. In the third step, a quality rating scale ($q_i$) is calculated by using the following Eq. (3):

$$q_i = \left( \frac{C_i}{S_i} \right) \times 100.$$  \hspace{1cm} (3)

It is computed by dividing each water sample concentration to its own standard (WHO 2004) and multiplying by 100. The quality ranking is represented by ($q_i$). $C_i$ is the measured concentration of each physicochemical parameter, and $S_i$ is the WHO standard value for the same parameter in mg/l. Finally, for the calculation of WQI, the sub-index ($SI_i$) of the ith parameter is first determined using the following Eq. (4). Finally, Eq. (5) is utilized to calculate the WQI.

$$SI_i = (W_i \times q_i),$$  \hspace{1cm} (4)

$$WQI = \sum_{i=1}^{n} SI_i.$$  \hspace{1cm} (5)

**Entropy water quality index**

The theory of entropy has been applied in various fields of hydrology and water quality assessment studies. Shannon Claude was the first to propose the concept of information entropy in 1948 (Pei-Yue et al. 2010). Mathematically, it reports the degree of uncertainty and expresses an event’s randomness (Shyu et al. 2011). A negative correlation exists between entropy weight and the amount of information randomness (Shyu et al. 2011). A negative correlation exists between entropy weight and the amount of information randomness (Shyu et al. 2011). A negative correlation exists between entropy weight and the amount of information randomness (Shyu et al. 2011). A negative correlation exists between entropy weight and the amount of information randomness (Shyu et al. 2011).

The normalizing function is applied to eliminate the effect of different units of physicochemical parameters and converted to matrix $Y$, defined in Eq. (7).

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mn} \end{bmatrix}.$$  \hspace{1cm} (7)

The parameter index value for all groundwater samples is subsequently calculated using Eq. (8).

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m} (Y_{ij})}.$$  \hspace{1cm} (8)

The entropy ($e_j$) can be computed, using Eq. (9) used for all parameters.

$$e_j = \frac{1}{\ln m} \sum_{i=1}^{n} P_{ij} \ln P_{ij}.$$  \hspace{1cm} (9)

Then the entropy weight ($w_j$), can be calculated for each parameter using the following Eq. (10).

$$w_j = \frac{1 - e_j}{\sum_{i=1}^{n} \left( 1 - e_j \right)}.$$  \hspace{1cm} (10)

A qualitative grade ($q_j$) is computed for each groundwater quality parameter using Eq. (11).

$$q_j = \frac{C_j}{S_j} \times 100.$$  \hspace{1cm} (11)

$C_j$ represents all physicochemical parameters’ concentration, and $S_j$ is the permissible limit according to WHO standards. The entropy water quality index (EWQI) is computed using the following Eq. (12).

$$EWQI = \sum_{j=1}^{n} \left( \frac{w_j}{q_j} \right).$$  \hspace{1cm} (12)

Spatial disparity maps of WQI and EWQI from 2005 to 2016 were prepared to recognize the regions with poor groundwater quality. For a convenient comparison, all-time series data related to WQI and EWQI were categorized into five classes, presented in Table 2. (Vasanthavilagar et al. 2010; Jianhua et al. 2011; Tiwari et al. 2018).
Sample density and pattern analysis

The combination of geostatistical models (IDW interpolation) and the spatial autocorrelation index can be successfully employed as a decision-making tool to study the impact of sample density on the spatial variability patterns of groundwater quality parameters (Islam et al. 2017). The IDW interpolation model is considered the most extensively used model, and its cross-validation technique is employed to test the predictive behavior of the best fit model. The accuracy of interpolation surfaces and the predictive performance of the best-fitted model can be evaluated by calculating the root mean square error (RMSE) and coefficient of determination ($R^2$) (Zhang et al. 2015, 2018). A minimum RMSE and $R^2$ values close to unity exhibit suitable interpolation surfaces and best-fitted model (ESRI 2009). RMSE is applied to determine the accuracy of interpolation surfaces by comparing their errors, and the surface with the lowest error is considered the most suitable. Curve fitting between observed and predicted values was also established to determine the coefficient of determination and to study the effect of sample density on the accuracy of prediction surfaces (Rahman et al. 2018). The following Eq. (13) is used to determine the error:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}},$$

where $N$ is the total number of observations, $x_i$ is the observed value, and $\bar{x}$ is the predicted value.

Due to spatial heterogeneity in the groundwater sampling locations and changes in the density of sample data over the years in the study, an efficient autocorrelation model is required that can effectively identify the spatial patterns, resulting in varying sample densities (Islam et al. 2017). This model defines the three states of spatial patterns: dispersed, random, and clustered based on Moran’s I index value that describes the relationship between two neighboring quality parameter’s proximity in a geographic location (Anselin 1995; Griffith 2003). This index can also indicate the statistically significant spatial variation among the sampling sites and has been extensively used in mapping the spatial cluster analysis as compared to other spatial autocorrelation index models (Fu et al. 2014; Islam et al. 2017; Gorgij et al. 2017). Subsequently, this index has been utilized in this study and can be calculated by the following Eq. (14):

$$I = \frac{n \sum \sum w_{ab} [x_a - \bar{x}] [x_b - \bar{x}]}{w \sum [x_a - \bar{x}]^2},$$

where $\bar{x}$ is the mean value of $n$ number of samples, $x_a$ is the value of the parameter at site $a$, $x_b$ is the value of a parameter at site $b$, and $w_{ab}$ is the assigned weight based on the geographic location of $a$ and $b$, can be written in term of the inverse of the distance.

The value of Moran’s I index fluctuates between 1 and −1, where 1 represents positive spatial autocorrelation between the sample points exhibiting the cluster distribution pattern. While −1 represents the negative spatial autocorrelation indicating the dispersed distribution pattern. If there is no spatial interaction between the sample points, then the Moran’s I index value is equal to zero suggesting the spatial randomness (Islam et al. 2017).

Results and discussion

Groundwater quality variation can be best interpreted using spatial–temporal maps of different physicochemical parameters. This study has resulted in a comprehensive analysis of selected parameters pH, TDS, EC, TH, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, As, F, Fe, HCO$_3^-$, NO$_3^-$, and SO$_4^{2-}$ that have a maximum impact on Groundwater quality. Initially, the ionic charge balance was computed to verify the complete chemical analysis accuracy between the cations (Ca$^{2+}$, Mg$^{2+}$, Na$^+$, and K$^+$) and anions (HCO$_3^-$, Cl$^-$, SO$_4^{2-}$ and NO$_3^-$). It was observed that approximately 96% of the water samples were within this range of <10%. The measured relative standard deviation for all collected groundwater samples is within the range of ±2%.

Spatio-temporal analysis of individual physicochemical parameters

Spatio-temporal thematic maps of all the quality parameters were prepared by utilizing the best fit interpolation technique IDW. By considering the correlation matrix and % of the samples exceeding allowable limits prescribed by WHO standards, the maps of Ca$^{2+}$, Mg$^{2+}$, Cl$^-$, Fe, NO$_3^-$ and F are not shown here because all samples lie within the permissible limit. In this research, fifteen quality parameters were characterized into classes as per their minimum and maximum values to understand their possible trends considering WHO Standards. Finally, to show the data’s spatial and
temporal variation, thematic maps of all the quality parameters were created for further analysis for the study area.

**pH**

A relatively high percentage of 56% of samples in the year 2005 was observed to have pH values exceeding the prescribed limit of WHO guideline values for drinking water. Only a few samples showed a neutral pH value, and all the other samples had high pH values.

Spatio-temporal maps, as shown in Fig. 3, suggest that the area along the Hudiara drain (Wagah, Nishter and Iqbal zone) falls above the neutral value indicating the alkaline behavior of water. Hudiara drain originates from Batala (Punjab, India), and in Amritsar, after joining many tributaries, enters Pakistan near village Laloo (Yaseen et al. 2009). It flows through the border area between Pakistan and India, so domestic wastewater and industrial effluents are being discharged from both sides into this drain without treatment. Many industries, including textile mills, paper, electronic, plastic, paint, and pharmaceutical industries, are the primary source of pollution, which are the reasons for a cautious increase in pH value along the Hudiara Drain (Afzal et al. 2000).

Overall, it is clear from the maps generated for this study that there is no significant pH value change. However, all Lahore zones in different periods showed higher values (more than 7.99), with a maximum of 8.5 except for very few Lahore areas. The peak value of pH of 8.9 was observed in the Samanabad zone of Lahore. Overall, an increasing to neutral trend for most of years had been observed in pH values with a maximum value of 8.5 in 2013 Wagha Town, Lahore. A high concentration plume reappeared in 2014 in the Southwestern part of Lahore with a peak value of 8.7. For all the years the region along the Hudiara drain in the EW direction remained prominent with a progressive pH concentration in the temporal analysis.

**Total dissolved solids (TDS)**

The analysis shows a relatively high percentage (25% to 84%) of the samples having TDS values of more than
500 mg/l. Spatio-temporal distribution of TDS is demonstrated in Fig. 4.

This study shows a remarkable change in TDS concentration with the lowest value of 762 mg/l in 2005, and an increasing trend has been observed in the peak value of 2182 mg/l and 1875 mg/l in the years 2007 and 2009, respectively. In the years 2010 and 2011, a small reduction in TDS value was observed, then its value goes up to 2710 mg/l in the region. The most southern part (Nishter town) of the Lahore area, including few villages, exhibits TDS' highest values. In short, the spatio-temporal maps reveal that most of the research area under study showed the TDS value > 500 mg/l for all the years under consideration, hence making it unsuitable for drinking purposes. Based on this analysis, only a few Lahore areas meet WHO standards, which is quite alarming for the local authorities.

Anthropogenic activities like urban, agricultural, industrial, and sewage wastewater could be the primary sources of TDS in groundwater. Geogenic processes are also responsible for the high concentration of TDS in groundwater due to the mixing of Ca$^{2+}$, HCO$_3^-$, Cl$^-$, and SO$_4^{2-}$ (Subba Rao 2002; Muhammad and Zhonghua 2014; Mahmood et al. 2016). More specifically, the regions including the villages in the southern part (Nishter town) of the study area where sewage is not properly designed, and the domestic wastewater could be the reason for high TDS in the study area. While in western part (Iqbal Town) of Lahore, there is a rapid growth in industrial units over the last decade that could be the obvious reason for the anthropogenic contamination of groundwater in these areas as reported in the results.

**Electrical conductivity (EC)**

The electrical conductivity (EC) of groundwater could be due to numerous dissolved salts. Around 16% of the samples (see Table 3) were found to have electrical conductivity values exceeding the WHO-recommended standards. Figure 5 shows that EC values are in the desirable limit in most of the Lahore region, but moderate increases were observed in southern parts along the Hudiara drain, making it unsuitable for drinking purposes. Spatio-temporal analysis of electrical conductivity shows that the EC concentration varies from 1090 μS/cm to a peak value of 2870 μS/cm in the years 2005 and 2016, respectively. However, there are isolated peaks...
| No. | Quality parameter | No. of samples | No. of samples exceeding allowable limits | % age of samples exceeding allowable limits |
|-----|-------------------|----------------|------------------------------------------|-------------------------------------------|
| 1   | pH                | 1305           | 94                                       | 7.2                                       |
| 2   | TDS (mg/l)        | 1305           | 777                                      | 59.5                                      |
| 3   | EC (μS/cm)        | 1305           | 211                                      | 16.2                                      |
| 4   | TH (mg/l)         | 1305           | 181                                      | 13.9                                      |
| 5   | Ca (mg/l)         | 1305           | 69                                       | 5.3                                       |
| 6   | Mg (mg/l)         | 1305           | 73                                       | 5.6                                       |
| 7   | Na (mg/l)         | 1305           | 229                                      | 17.5                                      |
| 8   | K (mg/l)          | 1305           | 104                                      | 8.0                                       |
| 9   | Cl (mg/l)         | 1305           | 9                                        | 0.7                                       |
| 10  | As (μg/l)         | 1305           | 1112                                     | 85.2                                      |
| 11  | F (mg/l)          | 1305           | 91                                       | 7.0                                       |
| 12  | Fe (mg/l)         | 1305           | 39                                       | 3.0                                       |
| 13  | HCO₃²⁻ (mg/l)     | 1305           | 1290                                     | 98.9                                      |
| 14  | NO₃⁻ (mg/l)       | 1305           | 0                                        | 0.0                                       |
| 15  | SO₄²⁻ (mg/l)      | 1305           | 105                                      | 8.0                                       |

Fig. 5 Spatio-temporal map of EC in the Lahore region
with higher values observed in few regions where the agricultural activities are prominent.

**Total hardness (TH)**

Cations (Ca$^{2+}$, Mg$^{2+}$) and anions (Cl$^-$, HCO$_3^-$, CO$_3^{2-}$ and SO$_4^{2-}$) are the primary sources of water hardness (Ravikumar et al. 2011). According to the grading standards, groundwater hardness can be classified as soft water (TH < 150 mg/l), moderately hard water (150–300 mg/l), hard water (300–450 mg/l), and extremely hard water (TH > 450 mg/l) (Sadat-Noori et al. 2014). The spatial distribution map shows that most of the research area’s groundwater lies in the moderate category, which indicates the poor groundwater quality for drinking purposes. It is attributed to the untreated industrial effluents and sewage waste disposal, as shown in Fig. 6. The temporal distribution shows that hardness is average in the starting years. Still, a high concentration plume appeared in the Western part (Iqbal Town) of the Lahore region with a maximum of 1070 mg/l value. It is pertinent to mention that in this specific region a new industrial state was established named as Sundar Industrial state back in 2007. Due to this several new industrial units were installed in the region including paper, garments, pharmaceutical, dying, and others in the same year. This might have rapidly increased the TH values in the groundwater and in later years the proper measures were taken by many of these industrial units to minimize the groundwater contamination.

**Cations**

A relatively low percentage (up to 6% for calcium and magnesium) of the samples was found to have calcium and magnesium values exceeding the prescribed values of WHO-recommended limits for drinking water given in Table 3. For both ions, most of the research area is within the WHO standards’ permissible limit. In the temporal analysis, mostly the same areas of the Lahore region found prominent with a high concentration of calcium ions. From the year 2005 to 2016, most of the study area showed a random pattern of change in Ca$^{2+}$ concentration with a peak value of 440 mg/l

---

Fig. 6 Spatio-temporal map of total hardness in the Lahore region
in 2007 in Ravi town. Other regions with more significant values were found in Data Gunj Buksh (Anarkali) and Gulberg town. The spot of Anarkali was also identified for its deteriorating water quality by (Mahmood et al. 2016). Other than the Anarkali region in different years, different areas were highlighted for a high concentration of Ca$^{2+}$ and Mg$^{2+}$ leading to hard water problems.

Industrial effluents, improper sewage waste, and landfill leachate are the main reasons for groundwater hardness in the research area. (Ibtisam and Ghaffar 2012; Muhammad and Zhonghua 2014). Magnesium exhibits more or less similar distribution patterns in time series analysis. Results showed the highest values for the years 2007, 2014, 2015, and 2016 were 86 mg/l, 77 mg/l, 95 mg/l, and 75 mg/l, respectively. The region of Anarkali was found again prominent in 2007 for Mg concentration. Overall, the concentration of calcium ions was higher than the magnesium ion.

Spatial–temporal map, as shown in Fig. 7, depicts Na$^+$ distribution in the study area, which is within the range prescribed by WHO standards in the year 2005 then it starts increasing with a peak value of 460 mg/l in 2007 in the southern part (Nishtar Town) of Lahore region including western part (Iqbal Town) where most of the land is agriculture. Its concentration increases over time, with the highest value of 770 mg/l in 2015 in the same southern part of Lahore. As shown in Fig. 8, it was observed that most of the research areas having potassium ions within the WHO limit, except the Eastern and Southern parts of Lahore, which includes Wagha town and Nishtar town. They are prominent in most of the years of analysis. For the temporal window, only in the year 2005 peak concentration of potassium is within limits. In the remaining years, there is a progressive increase in the peak concentration of potassium. The maximum value was observed in 2007 in the Samanabad area of Lahore, 153.6 mg/l. A high concentration plume reappeared in 2015 near the Lahore zoo with a peak value of 145.9 mg/l.

Anions

Municipal or domestic sewage waste, fertilizers, and atmospheric precipitation are the primary sources of Cl$^-$ ions in groundwater (Mallick et al. 2018). Low percentages (up to

![Fig. 7 Spatio-temporal map of sodium in the Lahore region](image)
3%) of the samples were found to have Cl$^-$ values exceeding the WHO-recommended limits for drinking water, as given in Table 3. Its concentration in groundwater has more or less the same distribution pattern in the Lahore region for all the studied years. Chloride exceeded the WHO limit of 250 mg/l in years 2007, 2009, 2014, 2015 and 2016 with peak values of 276 mg/l, 278 mg/l, 268 mg/l, 379 mg/l and 528 mg/l, respectively. These peak values were observed in some areas of Nishter and the Data Ganj Baksh zone of Lahore.

Figure 9 shows the spatio-temporal distribution of bicarbonate in the area under study. According to spatial distribution maps, 97–100% of samples, as given in Table 3, are not in the permissible limit prescribed by the WHO Standards. In the year 2005, the concentration of the HCO$_3^-$ was quite average. In 2007, a peak value of 950 mg/l was observed in the southern side (Nishter zone) and the Western part (Allama Iqbal Zone) of the Lahore region. The same peak value of 990 mg/l reappeared in 2015 and spread again in Lahore’s southern part in the Nishter zone. Fertilizers, domestic and sewage wastewater are the primary sources of nitrate in the groundwater. The analysis reveals that NO$_3^-$ behavior is quite normal in all the years. Table 3 indicates that no samples were found to have nitrate values exceeding the WHO-recommended limits. The nitrate concentration in groundwater samples has more or less the same distribution pattern in the Lahore region for all the studied years.

Oxidative weathering of sulfide minerals such as pyrite is the primary source of sulfate in water. However, gypsum and anhydrite can also increase the sulfate ions in drinking water (Tiwari et al. 2018). Sulfates are also derived in minor quantities from air pollutants, fertilizers, and household wastewater (Farooqi et al. 2007). As shown in Table 3, sulfate concentrations of three to seventeen percent (3–17%) of the water samples are not within the permissible limit prescribed by WHO standards. Spatio-temporal distribution of sulfate ions, as shown in Fig. 10, exhibits more or less similar distribution patterns. It depicts that the elevated sulfate concentrations were recorded only in few samples every year. Simultaneously, the highest value was observed in 2015, with 800 mg/l located in Lahore’s Southern parts, mostly near industrial areas.
As per the temporal analysis 2005–2016, most of the research area showed sulfate concentration higher than WHO prescribed value of 250 mg/l except for 2005. In 2007 and 2009, a peak value of sulfate ions was observed in the Southern part of Lahore with a value of 470 mg/l and 603 mg/l in the Nishter zone and the Western part of Lahore Allama Iqbal Zone. A declining trend was observed in 2010, 2011, and 2013 with a maximum value of 263, 408, and 509 mg/l. Its high concentration reappeared with a peak value of 800 mg/l in 2015 again in the Southern parts of Lahore.

**Arsenic**

According to the previous studies, along the Ravi River, groundwater quality with respect to arsenic is satisfactory, but it slowly deteriorates in Lahore's Southern and southwestern parts. The Water and Sanitation Agency (WASA), Punjab Irrigation Department (PID), Pakistan Council of Research in Water Resources (PCRWR), and the Environmental Protection Agency (EPA) continuously monitor the groundwater quality and found it to be contaminated with Arsenic (Qureshi and Sayed 2014). Relatively high percentages (67–97%) of the samples in each year were observed to have Arsenic concentrations exceeding WHO standards given in Table 3. Figure 11 shows the spatio-temporal distribution of arsenic in the research area. From this time series analysis, most groundwater samples have higher arsenic concentrations than 10 ppb (WHO 2004). Spatial disparity maps of each year show the high arsenic concentration in groundwater, and there is no safe zone deducted as prescribed by WHO standards. A peak value of 180 ppb was observed in 2007 from the Northern side of Lahore. This peak value reappeared in the same area in 2010 and 2011 with the highest value of 128.66 ppb, and in the years 2013 and 2014, it appeared in the southern part of the research area with a peak value of 159.67 ppb. Kiln Factories are the major primary source of arsenic contamination, while fertilizers are considered secondary sources (Farooqi et al. 2007; Rasool et al. 2015). Hence, the new severe problems related to groundwater quality deterioration may arise very soon in the study area.

Fig. 9  Spatio-temporal map of Bicarbonates in the Lahore region
Fluoride

According to the spatio-temporal distribution, a relatively low percentage (up to 7%) of groundwater samples having a concentration of fluoride exceeding 1.5 mg/l (WHO 2004), as given in Table 3. In the temporal analysis, mostly the same areas of the Lahore region were prominent with a high concentration of fluoride ions. From 2005 to 2016, most of the study area showed a random concentration pattern with a peak value of 7.9 mg/l in 2007 in the Nishter zone. Other more significant values were found in the same area with a maximum concentration of 3.86 mg/l for 2010 and 2011 and a peak value of 4.56 mg/l for years 2013 and 2014. Peak values are also observed in the Allama Iqbal and Wagha zone in 2015 and 2016. The untreated, harmful industrial discharge and domestic sewage wastewater resulted in high fluoride concentration in drinking water (Azizullah et al. 2011; Brahman et al. 2013; Rasool et al. 2015).

Iron

The spatio-temporal distribution of Iron ions concentration reveals that most of the research area lies within WHO limits. In the temporal analysis, mostly the same areas of the Lahore region were prominent with a high iron ion concentration. Most of the study area from the year 2005 to the year 2016 showed more or less the same distribution pattern of Iron concentration with a peak concentration of 2.32 mg/l in the year 2005. The peak values of iron concentration in the area were only in the year 2007 is within the WHO standards limit other years exhibit high values in a few areas of Lahore. Low percentages (up to 5%) of the samples were found to have iron values exceeding the WHO-recommended limits for drinking water, as given in Table 3.
Index method (WQI and EWQI)

It is necessary to understand the relationship between the information entropy value ($e_j$), physicochemical parameters, and entropy weight ($w_j$) before creating the thematic maps of the EWQI and WQI. The entropy theory concept helps to identify the more critical physicochemical parameters affecting groundwater quality with the maximum entropy weight and the minimum information entropy value (Shyu et al. 2011; Jianhua et al. 2011; Gorgij et al. 2017). Table 4 presents the assigned weights ($w_i$) to all 15 quality parameters based on their perceived significance in the drinking water quality, relative weight ($W_i$), calculated information entropy and entropy weights. Results show that Fe (0.210), then $\text{NO}_3^-$ (0.203), K (0.164), F (0.065), $\text{SO}_4^{2-}$ (0.052), and As (0.024) have the leading role, determining the water quality in the Lahore region. All other parameters with a minimum entropy weight and maximum information entropy have less impact on groundwater quality.

The spatial disparity map of WQI and EWQI from 2005 to 2016 is shown in Figs. 12 and 13. Based on WQI, the...
research area has been categorized into excellent water (WQI < 50), very good water (50–70), good water (70–90), fair water (90–100), and poor water (WQI > 100). As per these Figs, in 2005, the standard of groundwater quality was reasonable compared to subsequent years, in which WQI of most of the groundwater samples of research areas fall in the good category. GIS-based models of EWQI and WQI reveal similar trends in identifying the low-water quality index zones in the research area. Degraded water quality was observed in the next years from 2007 to 2016. A number of spots in the study area were identified in the poor category. One significant plume is in the Northeastern portion of the research area, and others are in the Western and the southern regions near the industrial activities. In the next few years, the southern concentration diluted subsequently from "category poor to category good showing the temporary nature of contaminations. But it reappeared in 2013, and continuous spatial spreading was identified in the Western and Southern regions of Lahore with the poor category. As compared to WQI, the low values of EWQI are observed in most of the research areas, ranked as excellent to good category. Table 5 summarizes the spatio-temporal analysis of WQI and EWQI and enlists the affected Lahore regions concerning their zones.

Table 6 reveals that 100 is the critical limit for WQI and EWQI, indicating that 88% samples in WQI analysis and 96% of samples EWQI analysis are considered safe for drinking purposes. Muhammad and Zhonghua (2014) and Mahmood et al. (2016) have also studied the few Lahore areas for groundwater quality and identify the same regions with contaminated water. However, with time series analysis from 2005 to 2016 in the present study, the Nishter, Allama Iqbal, Data Gunj Baksh, and Wagha zones have been recognized as a persistent water contamination source (see Table 8). Demographic transition, industrialization, and urbanization exerted extreme pressure on the city area's groundwater in excessive pumping generating a deep and large cone of depression in Lahore's urban area. Because of this composite drawdown curve, the seepage water has to travel a long and deep vertical path to become a part of the local aquifer. The possibility that groundwater dissolves different solids, minerals, and ions while traveling through

![Fig. 12 Spatio-temporal map of WQI in the Lahore region](image-url)
the aquifer material are enhanced, reflecting the low-water quality index zones.

The groundwater table is getting lower by 0.84 m annually due to the increased abstraction rates and lateral expansion of Lahore city limits, leading to reduced permeable surfaces (Mahmood et al. 2016). In addition to geogenic activities, other anthropogenic factors are also reported in the research area, which is leachate of municipal solid waste, industrial activities, poor sanitation system. Overall, as per the detailed analysis, it is clear from Fig. 14 that a significant part of the Nisther, Allama Iqbal, and Wagha town falls in the poor water quality index. Other than these zones, small spots were also observed in Data Gunj Baksh, Samanabad, and Gulberg Zone. Groundwater quality deteriorates over

![Fig. 13 Spatio-temporal map of EWQI in the Lahore region](image)

| Sr # | Directions | Zones of Lahore                     | Affected areas of Lahore                          |
|------|------------|------------------------------------|--------------------------------------------------|
| 1    | North      | Ravi, Shalimar, and Data Ganj Baksh Town | Mazang, Anarkali                                  |
| 2    | South      | Nishter Town                       | Youhana abad, Pandoki, Bagharain, Gajju Matta, Jia Bagga, Ladeke |
| 3    | East       | Wagha Town                         | Lakhodair, Hudiara, Padhana                       |
| 4    | West       | Iqbal Town                         | Raiwind, Manga Mandi, Shahpur, Shamkey bhattian, Sultankay, Pajian |
| 5    | Central Part | Aziz Bhatti and Gulberg Town     |                                                   |
| 6    | Northwest  | Samanabad Town                     | Samanabad                                         |

Table 5 Summary areas with poor WQI and EWQI
time (the year 2005 to 2016), and the contaminated zones are consistently growing, which may pose severe threats to the local population.

**Statistical analysis**

After performing the spatio-temporal analysis, statistical properties were evaluated to visualize potential relationship between physicochemical parameters. The statistical measures of all physicochemical parameters like minimum and maximum values, mean, median, skewness, kurtosis, standard deviation, and WHO standard limits are presented in Table 7. The mean value of two quality parameters $\text{HCO}_3^-$ and As, are more than the recommended WHO standards as given in Table 7. After a preliminary assessment of the 15 physicochemical parameters under study, the correlation matrix was utilized to identify the correlation coefficients of parameters influencing groundwater’s standard quality.

As shown in Table 8, numerous physicochemical parameter pairs have a statistically significant positive correlation at $\alpha$ is equal to 0.05, indicating identical origin sources, which could be geogenic, or anthropogenic. EC and $\text{Na}^+$ exhibit a positive correlation, showing a geogenic source in the research area, from similar origin weathering and disintegration of soil-forming minerals. Simultaneously, insignificant correlations are also observed in other groundwater quality parameters, indicating an independent source of origin.

To further understand the impact of physicochemical parameters on groundwater quality, the normalized hydrochemical data were analyzed by employing the Principal Component Analysis (PCA). In this analysis, to recognize the number of principal component factors, an orthogonal Kaiser’s varimax rotation and scree plot were utilized (Davis 1986; Lattin et al. 2003). Physicochemical parameters clustered into five groups based on eigenvalues greater than 1. Five factors are representing a cumulative total variance of 69.312%.

The first Principal Component 1 (PC-1), explaining 36.56% of the total variance, is high loaded with EC, $\text{Na}^+$, $\text{HCO}_3^-$, $\text{SO}_4^{2-}$, TDS, and $\text{Cl}^-$ (see Table 9) and moderately loaded with TH and $\text{Mg}^+$. In contrast, in the PC-2, only three parameters ($\text{Ca}^{2+}$, $\text{Mg}^{2+}$, and TH) show moderate loading exhibiting a total variance of 10.5%, representing the same source of origin in the area under study. It is associated with significant ions resulting from geogenic activities, including geochemical processes and weathering of soil-forming minerals (Okiongbo and Douglas 2015; Pazand 2016). This study shows that $\text{Na}^+$ and

| Index method | WQI range | No. of samples | % age | Quality of water |
|--------------|-----------|----------------|-------|-----------------|
| WQI          | < 50      | 139            | 11    | Excellent       |
|              | 50–100    | 1002           | 77    | Good            |
|              | 100–200   | 160            | 12    | Poor            |
|              | 200–300   | 4              | 0     | Very poor       |
|              | > 300     | 0              | 0     | Not fit for drinking |
| EWQI         | < 50      | 824            | 63    | Excellent       |
|              | 50–100    | 426            | 33    | Good            |
|              | 100–200   | 51             | 4     | Poor            |
|              | 200–300   | 3              | 0     | Very poor       |
|              | > 300     | 1              | 0     | Not fit for drinking |

**Table 6** Percentage (%) calculation of groundwater samples based on WQI and EWQI

![Fig. 14 Summary spatial map of WQI and EWQI in the Lahore region](image-url)
HCO$_3^-$ are the most prominent ions among the analyzed parameters, grouped in the first PC, and strongly correlated. They may be originated from the natural weathering of calcium, magnesium carbonate, bicarbonate, and silicate minerals from the underlying geology. The reverse ion exchange process may be attributed to the high quantity of Na$^+$ ions in the research area’s groundwater samples. This study reveals that the high amount of Ca$^{2+}$, Mg$^{2+}$ is also observed in a few of the research area’s groundwater samples. Recharge water passes through the alluvial aquifer enriched in calcium and magnesium carbonate minerals, resulting in a high concentration of these ions in groundwater samples (Singh et al. 2008).

The third PC (PC-3) is only positively loaded of As with 7.99% of the total variance. The presence of arsenic reflects different and multiple sources, including natural activities like oxidation reactions and arsenic compound dissolution and human activities such as industrial effluents (Rasool et al. 2015). The fourth PC (PC-4) explains 7.547% of the total variance moderately loaded with pH, NO$_3^-$, and Fe, indicating an anthropogenic origin source. Their primary source is the sewage wastewater, industrial effluents, and

### Table 7

| Parameter       | Minimum | Maximum | Mean | Median | Skewness | Kurtosis | Std-dev. | WHO (2004) limit |
|-----------------|---------|---------|------|--------|----------|----------|----------|------------------|
| pH              | 6.6     | 8.9     | 7.5  | 7.5    | 0.6      | 4.0      | 0.2      | 8.5              |
| TDS (mg/l)      | 69.0    | 2710    | 565.6| 495.0  | 1.7      | 7.9      | 294.6    | 1000             |
| EC (μS/cm)      | 108.0   | 2870    | 983.9| 910.0  | 1.4      | 6.1      | 509.8    | 1400             |
| TH (mg/l)       | 20.0    | 1070    | 204.0| 186.8  | 1.6      | 10.1     | 92.1     | 500              |
| Ca (mg/l)       | 6.0     | 440     | 43.4 | 38.6   | 7.6      | 86.5     | 30.9     | 75               |
| Mg (mg/l)       | 0.5     | 95      | 25.1 | 22.9   | 1.2      | 5.1      | 14.2     | 50               |
| Na (mg/l)       | 2.0     | 770     | 135.5| 115.3  | 2.1      | 10.0     | 101.8    | 200              |
| K (mg/l)        | 0.1     | 304     | 6.4  | 11.4   | 198.0    | 13.8     | 12       |
| Cl (mg/l)       | 5.0     | 528     | 58.4 | 50.0   | 2.8      | 18.9     | 44.1     | 250              |
| As (μg/l)       | 3.1     | 180     | 24.5 | 23.1   | 3.4      | 26.6     | 16.3     | 10               |
| F (mg/l)        | 0.0     | 7.9     | 0.7  | 0.5    | 3.9      | 29.7     | 0.6      | 1.5              |
| Fe (mg/l)       | 0.0     | 2.32    | 0.1  | 0.0    | 7.7      | 82.4     | 0.1      | 0.3              |
| HCO$_3^-$ (mg/l)| 25.0    | 990     | 338.4| 325.5  | 1.0      | 5.3      | 135.7    | 120              |
| NO$_3^-$ (mg/l) | 0.0     | 32      | 0.8  | 0.3    | 11.0     | 184.9    | 1.7      | 45               |
| SO$_4^{2-}$ (mg/l)| 8.0   | 800     | 122.1| 98.4   | 2.1      | 10.3     | 91.1     | 250              |

### Table 8

| Parameter       | pH     | TDS    | EC     | TH     | Ca     | Mg$^{2+}$ | Na$^+$  | K$^+$  | Cl$^-$ | As     | F      | Fe     | HCO$_3^-$ | NO$_3^-$ | SO$_4^{2-}$ |
|-----------------|--------|--------|--------|--------|--------|-----------|---------|--------|--------|--------|--------|--------|-----------|-----------|-------------|
| pH              | 1      |        |        |        |        |           |         |        |        |        |        |        |           |           |             |
| TDS             | −0.12  | 1      |        |        |        |           |         |        |        |        |        |        |           |           |             |
| EC              | −0.04  | 0.89   | 1      |        |        |           |         |        |        |        |        |        |           |           |             |
| TH              | −0.02  | 0.44   | 0.49   | 1      |        |           |         |        |        |        |        |        |           |           |             |
| Ca$^{2+}$       | 0.03   | 0.03   | 0.04   | 0.26   | 1      |           |         |        |        |        |        |        |           |           |             |
| Mg$^{2+}$       | −0.01  | 0.54   | 0.61   | 0.86   | 0.15   | 1         |         |        |        |        |        |        |           |           |             |
| Na$^+$          | −0.14  | 0.75   | 0.77   | 0.16   | −0.09  | 0.29      | 1       |        |        |        |        |        |           |           |             |
| K$^+$           | −0.01  | 0.22   | 0.27   | 0.21   | 0.12   | 0.23      | 0.22    | 1      |        |        |        |        |           |           |             |
| Cl$^-$          | 0.01   | 0.71   | 0.78   | 0.41   | 0.04   | 0.45      | 0.53    | 0.26   | 1      |        |        |        |           |           |             |
| As              | 0.09   | −0.10  | −0.09  | −0.02  | 0.00   | −0.03     | 0.01    | −0.02  | −0.06  | 1      |        |        |           |           |             |
| F               | −0.10  | 0.27   | 0.24   | 0.07   | −0.03  | 0.15      | 0.26    | 0.10   | 0.19   | −0.16  | 1      |        |           |           |             |
| Fe              | 0.05   | −0.01  | −0.05  | 0.00   | −0.05  | 0.00      | −0.04   | 0.00   | −0.02  | −0.08  | 1      |        |           |           |             |
| HCO$_3^-$       | −0.09  | 0.75   | 0.75   | 0.41   | 0.01   | 0.52      | 0.81    | 0.26   | 0.45   | −0.04  | 0.24   | −0.03  | 1         |           |             |
| NO$_3^-$        | 0.10   | 0.16   | 0.11   | 0.08   | 0.06   | 0.12      | 0.06    | 0.10   | 0.08   | −0.11  | 0.07   | 0.06   | 0.17      | 1         |             |
| SO$_4^{2-}$     | −0.08  | 0.76   | 0.76   | 0.32   | 0.03   | 0.43      | 0.85    | 0.16   | 0.60   | 0.06   | 0.18   | −0.07  | 0.70      | 0.06      | 1           |

Bold values show significant positive correlation between physicochemical parameters.
excessive use of fertilizers. The observed iron concentration may also be related to inorganic sulfides' oxidation and weathering of iron-bearing minerals (Okiongbo and Douglas 2015).

The relationship between groundwater chemistry and aquifer lithology can be better understood using the Gibbs plot and Trilinear piper diagram (Gibbs 1970). Gibbs plot is dependent on three standard hydrogeochemical processes: evaporation, precipitation, and rock water interaction mechanism and a generalized classification that interprets these controlling processes (Okiongbo and Douglas 2015). Gibbs diagram for the research area is plotted for 2005–2016, as shown in Fig. 15.

Gibbs plot reveals that 91% of groundwater samples fall under the rock weathering dominance area while only 9% of samples from all years exist in the evaporation zone. Since the study area is a semi-arid region, evaporation is also a contributing factor to hydrogeochemical processes. The local geology of the research area and ionic chemistry of groundwater samples are measured to understand these processes.

Piper diagram is the most commonly used graphical method, plotted to examine the chemical histories and hydrochemical water types for groundwater classification. The Piper Trilinear diagram presented in Fig. 16, from 2005 to 2016, clearly depicts the concentration of cations and anions in the research area. The piper diagram outcome reveals that the combination of ionic parameters of water samples in the Lahore region consists of five major types: Ca–Mg–HCO3, Na–Cl, Ca–Mg–Cl, Na–HCO3, Ca–Na–HCO3. During this study, most groundwater samples fall in the Ca–Na–HCO3, Ca–Mg–HCO3 kind of facies. Some of them lie in mixed Ca–Mg–Cl type of facies.

Sample density analysis

The Moran’s I Index, RMSE, and R² of each groundwater quality parameter highlight that the sample density may impact the accuracy and precision of the data. For a precise approximation of the concentration of groundwater quality parameters, robust water sampling at close intervals is required, which is often difficult due to financial and time constraints. The sample size for the first two years data is marginal as compared to the rest of the years of study area, i.e., 16, 37 for the years of 2005...
and 2007, respectively. Therefore, the sample density analysis was performed for this study in addition to achieve the stated objectives of this research. The coefficient of determination ($R^2$), root mean square error (RMSE) and Moran’s I index of groundwater quality parameters of all years (2005–2016) were calculated to study the impact of sample density on patterns created by IDW interpolation surfaces. Moran’s I index shows how neighboring values of physicochemical parameters are spatially correlated and identifies the spatial patterns as a random, cluster or disperse. Table 10 reveals the Moran’s I values, RMSE, $R^2$, and spatial patterns of analyzed groundwater quality parameters concerning the sample size in the study area. The results of autocorrelation values shows that there is no significant variation in Moran’s I Index, RMSE, $R^2$, and Distribution values of most of the physicochemical parameters. There is a noticeable difference in the autocorrelation values in the case where the sample size is 16 for the data of the year 2005 as compared to the rest of the years where the sample size is reasonably good. Moreover, the pattern analysis for the distribution factor depicts that half of the analyzed sample points and groundwater quality parameters showed a spatial cluster pattern with positive spatial autocorrelation and half of them exhibited a spatially random pattern with negative spatial autocorrelation.

**Conclusions**

In this study GIS-based geostatistics tools, including water quality index, entropy water quality index, and multivariate statistical analysis were successfully utilized to analyze the regional-level spatio-temporal variation in the physicochemical parameters of groundwater quality. The study involves a large data set of 15 water quality parameters of the groundwater quality parameters (for a period of 2005–2016) of deep aquifers of Lahore region.

The basic statistical analysis shows that concentrations levels of majority of the parameters (pH, EC, TH, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, F, Fe, NO$^3$ and SO$^4$) in groundwater samples are within allowable limits prescribed by the WHO except As, HCO$^3$ and TDS. Overall, the groundwater quality is acceptable in the Northern part of the city near the Ravi River and become progressively worse in the South and Southwestern areas of Lahore. It was
also found that more than 85% of samples from Lahore City are contaminated with arsenic with a peak value of 159.67 ppb that was detected in the southern part of the research area, most likely due to anthropogenic activities like industrial growth and agricultural activities. No zone is declared a safe or arsenic-free zone, and it is quite costly and demanding to restore the quality of groundwater of Lahore city. A number of plumes with poor water quality index have been identified in different parts of the research area. Some of them have expanding trends, including Samanabad, Data Gunj Baksh, Wagh, Nishter,

Table 10 Results of Moran’s I index and error assessment of physicochemical parameters at different sample density

| No of valid samples | Quality parameter | Moran’s I | $R^2$ | RMSE | Distribution |
|---------------------|-------------------|-----------|-------|------|--------------|
| 16                  | pH                | −0.186    | 0.005 | 0.268 | Random       |
| 58                  | K                 | 1.097     | 0.920 | 1.727 | Cluster      |
| 113                 | 1.179             | 0.985     | 1.494 | Cluster |
| 133                 | 1.093             | 0.988     | 1.509 | Cluster |
| 135                 | 1.492             | 0.557     | 1.24  | Cluster |
| 145                 | 1.492             | 0.557     | 1.24  | Cluster |
| 205                 | 0.399             | 0.205     | 0.262 | Clustor |
| 209                 | 1.071             | 0.895     | 1.194 | Random |
| 291                 | −0.004            | 0.907     | 0.194 | Random |
| 16                  | TDS               | 0.044     | 0.000 | 166.027 | Random       |
| 58                  | 0.778             | 0.000     | 12.834 | Cluster |
| 113                 | 0.556             | 0.363     | 11.705 | Cluster |
| 133                 | 0.045             | 0.549     | 12.201 | Cluster |
| 145                 | 1.097             | 0.920     | 1.727 | Cluster |
| 205                 | 1.097             | 0.920     | 1.727 | Cluster |
| 209                 | 0.152             | 0.822     | 1.483 | Random |
| 291                 | 0.015             | 0.991     | 1.184 | Random |
| 16                  | EC                | −0.053    | 0.000 | 231.847 | Random       |
| 58                  | 0.779             | 0.563     | 73.423 | Cluster |
| 113                 | 0.994             | 0.486     | 72.977 | Cluster |
| 133                 | 0.353             | 0.665     | 71.838 | Cluster |
| 135                 | 0.533             | 0.826     | 200.077 | Random |
| 145                 | 0.533             | 0.826     | 200.077 | Random |
| 205                 | 1.071             | 0.895     | 1.194 | Random |
| 209                 | 0.144             | 0.679     | 69.527 | Random |
| 291                 | 0.091             | 0.734     | 67.818 | Random |
| 16                  | TH                | −0.010    | 0.009 | 71.993 | Random       |
| 58                  | 0.779             | 0.563     | 73.423 | Cluster |
| 113                 | 0.994             | 0.486     | 72.977 | Cluster |
| 133                 | 0.353             | 0.665     | 71.838 | Cluster |
| 135                 | 0.533             | 0.826     | 200.077 | Random |
| 145                 | 0.533             | 0.826     | 200.077 | Random |
| 205                 | 0.144             | 0.679     | 69.527 | Random |
| 209                 | 0.091             | 0.734     | 67.818 | Random |
| 291                 | 0.061             | 0.936     | 25.964 | Random |
| 16                  | Na                | 0.047     | 0.057 | 38.269 | Random       |
| 58                  | 1.027             | 0.862     | 39.246 | Cluster |
| 113                 | 0.860             | 0.608     | 39.341 | Cluster |
| 133                 | 0.324             | 0.890     | 38.793 | Random |
| 135                 | 0.340             | 0.901     | 39.499 | Random |
| 145                 | 0.962             | 0.665     | 39.554 | Cluster |
| 205                 | 0.321             | 0.837     | 34.894 | Random |
| 209                 | 0.211             | 0.767     | 34.350 | Random |
| 291                 | 0.055             | 0.945     | 26.929 | Random |

| No of valid samples | Quality parameter | Moran’s I | $R^2$ | RMSE | Distribution |
|---------------------|-------------------|-----------|-------|------|--------------|
| 16                  | K                 | −0.249    | 0.154 | 1.816 | Random       |
| 58                  | 1.097             | 0.920     | 1.727 | Cluster |
| 113                 | 2.259             | 0.973     | 1.786 | Cluster |
| 133                 | 1.179             | 0.985     | 1.494 | Cluster |
| 135                 | 1.093             | 0.988     | 1.509 | Cluster |
| 145                 | 0.531             | 0.976     | 1.518 | Cluster |
| 205                 | 0.099             | 0.993     | 1.446 | Random |
| 209                 | 0.152             | 0.822     | 1.483 | Random |
| 291                 | 0.015             | 0.991     | 1.184 | Random |
| 16                  | As                | −0.074    | 0.153 | 13.254 | Random       |
| 58                  | 0.778             | 0.000     | 12.834 | Cluster |
| 113                 | 0.556             | 0.363     | 11.705 | Cluster |
| 133                 | 0.045             | 0.549     | 12.201 | Cluster |
| 135                 | 0.009             | 0.597     | 12.427 | Random |
| 145                 | 0.795             | 0.231     | 12.289 | Cluster |
| 205                 | 0.287             | 0.778     | 12.193 | Cluster |
| 209                 | 1.645             | 0.704     | 6.727  | Cluster |
| 291                 | 0.195             | 0.004     | 11.283 | Random |
| 16                  | HCO3              | −0.009    | 0.012 | 74.646 | Random       |
| 58                  | 0.779             | 0.563     | 73.423 | Cluster |
| 113                 | 0.994             | 0.486     | 72.977 | Cluster |
| 133                 | 0.353             | 0.665     | 71.838 | Cluster |
| 135                 | 0.534             | 0.720     | 71.147 | Cluster |
| 145                 | 0.107             | 0.354     | 71.139 | Cluster |
| 205                 | 0.597             | 0.830     | 71.428 | Cluster |
| 209                 | 0.144             | 0.679     | 69.527 | Random |
| 291                 | 0.091             | 0.734     | 67.818 | Random |
| 16                  | SO4               | 0.005     | 0.037 | 28.360 | Random       |
| 58                  | 1.040             | 0.927     | 28.092 | Cluster |
| 113                 | 0.850             | 0.859     | 27.861 | Cluster |
| 133                 | 0.262             | 0.922     | 27.632 | Random |
| 135                 | 0.337             | 0.930     | 27.230 | Random |
| 145                 | 2.075             | 0.812     | 28.769 | Cluster |
| 205                 | 0.224             | 0.902     | 27.318 | Random |
| 209                 | 0.220             | 0.788     | 27.428 | Random |
| 291                 | 0.061             | 0.936     | 25.964 | Random |
and Allama Iqbal Zone, found as the persistent groundwater contamination source. In rural areas, groundwater quality is affected probably due to the mixing of contaminants generated from the unplanned industrial activities in the surrounding and which is responsible for the degradation of groundwater quality. Municipal solid waste leachate, weak sewerage system, discharge of municipal and industrial effluents in the Ravi River, and different drains passed through the city are other possible groundwater contamination sources.

The spatio-temporal study has enabled identifying the contaminated zones through GIS-based maps of WQI and EWQI. The entropy theory results reveal that Fe, then NO$_3^-$, K, F, SO$_4^{2-}$ and As have the leading roles in the water quality in the Lahore region. The remaining parameters, including pH, Na$^+$, Cl$^-$, Ca$^{2+}$, Mg$^{2+}$, EC, TDS, TH, and HCO$_3^-$ have a minimum influence on overall groundwater quality with the lowest entropy weight and showed a positive correlation with each other. The principal component analysis confirmed the entropy theory results. The parameters Fe, F, K, SO$_4^{2-}$ and As required significant attention because they have their independent origin designation.

The Gibbs diagram and piper trilinear outcome revealed that the concentration of these parameters (Na$^+$, Cl$^-$, Ca$^{2+}$, Mg$^{2+}$, EC, TDS, TH, and HCO$_3^-$) are mostly controlled by weathering of soil-forming minerals. These observations confirm that the groundwater quality and chemistry of the majority of the study area is influenced mainly by geogenic activities attributed to the weathering and disintegration of the soil-forming minerals and in few areas due to anthropogenic activities related to sewage and industrial wastewater, and agricultural activities.

The autocorrelation analysis was also performed which reveals that there is no significant impact of sample density on the spatio-temporal patterns of groundwater quality parameters except in the case of 2005 data, where the sample size is relatively low. In general, a slight increase in the coefficient of determination and reduction in RMSE values is observed in the analyzed data that indicates the improved spatio-temporal pattern of groundwater quality parameters of different years. Overall, the results of this analysis were found satisfactory as the sample density of the data of different years is reasonably good excluding 2005.

The study reveals the effectiveness of the combined utilization of GIS-based geostatistical methods, multicriteria statistical analysis and spatial autocorrelation at regional level with large data sets for the evaluation of groundwater quality parameters and their respective source characterization.

**Acknowledgements** The authors wish to thank University of Engineering and Technology, Lahore Pakistan for the financial and laboratory support to conduct this study.

**Author contributions** The corresponding author SI is a PhD scholar and main contributor to this research work. The second author, MFA, is her PhD research supervisor.

**Funding** This research work was partially funded by the University of Engineering and Technology, Lahore, Pakistan.

**Availability of data and materials** The authors have no reservations about sharing the data and material used for this study upon request.

**Code availability** No new codes were generated for this study.

**Declarations**

**Conflict of interest** The authors do not have any conflict of interest.

**Ethics approval** This article does not include any material that is already published elsewhere, so any formal approval is not required for the material used in this article. This is the authors' original work and found 8% plagiarism as Turnitin before the submission, which is within allowable limits.

**Consent for publication** Closed access publication.

**References**

Afzal S, Ahmad I, Younas M et al (2000) Study of water quality of Hudiara drain, India–Pakistan. Environ Int 26:87–96. https://doi.org/10.1016/S0160-4120(00)00086-6

Ahmad N, Ahmad M, Rafiq M, Iqbal N, Ali M, Sajjad MI (2002) Hydrological modeling of the Lahore aquifer, using isotopic, chemical and numerical techniques. Sci Vis 7(3–4):169–194

Al-Ahmadi ME (2013) Hydrochemical characterization of groundwater in wadi Sayyah, Western Saudi Arabia. Appl Water Sci 3:721–732. https://doi.org/10.1007/s13201-013-0118-x

Al-hadithi M (2012) Application of water quality index to assess suitability of groundwater quality for drinking purposes in Ratmao-Patrih Rao watershed, Haridwar District, India. Am J Sci Ind Res 3:396–402. https://doi.org/10.5251/ajisir.2012.3.6.395,402

Amiri V, Rezaei M, Sohrabi N (2014) Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. Environ Earth Sci 72:3479–3490. https://doi.org/10.1007/s12665-014-2355-0

Anderson TW (1958) An introduction to multivariate statistical analysis

Anselin L (1995) Local indicators of spatial association–LISA. Geogr Anal 27:93–115

APHA (1995) Standard methods for the examination of water and wastewater, 19th edn. American Public Health Association, New York

Azizullah A, Khattak MNK, Richter P, Häder DP (2011) Water pollution in Pakistan and its impact on public health—a review. Environ Int 37:479–497. https://doi.org/10.1016/j.envint.2010.10.007

Bairu A, Bureau WR, Tafesse NT (2020) Rainfall erosivity in southeastern Nigeria. Ethiop J Environ Stud Manag. https://doi.org/10.4314/ejesm.v5i2.12

Balakrishnan P (2011) Groundwater quality mapping using geographic information system (GIS): a case study of Guhargha City, Karnataka, India. Afr J Environ Sci Technol 5:1069–1084. https://doi.org/10.5897/ajest11.134
study in Gopalganj, Bangladesh. Exposure Health 10:167–188. https://doi.org/10.1007/s12403-017-0253-y

Rasool A, Xiao T, Baig ZT et al (2015) Co-occurrence of arsenic and fluoride in the groundwater of Punjab, Pakistan: source discrimination and health risk assessment. Environ Sci Pollut Res 22:19729–19746. https://doi.org/10.1007/s11356-015-5159-2

Ravikumar P, Somashekar RK, Angami M (2011) Hydrochemistry and evaluation of groundwater suitability for irrigation and drinking purposes in the Markandeya River basin, Belgam District, Karnataka State, India. Environ Monit Assess 173:459–487. https://doi.org/10.1007/s10661-010-1399-2

Sadat-Noori SM, Ebrahimi K, Liaghat AM (2014) Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer. Iran Environ Earth Sci 71:3827–3843. https://doi.org/10.1007/s12665-013-2770-8

Saeedi M, Abessi O, Sharifi F, Meraji H (2010) Development of groundwater quality index. Environ Monit Assess 163:327–335. https://doi.org/10.1007/s10661-009-0837-5

Selvam S, Venkatraman S, Chung SY, Singaraja C (2016) Identification of groundwater contamination sources in Dindugal district of Tamil Nadu, India using GIS and multivariate statistical analyses. Arab J Geosci. https://doi.org/10.1007/s12517-016-2417-7

Shabbir R, Ahmad SS (2015) Use of geographic information system and water quality index to assess groundwater quality in Rawalpindi and Islamabad. Arab J Sci Eng 40:2033–2047. https://doi.org/10.1007/s13369-015-1697-7

Shyu GS, Cheng BY, Chiang CT et al (2011) Applying factor analysis combined with kriging and information entropy theory for mapping and evaluating the stability of groundwater quality variation in Taiwan. Int J Environ Res Public Health 8:1084–1109. https://doi.org/10.3390/ijerph80410884

Singh AK, Mondal GC, Kumar S et al (2008) Major ion chemistry, weathering processes and water quality assessment in upper catchment of Damodar River basin, India. Environ Geol 54:745–758. https://doi.org/10.1007/s00254-007-0860-1

Srinivasamoorthy K, Chidambaram S, Prasanna MV et al (2008) Identification of major sources controlling groundwater chemistry from a hard rock terrain—a case study from Mettur taluk, Salem district, Tamil Nadu, India. J Earth Syst Sci 117:49–58. https://doi.org/10.1007/s12040-008-0012-3

Strassberg G (2005) A geographic data model for groundwater systems. Philosophy

Subba Rao N (2002) Geochemistry of groundwater in parts of Guntur district, Andhra Pradesh, India. Environ Geol 41:552–562. https://doi.org/10.1007/s002540100431

Thapa R, Gupta S, Guin S, Kaur H (2017) Assessment of groundwater potential zones using multi-influencing factor (MIF) and GIS: a case study from Birbhum district, West Bengal. Appl Water Sci 7:4117–4131. https://doi.org/10.1007/s13201-017-0571-z

Tiwari AK, Singh AK, Singh AK, Singh MP (2017) Hydrogeochemical analysis and evaluation of surface water quality of Pratappargarh district, Uttar Pradesh, India. Appl Water Sci 7:1609–1623. https://doi.org/10.1007/s13201-015-0313-z

Tiwari AK, Singh AK, Mahato MK (2018) Assessment of groundwater quality of Pratappargarh district in India for suitability of drinking purpose using water quality index (WQI) and GIS technique. Sustain Water Resour Manag 4:601–616. https://doi.org/10.1007/s40899-017-0144-1

Vasanthavigar M, Srinivasamoorthy K, Vijayaragavan K et al (2010) Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environ Monit Assess 171:595–609. https://doi.org/10.1007/s10661-009-1302-1

WHO (2004) Guidelines for drinking water quality: training pack. WHO, Geneva

Wishrat J (1928) The generalized product moment distribution in samples from a normal multivariate population. Biometrika 20(1/2):32–52. https://doi.org/10.2307/2331939

Yan H, Zou Z (2014) Water quality evaluation based on entropy coefficient and blind number theory measure model. J Netw 9(7):1868–1874. https://doi.org/10.4304/jnw.9.7.1868-1874

Yaseen M, Salik M, Khan A et al (2003) Studies on heavy metals status and their uptake by vegetables in adjoining areas of Hudiara drain in Lahore. Plant Nutrition View project digitization of ground water quality in Punjab View project Studies on heavy metals status and their uptake by vegetab. Soil Environ 28:7–12

Zhang Z, Yu D, Shi X et al (2015) Priority selection rating of sampling density and interpolation method for detecting the spatial variability of soil organic carbon in China. Environ Earth Sci 73:2287–2297. https://doi.org/10.1007/s12665-014-3580-3

Zhang Z, Zhou Y, Wang S, Huang X (2018) Spatial distribution of stony desertification and key influencing factors on different sampling scales in small karstwatersheds. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph15040743

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.