Multivariate Statistical Analysis for Assessment of Ground Water Quality in Hussainsagar Catchment Area, Hyderabad, India

M. Chandra Sekhar[1], S. Manvitha[2]

Journal of Water Engg. and Management

ISSN 2582 6298
Volume-03
Number- 02

Jr. of Water Engg. and Mgt.
2022, 3(2) : 20-30

Research Paper

Multivariate Statistical Analysis for Assessment of Ground Water Quality in Hussainsagar Catchment Area, Hyderabad, India

M. Chandra Sekhar[1], *S. Manvitha*[2]

[1]Professor, Civil Engineering Department, National Institute of Technology, Warangal, Telangana, India
[2]Former M.Tech Student, National Institute of Technology, Warangal, Telangana, India

Received August 25, 2022 Revised October 20, 2022
Accepted November 11, 2022 Published December 20, 2022

ABSTRACT

Groundwater quality mapping is often a complex problem due to uneven distribution of monitoring stations, variation in sampling frequency, hydrology, water quality, soil chemistry, etc. Numerical modeling is commonly used for mapping groundwater quality. As an alternative, in the present study, multivariate statistical approach is used to interpret ground water quality of Hussainsagar catchment area of Hyderabad, India. Ground water quality data was collected from Telangana State Ground Water Department which consists of analytical results from 2010 to 2016. Fourteen quality parameters are considered for 25 sampling sites for both pre monsoon and post monsoon period separately. The dataset was analysed using Cluster Analysis (CA) and Principle Component Analysis (PCA). CA formed three groups of similarity among the sampling sites representing the different physicochemical characteristics of the study area. Results of PCA showed that, four factors were responsible for the data structure explaining 78% of the total variance of the dataset in pre monsoon period and are identified as soil leaching, industrial, natural and anthropogenic Pollution. Similarly, principle components were identified as responsible for the data structure explaining 80% of the total variance in the post monsoon period and are described as soil leaching, industrial, natural pollution and weathering.

Keywords: Multivariate statistical analysis, Cluster Analysis, Principal Component Analysis, Groundwater quality.

INTRODUCTION

Ground water quality is very important and is of great environmental concern. In India, 50 percent of urban water requirement and 85 percent of domestic water requirements is met by ground water in rural sector. In Hyderabad, 25-30 percent of total water requirement is being met by ground water. Increasing water demands due to rapidly growing urbanization and industrialization is shifting the focus to ground water quantity and quality. Groundwater contamination is primarily caused by disposal of untreated / partially treated domestic and industrial wastewater. Also according to the survey done by Central Ground Water Board (2015-16), Hyderabad has fallen into over exploited category of ground water (Ground water year book, 2015-16). This necessitates shifting the focus on ground water quality analysis. In order to analyse the ground water quality, different approaches have been developed which include development of various water quality indices, mass balance models etc. But, results from these methods do not represent the entire dataset and do not give complete picture of the ground water quality and its pollution sources. These cannot be used to suggest the needed management measures for pollution control. This increased the necessity for developing statistical methods that can process complex dataset providing reliable results.For interpretation of groundwater quality in large basins, different statistical techniques such as cluster analysis (CA) and principal component analysis (PCA) are popularly used. These statistical analyses result in finger printing possible sources of pollution and hence in effective management of water resources (Sridhar et al., 2015). This paper demonstrates the use of multivariate statistical analyses for assessment of groundwater quality in Hussainsagar catchment in Hyderabad, India.

METHODOLOGY

Description of Study Area

The study area considered is Hussainsagar catchment area which is located in Hyderabad district of Tealngana state, India. Catchment area of the Hussainsagar is about 287 km^2, which is divided into six sub catchments namely Kukatpally, Dulapally, Bowenpally, Hussainsagar downstream, Banjarahills and Yusufguda. Most of the area has a rugged terrain and is underlain by pink and grey granites and granite gneisses. In the study area, groundwater levels in pre monsoon to post monsoon period vary from 4 to 15 m. The annual mean maximum and minimum temperature vary from 40 to 14°C and the normal rainfall is 786 mm. Figure 1 indicates the location of study area.

Data Base

Ground water quality data of Hussainsagar catchment area was collected from Telangana State Ground Water Department (TSGWD). The dataset consist of analytical results from 2010 to 2016 monitoring 25 sampling sites. A total of 14 quality parameters were analysed for both pre-monsoon and post-monsoon period separately.

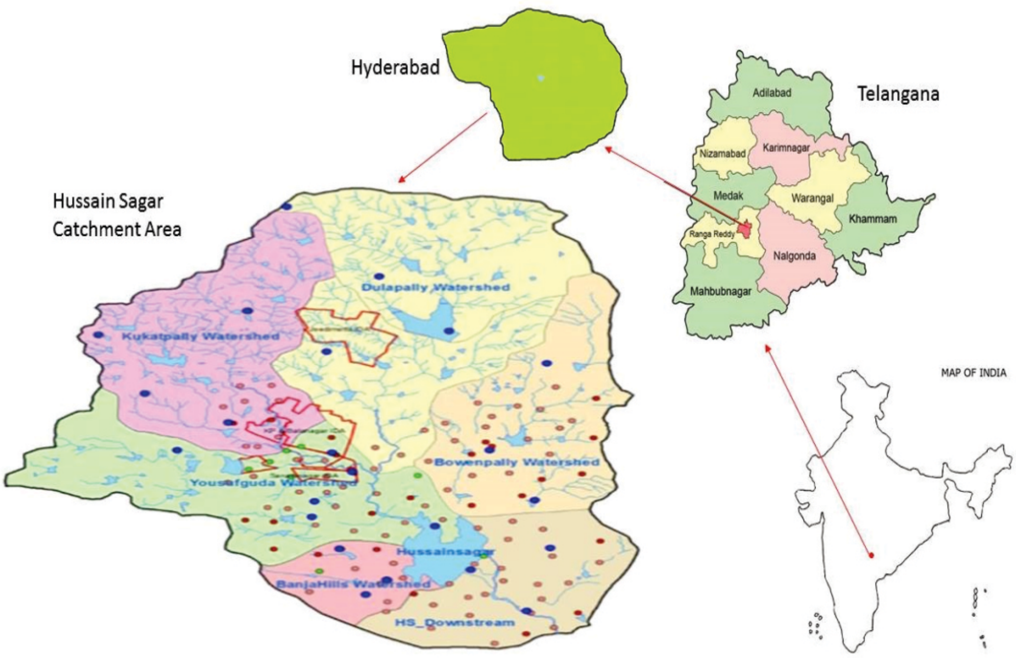

Fig. 1 Index map of study area with sampling locations

Data Processing

Collected data was treated to find the missing values. Based on the percentage and pattern of missing data, different methods were used to fill up the missing values. In order to reduce the effect of large dimensional variability, the concentrations of the dataset are standardized by making the mean of the observations to zero and the standard deviation equal to 1. In brief, standardization of variable concentration as Zij was done by using the formula $Zij = (Xij – X) / (\sigma j)$, where, Xij is the measured concentration of variable j in sample i. X is the arithmetic mean concentration of variable j, and σj is the standard deviation of variable j for all samples included in this analysis (Liu et al., 2003).

Multivariate Statistical Analysis

Cluster analysis is a technique which can be used to group the individuals based on their similarities and dissimilarities. Each group represents a cluster of variables or objects with homogeneous properties separated from a heterogeneous large group (Kim et al., 2005; Singh et al., 2005; Sridhar Kumar et al., 2004). Clustering can be hierarchical or non-hierarchical. Hierarchical method was adopted in this study. For cluster analysis during the study, normalised data is generally used. Ward's method uses squared Euclidean distancesas a measure of similarity.

Principal Component Analysis is a popular pattern recognition method used for reduction of multi dimensionality in dataset (Lucas and Jauzein., 2008; Nair et al., 2010, Shrestha and Kazama, 2006). Here, observed inter correlated variables are transformed to a new set of uncorrelated variables organized into order of decreasing importance which are called Principal components (PCs). The technique uses correlation matrix to characterize dataset which gives a clear picture of how each ground water quality parameter is related to the other quality parameter. Eigen values will be derived for this correlation matrix developed. In PCA, eigen values of magnitude more than 1 were considered. The resulting PCs are sometimes not interpreted readily. So, varimax rotation was performed to generate a new rotated component matrix from the original component matrix. The factor loadings obtained illustrates the correlation between the variables and the factors. The objects with higher loading in each factor were interpreted as significant pollution source (Deepesh et al., 2015).

RESULTS AND DISCUSSIONS

Data Standardization

Before proceeding for the analysis, the dataset was treated for finding the missing data and was filled accordingly. In the pre monsoon data, only one data point was missing. So it was replaced by the geometric mean of the respective variable. In the post monsoon data, missing data percentage was higher. So it was replaced by interpolation techniques. Data was standardized by z-transformation method by making the mean of observations to 0 and standard deviation of the observations to 1.

Cluster Analysis

Cluster Analysis was performed using SPSS 24 (Trial version) software. The dendrogram presented in Fig. 2 indicates the grouping oftwenty-five sampling locations in the study area into three clusters in such a way that the locations in these clusters have identical characteristics. For quick assessment of groundwater quality, one sampling location from each cluster may be considered which excludes considering all monitoring sites. It serves as representative sample for the entire spatial assessment. It is very clear that the cluster analysis is useful in providing reliable categorization of of groundwater in the study area. It can also help in design of optimal sampling strategy by decreasing the number of sampling locations in the study area. Groundwater quality assessment can be less expensive if the sampling locations are reduced without defeating the outcome.

Principal Component Analysis

PCA was attempted for the standardized dataset of 14 variables for pre-monsoon and post monsoon seasons. PCA was applied to find significant principle components and to decrease the contribution of variables with low significance. For PCA, correlation matrix was developed using standardized data and is presented in Tables 1 and 2. Correlation explains the correlation of one parameter with the other. Eigen values are presented in Tables 3 and 4. Eigen values are considered as important tools for determining the number of significant factors. Greater the Eigen value, higher is the significance of respective factors. If Eigen valueis greater than 1, it indicates good correlation (Pekey et al., 2004). For pre monsoon data, four Eigen values 6.726, 1.670, 1.487 and 1.026 are considered as PCs. Similarly for post monsoon matrix, four Eigen values 7.051, 1.896, 1.253 and 1.024 are considered as PCs.

These PCs are subjected to varimax rotation generating factors and the results in Tables 3 and 4 explain 78 % and 80 % of the total variance in pre-monsoon and post monsoon seasons respectively. Similar results were reported by Deepesh et al. (2015) with variance of 75-80% in both the pre-monsoon and post monsoon periods. Varimax rotation is performed to get the factor loadings of each ground water quality parameter and presented in Table 5. Variables displaying higher loading are interpreted contributing to significant pollution source. Component loading values higher than 0.7 are considered as strong; while 0.7 - 0.5 as moderate and 0.5 - 0.3 as weak.

Determination of principal components

PCA results presented in Tables 6 were used to identify the probable sources of pollution in the study area. Four PCs with Eigen values greater than 1 are identified in both the seasons (pre and post monsoon). These PCs are able to explain 78 % and 80 % of the total cumulative variance in the groundwater quality data of pre and post monsoon seasons respectively.

PC 1 exhibits about 41 % and 45 % of the total variance for pre and post monsoon respectively, with strong positive loadings on SC, TDS, TH, Ca, Mg, Cl, and SO_4. The other lodgings are not very significant (Table 5). SC refers to occurrence of dissolved solids. The possible sources of hardness are dissolved multivalent cations from sedimentary rocks, infiltration and run off (Chen et al., 2013). Ca and Mg are present in many sedimentary rocks.

SO4 can also be found from soils and rocks. Sources of hardness and total dissolved solids can also be attributed to industrial pollution. Thus, this component can be interpreted as one kind of mixed pollution, i.e. soil leaching and industrial pollution.

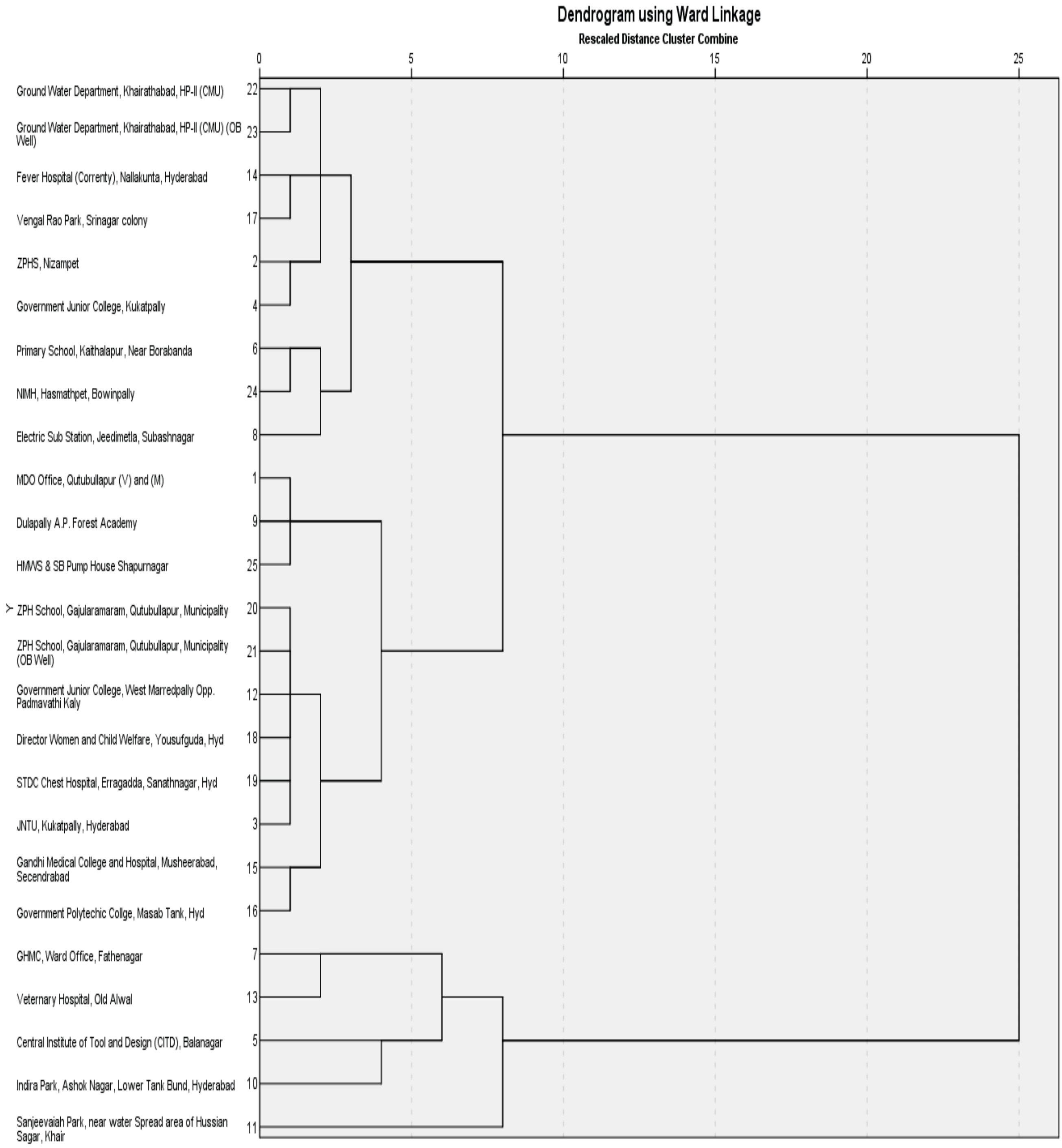

Fig. 1 Dendogram showing clustering of sampling site

Table 1 Correlation matrix for Pre-Monsoon Period

	pH	SC	TDS	CO_3^{2-}	HCO_3^-	Cl^-	F^-	NO_3^-	SO_4^{2-}	Na^+	K^+	Ca^{2+}	Mg^{2+}	TH
pH	1.000	-0.477	-0.477	0.696	-0.501	-0.419	0.151	-0.065	-0.280	-0.208	0.151	-0.499	-0.364	-0.532
SC	-0.477	1.000	1.000	-0.298	0.571	0.926	-0.087	0.309	0.695	0.742	0.188	0.698	0.777	0.868
TDS	-0.477	1.000	1.000	-0.298	0.571	0.926	-0.087	0.309	0.695	0.742	0.188	0.698	0.777	0.868
CO_3^{2-}	0.696	-0.298	-0.298	1.000	-0.349	-0.277	0.099	-0.148	-0.230	-0.080	0.130	-0.321	-0.283	-0.364
HCO_3^-	-0.501	0.571	0.571	-0.349	1.000	0.318	0.121	-0.110	0.053	0.560	-0.019	0.268	0.431	0.394
Cl^-	-0.419	0.926	0.926	-0.277	0.318	1.000	-0.154	0.215	0.677	0.614	0.203	0.705	0.748	0.859
F^-	0.151	-0.087	-0.087	0.099	0.121	-0.154	1.000	0.034	-0.248	0.047	-0.009	-0.176	-0.063	-0.157
NO_3^-	-0.065	0.309	0.309	-0.148	-0.110	0.215	0.034	1.000	0.270	0.135	0.151	0.275	0.282	0.331
SO_4^{2-}	-0.280	0.695	0.695	-0.230	0.053	0.677	-0.248	0.270	1.000	0.485	0.127	0.564	0.469	0.628
Na^+	-0.208	0.742	0.742	-0.080	0.560	0.614	0.047	0.135	0.485	1.000	0.124	0.127	0.495	0.321
K^+	0.151	0.188	0.188	0.130	-0.019	0.203	-0.009	0.151	0.127	0.124	1.000	0.026	0.145	0.085
Ca^{2+}	-0.499	0.698	0.698	-0.321	0.268	0.705	-0.176	0.275	0.564	0.127	0.026	1.000	0.381	0.904
Mg^{2+}	-0.364	0.777	0.777	-0.283	0.431	0.748	-0.063	0.282	0.469	0.495	0.145	0.381	1.000	0.739
TH	-0.532	0.868	0.868	-0.364	0.394	0.859	-0.157	0.331	0.628	0.321	0.085	0.904	0.739	1.000

Table 2 Correlation matrix for Post-Monsoon Period

	pH	SC	TDS	CO_3^{2-}	HCO_3^-	Cl^-	F^-	NO_3^-	SO_4^{2-}	Na^+	K^+	Ca^{2+}	Mg^{2+}	TH
pH	1.000	-0.382	-0.382	0.422	-0.529	-0.244	-0.166	-0.072	-0.265	-0.205	-0.064	-0.373	-0.332	-0.391
SC	-0.382	1.000	1.000	-0.137	0.448	0.931	-0.187	0.342	0.850	0.712	0.177	0.866	0.825	0.929
TDS	-0.382	1.000	1.000	-0.137	0.448	0.931	-0.187	0.342	0.850	0.712	0.177	0.866	0.825	0.929
CO_3^{2-}	0.422	-0.137	-0.137	1.000	-0.265	-0.060	0.015	-0.181	-0.194	-0.029	-0.009	-0.217	-0.021	-0.163
HCO_3^-	-0.529	0.448	0.448	-0.265	1.000	0.149	0.261	-0.013	0.159	0.626	0.043	0.183	0.331	0.255
Cl^-	-0.244	0.931	0.931	-0.060	0.149	1.000	-0.294	0.277	0.812	0.535	0.183	0.891	0.787	0.932
F^-	-0.166	-0.187	-0.187	0.015	0.261	-0.294	1.000	-0.081	-0.290	0.077	-0.313	-0.280	-0.212	-0.279
NO_3^-	-0.072	0.342	0.342	-0.181	-0.013	0.277	-0.081	1.000	0.213	0.183	0.171	0.278	0.396	0.348
SO_4^{2-}	-0.265	0.850	0.850	-0.194	0.159	0.812	-0.290	0.213	1.000	0.540	0.099	0.822	0.634	0.825
Na^+	-0.205	0.712	0.712	-0.029	0.626	0.535	0.077	0.183	0.540	1.000	0.067	0.326	0.450	0.402
K^+	-0.064	0.177	0.177	-0.009	0.043	0.183	-0.313	0.171	0.099	0.067	1.000	0.167	0.156	0.178
Ca^{2+}	-0.373	0.866	0.866	-0.217	0.183	0.891	-0.280	0.278	0.822	0.326	0.167	1.000	0.648	0.958
Mg^{2+}	-0.332	0.825	0.825	-0.021	0.331	0.787	-0.212	0.396	0.634	0.450	0.156	0.648	1.000	0.839
TH	-0.391	0.929	0.929	-0.163	0.255	0.932	-0.279	0.348	0.825	0.402	0.178	0.958	0.839	1.000

Table 3 Total Variance Explained (Pre-Monsoon)

Component Number	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.726	48.045	48.045	5.686	40.612	40.612
2	1.670	11.926	59.971	2.227	15.905	56.517
3	1.487	10.624	70.595	1.920	13.716	70.233
4	1.026	7.332	77.927	1.077	7.694	77.927
5	0.829	5.922	83.849			
6	0.765	5.462	89.310			
7	0.550	3.930	93.240			
8	0.476	3.397	96.637			
9	0.248	1.775	98.412			
10	0.218	1.557	99.969			
11	0.004	0.031	100.000			
12	3.462E-6	2.473E-5	100.000			
13	1.009E-7	7.207E-7	100.000			
14	3.785E-8	2.704E-7	100.000			

Table 4 Total Variance Explained (Post-Monsoon)

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.051	50.367	50.367	6.327	45.191	45.191
2	1.896	13.545	63.912	1.944	13.885	59.075
3	1.253	8.953	72.865	1.551	11.077	70.153
4	1.024	7.317	80.182	1.404	10.029	80.182
5	0.922	6.585	86.767			
6	0.684	4.889	91.656			
7	0.493	3.522	95.178			
8	0.300	2.145	97.323			
9	0.223	1.592	98.914			
10	0.152	1.085	100.000			
11	3.569E-5	0	100.000			
12	4.962E-7	3.544E-6	100.000			
13	5.769E-8	4.121E-7	100.000			
14	2.863E-16	2.045E-15	100.000			

PC 2 exhibits about 16 % (pre-monsoon) and 14 % (post monsoon) of the total variance, with significant positive loadings on Sodium and Bicarbonates (Table 5). Rock minerals like sodalite, glaucophane, albite, nepheline, aegirine, etc were the main contributors of Na in groundwater. Erosion of salts and rock mineral, precipitation, wastewater discharges supplement Na in groundwater. Base Exchange reactions in clay soils and zeolites also contribute Na in groundwater. Bicarbonates in groundwater are caused by dissolution of CO_2 in rain water. Decomposing organic matter in soil also contributes CO_2 during infiltration of water in soil. Solubility of CO_2 is also influenced by temperature and pressure like the other gases. Carbonate equilibrium in soil results in carbonates in soil as well (Ground water year book, 2015-16). These are interpreted as natural pollution in the study area.

PC 3 accounts for 14 % (pre-monsoon) and 11 % (post monsoon) of total variance, with significant positive loadings on pH, Carbonates in pre-monsoon season and Potassium in post monsoon season. Reactions of CO_2 in soil and solubility of silicate contribute to carbonates and pH in groundwater. CO_2 and H_2CO_3 concentrations influence the solubility of minerals during interaction of water with soil and water minerals. Magnitude of partial pressure of carbonic acid and bicarbonate affects the saturation of minerals in groundwater. Reduction in partial of CO_2 and H_2CO_3 was reported to result in higher carbonate and pH levels (Deepesh at al., 2015). Accordingly, this component was attributed to pollution by soil leaching.

Table 5 Varimax Rotated Factor Loadings

Pre Monsoon

Rotated Component Matrix				
	1	2	3	4
pH	-0.375	-0.209	**0.785**	0.048
SC	**0.865**	0.484	-0.084	0.057
TDS	**0.865**	0.484	-0.084	0.057
CO_3^{2-}	-0.263	-0.043	**0.788**	-0.093
HCO_3^-	0.159	**0.803**	-0.395	0.080
Cl^-	**0.884**	0.327	-0.032	-0.058
F^-	-0.301	0.274	0.051	**0.772**
NO_3^-	0.501	-0.341	0.051	**0.619**
SO_4^{2-}	**0.806**	0.045	0.059	-0.169
Na^+	0.433	**0.766**	0.177	0.034
K^+	0.276	0.036	0.520	0.148
Ca^{2+}	**0.795**	-0.082	-0.342	-0.025
Mg^{2+}	**0.690**	0.398	-0.074	0.131
TH	**0.898**	0.124	-0.283	0.042

Post Monsoon

Rotated Component Matrix				
	1	2	3	4
pH	-0.240	-0.372	**0.716**	0.001
SC	**0.930**	0.320	-0.080	0.156
TDS	**0.930**	0.320	-0.080	0.156
CO_3^{2-}	-0.051	0.021	**0.881**	-0.069
HCO_3^-	0.154	**0.859**	-0.303	0.030
Cl^-	**0.958**	0.049	0.015	0.134
F^-	-0.308	0.519	-0.103	-0.480
NO_3^-	0.281	-0.012	-0.161	0.438
SO_4^{2-}	**0.893**	0.024	-0.098	0.040
Na^+	0.500	**0.733**	0.143	0.077
K^+	0.020	0.061	0.022	**0.910**
Ca^{2+}	**0.911**	-0.067	-0.239	0.098
Mg^{2+}	**0.799**	0.209	-0.027	0.205
TH	**0.951**	0.031	-0.181	0.147

PC 4 accounts for 8 % (pre-monsoon) and 10 % (post monsoon)of the total variance, with significant positive loadings Fluorides and Nitrates in pre-monsoon and potassium in post monsoon season. Fluoride in groundwater is attributed to contribution of fluoride by rock minerals, long residence time due to deeper groundwater table, intensive irrigation practices and indiscriminate use of fertilizers. Phosphate fertilizer industries in the study area also contribute additional fluoride in groundwater. Nitrates are contributed by treated / partially treated wastewater effluents, leakage and spillages wastewater, septic tank effluents and seepage of agricultural runoff with fertilizers (Kumar et al., 2014). Eventually, this PC wasascribed to anthropogenic pollution. However, contribution of potassium in post monsoon season could be either weathering or anthropogenic. Further investigations are essential to exactly pin point the sources of potassium and fluorides.

Table 6 Source Identification based on Principal Component Analysis

Principal Component	Typical Loadings	Interpretation
Pre Monsoon Season		
PC 1	SC, TDS, TH, Ca, Mg, Cl, SO_4	Soil Leaching + Industrial Pollution
PC 2	Na, HCO_3	Natural Pollution
PC 3	pH, CO_3, K	Soil Leaching
PC 4	F, NO_3	Anthropogenic Pollution
Post Monsoon Season		
PC 1	SC, TDS, TH, Ca, Cl, SO_4	Soil Leaching + Industrial Pollution
PC 2	Na, HCO_3	Natural Pollution
PC 3	pH, CO_3	Soil Leaching
PC 4	K	Weathering, Anthropogenic Pollution

CONCLUSIONS

Cluster analysis and principal component analysis is successfully demonstrated for studying the ground water quality of the study region. Twenty five monitoring stations in the study area are grouped into three categories based on variation in groundwater quality. To reduce the cost of data collection, optimum no of sapling sites can be worked out in each cluster for monitoring groundwater quality. By using PCA, major pollution contributors are identified without eliminating any data and parameters. Four potential pollution types identified were soil leaching, industrial pollution, natural pollution and weathering. This study showed the reliability of these methods in ground water quality analysis.Also these results can be used in implementation of optimum sampling strategies and effective ground water management techniques. Soil chemistry and geological investigations can improve the source interpretation in the study area.

REFERENCES

Chen, H., Teng, Y., Yue, W. and Song, L. (2013). Characterization and source apportionment of water pollution in Jinjiang River, China. Environ. Monit. Assess. 185: 9639–9650.

Machiwal, D. and Jha, M. K. (2015). Identifying sources of groundwater contamination in a hard-rock aquifer system using multivariate statistical analyses and GIS-based geostatistical modelling techniques, Journal of Hydrology: Regional Studies, 4: 80-110.

Pekey, H., Karakas, D., Bakoglu, M. (2004). Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses. Marine Pollution Bulletin, 49: 808-818.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and Fernandez, I. (2000).Temporal evaluation of ground water composition in an alluvial aquifer (Pisuerga River, Spain) by Principal component analysis. Water Research, 34: 807-816.

Kim, J.H., Kim, R. H., Lee, J., Cheong, T.J., Yum, B.W. and Chang, H.W. (2005). Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. Hydrol. Process. 19, 1261–1276.

Kumar, M., Kumari, K., Singh, U.K. and Ramanathan, A. L. (2009). Hydrogeochemical processes in the groundwater environment of Muktsar, Punjab: conventional graphical and multivariate statistical approach, Environmental Geology, 57 (4): 873-884.

Lucas, L. and Jauzein, M. (2008). Use of principal component analysis to profile temporal and spatial variations of chlorinated solvent concentrations in groundwater. Environmental Pollution, 151:205-212.

Nair, I.V., Singh, K., Arumugam, M., Gangadhar, K. and Clarson, D. (2010). Trace metal quality of Meenachil river at Kottayam, Kerala (India) by Principle component analysis, World applied science journal, 9(10): 1100-1107.

Shrestha, S. and Kazama, F. (2006). Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ. Model, 464–475.

Singh, K.P., Malik, A., Mohan, D. and 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 Res., 38:3980–3992.

Kumar, S. A., Shankaraiah, K., Rao, P. L. K. M. and Sathyanarayana, M. (2014). Assessment of water quality in Hussainsagar lake and its inlet channels using multivariate statistical techniques. International Journal of Scientific & Engineering Research, 5(9): 327-333.

❑❑❑