Pedology and Environment 2024, 34 (02), 179-195
doi.org/10.47114/j.pedoenv.2024.dec 6

Characterization of groundwater quality using multivariate statistical analysis in Jakham River Basin of Rajasthan, India

V. K. Gautam[1]*, K. K. Yadav[2], M. Kothari[3] and P. K. Singh[3]

[1*] *School of Natural Resource Management, College of Post Graduate Studies in Agricultural Sciences, Central Agricultural University (Imphal), Umiam, Meghalaya-793103*
[2]*Department of Soil Science and Agricultural Chemistry, RCA, MPUAT, Udaipur, Rajasthan-313001*
[3]*Department of Soil and Water Engineering, CTAE, MPUAT, Udaipur, Rajasthan 313001*

Abstract : In this study, groundwater quality in the Jakham River Basin, located in southern Rajasthan, India, was characterised using a multivariate statistical approach and a geographic information system (GIS). A total of 76 groundwater samples were collected from different sites in the area. Two multivariate statistical approaches, i.e., principal component analysis (PCA) and hierarchical cluster analysis (HCA), were applied to pre-monsoon and post-monsoon groundwater quality data to identify the most critical factors controlling groundwater quality. Spatial maps of groundwater quality parameters were developed using GIS. During the pre-monsoon and post-monsoon seasons, the computed values of the Water Quality Index (WQI) ranged from 28.28 to 116.74 and from 29.49 to 111.98, respectively. Based on values of GIS-based WQI, 63.42 and 42.02% of the groundwater samples were classified as 'good' during pre-monsoon and post-monsoon seasons, respectively. The interpretation of PCA results revealed the impact of geological and human interventions on increased levels of electrical conductivity, total dissolved solids, sodium, chloride, bicarbonate, fluoride, and sulphate. In this study, Ward's method was used for clustering of samples in HCA. The findings of the HCA indicated that there were four distinct groundwater quality groups within the basin. The findings of this study provide a reference database for the groundwater quality, facilitating further development and management of groundwater resources in the study area. Moreover, PCA and HCA were recommended as suitable tools for simplification of the evaluation process in groundwater quality analysis.

Keywords: *Geographic information system; Jakham river basin; Principal component analysis; Hierarchical cluster Analysis; Water quality index.*

Introduction

One of the most pressing problems in today's world is ensuring an adequate supply of good-quality water. This leads to sustainable management of water resources at stakeholder end (Das et al., 2019), and to achieve this cutting-edge concept, a dedicated approach is required. In India, the majority of the population relies on rivers and lakes for fulfilling their water requirements; however, the management of sewage and discharged water is very poor. As a result, the available water is prone to contamination, which has a negative impact on human health. Therefore, it is necessary to monitor available water resources (Noshadi & Ghafourian, 2016). According to a United Nations report, 22% of human deaths are directly linked to waterborne diseases (Sharma et al., 2021). This emphasises the need to characterise the groundwater quality.

Characterisation of groundwater quality (GWQ) refers to the physico-chemical and statistical analysis of groundwater, based on different standardised approaches

*Corresponding author: (Email: drvgautam95@gmail.com)

(Gautam et al., 2022a). GWQ is a significant aspect for the sustainable growth of water resources, as many sectors, such as irrigation, drinking, and industrial, rely on it (Ravikumar et al., 2011; Mohamed & Elmahdy, 2015). As all geological formations impact groundwater quality in the aquifer system, the quality of groundwater is dependent on the combination of various hydrogeological activities (Lee & Song, 2007; Batayneh & Al-Taani, 2016; Abdelaziz et al., 2020). The assessment of GWQ is a challenging task that involves handling a large number of variables, each of which has the potential to exert a specific impact on the overall GWQ (Bodrud-Doza et al., 2016). Additionally, GWQ exhibits spatio-temporal variability that is commonly influenced by topography, industrial effluents, open sewage dumping, and agricultural waste (Zavareh & Maggioni, 2018; Barkat et al., 2022; Mohseni et al., 2022). In southern Rajasthan of India, GWQ is rarely characterised using multiple and integrated approaches (Machiwal et al., 2011; Gautam et al., 2022a). The GWQ indexing is a new approach for determining the potability of water for human consumption, with an index varying on a 0-100 scale that represents the suitability of groundwater both in space and time (e.g., Gautam et al., 2023). The spatial maps of groundwater quality can also be generated through the integrated application of geographic information systems (GIS) and spatial interpolation methods such as inverse distance weighting (IDW), as an efficient tool for interpreting the spatial variation of GWQ (Tirkey et al., 2017; Gauns et al., 2020). The water quality index (WQI) is nowadays employed as a customary approach for evaluating the GWQ (Khan & Jhariya, 2017; Kawo & Karuppannan, 2018). The WQI has been employed to assess the spatial and temporal variability of aquifers' water quality in many states of India, including Rajasthan, Karnataka, and Delhi, among others, as it provides a single value and is easy to comprehend (Mohamed & Elmahdy, 2015; Shil et al., 2019).

The correlation matrix establishes relationships among the various GWQ parameters under the influence of geological and chemical processes (Das et al., 2019). It helps to detect the key influencing groups of water quality parameters and their sources of origin. Bodrud-Doza et al. (2016) monitored and analysed the GWQ of 60 open wells in Faridpur district of central Bangladesh. The outcomes of the correlation matrix were consistent with the results of other statistical analyses (e.g., Roy et al., 2021). The spatio-temporal distributions of GWQ parameters were also estimated using geostatistical analysis (e.g., Ram et al., 2021). In literature, few statistical methods, such as the projection approach, have been utilised for evaluating the WQI (Iqbal et al., 2021). However, PCA has emerged as a vital tool to deal with multivariate datasets concerning different GWQ parameters (Shil et al., 2019; Mohseni et al., 2022). In general, principal component analysis (PCA) has the primary function of simplifying the processing of big data input variables by isolating the most important aspects of a massive dataset (Vishwakarma & Thakur, 2012). As a part of PCA, grouped variables, termed as principal components (PCs), may be transformed using factor analysis (FA) (Jankowska et al., 2017). Hence, in this study, the influential GWQ parameters of the local aquifer were assessed using PCA.

Hierarchical cluster analysis (HCA) displays the clustering of different GWQ parameters by identifying their proximity with respect to particular features (Omo-Irabor et al., 2008; Teixeira et al., 2021). It facilitates in physico-chemical understanding of GWQ chemistry. In this study, a dendrogram was generated using Ward's method, with squared Euclidean distance serving as the similarity metric. Teixeira et al. (2021) employed HCA to examine the homogeneity and characteristics of GWQ, resulting in a better understanding of the geological action on the aquifer system.

The present study aimed at characterising groundwater quality in the Jakham River Basin by employing multivariate statistical techniques in a GIS environment. This study applies PCA and HCA (Loganathan & Ahamed, 2017; Mohseni et al., 2022) to examine the concentration of critically influencing GWQ parameters and the hazardous concentrations of fluoride, nitrate, and sulphate (Herojeet et al., 2016; Sharma et al., 2021). As per reports of the Central Ground Water Board (CGWB, 2017), two blocks of the study area, namely Pratapgarh and Chhoti Sadri, have been characterised as a sub-critical stage, which is due to high fluctuations in groundwater level and critical GWQ in pre-monsoon and post-monsoon seasons (Bouteraa et al., 2019; Gautam et al., 2022c). It is expected that the findings of this study will aid in ensuring GWQ and maintaining an environment that permits the sustainable use of groundwater resources in the study area.

Materials And Methods

Overview of Study Area

Jakham River basin is located between the latitudes of 24.451°N - 23.988°N and the longitudes of 74.501°E - 74.802°E, covering an area of 953 km^2 in the upper reaches of the Mahi River basin (Fig. 1). The south-west portion of the basin evidences forest and hilly area, while northeast-southeast portion is covered with urban and agricultural lands, which reflects the land use diversity within the basin (Gautam et al., 2021; Gautam et al., 2022a). The soil characteristics of the basin resemble those of black, fertile soil, and the climate of the basin varies from sub-humid to semi-arid, with a moderate average annual rainfall of 700 mm.

The geology of the basin is composed of various rock formations, which are predominantly igneous and meta-sedimentary in nature, and influence the dynamics of groundwater in the subsurface strata. Approximately 45% of the southern portion of the study area consists of basaltic geologic formations, while the remaining portion features a diverse range of geological formations (Gautam et al., 2022b). However, these types of rock formations are not considered good aquifers. Moderate groundwater potential occurs within the contact zone of basalt and other lithological units. The basin also exhibits a moderate evaporation rate, i.e., approximately 11.20 mm/day in the summer season. The major part of the basin is involved in the cultivation of opium and processing of its by-products, which leads to overuse of fertilisers and saline chemicals. Hence, the north-to-central part of the basin has a groundwater salinity problem. In this basin, groundwater is only found in semi-confined to unconfined conditions. However, phreatic aquifers are the primary sources of water for both human consumption and agricultural activities. The locations of the groundwater sampling sites are shown in Figure 1.

Data Collection

Groundwater samples were collected manually from a total of 76 sites in the basin during both the pre-monsoon and post-monsoon seasons during 2019-2020 using a well-established stratified sampling technique. The groundwater level in the study area varies from 15-30 m below ground level (bgl) and 25-40 m bgl in pre-monsoon and post-monsoon seasons, respectively. In addition, GWQ data for the last 13 years (2006-2018) were collected from the Ground Water Department (GWD), Jaipur and CGWB, Jaipur, for statistical analyses. High-quality, sealed polyethene bottles (250 mL) were used for sample collection. Some GWQ parameters, such as electrical conductivity (EC), pH, total dissolved solids (TDS), temperature, fluoride (F), and residual chloride (Cl), were analysed on the spot usinga low-cost Rapid Water Quality Testing Kit (RaQKT). This low-cost RaQKT kit was developed for the on-site determination of drinking and irrigation water quality, which helps reduce the workload in laboratory work (Gautam et al., 2022c). The remaining parameters were analysed in the Groundwater Laboratory of the College of Technology and Engineering (CTAE), Udaipur, India. The regulations and recommendations described by Bureau of Indian Standards (BIS), New Delhi (BIS, 2012) were used to analyse the major cations, i.e., calcium (Ca), magnesium (Mg), sodium (Na), potassium (K) and anions, i.e., bi-carbonate (HCO_3), chloride (Cl), sulfate (SO_4) and nitrate (NO_3). We used the Charge Balance Error (CBE) method to ensure the precision of our analysed samples (CBE) (Freeze & Cherry, 1979):

$$CBE\ ratio = \frac{(\sum Cations - \sum Anions)}{(\sum Cations + \sum Anions)} \times 100\% \tag{1}$$

Most of the investigated samples had concentrations of CBE below 10%.

The flowchart of the methodology, which illustrates the processes involved in characterising GWQ, is shown in the methodology section. This flowchart outlines the step-by-step process for spatial mapping of the WQI index and statistical analyses of GWQ parameters using PCA, HCA, and a correlation matrix.

Calculation of Water Quality Index

The WQI is a common approach to express significant quantities of GWQ data as a single numerical value. It indicates the parameters whose index values express the complete water quality. This indexing is suitable for analysing the quality of drinking and irrigation water. The criteria for WQI mapping are presented in Table 1. Brown et al. (1972) formulated the weighted arithmetic indexing (WAI) method, which is often used for the computation of WQI. The following expression has been applied for the calculation of WQI:

$$WQI = \frac{\sum_{i=1}^{n} qiWi)}{(\sum_{i=1}^{n} Wi)} \tag{2}$$

Table 1: Criteria for WQI mapping (Prashanthi et al., 2004)

S. No.	WQI index class	Status
1.	I (0-25)	Excellent
2.	II (26-50)	Good
3.	III (51-75)	Poor
4.	IV (76-100)	Very Poor
5.	V (> 100)	Unfit for consumption

Multi-variate Statistical Techniques for Interpretation of Groundwater Quality

In this study, box and whisker plots of GWQ parameters were created, and multivariate statistical methods, such as PCA and HCA, were applied to determine the chemical properties of groundwater. The relationship among quality indicators of groundwater samples was represented graphically with these tools. These analyses were performed using SPSS 26.0 (IBM, 2020) and XLSTAT software (XLSTAT, 2020).

Principal component analysis

Principal Component Analysis (PCA) is the most commonly used multivariate statistical method for interpreting GWQ parameters. It is typically used to compress water quality datasets spanning multiple dimensions, reducing jitter and redundancy to facilitate efficient analysis. PCA approaches datasets consisting of several correlated components by portraying them as smaller sets of independent, uncorrelated variables. It incorporates data in a correlation matrix and reorganises it in a system that can improve the interpretability of the underlying data structure. The process of PCA commences with the generation of a novel set of GWQ variables (called principal components or PCs) based on linear combinations of variables belonging to the original datasets. Generally, the entire PCA process can be divided into two steps, viz., standardisation of data and extraction of PCs. Initially, measured water quality data (X_{ji}) were standardised by Z-scale transformation using the following formula (Kawo & Karuppannan, 2018):

$$Z = \frac{X_{ji} = \overline{X}_j}{S_j} \tag{3}$$

Where, X_{ji} is a value of j^{th} GWQ parameters measured at i^{th} location; X_j is the mean value of j^{th} parameter; and S_j is the standard deviation of the j^{th} parameter.

Correlation matrix

The variance proportion of one GWQ parameter explained by its relationship with another GWQ parameter may be measured in terms of a correlation coefficient. The correlation coefficient varies between -1 and +1, demonstrating extreme dissimilarity and similarity, respectively, while a correlation coefficient of 0 denotes the absence of any relationship between the variables. This study developed a correlation matrix for all GWQ parameters to understand their inter relationships.

Hierarchical cluster Analysis

Cluster analysis (CA) involves dividing an observed dataset into different clusters or groups based on their similarities, as measured by their respective correlation coefficients (Davis, 2002). CA is a popular tool for evaluating the potential to group GWQ variables across various samples based on their similarities in hydrochemical characteristics (Cloutier et al., 2008; Zaki et al., 2018). The water quality variables typically used in CA include percentages of major ions, pH, and/or salinity levels (Van & Hodgson, 1986; Ground & Groeger, 1994). In this study, Hierarchical Cluster Analysis (HCA) was used as a data classification tool to group similar chemical parameters. According to the literature, HCA is the most commonly employed method among the different clustering techniques used in environmental sciences (Davis, 2002). The primary objective of applying HCA was to cluster multiple parameters into a single group. The Euclidean distance method was used to measure similarities and differences among selected variables, i and j, which were calculated as follows (Davis, 2002):

$$d_{ij}^2 = \sum_{k=1}^{m} (Z_{(i,k)} - Z_{(j,k)}) \tag{4}$$

Where, d_{ij} is the Euclidean distance; Z_{ik} and Z_{jk} are the variables, k for objects i and j, respectively, and m is the number of variables.

The clustering is depicted graphically by a dendrogram derived from the application of the Euclidean distance approach and Ward's method. The Ward's method evaluates distances between clusters using an analysis of variance technique, aiming to reduce the sum of squares of any two clusters (hypothetical), which may be computed at each stage (Machiwal et al., 2011). A short distance indicates the closeness of two chemical parameters, while a high distance indicates dissimilarity between the parameters.

Results And Discussion

GIS-based Water Quality Indexing

WQI maps of pre-monsoon and post-monsoon seasons were generated using ArcGIS 10.5 software based on pre-selected GWQ parameters and the same were classified into different GWQ categories, i.e., 'excellent' (Class I), 'good' (Class II), 'poor' (Class III), 'very poor' (Class IV) and 'unfit for consumption' (Class V) for all individual sites (Fig. 2a). The WQI for Jakham River basin in pre-monsoon season with 76 sites was analysed and categorized, as per BIS and/or WHO standards. The computed WQIs ranged from 30 to 105 within the basin during the pre-monsoon season. The analysis indicated that none of the sampling sites belonged to the Class I category. The GWQ maps highlighted that during the pre-monsoon season, 603.705 km^2, i.e., 63.42% of the total study area, possessed 'good' water quality, followed by 326.02 km^2 (34.21% of the total study area) falling under the 'poor' category of GWQ. Furthermore, it was found that a minor proportion (2.21% of the total study area) of the area possessed the WQI under the 'very poor' category and a negligible portion (0.161% of the total study area) under the category of 'unfit for consumption'. The TDS in groundwater was identified as the most active GWQ parameter, followed by pH, EC, and Na during the pre-monsoon season.

The GWQ, during the post-monsoon season, is presented in Fig. 2(b). Similar to the pre-monsoon season, there was no sampling site found under the 'excellent' category during the post-monsoon season. About 490.89 km^2 (51.51%) of the basin was observed under 'poor'

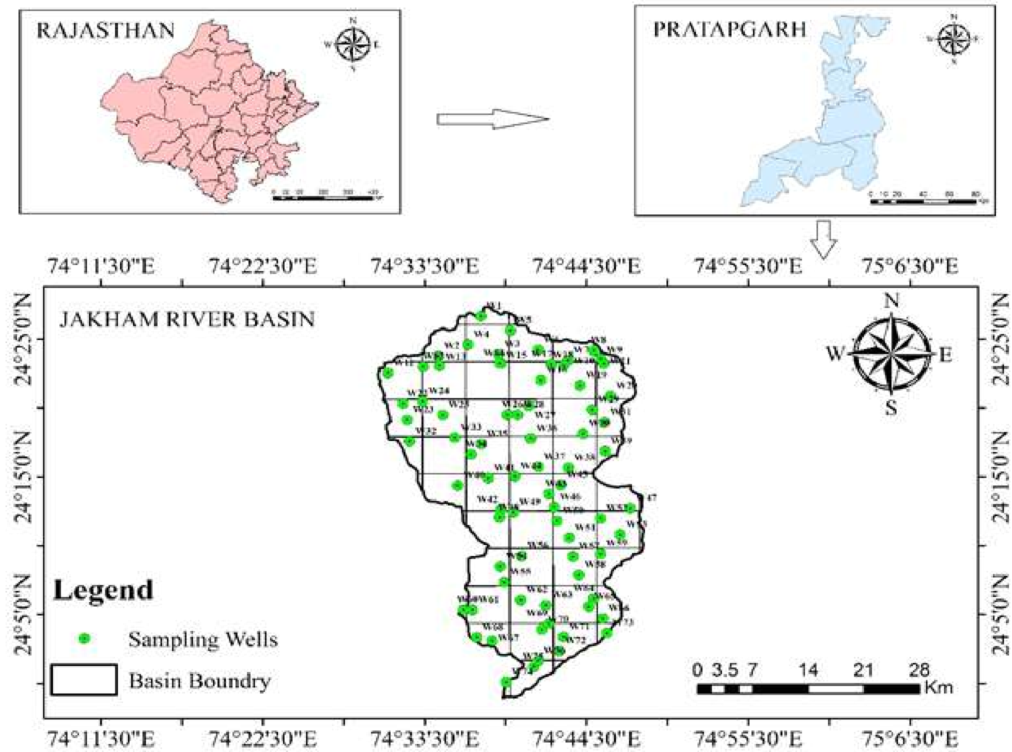


Fig. 1: Location map of Jakham River Basin in southern Rajasthan

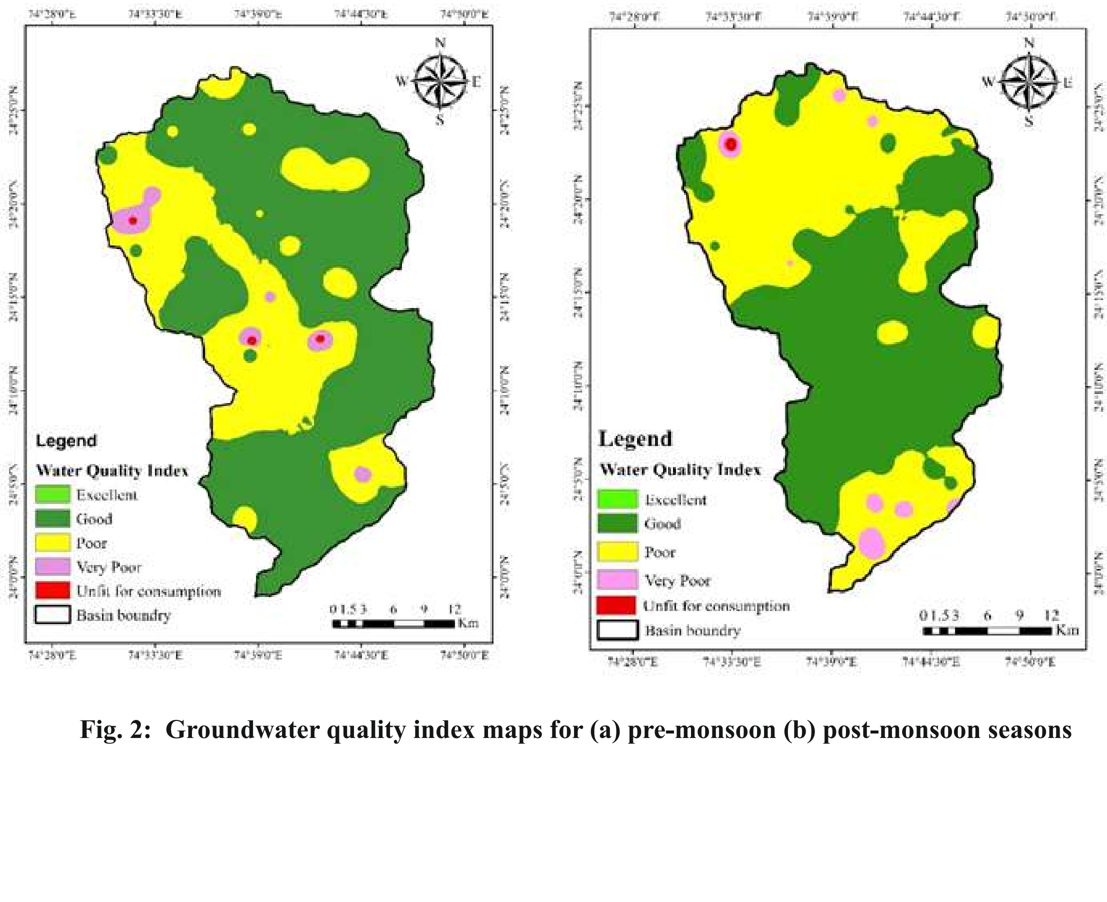


Fig. 2: Groundwater quality index maps for (a) pre-monsoon (b) post-monsoon seasons

Table 2: Inter-correlations matrix of different water quality parameters for pre-monsoon season

Parameter	pH	EC	TDS	Na	K	Ca	Mg	Cl	SO_4	CO_3
pH	1									
EC	-0.011	1								
TDS	-0.090	**0.913****	1							
Na	0.020	**0.895****	**0.885****	1						
K	0.032	0.678**	**0.811****	0.660**	1					
Ca	-0.299	**0.724****	**0.898****	0.668**	0.743**	1				
Mg	0.083	**0.848****	**0.851****	**0.777****	0.725**	0.614*	1			
Cl	0.549	**0.851****	**0.913****	**0.880****	**0.853****	**0.768****	**0.830****	1		
SO_4	0.661	0.490	0.608*	0.399	0.484	0.661**	0.421	0.545*	1	
CO_3	0.269	-0.005	0.186	0.135	0.382	0.123	0.112	0.173	0.212	1
HCO_3	-0.470	**0.774****	**0.783****	0.727**	0.470	0.732**	0.677**	0.582*	0.182	-0.132
NO_3	-0.015	0.694**	**0.797****	0.586*	0.732**	0.739**	0.710**	0.714**	0.633*	0.450
F	-0.319	-0.171	-0.329	-0.180	-0.265	-0.310	-0.179	-0.355	-0.646**	-0.523*

Note: ** Correlation is significant at the 0.05 level (2-tailed) and **Correlation is significant at the 0.01 level (2-tailed). The significance of "bold" emphasis that values are showing strong correlation between the corresponding parameters.*

GWQ, followed by 400.54 km^2 (42.02%) area under 'good' GWQ. A smaller proportion of the study area is found to be under the GWQ categories of 'very poor' (6.30%) and 'unfit for consumption' (0.16%). The central portion of the basin appears to have the highest concentration of minerals of desirable quality. The final WQI indicates a larger area (63.42%) under a 'good' GWQ rating in the pre-monsoon season, which further decreased to 42.02% in the post-monsoon season, due to the combined effects of geogenic and climatic factors. The Na and Ca contents were found to be active parameters for drinking purposes in the post-monsoon season.

Multi-variate Statistical Analysis

A correction matrix of 13 GWQ parameters was generated to understand the relationships among the selected parameters (Table 2), which were used as independent variables in modelling while characterising the GWQ. The TDS was found to be strongly correlated (correlation coefficient over 0.9) with Cl and EC (Table 2). Likewise, GWQ parameters such as TDS showed good correlations with Na, K, Ca, Mg, and HCO_3. Additionally, Cl had a strong correlation with Na, K, Ca, and Mg, with correlation coefficients exceeding 0.75. Conversely, pH was poorly correlated with Cl and SO_4. It was difficult to organise the parameters into components and assign some physical significance at this stage, since some parameters, such as F, SO_4, and NO_3, did not have significant correlations with any other parameters. Therefore, the correlation matrix was further analysed using PCA.

During the post-monsoon season, TDS was found to have a good correlation with Na, Ca, and Mg (Table 3). Additionally, EC showed a good correlation with Cl. Similar to the pre-monsoon season, it was not easy to classify the GWQ parameters into components during the post-monsoon season and assign some physical significance at this stage, as some parameters, such as pH, F, SO_4, and NO_3, HCO_3, and CO_3, did not have significant correlations with any other parameters. Therefore, the results of the correlation matrix were subjected to PCA.

Principal Component Analysis

Adequacy of the GWQ data, prior to PCA, was verified with the help of K-M-O and Barlett's tests (Table 4). MATLAB 2020 software was used for the above analysis. In the pre-monsoon and post-monsoon seasons, the test-statistic values were found to be 0.702 and 0.698, respectively, which are within the acceptable limit, indicatingthat the data are adequate for PCA.

The PCA was applied to the correlation matrix of GWQ parameters, which consisted of 13 physico-chemical parameters. Its purpose was to determine the individual PC loadings of each of the 13 variables that affect GWQ characteristics. Eigenvalues (i_g) are frequently employed in order to derive the significant principal components (PCs). The i_g value of a relevant variable defines its peak value. The Eigenvalues of a magnitude more than 1 indicate the significantly important PCs having considerable contributions to the total variations of the system. The PCs, having i_g less than 1, were discarded from further analysis because of their lack of significance (Muangthong & Shrestha, 2015). As of the 5^{th} i_g value, the slope of the scree plot during both seasons gradually becomes flatter. Hence, only the first four PCs were considered significant in this study, explaining 91.30% of the cumulative variance during the pre-monsoon and 70.44% of the cumulative variance during the post-monsoon season.

During the pre-monsoon season, the four PCs, extracted based on eigenvalues greater than 1, explained 57.92%, 17.38%, 7.77%, and 8.23% of the total variance, respectively (Table 5). Each PC contains some strong positive, negative and near-zero factor loadings. The first PC explains approximately 57.92% of the total variance and has strong factor loadings for EC, TDS, Na, Mg, and Cl, a moderate loading for K, HCO_3, NO_3, and Ca, and weak loading values for SO_4. Hence, the PC1 of the 1^{st} PC is referred to as the 'salinity' factor with reference to strong loadings between Na and Cl ions. The moderate factor loadings between K and NO_3indicate the incorporation of chemical fertilisers and animal waste into agricultural activities in the study area (Adam et al., 2001). It was observed that the sanitation network was not seen during the sampling visits in the study area, and

Table 3: Inter-correlations matrix of different water quality parameters for post-monsoon season

Parameter	**pH**	**EC**	**TDS**	**Na**	**K**	**Ca**	**Mg**	**Cl**	**SO_4**	**CO_3**	**HCO_3**	**NO_3**	**F**
pH	1												
EC	-0.125	1											
TDS	0.071	0.404	1										
Na	-0.018	-0.052	**0.785***	1									
K	0.181	-0.332	-0.225	-0.004	1								
Ca	0.003	0.266	**0.821***	0.699*	0.363	1							
Mg	-0.104	0.548*	0.**776***	0.**740***	-0.509	0.075	1						
Cl	0.135	**0.809****	**0.717***	0.**803****	-0.165	0.693*	0.544	1					
SO_4	0.206	0.791**	-0.479	0.098	0.694*	-0.165	-0.694**	-0.584*	1				
CO_3	-0.069	-0.274	0.609	0.037	0.156	0.111	-0.061	-0.219	0.229	1			
HCO_3	-0.096	-0.389	-0.052	0.620	0.004	0.094	-0.131	-0.286	0.355	0.215	1		
NO_3	-0.144	0.208	0.568	0.150	0.017	0.387	0.150	0.215	-0.233	-0.575*	-0.181	1	
F	-0.004	-0.340	-0.253	0.059	-0.004	-0.755**	0.027	-0.613*	0.172	0.247	-0.196	-0.474	1

Table 4: Results of K-M-O and Barlett's tests

K-M-O and Bartlett test		Pre-monsoon	Post-monsoon
K-M-O adequacy		0.702	0.698
Bartlett's test spherecity	Chi-square	855.23	789.95
	Degree of freedom	59	59
	Significance	0	0

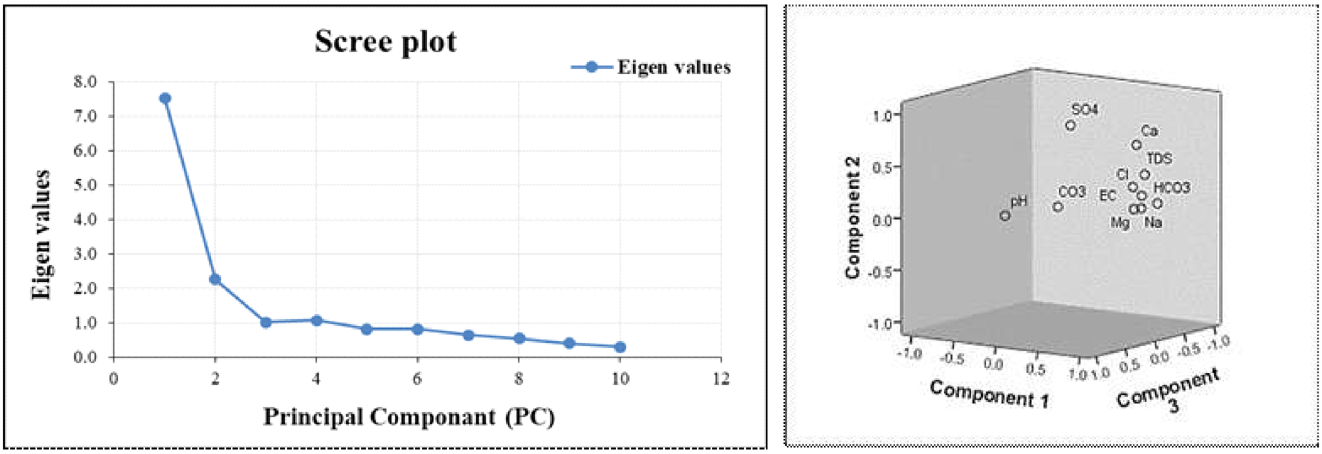
Table 5: Factor loadings of significant principal components for pre-monsoon season

Water quality parameter	Factor loadings of principal component			
	1	2	3	4
pH	-0.003	0.215	-0.953	0.096
EC	0.942	0.176	0.034	-0.097
TDS	0.912	0.338	0.184	0.100
Na	0.925	0.079	-0.018	0.030
K	0.772	0.204	0.057	0.401
Ca	0.697	0.500	0.460	0.056
Mg	0.919	0.080	-0.080	0.071
Cl	0.896	0.278	-0.017	0.129
SO_4	0.377	0.880	-0.151	0.003
CO_3	0.047	0.179	-0.123	0.947
HCO_3	0.779	-0.045	0.496	-0.173
NO_3	0.665	0.416	0.151	0.412
F	-0.046	-0.795	0.170	-0.376
Eigenvalue	7.53	2.26	1.01	1.07
Variance (%)	57.92	17.38	7.77	8.23
Cumulative variance (%)	57.92	75.30	83.07	91.30

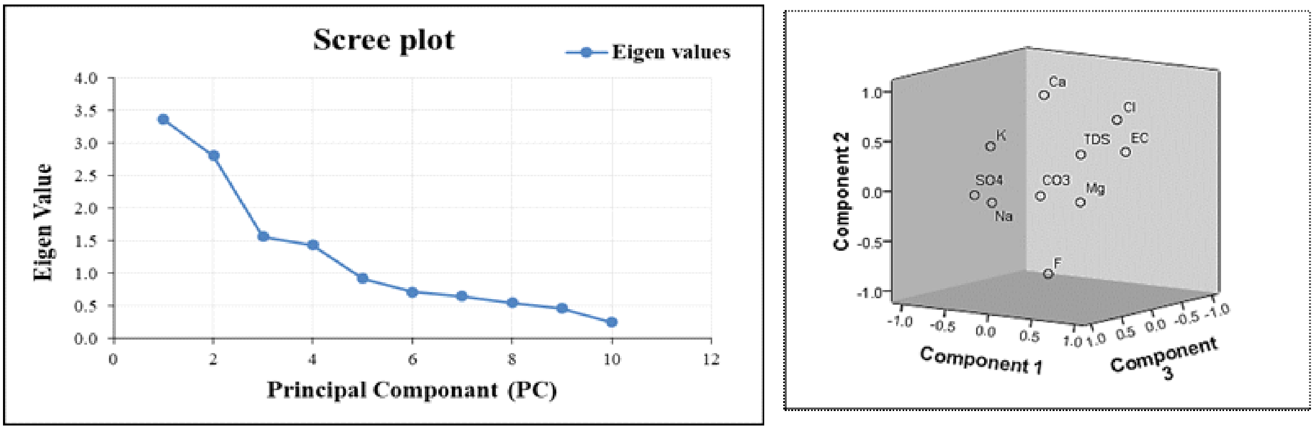
hence, there were instances in which untreated domestic sewage was directly discharged into the aquifers.

The 2nd PC, or PC2, explained nearly 17.38% of the total variance and was attributed to a strong positive loading for SO4, which corresponded to anthropogenic activities in the field, such as crop fertilisation and other land-use activities, and a weak loading for NO_3. Likewise, the 4th PC, or PC4, explains 8.23% of the total variance, along with strong positive loadings for TDS and CO_3. The PCA highlighted the order of importance of parameters, *viz,.* TDS> pH> EC> Na> Ca> Cl > Mg > CO_3> SO_4> HCO_3 in the pre-monsoon season.

In the post-monsoon season, the first four PCs explained 25.85%, 21.55%, 12.02%, and 11.02% of the



(a)



(b)

Fig. 3: PCA of groundwater quality parameters forpre-monsoon (a) and post-monsoon (b)

total variance, with a total cumulative variance value of 70.44% (Table 6). The results of the PCA further indicated that PC1 was found associated with strong loading values for Mg and NO_3, moderate loading values for EC and TDS, and weak loading values for Cl (Table 5), which were mostly distributed between the upper and central portions of the study area. The PC1 revealed that rock-water interaction with ion exchange was responsible for the geogenic hydro-geochemical evolution of groundwater (Bodrud-Doza et al., 2016). Das et al. (2019) have also reported that the origin of salinity (presence of Na-Cl) in croplands was primarily due to the use of chemical fertilisers, animal waste and industrial pollutants. Moreover, this factor also has moderate loadings for K and NO_3, indicating the cultivation activities and industrial disposals occurring near the water source in the study area.

The 2nd PC, or PC2, explained nearly 21.55% of the entire variance (Table 5). The Ca content exhibited strong positive loading values, which indicated the influence of 'hardness' associated with the presence of carbonates in groundwater. Hence, the groundwater was rendered as unfit for drinking and irrigation purposes, and it was also responsible for weak loading of NO_3. The 3rd PC had positive factor loadings for Na and HCO_3, which explained only 12.02% of the entire variation. This PC is usually associated with the seepage of untreated sewage water into the groundwater. The 4th PC explained 11.02% of the entire variation, along with a strong positive loading for CO_3.

Hierarchical Cluster Analysis

Dendrograms illustrating results of HCA for groundwater quality parameters during pre-monsoon and post-monsoon seasons are depicted in Figs.4 and 5, respectively. In the HCA, groundwater quality parameters with higher degrees of similarity were assigned to the first cluster. Based on Fig. 4, it is evident that two main clusters were formed for classifying the groundwater quality parameters during the pre-monsoon season. The first cluster consisted of four parameters, namely EC, TDS, Na, and Ca, which may be influenced by various sources, including over-pumping of groundwater, dissolution of alkaline rocks, and leaching of fertilisers from the soil horizon to the aquifer. The second cluster, which consisted of three parameters, i.e., K, Cl and Mg, is characterised by anthropogenic sources such as agricultural operations, sewage waste and drainage water infiltration from bleaching industries. It also became apparent that pH, F and SO_4 could not be clustered properly with other clusters during the pre-monsoon season.

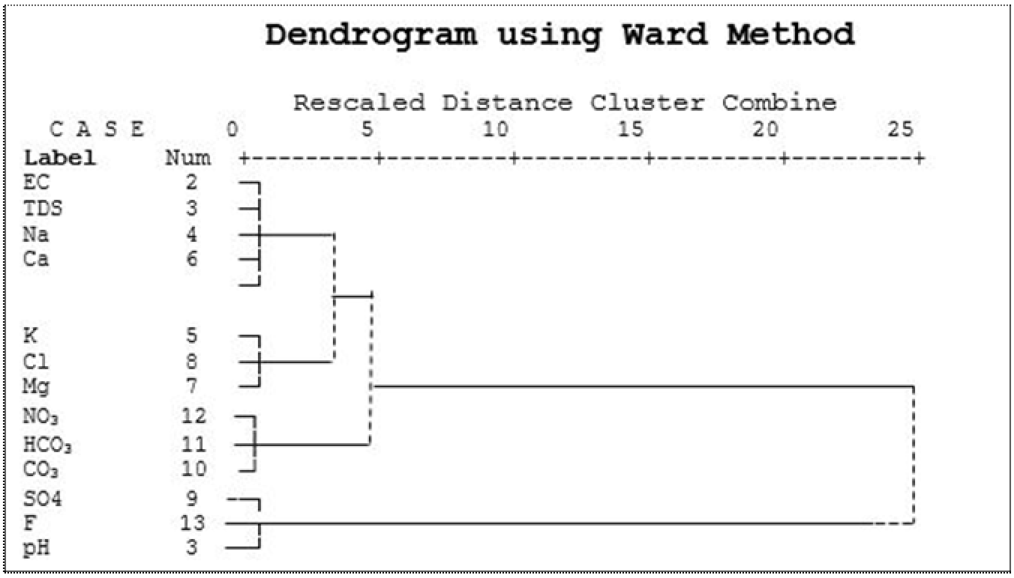


Fig. 4: Dendrogram presenting clustering of groundwater quality parameters for pre-monsoon

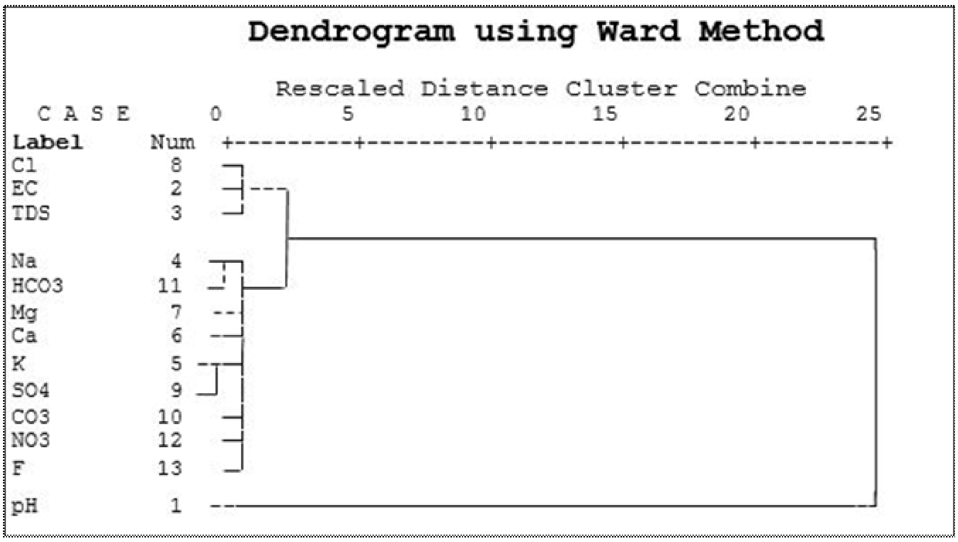


Fig. 5: Dendrogram presenting clustering of groundwater quality parameters for Post-monsoon

Table 6: Factor loadings of significant principal components for post-monsoon season

Water quality parameter	Factor loadings of principal component			
	1	2	3	4
pH	-0.182	0.116	-0.245	0.176
EC	0.761	0.360	-0.307	-0.166
TDS	0.556	0.558	0.205	0.342
Na	0.149	0.020	0.852	-0.031
K	-0.717	0.291	-0.065	0.104
Ca	-0.035	0.953	0.065	0.056
Mg	0.857	0.055	0.397	0.017
Cl	0.476	0.641	-0.449	-0.095
SO_4	-0.898	-0.172	0.185	0.139
CO_3	-0.100	0.031	0.149	0.925
HCO_3	-0.338	0.101	0.619	0.164
NO_3	0.100	0.439	0.181	-0.734
pH	-0.182	0.116	-0.245	0.176
Eigenvalue	3.36	2.80	1.56	1.43
Variance (%)	25.85	21.55	12.02	11.02
Cumulative variance (%)	25.85	47.40	58.42	70.44

Similarly, during the post-monsoon season, two main clusters emerged among the groundwater quality parameters. The first cluster consisted of three parameters, viz., EC, TDS and Cl, whereas the second cluster consisted of nine parameters among which Na and HCO_3, K and SO_4 were found to be closely related (Fig. 5). Based on the observations made through both dendrograms and correlation matrix, the TDS and EC parameters were found to be strongly correlated with respect to Ca, Na and Cl. The pH parameter, however, did not demonstrate any significant association with any of the groundwater quality parameters in the post-monsoon season. The factors that primarily influenced the first cluster include salinity due to mineral dissolution, high evaporation, and flushing of evaporated minerals from sedimentary rocks. In contrast, the second cluster can be attributed to the dissolution of agricultural waste, such as inorganic fertilisers and anthropogenic activities, in the study area (Loganathan & Ahamed, 2017).

Conclusions

This study attempts to characterise and interpret the groundwater quality of the Jakham River basin in southern Rajasthan, India, using an integrated approach of multivariate statistical analyses and geographic information system (GIS). Results of GIS-based water quality index (WQI) revealed that 63% of the groundwater samples of the pre-monsoon season and 42% of the post-monsoon season were classed under the 'good' category and found satisfactory for human consumption. Nevertheless, imparting a spatial sense to the water quality of local aquifers through spatial mapping of WQI highlighted the higher WQI values in the northern and central portions of the study area, whereas poor groundwater quality was observed in the lowland areas. Results of principal component analysis (PCA) indicated that four principal components were identified as significant, accounting for 91.30% and 70.44% of the total variance in the pre-monsoon and post-monsoon seasons, respectively; these four components acted as significant quality control factors. The findings of the PCA showed that the changes in the physicochemical properties of the groundwater strata are caused by both anthropogenic (i.e., excess fertiliser application and industrial waste) and geogenic factors (i.e., rock-water interaction). In addition, the outcomes of the correlation matrix showed a strong stake to the conclusions generated by PCA and hierarchical cluster analysis. Moreover, the findings of this study may be of service to water resource planners and policymakers in prioritising and safeguarding the groundwater supply from contamination, as well as in developing technology that maintains groundwater quality suitable for drinking.

Disclosure statement
No conflict of interest to declare.

Funding: Not applicable

Authors' contributions:
Vinay Kumar Gautam: Conceptualization, Investigation, Resources, Data Curation, Writing Original Draft, Formal Analysis, and development of methodology.
K.K Yadav: Supervision, writing review editing.
Mahesh Kothari: Supervision, writing review editing.
P. K. Singh: Supervision, writing review editing.
Ethics approval: Not applicable
Consent to participate: Not applicable.
Consent for publish: Not applicable.
Competing interests: The authors declare no competing interests.

References

Abdelaziz S., Gad M. I. and El Tahan A. H. M. (2020). Groundwater quality index based on PCA: wadi El-Natrun, Egypt. *Journal of African Earth Sciences*, *172*, 103964.Acikel, S. & Ekmekci, M. (2018). Assessment of groundwater quality using multivariate statistical techniques in the Azmak Spring Zone, Mugla, Turkey. *Environmental earth sciences*, *77(*22), 753.

Barkat A., Bouaicha F., Mester T., Debabeche M. and Szabó G. (2022). Assessment of spatial distribution and temporal variations of the phreatic groundwater level using geostatistical modelling: the case of oued souf valley—southern East of Algeria. *Water*, *14*(9), 1415.

Batayneh A. T. and Al-Taani A. A. (2016). Integrated resistivity and water chemistry for evaluation of groundwater quality of the Gulf of Aqaba coastal area in Saudi Arabia. *Geosciences Journal*, *20*, 403-413.

Bodrud-Doza M. D., Islam A. T., Ahmed F., Das S., Saha N. and 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.

Bouteraa O., Mebarki A., Bouaicha F., Nouaceur Z. and 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*, 796-814.

Brown R. M., McClelland N. I., Deininger R. A. and O'Connor M. F. (1972). A water quality index—crashing the physiological barrier. Indic Environ Qual.

BIS (2012). Indian Standard Drinking Water Specification,1-100.

CGWB (2017). *Concept Note on Geogenic Contamination of Ground Water in India.* Report of the Central Ground Water Board (CGWB). Ministry of Water Resources, Government of India, pp. 78. https://cgwb.gov.in/cgwbpnm/public/uploads/documents/1686055710748531399 file.pdf (accessed on 12 Oct 2023).

Cloutier V., Lefebvre R., Therrien R. and Savard M. M. (2008). Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*, *353*(3-4), 294-313.

Das N., Mondal P., Ghosh, R. and Sutradhar S. (2019). Groundwater quality assessment using multivariate statistical technique and hydro-chemical facies in Birbhum District, West Bengal, India. *SN Applied Sciences*, *1*, 1-21.

Davis J.C. (2002). *Statistics and Data Analysis in Geology*. John Wiley & Sons Inc., NY.

Freeze R. A. and CherryJ.A. (1979). *Groundwater*. Prentice-Hall Inc., Englewood Cliffs , NJ.

Gauns A., Nagarajan M., Lalitha R and Baskar M. (2020). GIS-based assessment of groundwater quality for drinking and irrigation by water quality index. *Int J CurrMicrobiolApplSci*, *9* (1), 2361–2370.https:// doi. org/ 10. 20546/ ijcmas. 2020. 903. 269.

Gautam V. K., Kothari M., Singh P. K., Bhakar S. R. and Yadav, K. K. (2021). Determination of geomorphological characteristics of Jakham River Basin using GIS technique. *Indian Journal of Ecology*, *48*(6), 1627-1634.

Gautam V.K., Mahesh K., Singh P. K., Bhakar S. R. and Yadav K. K. (2022a). Analysis of groundwater level trend in Jakham river basin of southern Rajasthan, *J. Groundwater. Sci. Eng, 10* (1), 1–9.

Gautam V. K., Kothari M., Singh P. K., Bhakar S. R. and Yadav K. K. (2022b). Spatial mapping of groundwater quality using GIS for Jakham River basin of Southern Rajasthan. *Environment Conservation Journal*, *23*(1&2), 234-243.

Gautam V. K., Kothari M., Singh P. K., Bhakar S. R. and Yadav K. K. (2022c). Decadal groundwater level changes in Pratapgarh district of southern Rajasthan, India. *Ecology Environment & Conservation*, *28*(1), 283-289.

Gautam V. K., Pande C. B., Kothari M., Singh P. K. and Agrawal A. (2023). Exploration of groundwater potential zones mapping for hard rock region in the Jakham river basin using geospatial techniques and aquifer parameters. *Advances in Space Research*, *71*(6), 2892-2908.

Ground T. A. and Groeger A. W. (1994). Chemical classification and trophic characteristics of Texas reservoirs. *Lake and Reservoir Management*, *10*(2), 189-201.Hamma, B., Alodah, A., Bouaicha, F., Bekkouche, M. F., Barkat, A. & Hussein, E. E. (2024). Hydrochemical assessment of groundwater using multivariate statistical methods and water quality indices (WQIs). *Applied Water Science*, *14*(2), 33. https://doi.org/10.1007/s13201-023-02084-0.

Herojeet R., Rishi M. S., Lata R. and Sharma R. (2016). Application of environmetrics statistical models and water quality index for groundwater quality characterization of alluvial aquifer of Nalagarh Valley, Himachal Pradesh, India. *Sustainable water resources management*, *2*, 39-53.

IBM (2020). IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp.

Iqbal J., Su C., Rashid A., Yang N., Baloch M. Y. J., Talpur S. A. and Sajjad M. M. (2021). Hydrogeochemical assessment of groundwater and suitability analysis for domestic and agricultural utility in Southern Punjab, Pakistan. *Water*, *13*(24), 3589.

Jankowska J., Radzka E. and Rymuza K. (2017). Principal component analysis and cluster analysis in multivariate assessment of water quality. *Journal of Ecological Engineering*, *18*(2). https:// doi.org/ 10. 12911/ 22998993/6814.

Kawo N. S. and 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. doi.org/ 10. 1016/j. jafrearsci. 2018. 06. 034.

Khan R. and Jhariya D. C. (2017). Groundwater quality assessment for drinking purpose in Raipur city, Chhattisgarh using water quality index and geographic information system. *Journal of the geological society of India*, *90*, 69-76.https:// doi. org/ 10. 1007/ s12594- 017- 0665-0.

Lee J. Y. and Song S. H. (2007). Groundwater chemistry and ionic ratios in a western coastal aquifer of Buan, Korea: implication for seawater intrusion. *Geosciences Journal*, *11*, 259-270.https ://doi.org/10.1007/bf029 13939.

Loganathan K. and Ahamed A. J. (2017). Multivariate statistical techniques for the evaluation of groundwater quality of Amaravathi River Basin: South India. *Applied Water Science*, *7*, 4633-4649.https://doi.org/10.1007/s13201-017-0627-0.

Machiwal D., Jha M. K. and Mal B. C. (2011). GIS-based assessment and characterization of groundwater quality in a hard-rock hilly terrain of Western India. *Environmental monitoring and assessment*, *174*, 645-663.

Mohamed M. M. and Elmahdy S. I. (2015). Natural and anthropogenic factors affecting groundwater quality in the eastern region of the United Arab Emirates. *Arabian Journal of Geosciences*, *8*, 7409-7423. https ://doi.org/10.1007/s1251 7-014-1737-8.

Mohseni U., Patidar N., Pathan A. I., Agnihotri P. G. and Patel D. (2022). An Innovative Approach for Groundwater Quality Assessment with the Integration of Various Water Quality Indexes with GIS and Multivariate Statistical Analysis—a Case of Ujjain City, India. *Water Conservation Science and Engineering*, *7*(3), 327-349.

Muangthong S. and Shrestha S. (2015). Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. *Environmental monitoring and assessment*, *187*, 1-12. https:// doi. org/ 10. 1007/ s10661- 015- 4774-1.

Noshadi M. and Ghafourian A. (2016). Groundwater quality analysis using multivariate statistical techniques (case study: Fars province, Iran). *Environmental monitoring and assessment*, *188*, 1-13.

Omo-Irabor O. O., Olobaniyi S. B., Oduyemi K. and Akunna J. (2008). Surface and groundwater water quality assessment using multivariate analytical methods: a case study of the Western Niger Delta, Nigeria. *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(8-13), 666-673.https://doi. org/10.1016/j.pce.2008.06.019.

Prashanthi M., Ravichandra M. and Veera Bhadram K. (2004). Evaluation of water quality index at Vishapatnam city, Andhra Pradesh. *Nature Env. & Poll. Technology*, *2*: 65-68.

Ram A., Tiwari S. K., Pandey H. K., Chaurasia A. K., Singh S. and Singh, Y. V. (2021). Groundwater quality assessment using water quality index (WQI) under GIS framework. *Applied Water Science*, *11*, 1-20. https:// doi. org/ 10. 1007/s13201- 021- 01376-7.

Ravikumar P., Somashekar R. K. and Angami M. (2011). Hydrochemistry and evaluation of groundwater suitability for irrigation and drinking purposes in the Markandeya River basin, Belgaum District, Karnataka State, India. *Environmental monitoring and assessment*, *173*(1), 459-487. https://doi.org/10.1007/s1066 1-010-1399-2.

Roy B., Roy S., Mitra S. and Manna A. K. (2021). Evaluation of groundwater quality in West Tripura, Northeast India, through combined application of water quality index and multivariate statistical techniques. *Arabian Journal of Geosciences*, *14*, 1-18. https:// doi. org/ 10. 1007/s12517- 021- 08384-6.

Sharma G., Lata R., Thakur N., Bajala V., Kuniyal J. C. and Kumar K. (2021). Application of multivariate statistical analysis and water quality index for quality characterization of Parbati River, Northwestern Himalaya, India. *Discover Water*, *1*, 1-20.

Shil S., Singh U. K. and Mehta P. (2019). Water quality assessment of a tropical river using water quality index (WQI), multivariate statistical techniques and GIS. *Applied water science*, *9*, 1-21.

Teixeira de Souza A., Carneiro L. A. T., da Silva Junior O. P., de Carvalho S. L. and Américo-Pinheiro J. H. P. (2021). Assessment of water quality using principal component analysis: a case study of the Marrecas stream basin in Brazil. *Environmental technology*, *42*(27), 4286-4295.

Tirkey P., Bhattacharya T., Chakraborty S. and Baraik S. (2017). Assessment of groundwater quality and associated health risks: a case study of Ranchi city, Jharkhand, India. *Groundwater for sustainable development*, *5*, 85-100. https:// doi. org/ 10. 1016/j. gsd. 2017. 05. 002.

Van Tonder and Hodgson F.D. (1986). Interpretation of hydrogeochemical facies by multivariate statistical methods. *Water SA.12*: 1-6.

Vishwakarma V. and Thakur L. S. (2012). Multivariate statistical approach for the assessment of groundwater quality in Ujjain City, India. *J Environ Sci Eng*, *54*, 533-543.

XLSTAT (2007.) Statistical Software for Excel.https:// www.xlstat.com

Zaki S. R., Redwan M., Masoud A. M. and Abdel Moneim A. A. (2019). Chemical characteristics and assessment of groundwater quality in Halayieb area, southeastern part of the Eastern Desert, Egypt. *Geosciences Journal*, *23*, 149-164. https: //doi.org/10.1007/s1230 3-018-0020-5.

Received: September 2024 Accepted: December 2024