Factor analysis as a tool for evaluation of spatial and temporal variations in groundwater quality: A case study

S. Krishnaraj[1], Sanjiv Kumar[2] and K. P. Elango[3]

[1]Department of Chemistry, R.V.S College of Engineering & Technology, Coimbatore, India
[2]School of Sciences, IGNOU, New Delhi, India
[3]Department of Chemistry, Gandhigram Rural Institute-Deemed University, Gandhigram, India

Abstract: *In this case study, factor analysis was applied for evaluation of temporal/spatial variations in the groundwater quality of Aravakurichi block, Karur district, Tamil Nadu, India. This statistical technique was employed for the better interpretation of large complex water quality data set obtained from twenty five groundwater locations in four seasons during the year 2012. The water samples were characterized for the physico-chemical parameters such as pH, total alkalinity, electrical conductivity, total hardness, calcium ions, magnesium ions, total dissolved solids, fluorides, chlorides and sulphates. Factor analysis indicated four factors initially and when rotation of the factor axis was executed, it yielded two factors with clear indication of high loadings for some variable and low loadings for others, facilitating data interpretation in terms of original variables. Overall, this case study demonstrated the effectiveness of factor analysis to identify marker variables for assessing the chemistry of groundwater besides earmarking representative sampling stations to undertake suitable water quality management in a shortest possible time.*
Key words: *factor analysis, varimax rotation, eigen value, factor score, marker*

I. Introduction

The chemistry of water is an important factor that determines its application as a resource (Kumar and Singh, 2010). Groundwater quality depends on the chemistry of recharged water, atmospheric precipitation, in-land surface water and on sub-surface geochemical processes [1]. The assessment of groundwater quality status is pivotal for socio-economic development of any region of the world [2]. Despite its importance, contamination from natural and man-made sources has affected the use of groundwater as a source for human consumption. The analysis of groundwater quality is an important and sensitive issue. Due to spatial and temporal variations in water chemistry, a monitoring program that provides a representative and reliable estimation of groundwater quality in a shortest possible time has become an important necessity. The search for an indicator that best characterizes a water source under study requires the use of appropriate statistical techniques. Among the multivariate techniques, factor analysis permit to define more sensitive indicators which can facilitate for evaluation of changes that have taken place in the water resources [3]. With this background, this case study was undertaken to evaluate the effectiveness of factor analysis as a tool to understand the factors affecting the groundwater quality.

II. Materials and Methods

2.1. Study area

The Aravakurichi block of Karur district, Tamil Nadu, India is located at 10.95º N and 78.08 ºE with an elevation of 122 m above mean sea level. The average annual rainfall is about 855 mm. This study area receives most of its seasonal rainfall from the north-east monsoon between late September to mid November.

2.2. Sample collection and monitoring parameters

A total of 25 water quality monitoring stations were identified and water samples were collected in the middle month of four seasons namely post-monsoon (January – March), summer (April –June), pre-monsoon (July – September) and monsoon (October – December) of the year 2012. The groundwater samples were analyzed for parameters which include pH, electrical conductivity, total dissolved solids, total alkalinity, total hardness, Ca (II), Mg (II), Na, K, fluorides, sulphates and chlorides using standard protocols [4] and the quality of the data was ensured through careful standardization.

2.3. Factor Analysis

Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable [5]. There are three stages in factor analysis [6] such as (i) For all the variables a correlation matrix is generated, (ii) Factors are extracted from the correlation matrix based on the correlation coefficients of the variables, (iii) To maximize the relationship between some of the factors and

variables, the factors are rotated. One of the most important steps of factor analysis is to determine the number of factors that need to be extracted for an accurate analysis of the data. In this regard, the rotation of the factor axis was executed to yield a simple structure such that factors that are clearly marked by high loadings for some variable and low loadings for others, facilitating data interpretation in terms of original variables. In order to determine the number of factors to be used, the variances and co-variances of the variable are computed. Then, the eigen values and eigenvectors were evaluated for the co-variance matrix and the data is transformed into factors.

III. Results and Discussions

The groundwater samples collected during the four seasons were analyzed and their descriptive statistical data is presented in Table 1 and 2.

Table 1. Descriptive statistical data of groundwater samples during post-monsoon and summer

Variable	Post monsoon				Summer			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
EC	756	2824	1749.9	542.3	665	2485	1538.9	478.0
TDS	558	2083	1288.2	395.0	472	1762	1093.3	339.2
TA	210	354	295.7	45.4	198	496	343.9	80.1
TH	88	284	157.7	54.6	113	268	190.3	45.2
Ca(II)	34	102	69.4	17.8	49	121	83.5	18.8
Mg(II)	37	121	63.0	26.8	11	129	65.9	26.7
pH	7.2	8.1	7.5	0.3	7.3	8.3	7.7	0.3
Na	110	208	154.9	26.4	127	215	156.9	24.7
K	21	49	36.3	7.9	12	53	37.1	7.7
F	0.3	1.91	0.8	0.4	0.34	1.92	0.8	0.4
SO$_4$	13	206	79.3	44.7	18	208	76.4	50.8
Cl	79	329	191.7	85.8	68	322	177.4	80.9

Table 2. Descriptive statistical data of groundwater samples during pre-monsoon and monsoon

Variable	Pre-monsoon				Monsoon			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
EC	1014	2814	1875.8	469.0	1114	3054	2037.5	525.1
TDS	598	1984	1099.4	367.6	657	2182	1203.5	399.3
TA	152	415	304.9	73.3	167	452	338.2	79.5
TH	83	232	152.8	45.1	93	258	171.4	49.8
Ca(II)	39	101	66.5	17.1	43	118	73.9	19.2
Mg(II)	18	102	55.0	19.0	24	96	54.6	15.6
pH	7.1	8	7.5	0.3	6.8	7.7	7.2	0.3
Na	127	215	156.9	24.7	120	208	149.4	24.0
K	18	48	34.4	6.5	22	51	35.9	7.5
F	0.42	2	0.8	0.3	0.48	1.7	0.8	0.3
SO$_4$	12	164	55.6	38.4	14	176	61.0	41.3
Cl	38	295	138.5	72.8	41	320	150.6	78.3

The results of the present study indicated that the water is alkaline in nature. Though it has no direct effect on human health, the recommended pH value for drinking purpose is 7.0 – 8.5 [7]. The average pH values varied from 7.2 - 8.1 (post-monsoon), 7.0 -8.3 (summer), 7.1 - 8.0 (pre-monsoon) and 6.8 - 7.7 (monsoon). The electrical conductivity (EC) is a valuable indicator of the amount of materials dissolved in water. The average EC values varied between 756 – 2824, 665 – 2485, 1014 – 2814, 1114 - 3054 μS during post-monsoon, summer, pre-monsoon and monsoon seasons, respectively. The average total dissolved solids (TDS) values varied from 558 - 2083 mg L^{-1} (post monsoon), 472 - 1762 (summer), 598-1984 (pre monsoon), and 657 - 2182 (monsoon). The total hardness (TH) and other parameters are found to be within the permissible limit except fluoride which was found to exceed the limit in certain sampling points and varies slightly with seasons.

The analysis generated four factors (eigen value greater than unity) which accounted for 66.4%, 67 % and 66 % of the total variance in post-monsoon, pre-monsoon and monsoon seasons, respectively. In the case of summer season, twelve factors were reduced to three factors were with 61.5% of total variance. In order to further reduce the number of factors and enhance the interpretability, the factors are rotated. The rotation usually increases the quality of interpretation of the factors [8].

After varimax rotation, each original variable tends to be associated with one (or a small number) of the factors and each factor represents only a small number of variable. Table 4 shows the factor pattern of water quality parameters after varimax rotation for the four seasons (2012).

Table 3. Eigen value (EV), percentage of variance (V) and cumulative percent (C)

Factor	Post-monsoon			Summer			Pre-monsoon			Monsoon		
	EV	V (%)	C (%)	EV	V (%)	C (%)	EV	V (%)	C (%)	EV	V (%)	C (%)
1	4.1	34.0	34.0	4.5	37.3	37.3	3.4	28.7	28.7	3.5	28.8	28.8
2	1.6	13.3	47.3	1.7	14.4	51.7	1.9	15.6	44.4	1.9	16.0	44.8
3	1.3	10.6	57.9	1.2	9.9	61.5	1.5	12.8	57.2	1.6	13.0	57.7
4	1.0	8.5	66.4	0.8	6.8	68.3	1.2	9.8	67.6	1.1	8.9	66.6
5	0.3	2.7	69.1	0.4	3.1	71.4	0.3	2.8	69.8	0.3	2.6	69.2
6	0.3	2.2	71.3	0.1	1.1	72.6	0.2	1.3	71.1	0.2	2.1	71.3

Table 4. Factor pattern after varimax rotation

Variables	Post-monsoon		Summer		Pre-monsoon		Monsoon	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
EC	0.778	0.469	0.783	-0.447	0.587	0.392	0.413	0.556
TDS	0.776	0.470	0.773	-0.462	0.948	0.096	0.180	0.954
TA	-0.607	-0.141	-0.497	0.253	-0.212	-0.605	-0.599	-0.163
TH	0.192	0.961	0.746	0.491	0.001	0.782	0.666	0.058
Ca	-0.026	0.712	0.502	0.384	-0.023	0.473	0.393	0.012
Mg	0.661	0.051	0.570	-0.182	0.337	0.101	-0.045	0.318
pH	-0.532	-0.337	-0.728	0.036	-0.012	-0.685	-0.752	0.153
Na	0.313	0.046	0.335	0.003	0.060	0.203	0.188	-0.032
K	0.017	0.202	-0.032	0.276	-0.158	0.142	-0.054	0.391
F	-0.690	0.233	-0.070	0.629	-0.679	-0.225	-0.401	-0.564
SO$_4$	0.724	-0.239	0.265	-0.671	0.841	-0.131	-0.065	0.875
Cl	0.654	0.396	0.691	-0.217	0.348	0.766	0.864	0.240
Variability (%)	32.181	19.314	31.536	15.591	22.666	21.324	22.313	22.244
Cumulative (%)	**32.181**	**51.495**	**31.536**	**47.127**	**22.666**	**43.990**	**22.313**	**44.557**

The parameters are grouped based on the factor loadings. Factor 1 exhibited 32.1% of the total variance of 51.4 % with strong positive loadings for EC and TDS in post-monsoon season, 31.5% of the total variance of 47.1 % with strong positive loadings for EC, TDS and TH in summer season, 22.6% of the total variance of 43.9 % with strong positive loadings for TDS and SO$_4$ in pre-monsoon and 22.3% of the total variance of 44.5% with strong positive loadings for Cl in monsoon season. These high loadings represent a relative high correlation between each other. Factor 2 exhibited 19.3% with strong positive loadings for TH in post-monsoon, 15.5% with moderate positive loadings for fluoride in summer, 21.3% with strong positive loadings for TH and Cl in pre-monsoon and 22.2% with strong positive loading for TDS and SO$_4$ monsoon season.

The factor scores were then calculated for all 25 monitoring stations and are shown in Table 5. In post-monsoon, the high scores for Factor 1 were observed at stations 3,4,5,6,9,10,12,14,16,17,22 and 23. During this season, the high scores for Factor 2 were observed at stations 2,6,8,12,19,21 and 22. In the case of summer season, the high scores for Factor 1 were observed at stations such as 3,4,6,9,10,11,12,18,20,23,24,25 and the high scores for Factor 2 were observed at stations 1,4,5,6,12, 14,16 and 22. In pre-monsoon season, the high scores for Factor 1 were observed at stations 4,6,12,20,23,24 and the high scores for Factor 2 were observed at stations 3,4,5,6,12,14,15,16,17 and 25. During the monsoon season, the high scores for Factor 1 were observed at stations 3,4,5,6,12,14,15,16,17, 25 and the high scores for Factor 2 are 4,6,12,20,23 and 24.

These findings reveal the fact that twelve water quality parameters listed in the present study can be represented by factors 1 and 2 and these could be used as an indicator for potential contamination. In other words, any arbitrarily selected parameter from factor 1 and 2 could be used as a 'marker' variable to detect potential contamination. For the present investigation, the probable candidates for this purpose could be any one of the easily measured parameters such as electrical conductivity or TDS for factor 1 and Total hardness (TH) for factor 2 in all the four seasons. Similar approach of deriving a marker variable through factor analysis has been reported [8].

Hence, this marker parameter selection might be used as a technical tool for designing groundwater quality monitoring program. Thus, marker parameter can be measured on a continuous basis on selected stations for assessing the groundwater quality.

Table 5. Factor scores after varimax rotation

Sampling Station	Post-monsoon		Summer		Pre-monsoon		Monsoon	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
1	-0.142	-0.693	-0.318	1.093	-0.188	-0.075	-0.121	-0.285
2	-1.053	1.717	-1.722	0.464	-1.364	0.187	-0.069	-1.016
3	1.268	-0.514	0.578	0.458	-0.503	1.145	0.694	-0.409
4	1.945	-0.444	0.848	1.626	0.694	1.081	0.803	0.796
5	0.663	0.451	-0.014	1.075	-0.198	1.781	1.816	-0.167
6	1.284	0.624	1.141	0.507	1.270	0.588	0.708	1.437
7	-1.349	0.468	-1.119	-0.962	-1.425	-1.305	-1.554	-1.121
8	-0.887	1.226	-1.702	0.467	-0.986	-1.049	-1.333	-0.792
9	0.640	-0.116	0.612	0.179	0.298	-0.271	-0.433	0.443
10	0.468	-1.199	0.652	-0.186	0.328	0.343	0.184	0.197
11	0.050	-1.021	0.925	-0.580	-0.492	-0.053	0.044	-0.757
12	1.628	0.929	1.028	1.722	1.357	1.271	1.158	0.869
13	-1.090	-0.009	-0.847	-1.068	0.055	-1.463	-1.054	0.013
14	0.872	0.131	-0.182	1.129	-0.050	0.826	0.963	-0.278
15	-0.353	-0.108	0.417	-0.297	0.265	0.559	0.954	-0.221
16	0.312	-0.313	0.383	0.675	0.448	1.038	1.322	0.156
17	0.374	-0.118	0.016	0.352	0.222	0.506	0.920	-0.289
18	0.009	-0.278	0.860	-0.186	0.007	0.133	0.235	-0.149
19	-1.496	0.592	-1.226	-0.738	-0.649	-1.500	-1.538	-0.750
20	-1.070	-1.966	0.880	-1.736	2.436	-1.570	-1.453	2.262
21	-2.093	0.520	-1.125	-1.533	-0.588	-1.325	-0.964	-1.105
22	0.391	2.310	-2.025	1.117	-2.154	-0.084	-0.750	-1.634
23	0.507	-0.226	0.909	-1.096	0.557	-0.438	-0.388	0.644
24	-0.877	-1.737	0.509	-1.335	1.890	-1.176	-0.954	2.702
25	0.002	-0.225	0.523	-1.146	-1.231	0.849	0.808	-0.546

IV. Conclusion

In the present case study, the groundwater samples were collected from 25 monitoring stations during four seasons and its water quality were analyzed for twelve chemical parameters. In order to interpret the complex dataset for undertaking appropriate management/treatment besides designing periodical monitoring program, factor analysis was employed. This multivariate statistical technique proved to be effective in reducing twelve parameters into three to four factors initially and ultimately brought down to two important factors. Overall, the factor analysis paved way to identify marker variable that would serve as an important and reliable indicator for undertaking periodical assessment of groundwater quality.

References

[1]   M. Vasanthavigar, K. Srinivasamoorthy, K. Vijavaraghavan, R. Rajiv Ganthi, S. Chidambaram, P. Anandhan, R. Manivannan, and S. Vasudevan, Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamil Nadu, India, *Environ Monit Assess,* 2010, 121-132.

[2]   J.M. Ishaku, Assessment of groundwater quality index for Jimeta -Yola area, Northeastern Nigeria. *Journal of Geology and Mining Research, 3(9),* 2011, 219-231.

[3]   C. Coletti, R. Testezlaf, A.P. Tulio, P. Ribeiro, T. Renata, G. de Souza, and D. de A. Pereira, Water quality index using multivariate factorial analysis, *R. Bras. Eng. Agric. Ambiental,* 14, 2010, 517-522.

[4]   APHA, *Standard methods for the examination of water and wastewater* (American Water Works Association, Washington DC, 1985).

[5]   S. Yu, J. Shang, and J. Zhao, Factor analysis and dynamics of water quality of the Songhua River Northeast China. *Water, Air & Soil Pollution,* 144, 2003, 159 – 169.

[6]   A.K. Gupta, S.K. Gupta, and R.S. Patil, Statistical analyses of coastal water quality for a port and harbour region in India. *Environ Monit. Assess.,* 2005, 179-200.

[7]   WHO, Guidelines for drinking water quality (Vol. I) - Recommendations (World Health Organization, Geneva, 2010).

[8]   H. Boyacioglu, H. Boyacioglu, and O. Gunduz, Application of factor analysis in the assessment of surface water quality in Buyank Menderes river basin, *European Water*, 2005, 43-49.