Assessment of groundwater salinity using principal component analysis (PCA): a case study from Mewat (Nuh), Haryana, India

G. Krishan[ORCID] · A. Bhagwat · P. Sejwal ·
B. K. Yadav · M. L. Kansal · A. Bradley · S. Singh ·
M. Kumar · L. M. Sharma · M. Muste

Abstract In the present study, principal component analysis (PCA) is used to investigate the processes controlling groundwater salinity in the Mewat (Nuh) district, Haryana, India. Twenty groundwater samples were collected from salinity-affected areas in the March–April months of years 2018 and 2019 and were analyzed for chemical variables pH, EC, Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, Cl^-, SO_4^{2-}, NO_3^-, TDS, and total hardness. Three principal components were selected based on the eigen value, which explains 79.58% and 85.08% of the total variation in the years 2018 and 2019, respectively. The first principal component (PC-1) is identified with salinity, the second principal component (PC-2) with alkalinity, and the third principal component (PC-3) described the pollution. When the yearly comparison was made, the samples collected in 2019 were found to have an increased salinity compared to 2018, which shows an increased vulnerability to the aquifer of Mewat on account of the decline in rainfall recharge. It was also evident that declining recharge also triggered the recharge from other sources; thus, the impact of pollution is more pronounced in 2019 compared to 2018.

Introduction

Water scarcity has resulted in many regions of the world due to increased demand for water with time. In rural India, groundwater provides 80 percent of the domestic water (Malik & Bhagwat, 2021, Soujanya et al., 2020; Aravindan & Shankar, 2011; Foster & Chilton, 2003). Out of this, about 30 to 84 percent of total groundwater consumed in various states of India is brackish (Krishan et al., 2020a; Prusty & Farooq, 2020). Yao and Lund (2021) pointed out that groundwater salinization is increasing globally and poses a major challenge to both water and agricultural management.

Salinity is considered as a major natural source of groundwater pollution (Heydarirad et al., 2019). Groundwater salinity is controlled by many factors such as rainfall, the composition of aquifer material, topography, hydrologic fluctuation, and climate, and it varies spatially and temporally as a result of these intricate interactions (Mohapatra et al., 2011; Venkateswaran et al., 2012;

G. Krishan (✉) · A. Bhagwat · P. Sejwal · S. Singh ·
M. Kumar
National Institute of Hydrology, Roorkee 247667,
Uttarakhand, India
e-mail: drgopal.krishan@gmail.com

B. K. Yadav · M. L. Kansal
Indian Institute of Technology, Roorkee 247667,
Uttarakhand, India

A. Bradley · M. Muste
The University of Iowa, Iowa City, IA 52242, USA

L. M. Sharma
Sehgal Foundation, Gurgaon, Haryana, India

Kawo & Karuppannan, 2018; Panneerselvam et al., 2020b; Li et al., 2020; Haji et al., 2021a, b; Muthusamy et al., 2022). The aquifer's variability is correlated to a variety of biological, physical, and chemical processes, including weathering, aerobic respiration, mineral dissolution and precipitation, cation exchange, and other anthropogenic sources (Aravinthasamy et al., 2021; Panneerselvam et al., 2020a; Mohapatra et al., 2011; Panneerselvam et al., 2020c, d, 2021, 2022; Ravi et al., 2020). Thus, for identifying the main causes of increases groundwater pollution, different source apportionment techniques like chemical characterization (Panneerselvam et al., 2020b; Li et al., 2022; Liu et al., 2019; Ravi et al., 2020), tracers (Haji et al., 2021a, b; Krishan et al., 2021a, 2022; Panda et al., 2022), statistical techniques (Panneerselvam et al., 2020a, El-Kholy et al., 2022), and modeling are employed. Techniques like tracers and modeling can only be successfully used for contaminant plume detection when the source is already known and the selected tracer changes its characteristics in accordance with the changing nature of the pollutant. Based on the chemical characterization, areas showing similarities in pollution are clustered by many researchers (El-Kholy et al., 2022, Krishan et al., 2021b), but this clustering does not represent the dominant factor responsible for the existing physio-chemical characteristic of groundwater. To give a quantifiable measure of correspondence of water quality metrics and to identify the underlying human and natural processes in groundwater, multivariate statistical techniques such as principal component analysis (PCA) have been utilized (Bonelli & Manni, 2019, Chandrasekar et al., 2021; Elemile et al., 2021; Karunanidhi et al., 2021; Panneerselvam et al., 2020c). The principal component analysis identifies the key features or parameters that determine the data structure, as well as group observatories based on sample similarities. Multivariate analysis approaches such as PCA have been proven to be effective for identifying significant water quality concerns and potential contamination sources/processes by evaluating and interpreting groundwater quality data sets (Kavitha et al., 2019a, b; Rao et al., 2010; Shankar & Kawo, 2019; Singh et al., 2004, 2005, 2009).

Mewat (Nuh) district is one of the seven districts of Haryana, India which registers a very high groundwater salinity. Most of the groundwater wells in the study area show a high TDS of 35,000 mg/l (Krishan, 2019; Krishan et al., 2020c) making it unfit for drinking and agricultural uses. Therefore, to ensure sustainable development, it becomes essential to understand the relationship between the physio-chemical parameters and to identify the regional processes which determine the groundwater quality. Though various methods for source identification like chemical characterization (Krishan et al., 2021b) and isotopic analysis (Krishan et al., 2020b) have been performed in the study area, no attempt has been made to understand the dominant phenomena of salinization through a statistical method like PCA in the study area. Thus, the present study attempts to apply PCA over the physicochemical parameters of groundwater to extract key mechanisms influencing groundwater quality and salinity control in Mewat. Identification of the salinity sources in the study area will be the first step towards achieving SDG-6, improving water quality, and reducing pollution.

Study area

Location, climate, and geomorphology

Mewat region of Haryana state is located between 27°40′00″N and 28°20′00″N latitude and between 76°50′00″E and 77°20′00″E longitudes (Fig. 1) having a population of 1,089,263. Firozpur Jhirka, Nuh, Nagina, Tauru, and Punahana are the five blocks that cover this region (Krishan et al., 2021a, b). The daily maximum temperature in the study region is 40 °C (in May and June) and the daily lowest temperature is 5.1 °C (January).

The elevation of the study area is from 100 m to ~550 m (Fig. 2a). Higher elevations are found in the Aravalli ranges. The study area found two major soil types, vertisols and salanchalks. These soil types generally have medium-textured loamy sand. Due to the recurring lack of rainfall during the monsoon season and the scarcity of freshwater resources, farmers are forced to plant crops that require less water, such as wheat, millet, and mustard. The district has a rolling topography with an urn-shaped structure. The geomorphology of the study area is shown in Fig. 2b. Topographically, the study area is mainly made up of alluvium of Quaternary and Paleoproterozoic age groups. The most dominant is the Quaternary age group alluvium, with a polycyclic sequence composed of sand, silt, and clay with kankar. The alluvium thickness varies from 90 to 300 m bgl, and the depth to water level varies from 2 to 32 m bgl. The major
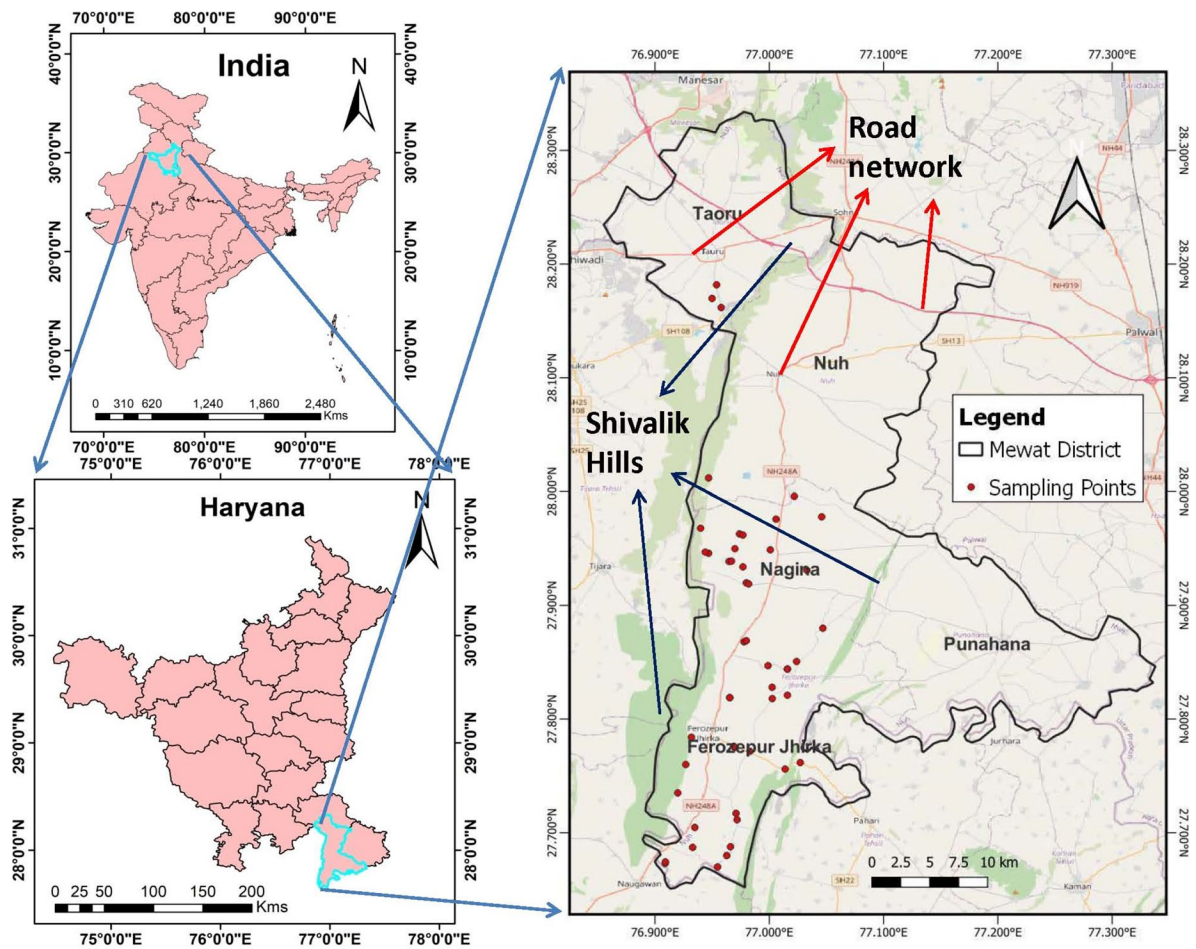
Fig. 1 Study area showing sampling locations

lithology is quartzite, schist, and phyllites of the Alwar and Ajabgarh groups. The weathering of these Proterozoic mobile belts forms the overlying younger quaternary deposits of the Pliestocence-holocene period, deposited as aolian deposits in the form of loose oxidized sand and silt. There are no natural drains in the district with no natural lakes and ponds.

Land use and land cover

Mewat district consists of 64 percent agricultural land, 12 percent built-up area (human settlements including the residential sites), about 6 percent falls under fallow land, 1 percent underwater bodies, and 18 percent under hills. (Fig. 3). The land use and land cover show that the agricultural activities utilize maximum land; since the water sources available do not fulfill the need for irrigation, the study area's current and future water usage is stressed.

Data collection and analysis

Groundwater samples from each of the 20 selected sites were collected in March–April of 2018 and 2019 from hand pumps, open wells, and bore wells with a depth range of 4–92 m. The water level of the open well in meters was recorded using a water-level indicator. GPS readings were taken to record latitude and longitude. Samples were taken in 125 mL capacity acid-washed Tarson bottles. The samples collected were analyzed for cations (Ca^{2+}, Mg^{2+}, Na^+, and K^+) and anions (HCO_3^-, SO_4^{2-}, Cl^- and NO_3^-) at the water quality laboratory of the groundwater division of
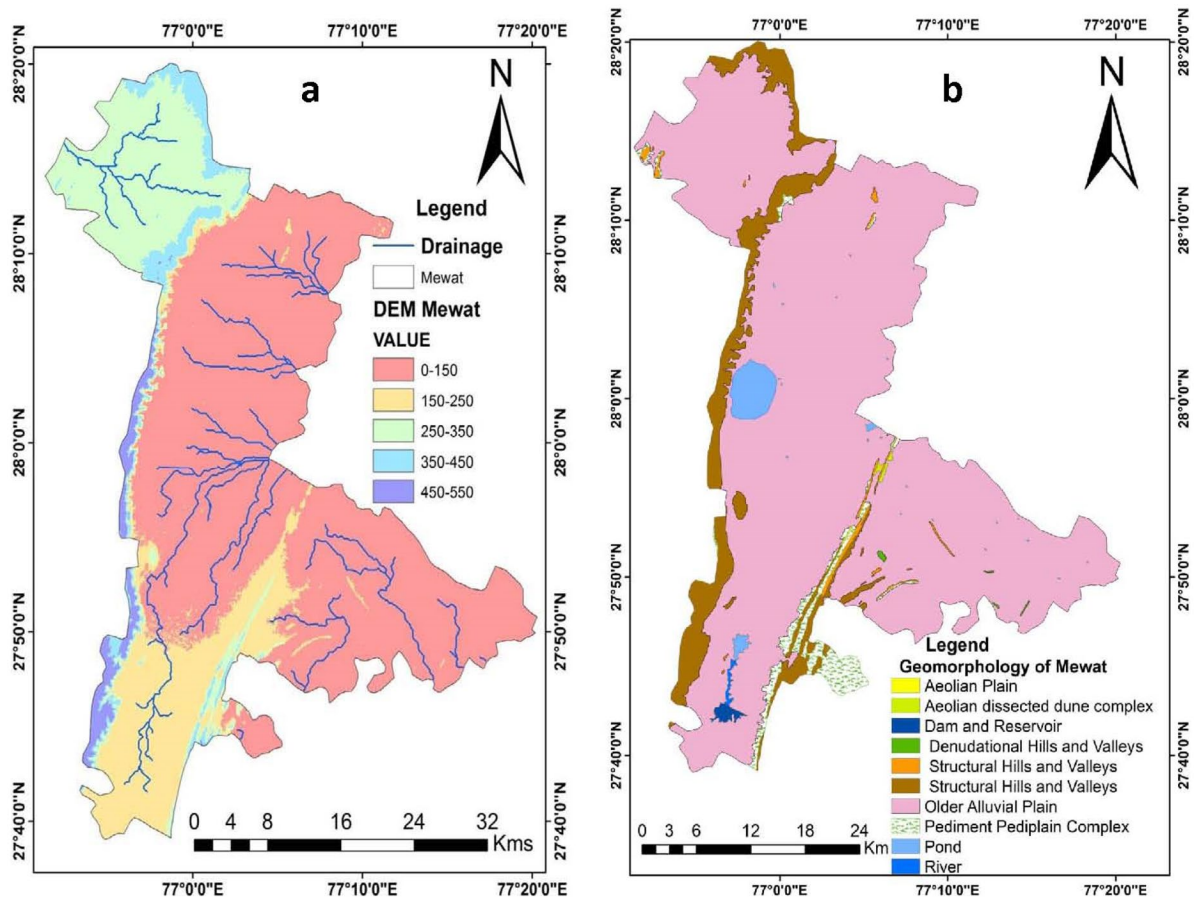
Fig. 2 a Digital Elevation Model. **b** Geomorphology of the Mewat district

NIH Roorkee, as per standard methodology (APHA, 2005).

Electrical conductivity (EC) and pH were measured using a portable handheld Hach, HQ30d EC meter, and total dissolved solids (TDS) were estimated from measured EC values and expressed in mg/l. Ca^{2+} and Mg^{2+} were determined titrimetrically using standard EDTA. For determining the Cl^- concentrations, the Argentometric method was used in which potassium dichromate was used as an indicator and titrated with silver nitrate solution, which was titrant. HCO_3^- was determined by titration with HCl. Na^+ and K^+ were measured by flame photometry at a wavelength of 589 nm and 766 nm, respectively, and SO_4^{2-} and NO_3^- by spectrophotometric turbidimetry by a HACH spectrophotometer. All concentrations are expressed in milligrams per liter (mg/l).

PCA analysis

Principal component analysis (PCA) is used to create new variables called principal component, which illustrates the relationship among different chemical variables. The new variables are made up of a linear combination of the old ones. In general, two or three principal components are utilized, which explains a reasonable percentage of variance based on the eigen value. But, before applying PCA to determine the existing relationship, it is important to determine whether the sampling points selected are adequate for evaluating the underlying relationship. Thus, the Kaiser–Meyer–Olkin (KMO) test is performed. The KMO value of more than 0.5 indicates the sample is adequate for factor analysis. Furthermore, to ensure that the F-test does not remain liberal in determining
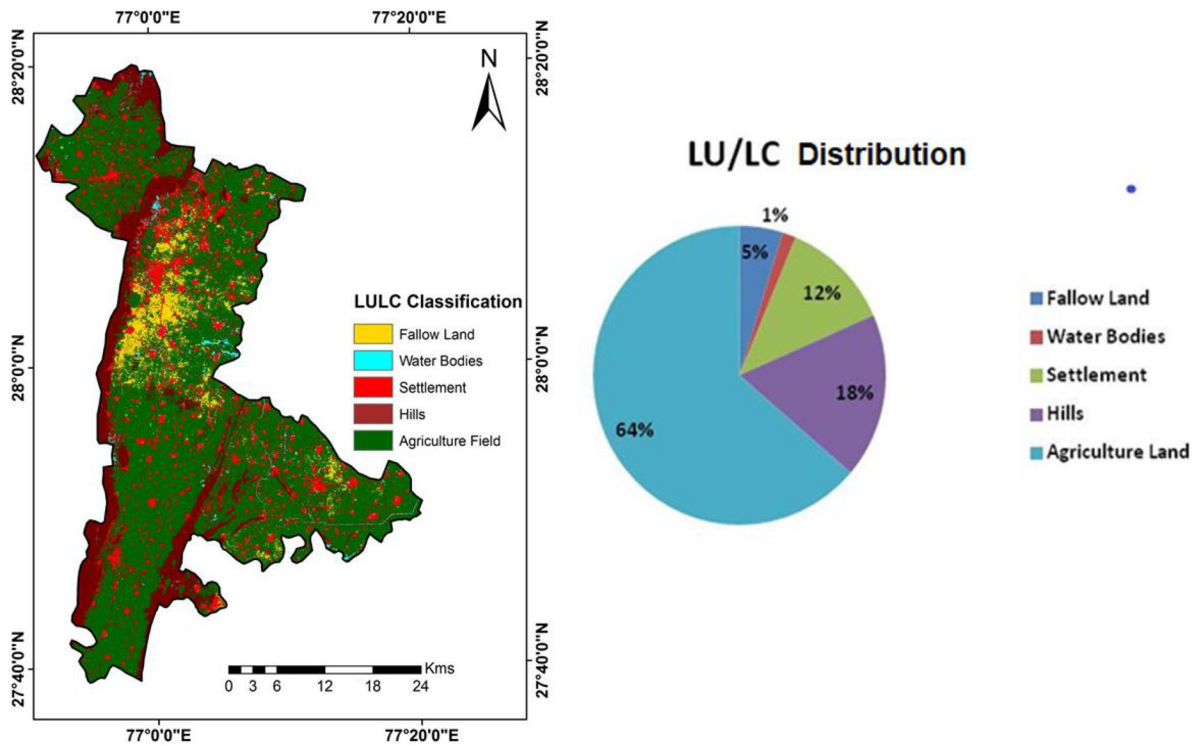
Fig. 3 Landuse/land cover map of the Mewat district

the relationship, Barlett's sphericity is performed. The Bartlett's test of sphericity tests the null hypothesis; if the significance value is less than 0.05 then the null hypothesis is rejected and PCA is valuable for the data. If the observed significance level is rejected, it indicates that there are significant relationships among variables.

In the present study, three principal components for both the years were utilized, and the selection was made based on eigenvalues greater than unity (Kaiser, 1958). The varimax rotation was applied to obtain score plots. The principal component scores describe the intensity of chemical processes in relation to each principal component. Lithological and anthropogenic controlled variables were identified from the principal component scores, and the influence of these factors at the sampling locations was then drawn through two sets of plots. These are (a) scatter plots between the principal component loadings and (b) scatter plots of principal component scores obtained through Varimax rotation.

Results and discussion

General

In the month of March–April, twenty samples were collected from different sites in the study area for the years 2018 and 2019 and analyzed according to the prescribed methods, and a data set was used to conduct PCA analysis using SPSS.

The pH varies from 6.8 to 8.3 in 2018 and 6.98 to 8.7 in 2019 indicating increased groundwater acidity. The concentration of TDS is in the range of 291 mg/l to 20,904 mg/l for the year 2018 and 52.93 mg/l to 13,393 mg/l for the year 2019. On comparing cations concentrations (Ca^{2+}, Mg^{2+}, Na^+, and K^+), Na^+ concentration (2.4 mg/l to 10,286 mg/l for 2018 and 31.68 mg/l to 11,635 mg/l) is found to be dominant followed by Ca^{2+} (64 mg/l to 1340 mg/l) and Mg^{2+} (6 mg/l to 878 mg/l) for the year 2018 whereas the trend for Ca^{2+} and Mg^{2+} gets reversed with Mg^{2+} (23 mg/l to 1030 mg/l) and Ca^{2+} (80 mg/l to 1280 mg/l) for the

years 2019. Among the anions (HCO_3^-, SO_4^{2-}, Cl^-, and NO_3^-), chloride concentration (135 mg/l to 8041 mg/l for 2018 and 333 mg/l to 17,042 mg/l for the year 2019) predominates for both the years, followed by HCO_3^- (34 mg/l to 540 mg/l for 2018 and 109 mg/l to 437 mg/l), followed by NO_3^- (1 mg/l to 11 mg/l for 2018 and 1 mg/l to 20 mg/l for the year 2019), followed by SO_4^{2-} (13 mg/l to 450 mg/l for 2018 and 40 mg/l to 4280 mg/l for the year 2019). Thus, the groundwater is generally characterized by sodium chloride-type water.

KMO and bartlett's test

To ensure that the interpretation made through principal components remains rigorous, two pre-PCA tests KMO and Bartlett's were performed, and the results so obtained are represented in Table 1.

High values of the KMO test (close to 1.0) generally indicate that a PCA is useful for the data. If the value is < 0.50, the results of the PCA are not considered valid (Field, 2013). In 2018, the KMO value is found to be more than 0.50 ($KMO = 0.623 > 0.50$) (Table 1). Therefore, it can be inferred that the results of the PCA are considered to be valid. If the significance value is less than 0.05 then the null hypothesis is rejected and PCA is valuable for the data ($significance = 0.00 < 0.05$). Similarly, in 2019, for the months of March and April, the KMO value and Bartlett's test for sphericity were found to be 0.528 and 0.000 (Table 2), respectively. Therefore, in both years, the data can be considered valid for PCA analysis.

Table 1 KMO and Bartlett's test for the samples collected in the year 2018 and 2019

KMO and Bartlett's test (2018)	
Kaiser–Meyer–Olkin measure of sampling adequacy	0.623
Bartlett's test of sphericity	
Approx. chi-square	324.1
Df	66
Sig.	0.000
KMO and Bartlett's test (2019)	
Kaiser–Meyer–Olkin Measure of Sampling Adequacy	0.582
Bartlett's Test of Sphericity	
Approx. chi-square	89.17
Df	66
Sig.	0.030

Principal component analysis

Tables 2 and 3 represent the eigen value, and percentage variances for the sampling done in the year 2018 and 2019, respectively through which PCA solution is obtained. The output of the final loading matrix obtained from the data for 2018 and 2019 indicates three principal components which explain 79.58% and 85.08% of the total variation, respectively. The principal component-I (F1) contributed 57.299% in the year 2018 and 61.7% in 2019; principal component-II (F2) contributed 13.11% and 13.63% for 2018 and 2019, respectively, and the principal component-III (F3) contributed 9.17% in the year 2018 and 9.75% in 2019 with 12 chemical variables. Thus, each principal component can be used to identify the specific hydrogeochemical processes and the processes which undergo seasonal variations by examining their loadings.

Principal component loadings

To find out the variation in parameter loadings from positive to negative scatter plots of principle component loadings between PC-1, PC-2, and PC-3 and also PC-2 and PC-3 are illustrated for the year 2018 in Fig. 4 and for 2019 in Fig. 5.

Table 2 Varimax orthogonal rotated factor loadings from PCA of standardized water quality data set 2018

	F1	F2	F3
TDS	0.975	0.053	⁻0.189
pH	⁻0.159	0.719	0.172
EC	0.977	0.003	⁻0.179
TH	0.656	⁻0.428	0.278
HCO_3^-	0.564	0.653	0.019
NO_3^-	0.589	⁻0.115	0.606
Mg^{2+}	0.523	0.188	0.495
Ca^{2+}	0.501	⁻0.610	0.063
Na^+	0.958	0.090	⁻0.213
K^+	0.923	0.015	⁻0.322
SO_4^{2-}	0.837	0.067	0.325
Cl^-	0.925	0.099	⁻0.234
Eigen value	6.876	1.573	1.101
Variability (%)	57.299	13.106	9.172
Cumulative (%)	57.299	70.405	79.577

Table 3 Varimax orthogonal rotated factor loadings from PCA of standardized water quality data set 2019

	F1	F2	F3
TDS	0.983	0.036	0.076
pH	‾0.326	0.761	0.171
EC	0.872	0.276	‾0.093
TH	0.974	‾0.125	0.035
HCO_3^-	0.286	0.674	0.227
NO_3^-	0.165	‾0.047	0.700
Mg^{2+}	0.971	‾0.017	0.108
Ca^{2+}	‾0.097	‾0.686	‾0.481
Na^+	0.962	‾0.171	‾0.049
K^+	0.883	‾0.089	0.218
SO_4^{2-}	0.941	‾0.209	‾0.094
Cl^-	0.985	‾0.130	0.026
Eigen value	7.404	1.635	1.170
Variability (%)	61.699	13.628	9.748
Cumulative (%)	61.699	75.328	85.076

Principle component-1

For the year 2018, the principal component PC-1 is dominated by TDS, Na^+, $Cl-$, SO_4^{2-} and K^+ variables whereas for 2019, principal component PC-1 is dominated by TDS, Na^+, $Cl-$, SO_4^{2-}, Mg^{2+}, TH, and K^+. The high domination of TDS, Na^+, $Cl-$, SO_4^{2-} and K^+ for 2018 and TDS, Na^+, $Cl-$, SO_4^{2-}, Mg^{2+}, TH, and K^+ for 2019 can also be concluded from the close plotting in the positive region of PC-1 as shown in Figs. 4(a, b) and 5(a, b), respectively. Thus, PC-1 is considered a salinity factor. For the year 2018, the salinity is directly controlled by Na^+ and $Cl-$ ions concentrations, whereas, for the year 2019, salinity is directly controlled by the concentrations of Na^+, Mg^{2+}, $Cl-$, and SO_4^{2-} ions. The hydrogeology of the region is dominated by silicate minerals feldspathic, gritty quartzite, amphibole, phyllite, and schist but lacks sodalite and chlorapatite minerals which are the
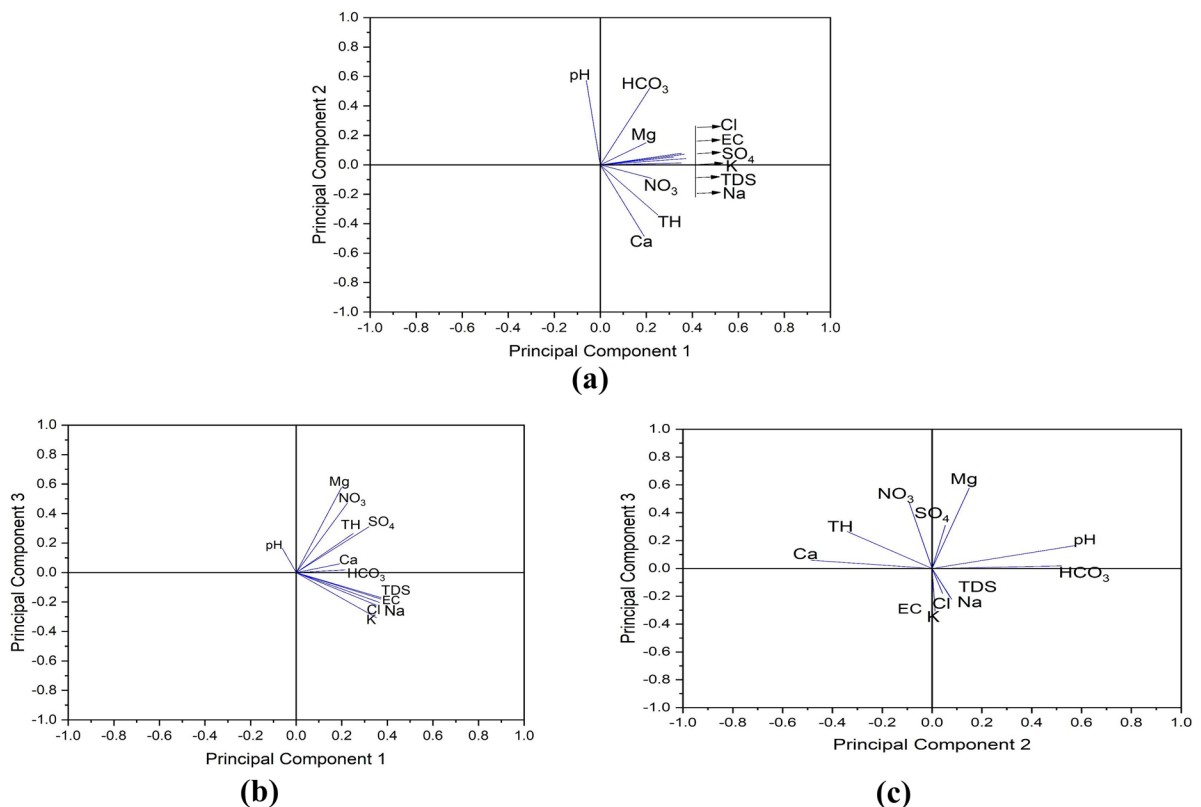

(a)

(b)

(c)

Fig. 4 Scatter plot of principal component loadings for the year 2018 **a** PC-1 vs PC-2, **b** PC-1 vs PC-3, and **c** PC-2 vs PC-3
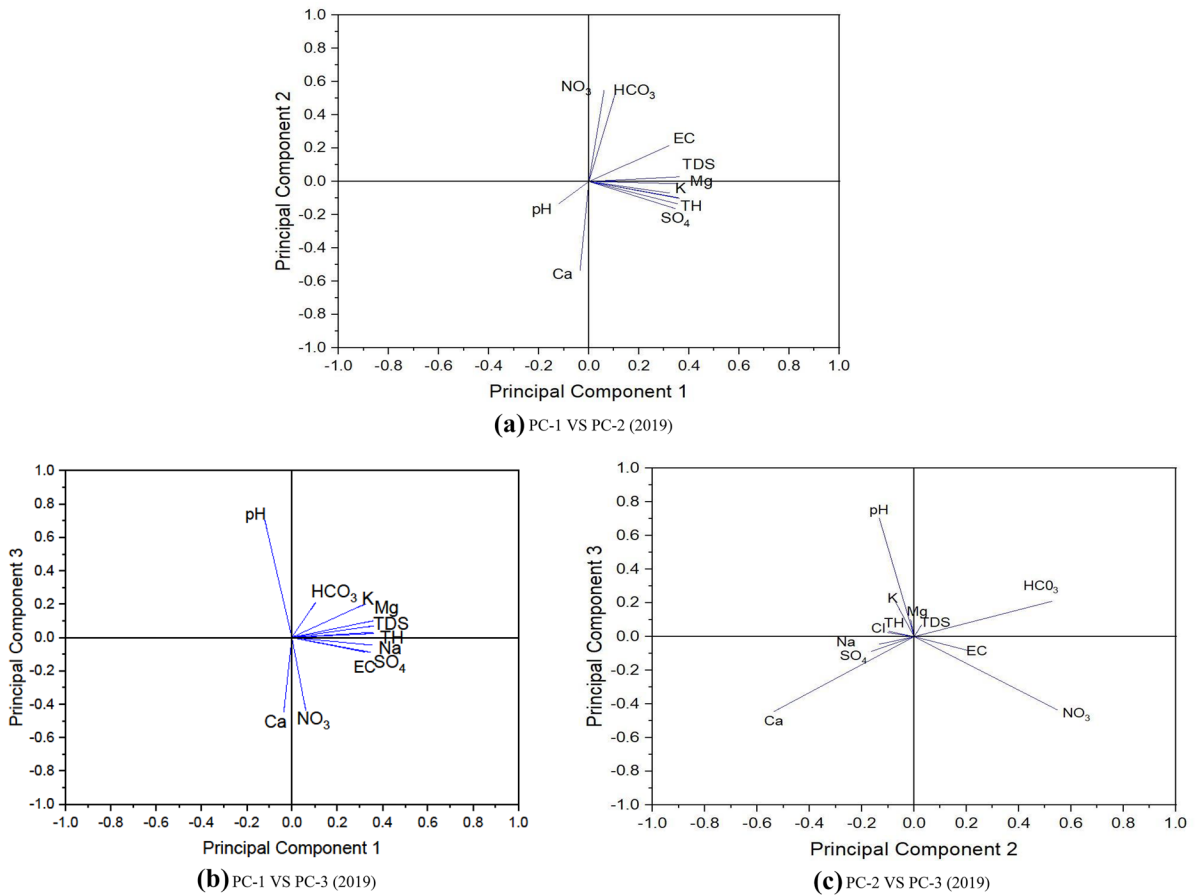
Fig. 5 Scatter plot of principal component loadings for the year 2019 **a** PC-1 vs PC-2, **b** PC-1 vs PC-3, and **c** PC-2 vs PC-3

main source of $Cl-$ and SO_4^{2-}. Thus, Na^+, $Cl-$ and SO_4^{2-} ions are contributed by anthropogenic sources (Krishan et al., 2021a) whereas, Mg^+ controlling the PC-1 of the 2019 is contributed by rock-water interaction. The change in controlling variables in both years can be associated with changing rainfall patterns. Year 2017 is a monsoon deficit year whereas 2018 is a year of normal rainfall since both the sampling was performed in the pre-monsoon season; it is an impression of the previous year which is reflected in the extent of rock-water interactions.

Furthermore, it is observed that TDS (salinity) shows a close positive relationship with Na^+, $Cl-$, SO_4^{2-}, and K^+ in the year 2018 and Na^+, $Cl-$, SO_4^{2-}, Mg^+, TH, and K^+ for the year 2019, thus the variation in TDS of the groundwater samples can be used to evaluate the temporal and regional influence of

the salinity factor and its modification. It is observed that out of the 20 samples taken for both the years, 14 had TDS > 1000 mg/l in the year 2018 and 15 had TDS > 1000 mg/l in the year 2019. Four samples showing freshwater environment (TDS < 1000 mg/l) in both years were taken from open wells located at Kotla, Bhoond, Ghata Basai, and Raja ka pul. The sample which shows a transition was located at Khedi Khurd and is a bore well thus it is the rock-water interaction that resulted in increased salinity. Nine out of 14 wells in 2018 and 12 out of 15 wells in 2019 have TDS > 2000 mg/l. Thus, the three wells which are vulnerable to salinity are Naharika, Kalakheda, and Ghagas village. Six samples in 2018 and eight samples in the year 2019 are brackish (TDS > 3500 mg/l). These brackish samples were taken from points located in the Delhi-Alwar subgroup. Since 2018 is

a deficit year causing an accelerated rate of groundwater extraction, deep groundwater carries an impression of rock-water interaction which is the leading cause of salinity in these regions. Thus, the two samples which are vulnerable to salinity on account of groundwater extractions are Gaghas SF and Naharika. Thus, rock-water interactions and agricultural return flow govern the PC-1.

Principal component-2

For both the years, the second principal component has higher positive loading for pH and HCO_3^- as shown in Figs. 4(b, c), and 5(b, c). Whereas, there is a negative plotting for Ca^{2+}, indicating that they come under the same process. PC-2 represents the process of alkalinity in which CO_2 dissolution plays an important role.

There were only 3 samples with pH > 8 in the year 2018, whereas there are 8 samples with pH > 8 in the year 2019. A total of 16 out of 20 samples is having a pH between 7 and 8 in 2018, whereas 8 out of 20 samples have a pH between 7 and 8. There is only one sample taken from Kansali, which has a pH = 6.8. The increase in pH and bicarbonate in the year 2019 over 2018 can be associated with the rainfall pattern. There was normal rainfall in the year 2018 compared with the deficit rainfall of 2017, which get translated as recharge in 2019 and 2018, respectively. The decay of organic matter and respiration of roots in the soil enhanced the CO_2 pressure, which combines with the recharging water and forms the bicarbonate and excess hydrogen ions, which elevates the pH as shown in Eq. (1).

$$CO_2 + H_2O \rightarrow H_2CO_3 \qquad (1)$$

$$H_2CO_3 \rightarrow H^+ + HCO_3^- \qquad (2)$$

Furthermore, the aquifer study reveals that Mewat is associated with clay and kankar ($CaCO_3$). The leaching of $CaCO_3$ from kankar increases the Ca^+ and HCO_3^- ions and increases the pH, whereas the precipitation of bicarbonate decreases both Ca^+ and HCO_3^- ions and consequently decreases the pH. Thus, aquifer properties play an important controlling factor for PC-2. PC-2 is therefore termed as a lithological controlled factor.

Principal component-3

The third principal component has a loading of 0.61 and 0.70 for nitrate for the years 2018 and 2019, respectively, as shown in Figs. 4(c) and 5(c). Thus, nitrate acts as an independent variable representing the pollution seeping into the groundwater. In 2018, there were 3 samples with NO_3^- whereas only one sample registers nitrate concentration > 10 mg/l for the year 2019. This indicates that in the case of deficit rainfall, groundwater recharge takes place through drains or other sources contaminated with agricultural or municipal waste. Since there is no source of nitrate pollution other than agricultural return flows or sewage waste, principal component-3 can be termed as the non-lithological controlling factor.

Principal component scores

To identify the regional influences responsible for salinity, principal component scores were plotted. For the year 2018, in PC-1, 6 samples have positive scores and the remaining have negative scores. In PC-2, 11 samples have positive scores, and in PC-3, 7 samples have positive scores. For the year 2019, 6 samples have positive scores and the remaining have negative scores. In PC-2, 10 samples have positive scores, and in PC-3, 11 samples have positive scores. The plot between PC-1 and PC-2 has 3 positive values for 2018 and 4 positive values for 2019. PC-2 vs PC-3 3 positive values for 2018 and 5 positive values from 2019, and the plot between PC-1 and PC-3 has 2 positive values for each year. The distribution of groundwater samples in the positive principal component is shown in Figs. 6 and 7 and in Table 4.

Thus, it can be concluded that rainfall plays a significant role in determining the regional factor which influences the salinity. In 2018, the salinity of Kansali, Jaitko, and Akhleempur was determined by both rock-water interaction and anthropogenic factors. But in 2019, apart from these factors, alkalinity also contributing to increased salinity. The salinity at Ghata Basin and Basai for both years is primarily determined by alkalinity. For Naharika, Kalakheda, and Mohmmad, alkalinity is dominating the high TDS value in the year 2018. However, pollution gets added to the groundwater of Naharika, Kalakheda, and Mohmmad in the year 2019. Kotla and Kameda,
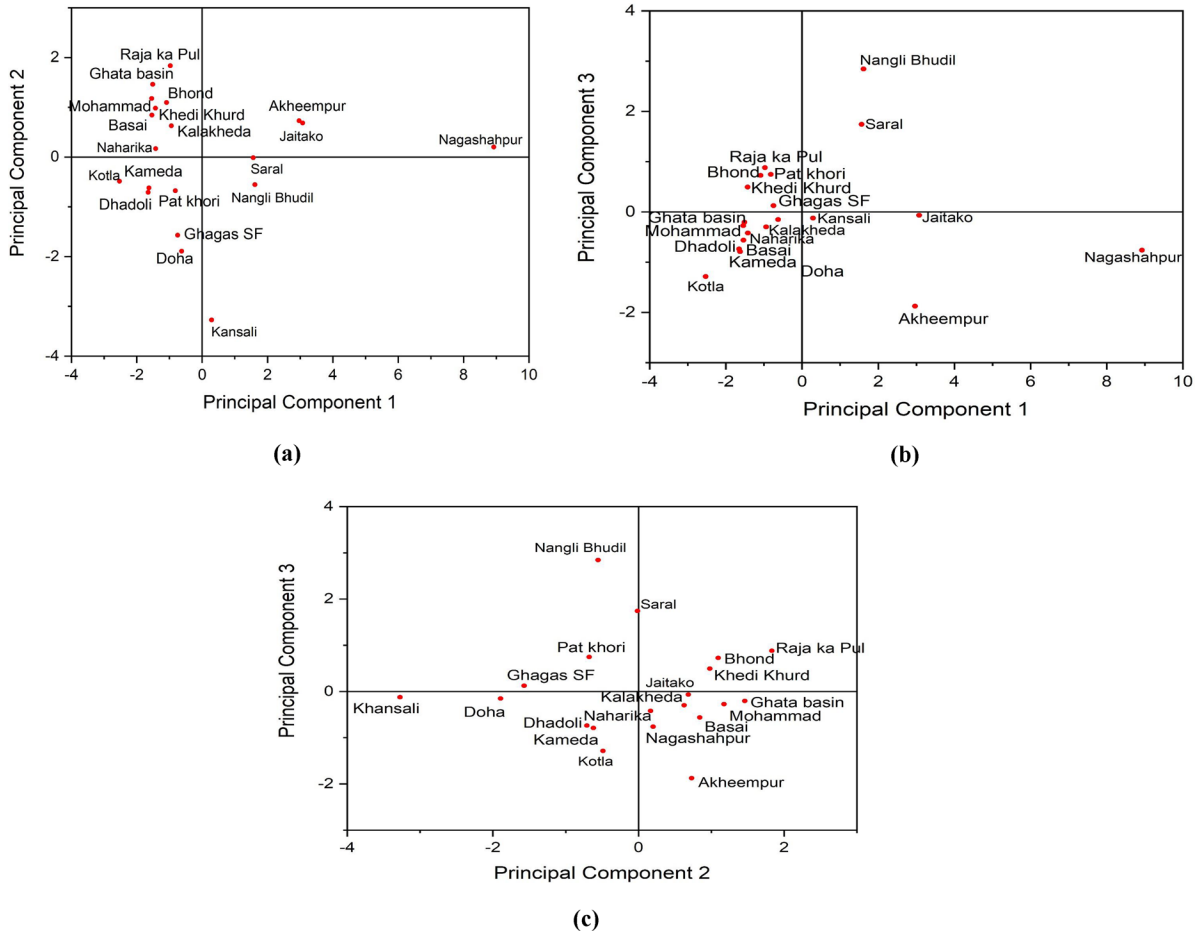
(a)

(b)

(c)

Fig. 6 Score plots for principal component loadings for the year 2018 **a** PC-1 vs PC-2, **b** PC-1 vs PC-3, and **c** PC-2 vs PC-3

a fresh water source in the year 2018, became saline in 2019 because of enhanced groundwater pollution. Dhadoli is the only sampling point that contains fresh water in both years. Pollution is the main cause of salinity at Ghagas SF and Pat Khori.

The groundwater quality at Nangli Bhudli is mainly determined by rock-water interaction; however, the alkalinity and pollution become determining factors based on their relative strength. Doha, which registered fresh groundwater in 2018, suffered from both salinity and alkalinity in 2019 under the influence of deficit rainfall and increased recharge. Bhond experienced both alkalinity and pollution in the year 2018, but the groundwater shows a major improvement with all parameters within permissible limits for the year 2019. Khedi khurd and Raja ka pul come under the influence of alkalinity and pollution. At Naglashahpur, rock-water interaction is the main cause of salinity, but it also experienced alkalinity in 2018 and pollution in 2019. At Saral, salinity is on account of rock-water interactions, but it comes under the influence of pollution in the year 2018. Thus, principal component analysis indicates the factors determining the groundwater quality in a specified location and helps identify the wells that are vulnerable to contamination.

Fig. 7 Score plots for principal component loadings for the year 2019 **a** PC-1 vs PC-2, **b** PC-1 vs PC-3, and **c** PC-2 vs PC-3

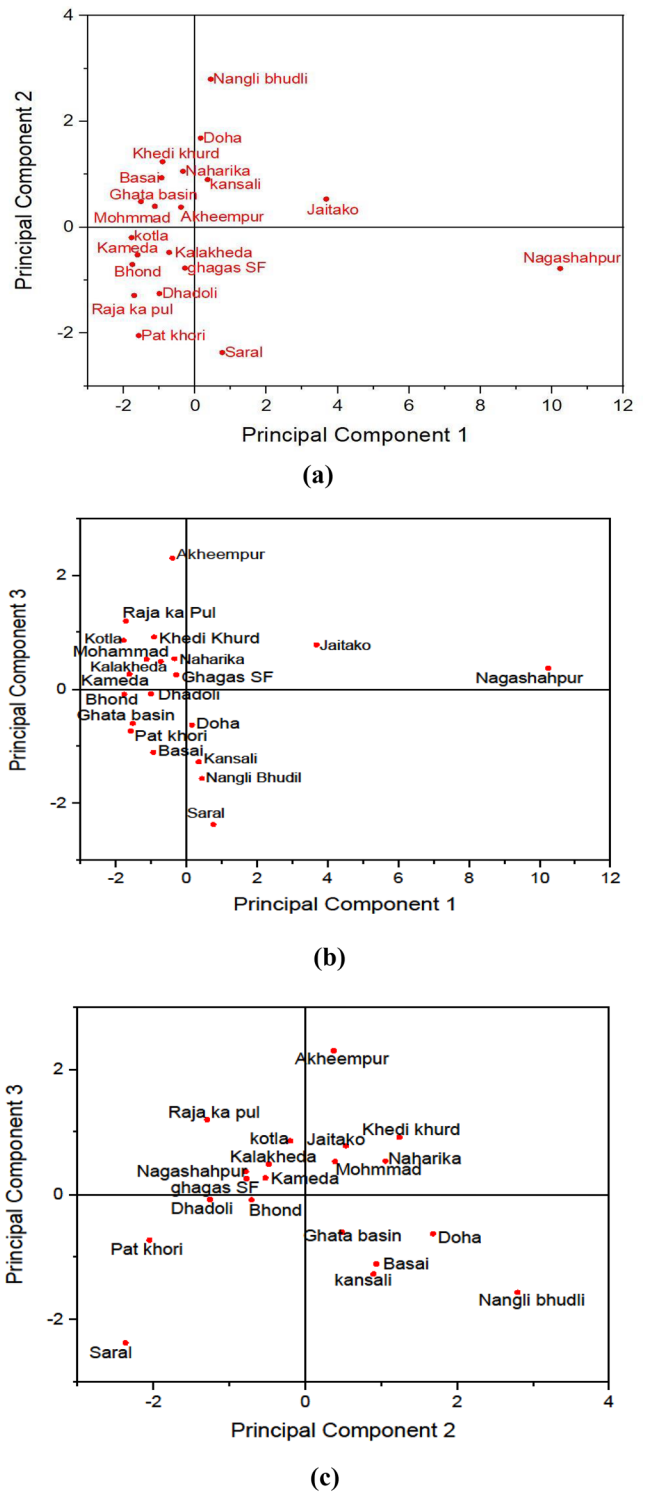**(a)**

(b)

(c)

Table 4. Distribution of the groundwater samples in the positive principal component score for the years 2018 and 2019

Positive Principal component score for the year 2018							
Sample Location	PC-1	PC-2	PC-3	PC-1 v/s PC-2	PC-1 v/s PC-3	PC-2 v/s PC-3	PC-1 v/s PC-2 v/s PC-3
Kansali	▫						
Kotla							
Ghagas SF			▫				
Nangli bhudli	▫		▫		▪		
Doha							
Pat khori			▫				
Dhadoli							
Bhond		▫	▫			▪	
Kameda							
Ghata basin		▫					
Basai		▫					
Naharika		▫					
Kalakheda		▫					
Mohmmad		▫					
Khedi khurd		▫	▫			▪	
Raja ka pul						▪	
Nagashahpur	▫	▫		▪			
Saral	▫		▫		▪		
Jaitako	▫						
Akheempur	▫						
Positive Principal component score for the year 2019							
Sample Location	PC-1	PC-2	PC-3	PC-1 v/s PC-2	PC-1 v/s PC-3	PC-2 v/s PC-3	PC-1 v/s PC-2 v/s PC-3
Kansali	▫	▫		▪			
Kotla			▫				
ghagas SF			▫				
Nangli bhudli	▫	▫		▪			
Doha	▫	▫		▪			
Pat khori			▫				
Dhadoli							
Bhond							
Kameda			▫				
Ghata basin		▫					
Basai		▫					
Naharika		▫	▫			▪	
Kalakheda		▫	▫			▪	
Mohmmad		▫	▫			▪	
Khedi khurd		▫	▫			▪	
Raja ka pul		▫	▫			▪	
Nagashahpur	▫		▫			▪	
Saral	▫						
Jaitako	▫	▫	▫	▪	▪	▪	▪
Akheempur		▫	▫			▪	

Conclusion

To assess the groundwater quality of Mewat district, Haryana 20 samples were collected and analyzed for $Cl-$, TDS, TH, HCO_3^-, EC, $Na+$, SO_4^{2-}, Mg^{+2}, and K^+ for the years 2018 and 2019. Based on these analyses, the following conclusions are drawn:

- 14 out of 20 samples in the year 2018 and 15 out of 20 samples in the year 2019 have TDS > 1000 mg/l. Thus, groundwater salinity and its expansion are evident. To identify the sources of groundwater contamination, principal component analysis is employed.

- KMO and Bartlett's test, for both the year the test, was found valid, and PCA is found to be suitable for the study area. Three principal components were selected based on the eigen value which explains 79.58% and 85.08% of the total variation in the years 2018 and 2019, respectively.

- The first principal component (PC-1) is identified with salinity and is governed by rock-water interactions and agricultural return flow. The second principal component (PC-2) with alkalinity and the third principal component (PC-3) described the pollution.

- When the yearly comparison was made, the samples collected in 2019 were found to have an increased salinity compared to 2018 which showed an increased vulnerability to the aquifer of Mewat on account of the decline in rainfall recharge. It was also evident that declining recharge also triggered the recharge from other sources, and thus pollution. The impact of pollution is more pronounced in 2019 compared to 2018.

Declarations

Conflict of interest The authors declare no competing interests.

References

APHA. (2005). Standard methods for the examination of water and wastewater, 21th ed. American Public Health Association: Washington, DC, USA; New York, NY, USA.

Aravindan, S., & Shankar, K. (2011). Ground water quality maps of Paravanar River sub basin, Cuddalore district, Tamil Nadu, India. *Journal of Indian Society of Remote Sensing, 39*(4), 565–658. https://doi.org/10.1007/s12524-011-0152-9

Aravinthasamy, P., Karunanidhi, D., Shankar, K., Subramani, T., Setia, R., Bhattacharya, P., & Das, S. (2021). COVID-19 lockdown impacts on heavy metals and microbes in shallow groundwater and expected health risks in an industrial city of South India. *Environmental Nanotechnology, Monitoring & Management, 16*, 100472. https://doi.org/10.1016/j.enmm.2021.100472

Bonelli, M. G., & Manni, A. (2019). Principal components analysis and spatial analysis integration for enhanced assessment of pollution emission sources. *IOP Conference Series: Earth and Environmental Science, 227*(6), 062013. IOP Publishing.

Chandrasekar, T., Keesari, T., Gopalakrishnan, G., et al. (2021). Occurrence of heavy metals in groundwater along the lithological interface of K/T boundary, peninsular India: A special focus on source, geochemical mobility and health risk. *Archives of Environmental Contamination and Toxicology, 80*, 183–207. https://doi.org/10.1007/s00244-020-00803-1

Elemile, O. O., Ibitogbe, E. M., Folorunso, O. P., Ejiboye, P. O., & Adewumi, J. R. (2021). Principal component analysis of groundwater sources pollution in Omu-Aran Community. Nigeria. *Environmental Earth Sciences, 80*(20), 1–16. https://doi.org/10.1007/s12665-021-09975-y

El-Kholy, R. A., Zaghlool, E., Isawi, H., Soliman, E. A., Khalil, M. M., El-Aassar, A. H. M., & Said, M. M. (2022). Groundwater quality assessment using water quality index and multivariate statistical analysis case study: East Matrouh, Northwestern coast. Egypt. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-022-19761-3

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). Sage.

Foster, S. S. D., & Chilton, P. J. (2003). Groundwater: The processes and global significance of aquifer degradation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1440), 1957–1972.

Haji, M., Karuppannan, S., Qin, D., Shube, H., & Kawo, N. S. (2021a). Potential human health risks due to groundwater fluoride contamination: A case study using multi-techniques approaches (GWQI, FPI, GIS, HHRA) in Bilate River basin of Southern Main Ethiopian Rift, Ethiopia. *Archives of Environmental Contamination and Toxicology, 80*, 277–293. https://doi.org/10.1007/s00244-020-00802-2

Haji, M., Qin, D., Guo, Y., Li, L., Wang, D., Karuppannan, S., & Shube, H. (2021b). Origin and geochemical evolution of groundwater in the Abaya Chamo basin of the main Ethiopian rift: Application of multi-tracer approaches. *Hydrogeology Journal, 29*(3), 1219–1238. https://doi.org/10.1007/s10040-020-02291-y

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*(24), 1–16.

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

Karunanidhi, D., Aravinthasamy, P., Deepali, M., Subramani, T., & Shankar, K. (2021). Groundwater pollution and human health risks in an industrialized region of southern india: Impacts of the COVID-19 lockdown and the monsoon seasonal cycles. *Archives of Environmental Contamination and Toxicology, 80*, 259–276. https://doi.org/10.1007/s00244-020-00797-w

Kavitha, M. T., Divahar, R., Meenambal, T., Shankar, K., VijaySingh, R., Haile, T. D., & Gadafa, C. (2019a). Dataset on the assessment of water quality of surface water in Kalingarayan Canal for heavy metal pollution, Tamil Nadu. *Data in Brief, 22*, 878–884. https://doi.org/10.1016/j.dib.2019.01.010

Kavitha, M. T., Shankar, K., Divahar, R., Meenambal, T., Saravanan, R. (2019b). Impact of industrial wastewater disposal on surface water bodies in Kalingarayan canal, Erode district, Tamil Nadu, India. *Archives of Agriculture and Environmental Science, 4*(4), 379–387. https://doi.org/10.26832/24566632.2019.040403

Kawo, N. S., & Karuppannan, S. (2018). Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia. *Journal of African Earth Sciences, 147*, 300–311. https://doi.org/10.1016/j.jafrearsci.2018.06.034

Krishan, G. (2019). Groundwater salinity. *Current World Environment: An International Research Journal of Environmental Sciences, 14*(2), 186–188.

Krishan, G., Bisht, M., Ghosh, N. C., Prasad, G. (2020a). Groundwater salinity in northwestern region of India: A critical appraisal. In: Singh, R., Shukla, P., Singh, P. (eds) Environmental Processes and Management. *Water Science and Technology Library, 91*. Springer, Cham. https://doi.org/10.1007/978-3-030-38152-3_19

Krishan, G., Prasad, G., Kumar, C. P., Patidar, N., Yadav, B. K., Kansal, M. L., & Verma, S. K. (2020b). Identifying the seasonal variability in source of groundwater salinization using deuterium excess-a case study from Mewat, Haryana. *India. Journal of Hydrology: Regional Studies, 31*, 100724. https://doi.org/10.1016/j.ejrh.2020.100724

Krishan, G., Ghosh, N. C., Kumar, C. P., Sharma, L. M., Yadav, B., Kansal, M. L., Singh, S., Verma, S. K., & Prasad, G. (2020c). Understanding stable isotope systematics of salinity affected groundwater in Mewat, Haryana, India. *Journal of Earth System Science, 129*, 109. https://doi.org/10.1007/s12040-020-1380-6

Krishan, G., Kumar, B., Sudarsan, N., Rao, M. S., Ghosh, N. C., Taloor, A. K., Bhattacharya, P., Singh, S., Kumar, C. P., Sharma, A., Jain, S. K., Sidhu, B. S., Kumar, S., & Vasisth, R. (2021a). Isotopes (δ18O, δD and 3H) variations in groundwater with emphasis on salinization in the State of Punjab. *India. Science of the Total Environment., 789*, 148051. https://doi.org/10.1016/j.scitotenv.2021.148051

Krishan, G., Sejwal, P., Bhagwat, A., Prasad, G., Yadav, B. K., Kumar, C. P., & Muste, M. (2021b). Role of ion chemistry and hydro-geochemical processes in aquifer salinization—A case study from a semi-arid region of Haryana. India. *Water, 13*(5), 617.

Krishan, G., Rao, M. S., Rajesh, V., Chaudhary, A., Singh, J., & Kumar, A. (2022). Isotopic assessment of groundwater salinity: A case study of Southwest (SW) region of Punjab. *India. Water, 14*, 133. https://doi.org/10.3390/w14010133

Li, C., Gao, X., Li, S., & Bundschuh, J. (2020). A review of the distribution, sources, genesis, and environmental concerns of salinity in groundwater. *Environmental Science and Pollution Research*, *27*(33), 41157–41174. https://doi.org/10.1007/s11356-020-10354-6

Li, C., Men, B.-H., & Yin, S.-Y. (2022). Analysis of groundwater chemical characteristics and spatiotemporal evolution trends of influencing factors in southern Beijing plain. *Frontiers in Environmental Science, 10*, 913542. https://doi.org/10.3389/fenvs.2022.913542

Liu, J., Jin, D., Wang, T., Gao, M., Yang, J., & Wang, Q. (2019). Hydrogeochemical processes and quality assessment of shallow groundwater in Chenqi coalfield, Inner Mongolia. China. *Environmental Earth Sciences, 78*(12), 1–13. https://doi.org/10.1007/s12665-019-8355-4

Malik, A., & Bhagwat, A. (2021). Modelling groundwater level fluctuations in urban areas using artificial neural network. *Groundwater for Sustainable Development, 12*, 100484. https://doi.org/10.1016/j.gsd.2020.100484

Mohapatra, P. K., Vijay, R., Pujari, P. R., Sundaray, S. K., & Mohanty, B. P. (2011). Determination of processes affecting groundwater quality in the coastal aquifer beneath Puri city, India: A multivariate statistical approach. *Water Science and Technology, 64*(4), 809–817.

Muthusamy, B., Sithu, G. D. S., Ramamoorthy, A., Shankar, K., Gnanachandrasamy, G., Monica, C., & Xiaozhong, H. (2022). Isotopic signatures, hydrochemical and multivariate statistical analysis of seawater intrusion in the coastal aquifers of Chennai and Tiruvallur District. *Tamil Nadu, India, Marine Pollution Bulletin,*. https://doi.org/10.1016/j.marpolbul.2021.113232

Panda, B., Chidambaram, S., Snow, D., Singh, D. K., & Ramanathan, A. L. (2022). Source apportionment and health risk assessment of nitrate in foothill aquifers of Western Ghats. South India. *Ecotoxicology and Environmental Safety, 229*, 113075. https://doi.org/10.1016/j.ecoenv.2021.113075

Panneerselvam, B., Kumar, P. S., Shankar, K., Nagavinothini, R., & Sajil Kumar, P. J. (2020a). Impact of climate and anthropogenic activities on groundwater quality for domestic and irrigation purposes in Attur region, Tamilnadu, India. *Desalination and Water Treatment, 208*, 172–195.

Panneerselvam, B., Kumar, P. S., & Shankar, K. (2020b). Dataset on the suitability of groundwater for drinking and irrigation purposes in the Sarabanga River region, Tamil Nadu. India. *Data in brief, 29*, 105255. https://doi.org/10.1016/j.dib.2020.105255

Panneerselvam, B., Karuppannan, S., & Muniraj, K. (2020c). Evaluation of drinking and irrigation suitability of groundwater with special emphasizing the health risk posed by nitrate contamination using nitrate pollution index (NPI) and human health risk assessment (HHRA). *Human and Ecological Risk Assessment: An International Journal*. https://doi.org/10.1080/10807039.2020.1833300

Panneerselvam, B., Paramasivam, S. K., Karuppannan, S., et al. (2020d). A GIS-based evaluation of hydrochemical

characterisation of groundwater in hard rock region, South Tamil Nadu. *India. Arab J Geosci, 13*(17), 1–22. https://doi.org/10.1007/s12517-020-05813-w

Panneerselvam, B., Muniraj, K., Pande, C., Ravichandran, N., Thomas, M., & Karuppannan, S. (2021). Geochemical evaluation and human health risk assessment of nitrate-contaminated groundwater in an industrial area of South India. *Environmental Science and Pollution Research.* https://doi.org/10.1007/s11356-021-17281-0

Panneerselvam, B., Muniraj, K., Duraisamy, K., Pande, C., Karuppannan, S., & Thomas, M. (2022). An integrated approach to explore the suitability of nitrate-contaminated groundwater for drinking purposes in a semiarid region of India. *Environmental Geochemistry and Health.* https://doi.org/10.1007/s10653-022-01237-5

Prusty, P., & Farooq, S. H. (2020). Seawater intrusion in the coastal aquifers of India-A review. *Hydro Research, 3*, 61–74. https://doi.org/10.1016/j.hydres.2020.06.001

Rao, Y. S., Keshari, A. K., & Gosain, A. K. (2010). Evaluation of regional groundwater quality using PCA and geostatistics in the urban coastal aquifer, East Coast of India. *International Journal of Environment and Waste Management, 5*(1–2), 163–180.

Ravi, R., Aravindan, S., Shankar, K., & Balamurugan, P. (2020). Suitability of groundwater quality for irrigation in and around the main Gadilam river basin on the east coast of southern India. *Archives of Agriculture and Environmental Science*, *5*(4), 554–562. https://doi.org/10.26832/24566632.2020.0504019

Shankar, K., & Kawo, N. S. (2019). Groundwater quality assessment using geospatial techniques and WQI in North East of Adama Town, Oromia Region, Ethiopia. *Hydrospatial Analysis, 3*(1), 22–36. https://doi.org/10.21523/gcj3.19030103

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research, 38*(18), 3980–3992.

Singh, K. P., Malik, A., Singh, V. K., Mohan, D., & Sinha, S. (2005). Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain. North India. *Analytica Chimica Acta, 550*(1–2), 82–91.

Singh, S., Singh, C., Kumar, K., Gupta, R., & Mukherjee, S. (2009). Spatial-temporal monitoring of groundwater using multivariate statistical techniques in Bareilly district of Uttar Pradesh. India. *Journal of Hydrology and Hydromechanics, 57*(1), 45.

Soujanya Kamble, B., Saxena, P. R., Kurakalva, R. M., & Shankar.K,. (2020). Evaluation of seasonal and temporal variations of groundwater quality around Jawaharnagar municipal solid waste dumpsite of Hyderabad city. India. *SN Appl. Sci., 2*, 498. https://doi.org/10.1007/s42452-020-2199-0

Venkateswaran, S., Karuppannan, S., & Shankar, K. (2012). Groundwater quality in Pambar Sub-Basin. *Tamil Nadu, India Using GIS, International Journal of Recent Scientific Research, 3*(10), 782–787.

Yao, Y., & Lund, J. (2021). Managing water and crops with groundwater salinity – A growing menace, CaliforniaWaterBlog.com