



MLR index–based principal component analysis to investigate and monitor probable sources of groundwater pollution and quality in coastal areas: a case study in East India

Chinmoy Ranjan Das · Subhasish Das · Souvik Panda

Received: 12 February 2023 / Accepted: 30 August 2023 / Published online: 6 September 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract Identifying groundwater contamination sources and supervising groundwater quality conditions are urgently needed to protect the groundwater resources of coastal areas like Contai of India, as communities here are heavily relying on groundwater which deteriorates progressively. So current research aims to address in detail about origins and influencing factors of groundwater contamination, status, and monitoring water quality by employing extremely useful leading technologies like principal component and factor analyses (PCA/FA), groundwater quality index (G_{WQI}), and multiple linear regression (MLR) that helps to simplify complicated works instead of the conventional methods. Eight groundwater quality parameters were evaluated here, such as pH, TH (total hardness), Tur (turbidity), EC (electrical conductivity), TDS (total dissolved solids), Mn (manganese), Fe (iron), and Cl (chloride) for 38 sites. Three principal components with ~81% of the total variance were extracted from the PCA/FA analysis. The origin

of maximum loadings of each factor is identified as a result of saline water, disintegration and leaching process, organic or else biogenic activities, and lithogenic or otherwise non-lithogenic links through percolating water. G_{WQI} results show that ~87% of the samples fall into the good category and ~13% of the samples fall into the poor to very poor category. A model consisting of Tur, Fe, EC, Mn, TH, and Cl as independent parameters is more feasible and is proposed to predict G_{WQI} obtained from MLR analysis. This MLR model also suggests that turbidity with the highest beta coefficient (0.820) is a key contributor relative to the entire groundwater class in this affected area. The findings relating to this research may support the designer and officials in monitoring and protecting coastal groundwater resources like selected areas.

Keywords Principal component analysis (PCA) · Factor analysis (FA) · Groundwater quality index · Multiple linear regression (MLR) · SPSS

C. R. Das · S. Das (✉)

School of Water Resources Engineering, Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: subhasish.das@jadavpuruniversity.in

C. R. Das

Civil Engineering Department, Global Institute of Science & Technology, Purba Medinipur, Haldia 721657, West Bengal, India

S. Panda

Ambuja Cement Ltd, Kolkata 700019, West Bengal, India

Introduction

The basic origin of water for large amounts of the populace all over the globe is groundwater, and that affects the stable socio-economic growth of each society. In the last few decades, the stability of this essential resource has been a warning sign in several segments of the world due to its overexploitation and

increasing demand (Abulibdeh et al., 2021; Sahour et al., 2020). Since about 60% population of the world is present in shoreside regions, the fast-growing population related to these regions leads to too much utilization of groundwater resources to fill the progressive water requirements (Motevalli et al., 2017). Groundwater quality in the coastal zone has been degraded due to seawater intrusion (Arslan, 2013; Arslan & Demir, 2013; Heydarirad et al., 2019) and several other factors such as man-made activities, disintegration of material in aquifers, pollution and salinity from irrigation, and agricultural operations (Ferchichi et al., 2018; Jayathunga et al., 2020). Identification of various pollution sources of groundwater is a basic need because extreme squeezing and different pollution sources of groundwater badly affect its quality related to drinking (Li et al., 2021). Routine-wise quality monitoring of groundwater is mandatory to take care of its quality for drinking. However, examining the several water quality variables is expensive, takes too much time, and is a monotonous exercise. Even assessing all the variables in consistent intervals is nonessential as it will not offer additional information about water quality features (Gulgundi & Shetty, 2018). Troudi et al. (2020) mentioned that monitoring groundwater quality conditions and finding multiple sources of pollution and the origin of influencing parameters related to groundwater contamination are initial requirements to find solutions to groundwater quality problems. An accurate and planned evaluation of the state of groundwater and, in addition, a correct forecast of groundwater quality are necessary to determine the optimal action plan for regional water resources management (Gholami et al., 2016).

Currently, several statistical techniques have been introduced to accurately examine and explain data due to the increase in the number of parameters for groundwater quality analysis (Taşan et al., 2022), considering that water class is commonly expressed in terms of many water quality parameters. PCA can be used to regroup complicated multivariate variables into a nominal and a reasonable number of components without compromising detail (Banda & Kumarasamy, 2020). PCA is an extremely used approach in data research studies and provides a true explanation of multicomponent estimates that allow a better perception of the configuration of groundwater classes (Tripathi & Singal, 2019). Conventional methods for explaining groundwater quality data using the usual

charts and graphs may not simultaneously express the uniformities between variables or samples. To overcome this incompleteness, the process of FA was introduced for efficient analysis to identify such similarities between samples or parameters (Patil et al., 2019). There is some research related to PCA/FA that is used to determine the origin of contamination that is prone to degrade groundwater quality. The PCA/FA approach was effectively applied by Bouteraa et al. (2019) and Nguyen et al. (2020) to identify sources of contamination and influencing parameters related to groundwater quality.

The main issue associated with groundwater is that when it is contaminated, it is very difficult to revive the groundwater class (Chen et al., 2018). A conventional approach, such as the analysis of individual parameters, is not immediately understandable, because each parameter has a different quality class, which makes it very difficult to clarify the results when more parameters are involved in the water quality assessment (Solangi et al., 2019). Accurate assessment of the variety and extent of water pollution is very difficult and complex. The main problem related to monitoring the degree of water quality is the complexity of examining many variables. For this reason, the use of the WQI approach is considered a very effective mechanism (Valentini et al., 2021). WQI helps to convert huge water quality data into an index score that directly and rationally tells about water quality (Awachat & Salkar, 2017). The WQI method is effectively employed to assess groundwater quality for drinking. Water quality sense and WQI assessment are essential to water quality management and control (Mahapatra et al., 2012). A model can be prepared with respect to WQI using a statistical method such as MLR, considering several variables that actually affect the water quality of a particular source. The inclusion of fewer variables in the WQI provides cost-effectiveness that can be used to evaluate WQI in a given region in the future (Valentini et al., 2021). If two or more parameters need to be considered simultaneously, the MLR approach should be used to find their interrelationship. It also plays an important role in identifying the variable that has the maximum impact on WQI. MLR model is applied to predict water quality for monitoring purposes (Wu et al., 2020).

Some progressive research studies are mentioned herein related to identifying pollution sources,

important parameters related to water quality, and forecasting water quality using different effective approaches. A few years ago, Zhang et al. (2020) and Li et al. (2021) identified the sources of groundwater pollution through positive matrix factorization (PMF) and PCA-based absolute principal component score multiple linear regression (APCS-MLR) models in their respective studies. Later, Mu et al. (2023) investigated the potential pollutants along with their variation in the Malian River through the PCA-APCS-MLR model. Hagnazar et al. (2022) compared the PMF and PCA-MLR receptor models and observed the main contributors associated with groundwater pollution. All the above studies effectively recognized the pollution sources and their impact on water quality in the respective study zone. Singha et al. (2021) acquired 91.7% accuracy in the prediction of WQI through the artificial neural network model approach. Alam and Singh (2023) measured the quality of groundwater by conducting statistical software-based multivariate statistics along with WQI and identified major variables that affected the water quality.

The communities have too much trust in groundwater sources within the Contai area in India for drinking and miscellaneous activities, but its quality has been degraded day after day (Halder et al., 2021; Maity et al., 2017, 2018). After examining 6 years of groundwater quality data, Chakraborty et al. (2020) stated that the concentration of groundwater variables such as pH, TH, Tur, EC, TDS, Mn, Fe, and Cl consistently increases within the Contai area. It makes groundwater in that zone unfit for drinking and domestic use. Recently, Das et al. (2022) observed an interrelationship between selected parameters using WQI, cluster, and regression analyses for the Dulalpur panchayat in Contai. Halder et al. (2021) declared that groundwater quality monitoring immediately requires for managing the vital resources here.

This research, therefore, focuses on water quality issues and the management of essential water resources in coastal areas like Contai by applying statistical approaches together with the groundwater quality index (G_{WQI}), although the concentration of several parameters is constantly increasing, which affects water quality. But to date, details of groundwater pollution sources, the most influential parameters, and the correct index model for water quality monitoring have not yet been identified for water resource management in this coastal region. Thus, it

is necessary to identify the probable sources of pollution and the degree of water quality and also to use feasible techniques to monitor the groundwater quality for managing the current water resources for the existing communities. The purpose of this analysis is therefore to determine in detail the likely sources of groundwater contamination, to identify the most influencing parameters for a specific location using PCA/FA, and to identify drinking groundwater quality using G_{WQI} . This study also aims to create new mathematical models for G_{WQI} using MLR analysis, to identify the influencing parameters related to water quality and assess the appropriate model for monitoring groundwater levels in this coastal zone.

Research area

Contai I is formally known as Kanthi I block. It is a municipal development block building directorate division in the Contai sub-division of Purba Medinipur district in West Bengal state of India. It is located at 21.75° N and 87.65° E with an elevation of 3 m (~10 ft). The total area of this block is 155.27 km². As reported by the census (2011), the population in this block is about 170,894. Regarding the availability of water quality data, Badalpur panchayat (B_p) and Raipur Pashimbarh panchayat (R_p) within this block were used for the study (Fig. 1). Selected area is abbreviated as BPRP area.

It is essentially a coastal area composed of systematic layers formed by the gradual deposition of mud, boulders, gravel, and cobblestones over many years following the continuation of rivers that continuously deposit alluvium. Soil salinity is high in the coastal region consisting of less than 60% clay soil. Agricultural zones are usually the outer side of the main town. The main food crop is paddy rice, which is grown here. In relation to the highest production of paddy rice, the selected district is considered second in the concerned state.

The land is extremely productive, and therefore, agriculture is also a key driver that sustains the economy of middle-class families in the area. However, sometimes agriculture is affected by recurring floods due to relentless monsoon rains and cyclones caused by depressions in the Bay of Bengal. Currently, salt-water intrusion is a threat to residents. People here are highly dependent on groundwater resources to meet

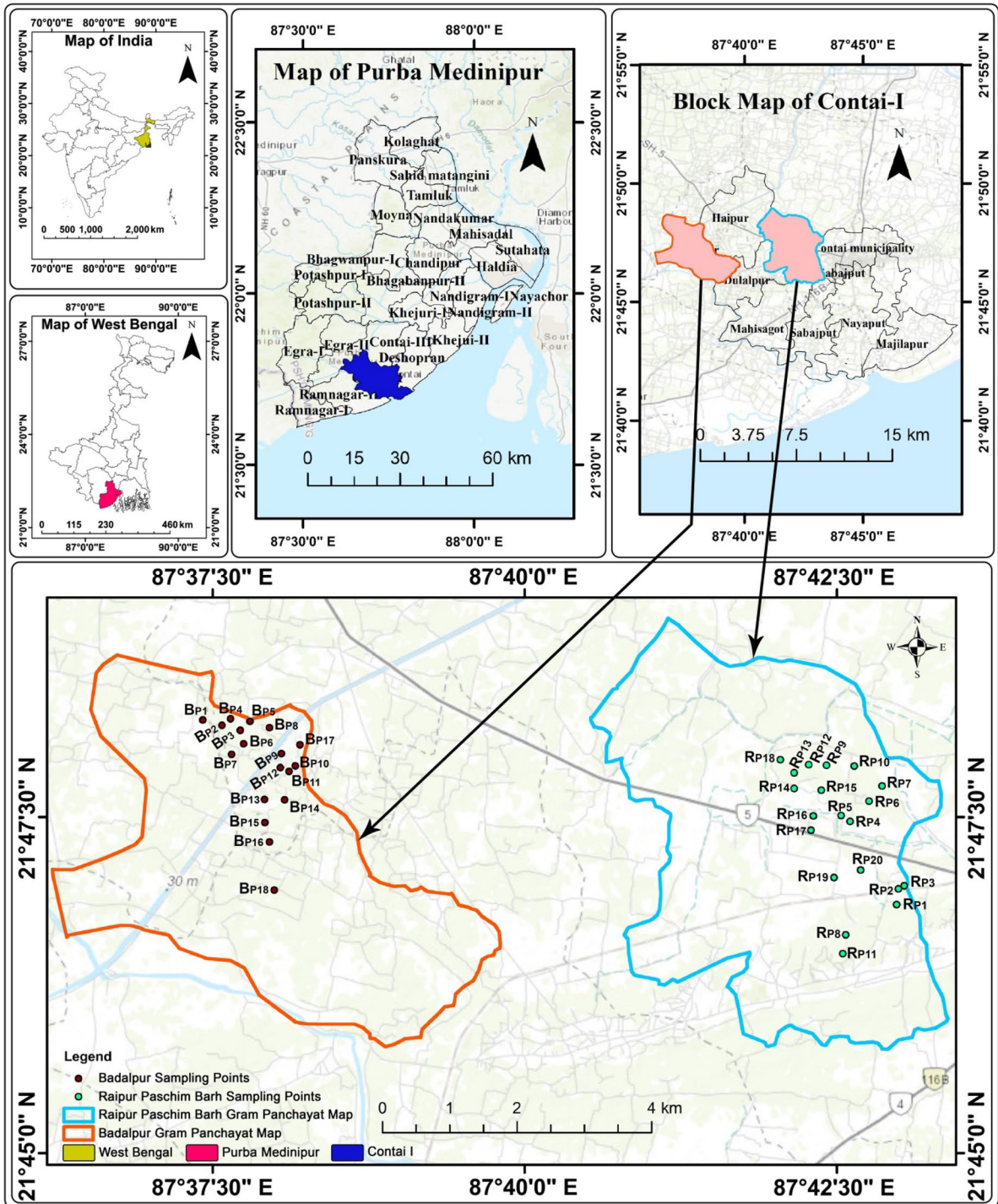


Fig. 1 Locus plot and sampling points of the research area

their essential requirements because of less rainfall and insufficient surface water resources. The soil here

is mainly of fresh alluvial type. The rivers that flow in the area are not enough to handle the necessary

water demand. The demand for water is gradually increasing every day due to urbanization, increase in population density, and industrial development. To offset these demands, groundwater is over-extracted, and for this reason, there is an internal movement of salt water into the region. Figure 1 describes the locus graph of the research area and the location of the sample collection. In notation, B_{pi} , $i=1$ to 18, and R_{pj} , $j=1$ to 20, indicate the number of sampling points corresponding to the respective panchayat.

Methodology

Sampling, analysis, data accumulation, and parameter selection

During August 2019, 38 groundwater samples were accrued from five different Mouzas of Badalpur panchayat ($B_p=18$) and Raipur Paschimbarh Panchayat ($R_p=20$) by Contai division specialists (CDS) of Public Health Engineering Directorate (PHED). Samples were taken at an approximate depth of 52 m (~170 ft) and 27 m (~90 ft) for Badalpur and Raipur Paschimbarh panchayats, respectively. CDS followed the APHA (2017) standard approach for sample accumulation, preservation, and analysis. The sampling method is that first a plastic bottle was taken and refined with purified water. Then, the sample was taken into this bottle from the tubewell after 10 min of pumping and then tightly closed with an inner and outer cap to prevent the passage of air and after properly labeling the dispatch to the laboratory. Cooling to 4–6 °C was performed for each sample to prevent microbiological degradation of solids. The samples at various locations were checked in detail for several water quality variables like pH, TH, Tur, EC, TDS, Mn, Fe, and Cl in the quality testing lab at Manoharchak in CDS PHED. pH and Tur were measured by the electrometric method. EC, TDS, TH, Cl, Fe, and Mn were measured by electromagnetic method, gravimetric method, EDTA titrimetric method, and argentometric method, using a spectrophotometer. Proper quality assurance (QA) and quality control (QC) of the samples were believed to provide the best satisfaction of the data set from the analytical procedures. The standard operating procedures were used for the day-to-day operation of any QA programme and for QC purposes. Special attention was paid to

sample collection and preservation, reagent standardization, equipment calibration, and blanks, which are designated as standard water approaches (APHA, 2017). All water quality datasets related to this study were collected from CDS PHED. The 38 samples and aforesaid parameters have been studied here as per existing information of the dataset from PHED.

All aforementioned eight parameters mentioned in IS 10500 (2012) and WHO (2011) are important in water quality assessment because their concentration above the permissible limit can significantly affect the quality of drinking water. The study area is located in the coastal zone and is subject to sea encroachment as identified by Chakraborty et al. (2020). EC, TDS, and Cl are major parameters as there are several soluble in seawater (Liu et al., 2003). Wang et al. (2020a) suggested that seawater intrusion influences Fe levels in groundwater. Gibrilla et al. (2011), Gulgundi and Shetty (2018), Wu et al. (2020), and Ram et al. (2021) considered the maximum of all, and Das et al. (2022) included all eight parameters mentioned in their study. Therefore, based on the codal perspective, literature review, and database availability, the eight parameters presented here are included for groundwater quality assessment as they are often reported by relevant researchers.

Normal statistical analysis

The collected data on groundwater quality parameters were statistically studied using the Statistical Package for the Social Sciences (SPSS, version 26). Initially, data files were prepared using Microsoft Excel media and then moved to the SPSS software package to determine normal statistics of quality parameters. It was performed to provide background information regarding water class data (Papaioannou et al., 2010). The functions used in this statistical approach are (a) minimum (min.), (b) maximum (max.), (c) mean, (d) standard deviation (Std. dev.), and (e) variance of the collected concentrations of groundwater quality variables.

Principal component/factor analysis (PCA/FA)

PCA/FA was used to share sources of contamination and also to select vital water class variables associated with these sources. PCA was performed on the standardized data set to remove key principal components,

and these components were then subjected to varimax rotation to create varifactors (VFAs). Initially, the relevance of PCA was examined using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity. Both checks were used to successively verify the sample efficiency and freedom of each parameter (Yang et al., 2020). Here, PCA was performed on the data set after the Z-score standardization of all parameters to minimize the impact of varieties on the dimensions used for quantification and variation and to provide a smaller unit of data (Gibrilla et al., 2011). Z-score normalization using MS Excel was performed with respect to Eq. (1).

$$Z = \frac{c - \bar{c}}{SD} \quad (1)$$

where, c , \bar{c} , and SD are the concentration, mean concentration, and Std. dev. of the variables.

In the PCA technique, the eigenvalues are generally used to find out the principal components (Zeinalzadeh & Rezaei, 2017). Principal components (PCs) with eigenvalues exceeding 1.0 were considered (Jayathunga et al., 2020). Eigenvalues imply the importance of the principal component, so a component with a larger eigenvalue is considered highly influential (Li et al., 2019). The first PC is considered for extremely impressive variance in the database, then the second component, and the rest. It is common to alternate components to achieve an optimal spread of variance in different components (Mahapatra et al., 2012). Varimax rotation was used to obtain an optimal distribution of variance in the various components. Varimax rotation can minimize the impact of nonsignificant variables in groundwater quality analysis because it naturalizes the loadings by accurately rotating the component axes to clarify the results (Islam et al., 2018). Here, a varimax rotation accompanied by Kaiser normalization was performed to find the rotated loadings of the various factors, and sources of contamination regarding the presentation of factor loadings were identified. A specific variable is how strongly it is associated with different factors as reflected by a rotated matrix of factor loadings (Li et al., 2019). Factor scores (F_1 , F_2 , and F_3) indicate the contribution of each factor for each sample location point (Narany et al., 2014), which are computed here, and it also helps to identify the more active variables for a specific sample site (Taşan et al., 2022).

Equation (2) is used to find the principal component (Li et al., 2019). Here, PCA and FA were performed through the SPSS suite of groundwater quality data sets of 38 sampling sites.

$$P_{IJ} = A_{I1}X_{1J} + A_{I2}X_{2J} + A_{I3}X_{3J} + \dots + A_{IM}X_{MJ} \quad (2)$$

where A , P , X , I , J , and M denote the component loading, component score, measured parameter value, number of components, number of samples, and the total number of parameters, respectively.

Groundwater quality index (GWQI) estimation

The WQI approach examines the respective contributions of water quality variables to the overall groundwater health risk and seeks to translate complex water quality data sets into a value that is commonly acceptable and usable in general (Asare et al., 2021). Here, G_{WQI} was calculated at different sampling sites by including variables such as pH, TH, Tur, EC, TDS, Mn, Fe, and Cl. Four consecutive phases were used to calculate the G_{WQI} .

The first step is to assign a weight (wt_i) to the selected eight water quality variables related to their respective importance for drinking water quality. A weight of five was assigned to maximum detrimental variables and a weight of one to nominal detrimental variables. The highest weight of five was assigned to TDS; weight of four was assigned to pH, Fe, Mn, Cl, and EC; weight of three was assigned to Tur; and weight of two was assigned to TH (Aminyan et al., 2018; Boateng et al., 2016). In the next step, by adopting Eq. (3), relative weight (RWT_{*i*}) was calculated.

$$RWT_i = wt_i / \sum_{i=1}^n wt_i \quad (3)$$

The quality rating (QR_{*i*}) was calculated using Eq. (4), where m_i and ST_i are the measured concentrations and prescribed Standard values related to each variable. The assigned weight, relative weight, and standard value related to each variable are shown in Table 1.

$$QR_i = \left(\frac{m_i}{ST_i} \right) \times 100 \quad (4)$$

In the last stage, the G_{WQI} was calculated using Eq. (5).

Table 1 Weight, relative weight, and the standard value of each parameter

Sl no	Parameters	Weight (wt _i)	Relative weight (RWT _i)	Standard value
1	pH	4	0.133	6.5–8.5 ^a
2	Tur (NTU)	3	0.100	5 ^a
3	TH (mg/l)	2	0.067	200 ^a
4	TDS (mg/l)	5	0.167	500 ^a
5	EC (μS/cm)	4	0.133	1500 ^b
6	Fe (mg/l)	4	0.133	0.3 ^a
7	Mn (mg/l)	4	0.133	0.1 ^a
8	Cl (mg/l)	4	0.133	250 ^a
		Σ=30	Σ=1	

^aIS 10500 (2012)

^bWHO (2011)

$$G_{WQI} = \sum_{i=1}^n RWT_i \times QR_i \tag{5}$$

The class of water according to the index score can be designated as class “A” or excellent ($G_{WQI} < 50$), class “B” or good (G_{WQI} lies between 51 and 100), class “C” or poor (G_{WQI} lies between 101 and 200), class “D” or very poor (G_{WQI} lies between 201 and 300), and class “E” or unacceptable ($G_{WQI} > 300$) for drinking (Haghnazar et al., 2022).

Multiple linear regression analysis

In this work, the variables used for the MLR study are the same as the variables used for G_{WQI} estimation. On a random sampling basis, 30 samples (80%) and eight samples (20%) of the total were taken for model preparation and model validation here. This statistical

tool is used to understand the relationship between a dependent parameter and various independent parameters (Wu et al., 2020). It is expressed by Eq. (6).

$$R = B_0 + B_1P_1 + B_2P_2 + \dots + B_M P_M + R_E \tag{6}$$

where R is the response parameter, $P_1 \dots P_M$ are the predictor parameters, $B_0 \dots B_M$ are the regression coefficients, and R_E is the random error. Here, step-wise MLR was performed to assess the association between groundwater quality parameters as independent variables and G_{WQI} as a dependent variable using IBM SPSS to obtain a favorable model. After finding the suitable model, index scores were calculated. Lastly, the significant variations were determined between the results of the original G_{WQI} and the new G_{WQI} using a paired samples t -test.

Results and discussion

Exploration of water quality through normal statistics of variables

The normal statistics of the individual values (Table 2) of water class variables sampled in this research area are shown here. The changes in the water quality variables for the BPRP area are illustrated by box plots (Fig. 2a–h). Basic statistics were calculated for the entire groundwater datasets to gain an overall view and to recognize dataset variations (Patil et al., 2019). Groundwater quality variables are considered the most essential basis for pointing out the character, class, and diversity of groundwater. According to IS 10500 (2012), the permissible pH of drinking water can be in the range of 6.5–8.5. The pH value measures the balance between the

Table 2 Normal statistics of groundwater quality variables (total observation $N=38$)

Sl no	Variables	Min	Max	Mean	Std. dev	Variance
1	pH	6.80	7.90	7.24	0.35	0.12
2	Tur (NTU)	1.34	64.90	8.72	16.07	258.26
3	TH (mg/l)	172.00	328.00	257.32	39.08	1527.57
4	TDS (mg/l)	319.00	735.00	397.95	135.57	18,378.05
5	EC (μS/cm)	534.00	1223.00	669.21	227.91	51,943.79
6	Fe (mg/l)	0.10	2.20	0.45	0.37	0.14
7	Mn (mg/l)	0.0020	0.0800	0.0266	0.0210	0.0004
8	Cl (mg/l)	11.00	139.00	56.34	27.87	776.99

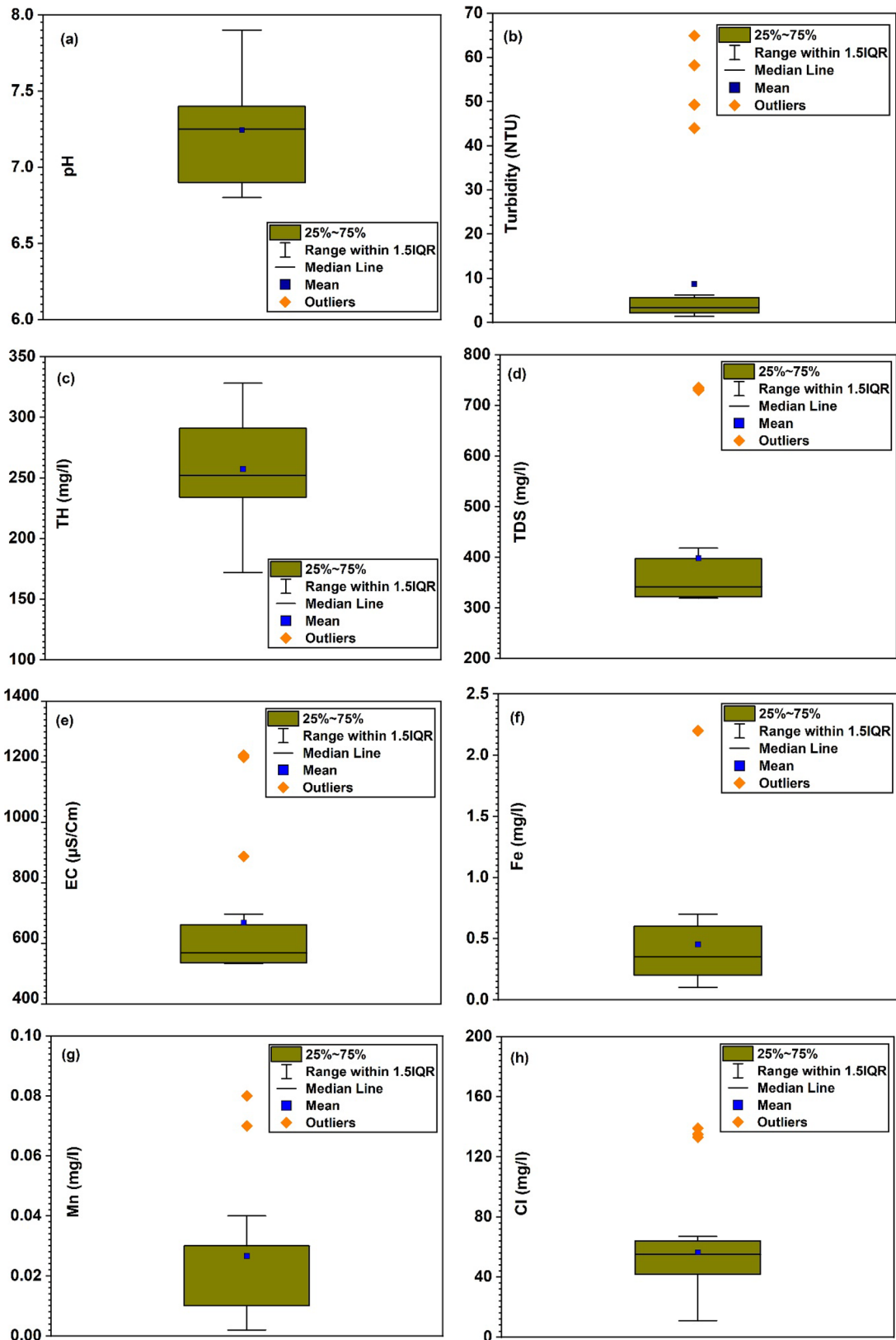


Fig. 2 Concentration representation of groundwater quality parameters such as **a** pH, **b** turbidity, **c** TH, **d** TDS, **e** EC, **f** Fe, **g** Mn, and **h** Cl through box plots

concentration of hydrogen ions [H^+] and the concentration of hydroxyl ions [OH^-] in water and indicates the basicity, or acidity, of a mixture. The groundwater pH of the BPRP area ranges from 6.80–7.90 with an average concentration of 7.24, indicating a marginally acidic to alkaline type. Tur concentration ranges from 1.34–64.90 NTU with an average concentration of 8.72 NTU, which is above the permissible value prescribed by IS 10500 (2012). Out of 38 samples, 11 samples (29%) have Tur concentration above the permitted limit. Here, the lower, higher, and mean values of TH of groundwater are 172, 328, and 257.32 mg/l, respectively. TH value (mg/l) between 150 and 300 and exceeding 300 indicates that the water is hard and very hard, respectively (Todd & Mays, 2005). In the BPRP area, the groundwater is classified as hard to very hard. Most of the sample has a TH value that exceeds the permissible limit (200 mg/l) specified by IS 10500 (2012) for drinking. Only two sample points B_{P6} and B_{P15} have an acceptable TH value for drinking.

The minimal, maximal, and average concentrations of TDS in groundwater in the BPRP area are 319, 735, and 397.95 mg/l, respectively. The maximum sampling point has a TDS value within the standard limits, and only sampling locations R_{P1} , R_{P7} , R_{P12} , R_{P13} , and R_{P20} exceed the permissible TDS for drinking. Groundwater at all sites is categorized as fresh because the TDS concentration at each site is less than 1000 mg/l (Todd & Mays, 2005). The least, highest, and mean levels of electrical conductivity (EC) in the BPRP area are 534, 1223, and 669.21 $\mu S/cm$, respectively. Results indicate that saltwater intrusion is less due to EC concentration below 1500 $\mu S/cm$ throughout the sites (Das et al., 2022). Groundwater samples with EC concentration are below the recommended drinking value in the BPRP area. Fe is a very crucial element that affects human health, and surplus or deficit of Fe ingestion can create several diseases (Wang et al., 2020a). The Fe range is between 0.10 and 2.20 mg/l with an average concentration of 0.45 mg/l in the BPRP area. Fifty percent of the samples exceeded the permissible limit of Fe for drinking given by IS 10500 (2012). Manganese appears in groundwater naturally, chiefly in anaerobic

conditions. Its concentration depends on the chemistry of rainfall, the lithologic condition of the aquifer, the flow courses of groundwater, etc. (Ram et al., 2021).

The concentration of Mn varies between 0.002 and 0.08 mg/l with a mean of 0.026 mg/l, and the concentration of Mn in each sampling point is within the acceptable level for drinking. Chloride levels range from 11–139 mg/l with an average concentration of 56.34 mg/l, and the Cl concentration at each sampling point is below the drinking limit.

Spotting the pollution sources through PCA/FA

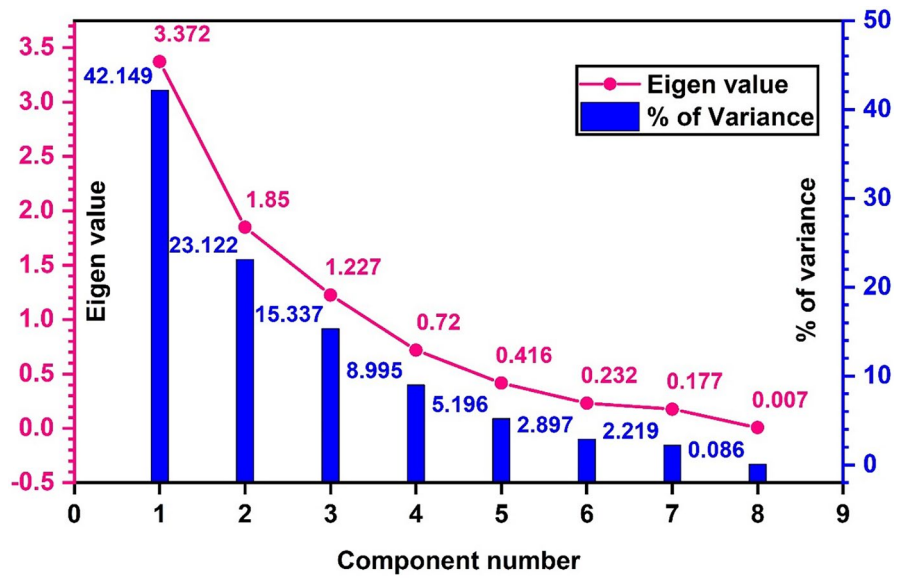
A KMO value above 0.5 and Bartlett's test value below 0.05 from water quality data are considered suitable for PCA (Yang et al., 2020). If the KMO value of any data set is found below 0.5, lies 0.5–0.7, and is above 0.7, the data set is considered unsatisfactory, adequate, and good for PCA (Ustaoglu et al., 2020). In this work, the KMO value (0.636) indicates that the data set is acceptable for PCA, and the parameters are significantly related to the significance level (0.00) obtained from Bartlett's test.

Eigenvalues for eight water quality parameters and their variances were calculated using SPSS. Based on PCA, the first three components have eigenvalues greater than one, and the other components have eigenvalues less than one. So the first three components (eigenvalues greater than 1) were considered, and the corresponding variance is shown in Fig. 3 using a scree plot which addresses the change in the eigenvalue curve.

It shows that the eigenvalues corresponding to components 1, 2, and 3 are above 1 too evidently. The inflection point of the curve (Fig. 3) appears at the third component, so it is convenient to select the first three components that are considered to be the principal components, and clearly, the three counts of the principal components (PC_1 , PC_2 , and PC_3) were achieved.

In this study, the first three PCs have 80.61% cumulative variance of the total variance, which explains that these PCs may represent the real eight water quality class variables. The first (PC_1), second (PC_2), and third (PC_3) principal components have eigenvalues and corresponding contributions shown in Fig. 3. The interrelationship between PC and selected parameters is indicated by factor loadings,

Fig. 3 Scree plot (pink line) and percent of variance (blue bars) corresponding to each component



and loadings with the highest positive or negative value determine the highest contribution (Arslan, 2013). Absolute loading values are more than 0.75, between 0.75 and 0.5, and between 0.5 and 0.3 referred to as strong, medium, and weak factor loadings separately (Zhang et al., 2020).

Dutta et al. (2018) stated in their research that the minimum factor loading standard used in several research papers is different for determining the decisive parameters. Here, factor loadings above 0.6 in bold (Table 3) were chosen to explain the results because they are significant for assessing the components, and loadings with a negative and positive sign (Table 3) indicate the direction of the effect.

Thus, a high negative loading value indicates that the factor is significantly and negatively affected by the parameter.

The PCA/FA approach shows that three key sources of groundwater contamination were extracted in parallel with three varimax factors (VFAs) (Table 3), which together illustrate 80.61% of the entire variance of the eight groundwater quality variables. The first varimax factor (VFA₁) explains 34.35% of the total variation of the eight groundwater quality parameters and is considered the most important factor. The second vital varimax factor (VFA₂) explains 28.14%, and the third varimax factor (VFA₃) explains 18.12% of the total variance.

Table 3 Factor loading and factor score coefficient of parameters after varimax rotation

Variables	Factor loading			Factor score coefficient		
	VFA ₁	VFA ₂	VFA ₃	VFA ₁	VFA ₂	VFA ₃
pH	0.256	-0.032	0.839	0.031	-0.148	0.616
Tur	0.826	0.137	-0.201	0.335	0.016	-0.243
TH	0.688	0.506	0.176	0.212	0.164	0.007
TDS	0.055	0.972	0.119	-0.080	0.462	-0.039
EC	0.129	0.969	0.090	-0.047	0.457	-0.068
Fe	0.090	-0.265	-0.762	0.128	-0.042	-0.551
Mn	0.872	-0.092	0.123	0.346	-0.143	0.027
Cl	0.860	0.118	0.236	0.311	-0.052	0.087
Eigenvalues	2.748	2.251	1.450			
% of variance	34.351	28.138	18.119			
Cumulative %	34.351	62.489	80.608			

Factor loadings exceeding 0.6 are marked as bold

VFA₁ consists of positive strong loadings of 0.826, 0.872, and 0.860 and medium loadings of 0.688 for Tur, Mn, Cl, and TH. The VFA₂ has strong positive loadings on TDS (0.972) and EC (0.969). VFA₃ has strong positive and negative loadings with pH (0.839) and Fe (−0.762). According to Liu et al. (2003), the main elements in seawater are Cl, TDS, and EC. Salinity is a very widespread form of groundwater pollution, specifically in coastal aquifers, and is measured by the rise in TDS (Rao et al., 2013). Parameters such as Cl, TDS, TH, and EC with significant positive factor loading values are an indication of the amalgamation of saline water with fresh groundwater (Akshitha et al., 2021). The above variables can cause salinity in groundwater, and Cl is recognized as a signal of saline water intrusion into groundwater resources (Taşan et al., 2022). Significant positive TH and Tur loading factors suggest that the origin is attributable to rock breakdown and leaching (Boateng et al., 2016). Perhaps, strong pH loading is predicted to be catalyzed by organic or otherwise biogenic activities (Reghunath et al., 2002). The strong Fe factor loading is probably caused by the liquefaction of non-lithogenic or otherwise lithogenic references via percolated water (Patil et al., 2019). The origin of Mn in groundwater is likely due to the weathering of manganese minerals in aquifers and may be induced by industrial wastewater and landfill leachate (Zhang et al., 2020). Perhaps, Mn originating from mineral sources can be released by the chemical disintegration of the parent material (Bodrud-Doza et al.,

2016). Here, the varimax factor (VFA₁) is influenced by parameters such as Tur, Mn, Cl, and TH. Factor 2 (VFA₂) is affected by TDS and EC. Factor 3 (VFA₃) is affected by pH and Fe. So the water in VFA₁ can be affected by salt water along with the decay and leaching process. Water in factor 2 (VFA₂) is an immixture of seawater together with clean groundwater. Factor 3 (VFA₃) can be affected by a combination of organic biogenic activities and lithogenic or otherwise non-lithogenic sources through percolating water.

Factor score coefficient (Table 3) is also determined for the entire variables that express the interpretation of a certain factor in a given sampling location (Patil et al., 2019). The factor scores of the 38 sampling sites are shown in Fig. 4. Absolute positive or negative scores (greater than +1 or less than −1) on any component indicate that the location is largely affected or unaffected by the parameters influencing the component, while a score close to zero defines a likely location moderately affected by the chemical action of that specific factor (Senthilkumar et al., 2008).

Sampling points R_{P2}, R_{P6}, R_{P9}, and R_{P11} with factor scores 3.013, 3.192, 2.719, and 1.106 are mostly affected by variables such as Cl, Tur, Mn, and TH, because the above sampling points have factor scores more than +1.

Sampling points R_{P1}, R_{P7}, R_{P12}, R_{P13}, and R_{P20} with factor scores 2.239, 2.256, 2.193, 2.632, and 2.397 are affected by parameters such as EC and TDS. Sampling points R_{P3}, R_{P4}, R_{P5}, R_{P8}, R_{P10}, and

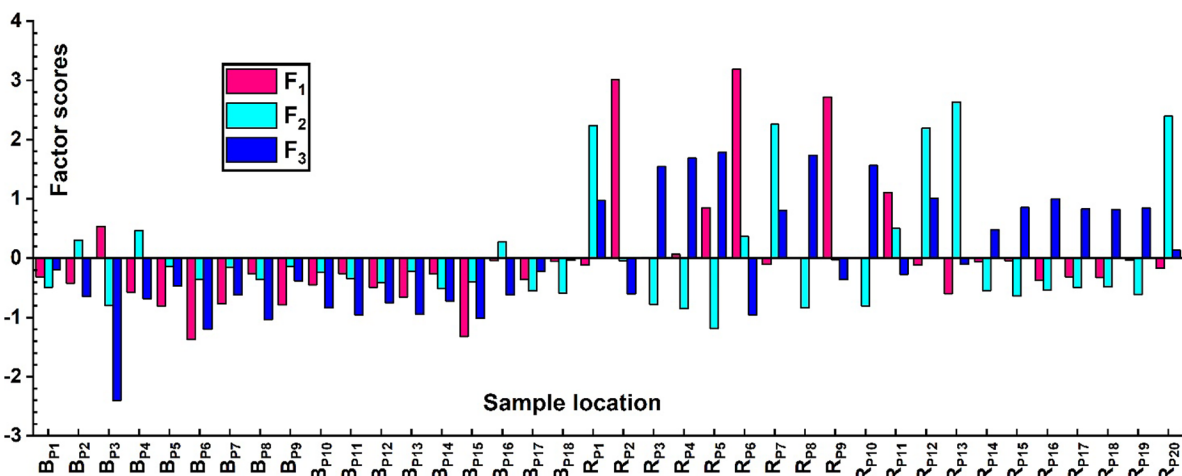


Fig. 4 Factor scores at different locations based on groundwater data

R_{P12} with factor scores 1.542, 1.687, 1.784, 1.733, 1.560, and 1.008 show that these sampling sites are most affected by pH and Fe. Biplots are more productive and instructive that can be used for graphical demonstrations of statistical analysis (Banda & Kumarasamy, 2020). Biplots were composed in three dimensions accompanied by the first three PCs as three axes. They ideally describe connections among parameters and PCs. Biplots point to a branch of highly correlated parameters using an estimate of the true multivariate space (Gradilla-Hernández et al., 2020). 3D biplots of this current study describing the association between extremely correlated parameters and the first three PCs are shown in Fig. 5.

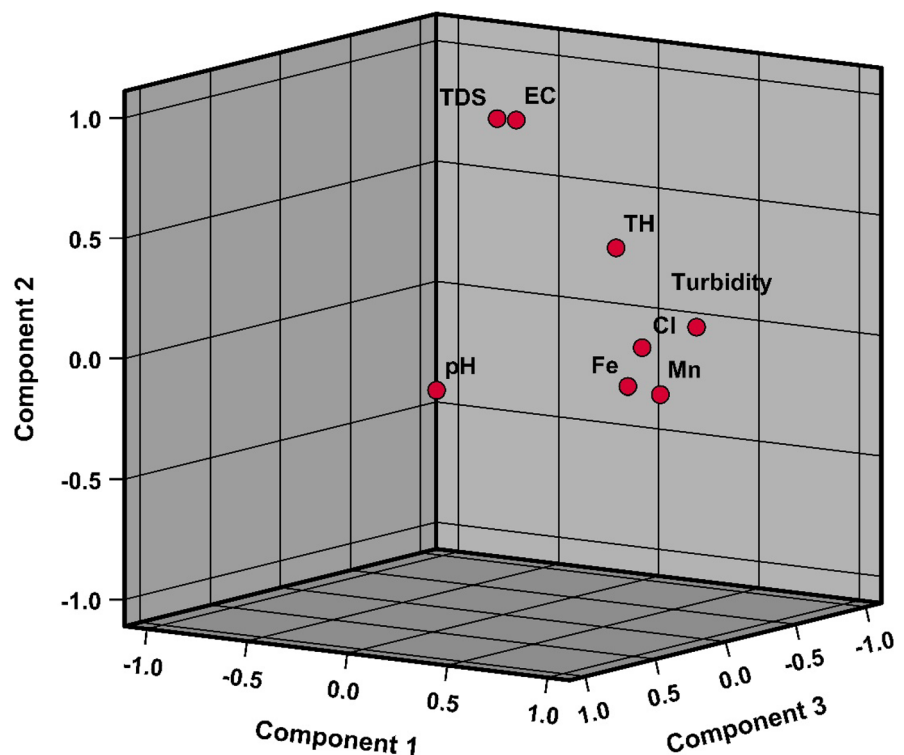
PCA/FA approach clearly indicates that the water quality in this selected zone is affected by salt water as mentioned in the earlier studies (Chakraborty et al., 2020; Halder et al., 2021; Maity et al., 2017, 2018) along with other several reasons like decay and leaching process, organic biogenic activities, and lithogenic and non-lithogenic sources from water percolation. This analysis also identifies the influencing parameters along with their probable sources in the selected sites. The groundwater quality can

be protected here by controlling the contamination sources. Some studies like Bouteraa et al. (2019), Patil et al. (2019), Nguyen et al. (2020), and Wu et al. (2020) investigated the potential water pollution sources through PCA/FA, and that confirms the precision and acceptance of this technique. So this approach provided priceless information by efficiently identifying the expected groundwater pollution sources including possible regulating parameters related to water quality for a specific site, and that helps the decision-maker restrain water pollution in that area.

Water quality rating by GWQI and identifying MLR-based new GWQI model

The water quality type was determined based on the G_{WQI} score. The measured G_{WQI} scores range from 52.33 to 204.84, and the water status related to the BPRP area is shown in Fig. 6. It is noted that 87%, 11%, and 2% of the total samples fall into the category of good, poor, and very poor condition for drinking, and none of the sampling sites fall into class E. In this study, a stepwise MLR approach

Fig. 5 3D biplot explaining the association between the extremely correlated parameters and the three PCs



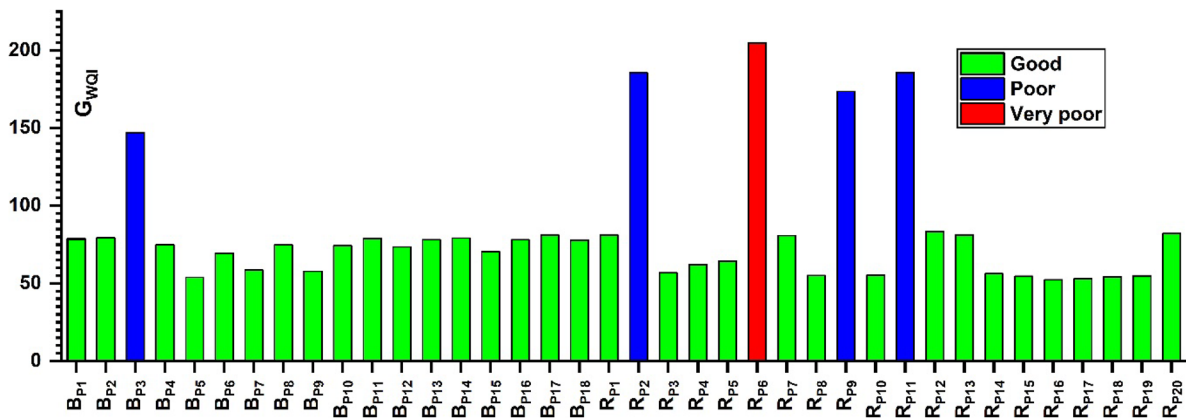


Fig. 6 Groundwater quality status at different locations according to G_{WQI}

was introduced to prepare new G_{WQI} equations. Eight models (Table 4) were obtained from this MLR study. The results of this analysis are shown in Tables 4 and 5. New G_{WQI} equations from this study are represented by Eqs. (7), (8), and (9) sequentially.

Three appropriate regression models Mod₆, Mod₇, and Mod₈ were considered for the unequal significance of the model variables (Table 5). Coefficients of determination (R^2) for MLR were also more promptly explained than coefficients of multiple correlations (R) as a scale of the level of relationship, because the multiple R^2 is equivalent to the portion of the total variance in the presence of parameter that is possibly attributable to the predictor parameter outcomes (Mondal et al., 2010).

The above three regression models Mod₆, Mod₇, and Mod₈ adjusted R^2 values equaling 1.0 (Table 4),

indicating that the overall significance level of these three models is high.

The model Mod₆ has six independent parameters that are significant (significance level < 0.05) (Mustapha et al., 2012) in describing the variation of water classes in the BPRP area. This regression model is expressed by Eq. (7). According to Mod₆, Tur has the highest standardized beta coefficient (0.820) among the variables measured using the stepwise MLR approach (Boateng et al., 2016; Uyanik & Güler, 2013), which means that Tur has the largest contribution to the entire groundwater class of in the BPRP area. The beta coefficient for Fe (0.404) is the second highest after EC (0.164), Mn (0.075), TH (0.035), and Cl (0.033). Cl has a minimal contribution in this model with the smallest beta coefficient (0.033).

The regression model Mod₇ consists of seven significant predictor parameters with a level of

Table 4 Overview of models from the MLR analysis

Eigen models	R	R^2	Adjusted R^2	Std. error of the estimate	Independent variables (predictors)
Mod ₁	0.918	0.842	0.836	17.3099	Predictors: (constant), Tur
Mod ₂	0.981	0.962	0.959	8.6893	Predictors: (constant), Tur, Fe
Mod ₃	0.996	0.992	0.991	4.0270	Predictors: (constant), Tur, Fe, EC
Mod ₄	0.999	0.999	0.999	1.6416	Predictors: (constant), Tur, Fe, EC, Mn
Mod ₅	1.000	0.999	0.999	1.2435	Predictors: (constant), Tur, Fe, EC, Mn, TH
Mod ₆	1.000	1.000	1.000	0.8940	Predictors: (constant), Tur, Fe, EC, Mn, TH, Cl
Mod ₇	1.000	1.000	1.000	0.4746	Predictors: (constant), Tur, Fe, EC, Mn, TH, Cl, TDS
Mod ₈	1.000	1.000	1.000	0.0030	Predictors: (constant), Tur, Fe, EC, Mn, TH, Cl, TDS, pH

G_{WQI} is the dependent variable

Table 5 Estimated models with coefficients

Model		Unstandardized coefficients		Standardized beta coefficients	t-test	Sig
		B	Std. error			
Mod ₆	(Constant)	10.989	1.588		6.921	0.000
	Tur	1.973	0.014	0.820	137.000	0.000
	Fe	43.594	0.470	0.404	92.782	0.000
	EC	0.028	0.001	0.164	33.193	0.000
	Mn	142.940	11.685	0.075	12.233	0.000
	TH	0.039	0.007	0.035	5.429	0.000
	Cl	0.047	0.010	0.033	4.841	0.000
Mod ₇	(Constant)	10.666	0.844		12.640	0.000
	Tur	1.990	0.008	0.827	250.297	0.000
	Fe	43.978	0.254	0.407	172.921	0.000
	EC	0.006	0.003	0.037	2.186	0.040
	Mn	143.562	6.203	0.076	23.143	0.000
	TH	0.036	0.004	0.033	9.590	0.000
	Cl	0.055	0.005	0.039	10.464	0.000
Mod ₈	(Constant)	0.009	0.015		0.611	0.548
	Tur	2.000	0.000	0.831	39,001.705	0.000
	Fe	44.442	0.002	0.412	26,173.001	0.000
	EC	0.009	0.000	0.051	484.591	0.000
	Mn	133.360	0.041	0.070	3260.263	0.000
	TH	0.033	0.000	0.030	1394.709	0.000
	Cl	0.053	0.000	0.038	1638.667	0.000
	TDS	0.033	0.000	0.115	1089.584	0.000
	pH	1.568	0.002	0.012	753.752	0.000

G_{WQI} is the dependent variable. All the measures of variables are in mg/l excluding Tur (NTU), pH, and EC (μ S/cm)

significance < 0.05 . This regression model Mod₇ is expressed by Eq. (8). The standardized beta coefficients in this model show that Tur with the maximal beta coefficient (0.827) has the highest contribution to the whole groundwater standard within the BPRP area, followed by Fe (0.407), TDS (0.129), Mn (0.076), Cl (0.039), EC (0.037), and TH (0.033). TH with the smallest beta coefficient (0.033) is considered the smallest contributor in this model.

$$G_{WQI} = 10.989 + 1.973Tur + 43.594Fe + 0.028EC + 142.940Mn + 0.039TH + 0.047Cl \pm 1.588 \tag{7}$$

$$G_{WQI} = 10.666 + 1.990Tur + 43.978Fe + 0.006EC + 143.562Mn + 0.036TH + 0.055Cl + 0.037TDS \pm 0.844 \tag{8}$$

$$G_{WQI} = 0.009 + 2.000Tur + 44.442Fe + 0.009EC + 133.360Mn + 0.033TH + 0.053Cl + 0.033TDS + 1.568pH \pm 0.015 \tag{9}$$

The regression model Mod₈ contains eight predictor parameters and is expressed by Eq. (9). The standardized beta coefficients related to this model show that Tur has the maximum contribution (similar observation with Uyanik & Güler, 2013) to the entire groundwater class in BPRP area due to its highest beta coefficient (0.831), followed by Fe (0.412), TDS (0.115), Mn (0.070), EC (0.051), Cl (0.038), TH (0.030), and pH (0.012). pH has a minimal contribution in relation to water quality due to the lowest beta value (0.012) in this model.

A suitable model for G_{WQI} prediction for the BPRP area was obtained using the MLR approach. The causative variables that are considered in the above three models are significant. In Mod₈, TDS and pH are attached, reducing the significance level of the constant. This suggests that Mod₈ is associated with the maximum uncertainty regarding the constant (Wu et al., 2020). So Mod₈ is not good compared to Mod₆ and Mod₇. In Mod₆, all the predictor variables including the constant are significant (0.00) compared to Mod₇. So, from the MLR study, the model Mod₆ which has Tur, Fe, EC, Mn, TH, and Cl as independent variables is more reliable and is designed for the appropriate projection of G_{WQI} (Eq. (7)) in this research area. To validate the proposed model, compare the original and new G_{WQI} of the 20% sample using a *t*-test. The results obtained from the new G_{WQI} are not significantly different from those of the original G_{WQI} . The *t*-test result confirms the effectiveness of the new G_{WQI} equation because the significance (0.873) is greater than 0.05.

G_{WQI} converts all the parameters into a solitary value which is very acceptable in general. Among all the samples, R_{P16} shows the better G_{WQI} score and indicates good for drinking, though the TH level is slightly higher than the permissible limit in that sample. On the other hand, R_{P6} shows a higher G_{WQI} score and suggests very poor for drinking, and though apart from Tur, TH, and Fe, the rest of the parameters are within permissible limits. The samples which are showing poor to very poor grade require some treatment prior to consumption. Before individual parameter analysis was done by Maity et al. (2017) and Chakraborty et al. (2020) in that zone, they proposed that the groundwater quality has been falling down due to elevated concentration of the above selected parameters progressively. But this conventional approach did not provide the entire water quality because of the different classifications of every parameter, and that is confusing the results as various parameters are included here to decide the water quality. So G_{WQI} comprehensively provides a better interpretation of the complete water quality scenario in chosen stretches.

The above study also implies that the new G_{WQI} MLR model with not many variables produces economic benefits, reduces the eclipse effect, offers good accuracy to predict the water quality, and suggests that Tur is the main contributor in connection with

groundwater pollution. So this will be very logical to monitor and manage the quality of existing water resources for the communities in this locality. Wang et al. (2020b), Banda and Kumarasamy (2020), Wu et al. (2020), and Valentini et al. (2021) forecasted the WQI through the MLR model approach and also identified the most dominant parameter linked to water quality, authenticate the significance of this approach, and certify its feasibility. The results obtained from the above-stated studies and our current study both prefer that the MLR model is a handy tool to predict water quality, and a model with lesser variables may provide greater performance than a model with more parameters.

The above investigation through PCA/FA, G_{WQI} , and MLR prescribes a lot of information in respect of groundwater quality in BPRP. In addition, here, we have taken into account the eight water quality variables in 38 sampling points as per information available about the dataset. Analysis of more parameters will be helpful to know about the other characteristics of water, and more representatives will furnish a better understanding about water quality processes and trends. As we have limited parameters for this analysis, the reliability of the statistical approaches may be enhanced by introducing more parameters.

Conclusions

The WQI, along with statistical approaches, has been used to provide a comprehensive picture of groundwater conditions associated with selected coastal areas, as groundwater quality is declining here. Therefore, here, the PCA/FA method was used to identify potential sources of groundwater contamination, and G_{WQI} was involved in identifying the current groundwater class in the BPRP area. The MLR analysis was used to develop a new appropriate mathematical model in G_{WQI} estimation for groundwater quality monitoring. Current research shows that groundwater is fresh, hard to very hard, and slightly acidic to alkaline.

Three principal components were extracted from the PCA/FA study, which explained 80.61% cumulative variance of the total variance. PCA along with FA revealed that VFA₁ may be affected by salt water along with the disintegration and leaching process, VFA₂ may be affected by salt water, and VFA₃ may

be affected by association with the organic or otherwise biogenic and lithogenic or otherwise non-lithogenic process. This study also reveals that the sampling points R_{P2} , R_{P6} , R_{P9} , and R_{P11} are mostly affected by variables such as Cl, Tur, Mn, and TH; sampling points R_{P1} , R_{P7} , R_{P12} , R_{P13} , and R_{P20} are mostly affected by the parameters such as EC and TDS; and sampling points R_{P3} , R_{P4} , R_{P5} , R_{P8} , R_{P10} , and R_{P12} are mostly affected by pH and Fe.

The G_{WQI} was evaluated for 38 sites, and the calculated index scores ranged from 52.33 to 204.84. It was observed that 87% of the sample falls in grade B, 11% of the sample falls in grade C, and 2% of the sample falls in grade D. None of the samples fall in grades A and E. Therefore, groundwater is possibly considered for drinking purposes in entire sampling locations. The MLR model containing prediction parameters such as Tur, Fe, EC, Mn, TH, and Cl is more reliable and is proposed to predict the G_{WQI} . The proposed MLR model also reveals that Tur is the highest contributor to the overall groundwater quality. This model also provides financial benefits because fewer variables are involved.

The WQI together with the statistical models is shown to be a favorable approach to identifying groundwater quality by converting the dataset into equivalent entity data and numerical index scores. Therefore, the assessment of groundwater quality using statistical and index approaches is not only a common study but allows to delineate the most unsafe area with regard to the class of groundwater that can adversely affect human health. This study can assist the designers and officials in observing and selecting solutions for groundwater contamination. Therefore, this type of analysis is very valuable for the modernization and permissible expansion of groundwater resources for the general public.

Acknowledgements The authors thank Mr. Haripada Maity, Lab. Chem., Drinking Water Test Dep., Contai Sub-div. PHED, Gov. West Bengal and Mr. Ganesh Dinda, Jr. Eng., I&WD, Digha Irrig. Sect., Gov. West Bengal, for providing relevant data.

Authors' contribution Conceptualization and methodology: Chinmoy Ranjan Das and Subhasish Das; data collection: Souvik Panda; formal analysis: Chinmoy Ranjan Das; investigation: Chinmoy Ranjan Das and Subhasish Das; writing—original draft preparation: Chinmoy Ranjan Das; writing (review and editing) and supervision: Subhasish Das.

Data availability All data analyzed during this study are included in this article.

Declarations

Ethical approval The authors declare that they will follow the journal's guidelines on the reliability of the scientific record.

Consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

- Abulibdeh, A., Al-Awadhi, T., Nasiri, N. A., Al-Buloshi, A., & Abdelghani, M. (2021). Spatiotemporal mapping of groundwater salinity in Al-Batinah, Oman. *Groundwater for Sustainable Development*, 12, 100551. <https://doi.org/10.1016/j.gsd.2021.100551>
- Akshitha, V., Balakrishna, K., & Udayashankar, H. N. (2021). Assessment of hydrogeochemical characteristics and salt-water intrusion in selected coastal aquifers of southwestern India. *Marine Pollution Bulletin*, 173, 112989. <https://doi.org/10.1016/j.marpolbul.2021.112989>
- Alam, A., & Singh, A. (2023). Groundwater quality assessment using SPSS based on multivariate statistics and water quality index of Gaya, Bihar (India). *Environmental Monitoring and Assessment*, 195, 687. <https://doi.org/10.1007/s10661-023-11294-7>
- Aminiyan, M. M., Aitkenhead-Peterson, J., & Aminiya, F. M. (2018). Evaluation of multiple water quality indices for drinking and irrigation purposes for the Karoon river. *Iran. Environmental Geochemistry and Health*, 40, 2707–2728. <https://doi.org/10.1007/s10653-018-0135-7>
- APHA. (2017). *Standard methods for the examination of water and wastewater* (23rd ed.). Washington.
- Arslan, H. (2013). Application of multivariate statistical techniques in the assessment of groundwater quality in seawater intrusion area in Bafra Plain, Turkey. *Environmental Monitoring and Assessment*, 185, 2439–2452. <https://doi.org/10.1007/s10661-012-2722-x>
- Arslan, H., & Demir, Y. (2013). Impacts of seawater intrusion on soil salinity and alkalinity in Bafra Plain, Turkey. *Environmental Monitoring and Assessment*, 185, 1027–1040. <https://doi.org/10.1007/s10661-012-2611-3>
- Asare, A., Appiah-Adjei, E. K., Ali, B., & Owusu-Nimo, F. (2021). Physico-chemical evaluation of groundwater along the coast of the Central Region, Ghana. *Groundwater for Sustainable Development*, 13, 100571. <https://doi.org/10.1016/j.gsd.2021.100571>
- Awachat, A. R., & Salkar, V. D. (2017). Ground water quality assessment through WQIs. *International Journal of Engineering Research and Technology*, 10, 318–322.

- Banda, D. T., & Kumarasamy, M. (2020). Application of multivariate statistical analysis in the development of a surrogate water quality index (WQI) for South African watersheds. *Water*, 12, 1584. <https://doi.org/10.3390/w12061584>
- Boateng, T. K. B., Opoku, F., Acquah, S. O., & Akoto, O. (2016). Groundwater quality assessment using statistical approach and water quality index in Ejisu-Juaben Municipality. *Ghana. Environmental Earth Sciences*, 75, 489. <https://doi.org/10.1007/s12665-015-5105-0>
- Bodrud-Doza, M., Islam, A. R. M. T., Ahmed, F., Das, S., Saha, N., & Rahman, M. S. (2016). Characterization of groundwater quality using water evaluation indices, multivariate statistics and geostatistics in central Bangladesh. *Water Science*, 30(1), 19–40. <https://doi.org/10.1016/j.wsj.2016.05.001>
- Bouteraa, O., Mebarki, A., Bouaicha, F., Nouaceur, Z., & Laignel, B. (2019). Groundwater quality assessment using multivariate analysis, geostatistical modeling, and water quality index (WQI): A case of study in the Boumerzoug-El Khroub valley of Northeast Algeria. *Acta Geochimica*, 38(6), 796–814. <https://doi.org/10.1007/s11631-019-00329-x>
- Chakraborty, S., John, B., Maity, P. K., & Das, S. (2020). Increasing threat on groundwater reserves due to seawater intrusion in Contai belt of West Bengal. *Journal of the Indian Chemical Society*, 97(5), 799–817.
- Chen, T., Zhang, H., Sun, C., Li, H., & Gao, Y. (2018). Multivariate statistical approaches to identify the major factors governing groundwater quality. *Applied Water Science*, 8, 215. <https://doi.org/10.1007/s13201-018-0837-0>
- Das, C. R., Das, S., & Panda, S. (2022). Groundwater quality monitoring by correlation, regression and hierarchical clustering analyses using WQI and PAST tools. *Groundwater for Sustainable Development*, 16, 100708. <https://doi.org/10.1016/j.gsd.2021.100708>
- Dutta, S., Dwivedi, A., & Kumar, M. S. (2018). Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. *Environmental Monitoring and Assessment*, 190, 718. <https://doi.org/10.1007/s10661-018-7100-x>
- Ferchichi, H., Hamouda, M. F. B., Farhat, B., & Mammou, A. B. (2018). Assessment of groundwater salinity using GIS and multivariate statistics in a coastal Mediterranean aquifer. *International Journal of Environmental Science and Technology*, 15, 2473–2492. <https://doi.org/10.1007/s13762-018-1767-y>
- Gholami, V., Khaleghi, R. M., & Sebghati, M. (2016). A method of groundwater quality assessment based on fuzzy network-CANFIS and geographic information system (GIS). *Applied Water Science*, 7(7), 3633–3647. <https://doi.org/10.1007/s13201-016-0508-y>
- Gibrilla, A., Bam, E. K. P., Adomako, D., Ganyaglo, S., Osa, S., Akiti, T. T., Kebede, S., Achoribo, E., Ahialey, E., Ayanu, G., & Agyeman, E. K. (2011). Application of water quality index (WQI) and multivariate analysis for groundwater quality assessment of the Birimian and Cape Coast granitoid complex: Densu River Basin of Ghana. *Water Quality, Exposure and Health*, 3, 63. <https://doi.org/10.1007/s12403-011-0044-9>
- Gradilla-Hernández, M. S., de Anda, J., Garcia-Gonzalez, A., Meza-Rodríguez, D., Montes, C. Y., & Perfecto-Avalos, Y. (2020). Multivariate water quality analysis of Lake Cajititlán. *Mexico. Environmental Monitoring and Assessment*, 192, 5. <https://doi.org/10.1007/s10661-019-7972-4>
- Gulgundi, M. S., & Shetty, A. (2018). Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques. *Applied Water Science*, 8, 43. <https://doi.org/10.1007/s13201-018-0684-z>
- Haghnazar, H., Johannesson, K. H., González-Pinzón, R., Pourakbar, M., Aghayani, E., Rajabi, A., & Hashemi, A. A. (2022). Groundwater geochemistry, quality, and pollution of the largest lake basin in the Middle East: Comparison of PMF and PCA-MLR receptor models and application of the source-oriented HHRA approach. *Chemosphere*, 288, 132489. <https://doi.org/10.1016/j.chemosphere.2021.132489>
- Halder, S., Dhal, L., & Jha, M. K. (2021). Investigating groundwater condition and seawater intrusion status in coastal aquifer systems of eastern India. *Water*, 13, 1952. <https://doi.org/10.3390/w13141952>
- Heydarirad, L., Mosaferi, M., Pourakbar, M., Esmailzadeh, N., & Maleki, S. (2019). Groundwater salinity and quality assessment using multivariate statistical and hydrogeochemical analysis along the Urmia Lake coastal in Azarshahr Plain, North West of Iran. *Environmental Earth Sciences*, 78, 670. <https://doi.org/10.1007/s12665-019-8655-8>
- IS 10500. (2012). Indian standard drinking water specification. Second Revision, Bureau of Indian Standards, New Delhi.
- Islam, A. R. M. T., Shen, S., Haque, M. A., Bodrud-Doza, M., Maw, W. K., & Habib, M. A. (2018). Assessing groundwater quality and its sustainability in Joypurhat district of Bangladesh using GIS and multivariate statistical approaches. *Environment, Development and Sustainability*, 20, 1935–1959. <https://doi.org/10.1007/s10668-017-9971-3>
- Jayathunga, K., Diyabalanage, S., Frank, A. H., Chandrajith, R., & Barth, J. A. C. (2020). Influences of seawater intrusion and anthropogenic activities on shallow coastal aquifers in Sri Lanka: Evidence from hydrogeochemical and stable isotope data. *Environmental Science and Pollution Research*, 27, 23002–23014. <https://doi.org/10.1007/s11356-020-08759-4>
- Liu, C., Lin, K., & Kuo, Y. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, 313(1–3), 77–89. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- Li, Q., Zhang, H., Guo, S., Fu, K., Liao, L., Xu, Y., & Cheng, S. (2019). Groundwater pollution source apportionment using principal component analysis in a multiple land-use area in southwestern China. *Environmental Science and Pollution Research*, 27, 9000–9011. <https://doi.org/10.1007/s11356-019-06126-6>
- Li, W., Wu, J., Zhou, C., & Nsabimana, A. (2021). Groundwater pollution source identification and apportionment using PMF and PCA-APCS-MLR receptor models in Tongchuan City, China. *Archives of Environmental Contamination and Toxicology*, 81, 397–413. <https://doi.org/10.1007/s00244-021-00877-5>

- Mahapatra, S. S., Sahu, M., Patel, R. K., & Panda, B. N. (2012). Prediction of water quality using principal component analysis. *Water Quality, Exposure and Health*, 4, 93–104. <https://doi.org/10.1007/s12403-012-0068-9>
- Maity, P. K., Das, S., & Das, R. (2017). Assessment of groundwater quality and saline water intrusion in the coastal aquifers of Purba Midnapur district. *Indian Journal of Environmental Protection*, 37(1), 31–40.
- Maity, P. K., Das, S., & Das, R. (2018). A geochemical investigation and control management of saline water intrusion in the coastal aquifer of Purba Midnapur district in West Bengal, India. *Journal of the Indian Chemical Society*, 95, 205–210.
- Mondal, N. C., Singh, V. P., Singh, V., & S. & Saxena, V. K. (2010). Determining the interaction between groundwater and saline water through groundwater major ions chemistry. *Journal of Hydrology*, 388, 100–111. <https://doi.org/10.1016/j.jhydrol.2010.04.032>
- Motevalli, A., Moradi, R. H., & Javadi, S. (2017). A comprehensive evaluation of groundwater vulnerability to salt-water up-coning and sea water intrusion in a coastal aquifer (case study: Ghaemshahr-juybar aquifer). *Journal of Hydrology*, 557, 753–773. <https://doi.org/10.1016/j.jhydrol.2017.12.047>
- Mu, D., Wu, J., Li, X., Xu, F., & Yang, Y. (2023). Identification of the spatiotemporal variability and pollution sources for potential pollutants of the Malian River water in Northwest China using the PCA-APCS-MLR receptor model. *Exposure and Health*. <https://doi.org/10.1007/s12403-023-00537-0>
- Mustapha, A., Aris, A. Z., Ramli, M. F., & Juahir, H. (2012). Temporal aspects of surface water quality variation using robust statistical tools. *The Scientific World Journal*, 2012, 294540. <https://doi.org/10.1100/2012/294540>
- Narany, T. S., Ramli, M., & F., Aris, A. Z., Sulaiman, W. N. A., & Fakharian, K. (2014). Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol-Babol Plain. *Iran. Environmental Monitoring and Assessment*, 186, 5797–5815. <https://doi.org/10.1007/s10661-014-3820-8>
- Nguyen, B. T., Nguyen, T. M. T., & Bach, Q. (2020). Assessment of groundwater quality based on principal component analysis and pollution source-based examination: A case study in Ho Chi Minh City. *Vietnam. Environmental Monitoring and Assessment*, 192, 395. <https://doi.org/10.1007/s10661-020-08331-0>
- Papaioannou, A., Mavridou, A., Hadjichristodoulou, C., Papatertgiou, P., Pappa, O., Dovriki, E., & Rigas, I. (2010). Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health. *Environmental Monitoring and Assessment*, 170, 87–97. <https://doi.org/10.1007/s10661-009-1217-x>
- Patil, V. B. B., Pinto, S. M., Govindaraju, T., Hebbalu, V. S., Bhat, V., & Kannanur, L. N. (2019). Multivariate statistics and water quality index (WQI) approach for geochemical assessment of groundwater quality—A case study of KanaviHalla Sub-Basin, Belagavi, India. *Environmental Geochemistry and Health*, 42, 2667–2684. <https://doi.org/10.1007/s10653-019-00500-6>
- Ram, A., Tiwari, S. K., Pandey, H. K., Chaurasia, A. K., Singh, S., & Singh, Y. V. (2021). Groundwater quality assessment using water quality index (WQI) under GIS framework. *Applied Water Science*, 11, 46. <https://doi.org/10.1007/s13201-021-01376-7>
- Rao, V. V. S. G., Rao, G. T., Surinaidu, L., Mahesh, J., Rao, S. T. M., & Rao, B. M. (2013). Assessment of geochemical processes occurring in groundwaters in the coastal alluvial aquifer. *Environmental Monitoring and Assessment*, 185, 8259–8272. <https://doi.org/10.1007/s10661-013-3171-x>
- Reghunath, R., Murthy, T. R. S. S., & Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydro-geochemical studies: An example from Karnataka. *India. Water Research*, 36(10), 2437–2442. [https://doi.org/10.1016/S0043-1354\(01\)00490-0](https://doi.org/10.1016/S0043-1354(01)00490-0)
- Sahour, H., Gholami, V., & Vazifedan, M. (2020). Comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. *Journal of Hydrology*, 591, 125321. <https://doi.org/10.1016/j.jhydrol.2020.125321>
- Senthilkumar, G., Ramanathan, A. L., Nainwal, H. C., & Chidambaram, S. (2008). Evaluation of the hydro geochemistry of groundwater using factor analysis in the Cuddalore coastal region, TamilNadu. *India. Indian Journal of Marine Sciences*, 37(2), 181–185.
- Singha, S., Pasupuleti, S., Singha, S.S., Singh, R., & Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, 276, 130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>
- Solangi, S. G., Siyal, A. A., Babar, M. M., & Siyal, P. (2019). Application of water quality index, synthetic pollution index, and geospatial tools for the assessment of drinking water quality in the Indus Delta. *Pakistan. Environmental Monitoring and Assessment*, 191, 731. <https://doi.org/10.1007/s10661-019-7861-x>
- Taşan, M., Demir, Y., & Taşan, S. (2022). Groundwater quality assessment using principal component analysis and hierarchical cluster analysis in Alaçam. *Turkey. Water Supply*, 22(3), 3431–3447. <https://doi.org/10.2166/ws.2021.390>
- Todd, K. D., & Mays, W. L. (2005). *Groundwater hydrology*. third ed. John Wiley and Sons.
- Tripathi, M., & Singal, S. K. (2019). Use of principal component analysis for parameter selection for development of a novel water quality index: A case study of river Ganga India. *Ecological Indicators*, 96, 430–436. <https://doi.org/10.1016/j.ecolind.2018.09.025>
- Troudi, N., Hamzaoui-Azaza, F., Tzoraki, O., Melki, F., & Zammouri, M. (2020). Assessment of groundwater quality for drinking purpose with special emphasis on salinity and nitrate contamination in the shallow aquifer of Guenniche (Northern Tunisia). *Environmental Monitoring and Assessment*, 192, 641. <https://doi.org/10.1007/s10661-020-08584-9>
- Ustaoglu, F., Tepe, Y., & Tas, B. (2020). Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. *Ecological Indicators*, 113, 105815. <https://doi.org/10.1016/j.ecolind.2019.105815>
- Uyanik, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral*

- Sciences*, 106, 234–240. <https://doi.org/10.1016/j.sbspro.2013.12.027>
- Valentini, M., dos Santos, G. B., & Vieira, B. M. (2021). Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul-Brazil. *SN Applied Sciences*, 3, 70. <https://doi.org/10.1007/s42452-020-04005-1>
- Wang, H., Chen, Q., Wei, J., & Ji, Y. (2020a). Geochemical characteristics and influencing factors of groundwater Fe in seawater intrusion area. *Water, Air, & Soil Pollution*, 231, 348. <https://doi.org/10.1007/s11270-020-04724-6>
- Wang, D., Wu, J., Wang, Y., & Ji, Y. (2020b). Finding high-quality groundwater resources to reduce the hydatidosis incidence in the Shiqu County of Sichuan Province, China: Analysis, assessment, and management. *Exposure and Health*, 12, 307–322. <https://doi.org/10.1007/s12403-019-00314-y>
- WHO (2011). Guidelines for drinking water quality. Fourth Edition, World Health Organization.
- Wu, J., Li, P., Wang, D., Ren, X., & Wei, M. (2020). Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. *Human and Ecological Risk Assessment: An International Journal*, 26(6), 1603–1621. <https://doi.org/10.1080/10807039.2019.1594156>
- Yang, W., Zhao, Y., Wang, D., Wu, H., Lin, A., & He, L. (2020). Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang River in Huangshan, China. *International Journal of Environmental Research and Public Health*, 17, 2942. <https://doi.org/10.3390/ijerph17082942>
- Zeinalzadeh, K., & Rezaei, E. (2017). Determining spatial and temporal changes of surface water quality using principal component analysis. *Journal of Hydrology: Regional Studies*, 13, 1–10. <https://doi.org/10.1016/j.ejrh.2017.07.002>
- Zhang, H., Cheng, S., Li, H., Fu, K., & Xu, Y. (2020). Groundwater pollution source identification and apportionment using PMF and PCA-APCA-MLR receptor models in a typical mixed land-use area in Southwestern China. *Science of The Total Environment*, 741, 140383. <https://doi.org/10.1016/j.scitotenv.2020.140383>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.