International Archive of Applied Sciences and Technology
Int. Arch. App. Sci. Technol; Vol 10 [2] June 2019 : 76-83
© 2019 Society of Education, India
[ISO9001: 2008 Certified Organization]
www.soeagra.com/iaast.html

IAAST
ONLINE ISSN 2277- 1565
PRINT ISSN 0976 - 4828

CODEN: IAASCA ORIGINAL ARTICLE

DOI: .10.15515/iaast.0976-4828.10.2.7683

Characterization of Groundwater Quality using Principal Component Analysis of Andipalayam and Neruperichel blocksof Tiruppur District

S.R. Naffees Gowsar[1], M. Nirmala Devi*, M. Radha[2], R. Kumaraperumal[3]and S. Vishnu Shankar[4]

[1,4] Agricultural Statistics, Department of Physical Sciences & IT, Tamil Nadu Agricultural University, Coimbatore.

[2]Assistant Professor in Agricultural Statistics, Department of Physical Sciences & IT, Tamil Nadu Agricultural University, Coimbatore.

[3]Assistant Professor in Remote Sensing and GIS, Tamil Nadu Agricultural University,Coimbatore

*Corresponding Author: Email Id: pnirmalah@rediffmail.com

ABSTRACT

Groundwater is the major source of water in both urban and rural India. In Tamil Nadu, Tiruppur district is known to be "Banian" city, where intense competition among users from industry, agriculture and domestic sectors is driving the groundwater table lower and the quality of groundwater is getting severely affected due to the widespread pollution of surface water. The present study involves 11 physico-chemicalparameters which have been analysed on groundwater samples collected from Andipalayam and Neruperichel village of Tiruppur district in Tamil Nadu. This study contemplates to assess the groundwater quality parameters using multivariate technique and to identify the variable whichis responsible formajor variation in ground water quality through principal components. Correlation analysis between original values and principal components is also performedto find the representation of variables within the PC's.

Keywords: PCA, dye effluent, groundwater quality, Correlation, cos2.

Received 12.02.2019 Revised 21.03.2019 Accepted 19.04.2019

CITATION OF THIS ARTICLE

S.R. Naffees Gowsar, M. Nirmala Devi, M. Radha, R. Kumaraperumal and S. Vishnu Shankar. Characterization of Groundwater Quality using Principal Component Analysis of Andipalayam and Neruperichel blocksof Tiruppur District.Int. Arch. App. Sci. Technol; Vol 10 [2] June 2019: 76-83

INTRODUCTION

Groundwater contributes to the major source of drinking water in most of the urban and rural areas. It is also an alternative source for irrigation, industrial and commercial purposes. During the past decades, groundwater usage in India has grown many folds and at present 80% of rural domestic needs and 65% of irrigation water requirement and 50% of industrial and urban water needs are sourced from scarce ground water resources. Over exploitation of ground water has also started affecting ground water quality in many areas.

Tamil Nadu is endowed with flowing rivers but not with perennial rivers. There are eight major flowing rivers in Tamil Nadu. Noyyal is one among those rivers, which flows in Tiruppur district. Due to the development of bleaching and dyeing industries in and around Tiruppur district, Noyyalriver has been polluted which in turn has affected the groundwater quality. The discharge of untreated waste water through bores and leachate from unscientific disposal of solid wastes also contaminates groundwater, thereby reducing the quality of water.

Groundwater is the main source of water for drinking and irrigation purposes in Tiruppur district. Physico-chemical parameters and biological parameters are responsible for quality of groundwater. Hence in this study, the physic-chemical parameter is alone taken in consideration to find the quality of groundwater. The most responsible parameter which determines quality of groundwater is identified using multivariate technique. This multivariate technique involves Principal Component of Analysis (PCA) method by which Principal Components (PC's) are tested with original variables for correlation, inorder to find the most contributing parameter in groundwater quality.

Study area:

Tiruppur is located at11.1075°N 77.3398°E on the banks of the Noyyal River. The groundwater quality datasets were collected from the report submitted by the Department of Environmental Science, Tamil Nadu Agricultural University to the Government of Tamil Nadu during the year 2013 [5].A total of 323 groundwater samples collected and analysed from Andipalayam (83 samples) and Neruperichel (240 samples) blocks of Tiruppur district.Thephysico-chemical properties of the groundwater samples aresubjected to statistical analysis.The parameters are: pH, EC, TDS, Na, Ca, Mg, SAR, K, Cl, SO_4 and HCO_3.

MATERIAL AND METHODS

Descriptive statistics

Descriptive statistics is used to know about the basic features of the data in the study. It provides simple summaries about the sample and the measures [4].

Pearson's correlation coefficient:

Pearson's correlation coefficient measures the strength of linear relationship between two variables. Pearson's r can range from -1 to 1, r = -1 indicates perfectly negative correlation between variables; r = 1 indicates perfectly positive correlation between variables; r = 0 indicates there is no correlation between variables. Correlation analysis is performed to know the magnitude of relationship between the parameters and the value of pearson'scorrelation coefficient is classified as poor (0.0-0.3), Moderate (0.4-0.6) and strong(0.7-1.0) [8][12].

Principal component analysis:

PCA method is useful when the variables within the dataset are highly correlated. Correlation indicates that there is redundancy in the data. Due to this redundancy, PCA can be used to reduce the original variables into a smaller number of new variables (Principal Components) which explains most of the variance of original variables. The main purpose of principal component analysis is to (1) identify hidden pattern in a dataset, (2) reduce the dimensionality of the data by removing the noise and redundancy in the data and (3) identify the correlated variables

The PCs are defined as a linear combination of the data's original variables, for example two dimensional (2D), PC1 will be PC1= $x/\sqrt{2} + y/\sqrt{2}$. These coefficients will be stored in a PCA loading matrix, which can be interpreted as a rotation matrix that rotates the data such that the projection with greatest variance goes along the first axis. At first glance, PC1 closely resembles the linear regression line. However, PCA differs from linear regression in that PCA minimizes the perpendicular distance between a data point and the principal component, whereas linear regression minimizes the distance between the response variable and its predicted value. In multivariate analysis technique, Principal component analysis have been reported as one of the effective method for the characterization and evaluation of water quality parameters [3][9].

Statistical analysis:

The descriptive statistics of the groundwater samples are presented in Table 1 and it could be used to determine the quality of groundwater. It is evident that, EC and TDS parameter is highly significant with Na, Ca, Mg, Cl, SO_4,, HCO_3($p > 0.0001$)whereas EC is significant with pH and K ($p > 0.05$)(Table 2).

Table 1: Descriptive Statistics of groundwater samples

	pH	*EC*	*TDS*	*Na*	*Ca*	*Mg*	*SAR*	*k*	*Cl*	*so4*	*Hco3*
Mean	7.37	2.55	1453.42	67.31	199.37	112.08	1.52	12.17	1303.69	73.06	712.70
SE	0.02	0.10	57.81	4.42	5.77	4.17	0.54	1.30	36.30	3.89	16.47
Median	7.32	2.01	1210.00	46.00	176.00	91.00	0.65	5.00	1136.00	53.00	683.00
Mode	7.45	2.10	1110.00	33.00	176.00	67.00	0.40	5.00	994.00	49.00	756.00
SD	0.36	1.77	1038.97	79.46	103.68	74.94	9.63	23.42	652.32	69.95	295.98
Variance	0.13	3.14	1079467.56	6314.33	10750.26	5615.83	92.81	548.56	425521.12	4892.34	87605.48
Kurtosis	2.65	12.96	17.80	29.75	6.85	4.38	314.56	12.61	7.54	15.33	1.53
Skewness	0.42	2.87	3.46	4.70	2.24	1.84	17.62	3.42	2.20	3.33	0.88
Range	3.20	15.14	9320.00	768.00	720.00	475.00	172.98	162.00	5059.00	540.00	1865.00
Minimum	5.58	0.44	180.00	1.00	72.00	5.00	0.02	0.00	337.00	0.00	2.00
Maximum	8.78	15.58	9500.00	769.00	792.00	480.00	173.00	162.00	5396.00	540.00	1867.00

Note: EC=dSm^{-1}. All values are in mg L^{-1} (except SAR)

Table 2: Pearson's Correlation for Physico-chemical Parameters

	pH	*EC*	*TDS*	*Na*	*Ca*	*Mg*	*SAR*	*k*	*Cl*	*so4*	*HCO3*
pH	1										
EC	0.0909*	1									
TDS	0.1464**	**0.9151****	1								
Na	0.0752*	0.1662**	0.1948**	1							
Ca	-0.1566	0.2430**	0.2087**	-0.0797	1						
Mg	-0.0395	**0.4744****	0.3929**	-0.0218	0.3472**	1					
SAR	-0.0648	-0.0212	-0.0140	0.1267	-0.0268	-0.0635	1				
K	-0.0701	0.0880*	0.0406	-0.1168	-0.1326	-0.0347	-0.0249	1			
Cl	-0.0001	**0.6374****	**0.5112****	0.0727*	0.2294**	0.3707**	0.0068	0.0208	1		

so4	HCO3
0.0554*	0.1082**
0.4481**	0.1874**
0.3623**	0.2194**
-0.0106	0.1494*
0.0967*	0.0541*
0.3606**	0.2412**
-0.0324	0.0682*
0.0395	0.1310**
0.4605**	0.1223**
1	0.0963*
	1

pH=pH, EC= Electrical Conductivity, TDS= Total Dissolved Solids, Na= Sodium, Ca= Calcium, Mg= Magnesium, SAR=Sodium Adsorption Ratio, K= Potassium, Cl= Chloride, S04= Sulphur Dioxide, HCO3= Bicarbonate, * = Significant $p > 0.05$, ** = Highly significant $p > 0.0001$.

EC and TDS are strongly correlated (0.9),whereas moderate correlation exits between Cl and EC (0.6)&Cl and TDS (0.5). pH is highly significant with TDS and HCO_3($p > 0.0001$) whereas pH is significant with Na and SO_4. Na is highly significant with EC and TDS but it is significant with Cl and HCO_3. Ca is highly significant with Mg and Cl; significant with SO_4 and HCO_3. Mg is highly significant with EC, TDS, Cl, SO_4, HCO_3. SAR is significant with HCO_3.

Potassium (K) is highly significant with HCO_3 and it's significant with EC. Cl is highly significant with EC, TDS, Ca, Mg, SO_4, HCO_3 and it's significant with Na. SO_4 is highly significant with EC, TDS, Mg, Cl and its significant with pH and Ca. HCO_3 is highly significant with pH, EC, TDS, Mg, K, Cl and its significant with Na, Ca, SAR, SO_4.

PCA method:

As a first step in Principal Component Analysis, variables are measured in different scales and standardizing the data into same scale is necessary. Since FactoMineR package of R version 3.1.1 is used, by which default function standardizes the data automatically during PCA hence transformation is not required.

Table3: Eigenvector of principle components.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
pH	0.0500	-0.4730	0.1700	**0.5370**	-0.2400	**0.3970**	-0.3370	0.2530
EC	**0.5080**	-0.0830	0.0440	0.0300	0.1540	-0.1560	-0.2300	-0.2000
TDS	**0.4740**	-0.1630	0.0120	0.0630	0.0930	-0.1870	-0.3170	-0.2790
Na	0.0920	-0.5400	-0.4100	-0.0490	0.0530	-0.4980	**0.3000**	**0.3140**
Ca	0.2110	**0.4800**	-0.3560	-0.0510	-0.3100	-0.0690	-0.3690	**0.5800**
Mg	**0.3670**	**0.2510**	-0.0630	-0.0430	-0.3200	0.1300	0.2160	-0.3660
SAR	-0.0150	-0.2380	-0.4030	-0.5480	**0.2880**	**0.5600**	-0.2570	-0.0450
K	0.0250	-0.0530	**0.6860**	-0.5020	0.0650	-0.1900	-0.2050	0.2960
Cl	**0.4200**	0.0630	0.0060	0.0010	**0.2680**	0.0530	0.0720	0.1960
So4	0.3430	0.0730	**0.1780**	0.0870	0.2490	0.3860	**0.5500**	**0.3470**
HCO3	0.1820	-0.2990	0.0720	-0.3730	-0.6980	0.1900	**0.2090**	-0.0270

First eight principal components accounts for 90.48 percent of total variation (Table 3). Considering the most significant variables in evaluating the component, the first PC explains information about EC, TDS, Mg, Cl variables and it accounts for 29.4%. EC showed lowest value of 0.44 dSm^{-1} to highest value of 15.58 dSm^{-1} with a mean of 2.55 dSm^{-1}, TDS showed lowest value of 180 mg $L^{-1}$to highest value of 9500mg L^{-1}, with a mean of 1453.42mg L^{-1}.In the second principle component, it explains 10.63% of variance and gives information about Ca and Mg. The Ca showed minimum value of 72 mg $L^{-1}$to maximum of 792 mg $L^{-1}$and mean value of 199.36mg L^{-1}. Mg ranged between 5 mg $L^{-1}$and 480 mg L^{-1}(Table 1).

Third component contributed about 10.63% of variance and indicates thatmajor influence are from K and SO_4.Fourth principal component explains about 9.99% of variance which was majorly influenced by pH. Likewise fifth PC explains about 8.8% of variance which was majorly imparted by SAR and Cl.

Table 4: Eigen value of Principal Components.

PC	Eigen value	Percentage of Variance	Cumulative Percentage
1	3.235	29.407	29.40
2	1.352	10.633	41.69
3	1.170	10.633	52.32
4	1.099	9.990	62.31
5	0.968	8.800	71.11
6	0.841	7.643	78.76
7	0.738	6.705	85.46
8	0.553	5.023	90.48
9	0.532	4.839	95.32
10	0.449	4.085	99.41
11	0.065	0.587	100.00

Eigenvalues measures the amount of variation retained by each principal component. Eigenvalue is large for the first PCs and it's less for the subsequent PCs. It can be used to determine the number of principal components to retain after PCA [7]. Eigenvalue >1 for a particular Principal Component indicates that PC has accounted for more variance than accounted by one of the original variables in standardized data, this acts as a cutoff point for which PCs are retained. From the Table4, it is evident that first four Principal components whose eigenvalues are greater than 1. Hence first four principal components retain from the overall observations.

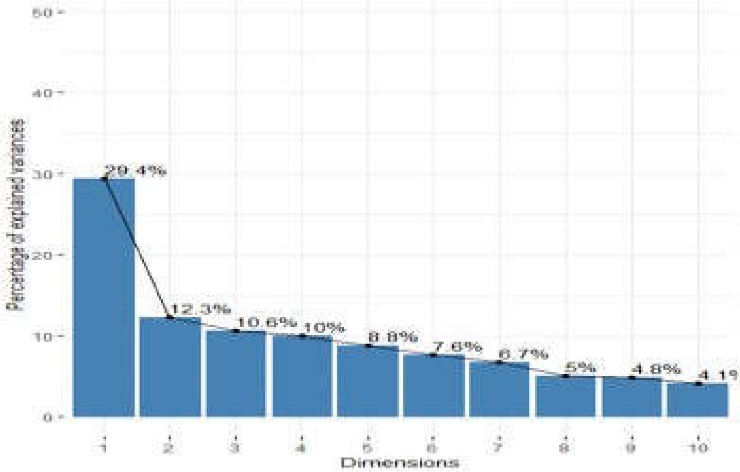

Figure1: Scree Plot of PCA.

Scree Plot, the plot of eigenvalues ordered from largest to smallest. The number of component is determined at a point beyond which eigenvalues are all iteratively small. Below 7 principal components the eigenvalues are iteratively reducing. Hence first seven principal components are taken into account for further investigations. Using eigenvalue first four components retained by PCA but by scree plot it results that 7 PCs. Since the observations are heterogeneous in nature, the variations in the PCs are scattered. If four PCs are determined as per eigen value, it retains only 62.32% of variation from the original observation. When Seven PCs are determined as per scree plot, it contributes 85.46% of total variation from the original observation.

Correlation Circle:

The correlation between a variable and a principal component (PC) is used as the coordinates of the variable on the PC. The observations are represented by the projections, whereas the variables are represented by their correlations [1].

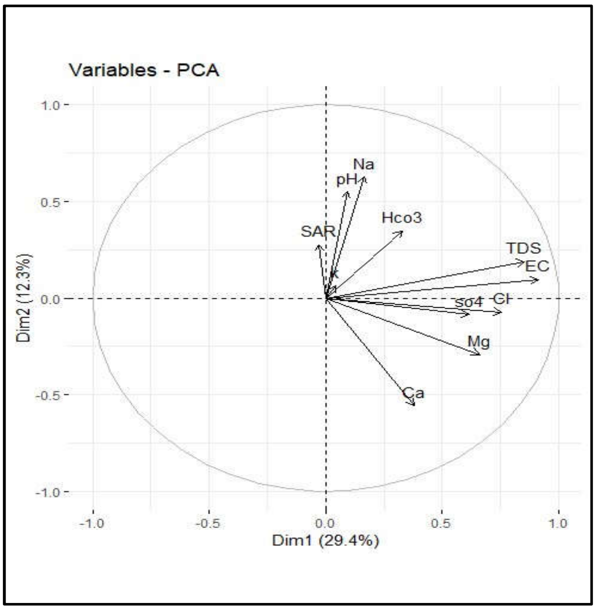

Figure 2: Factor Map of physico-chemical parameters

This plot is also known as variable correlation plots (Figure 2) which depicts the relationships between all the variables. The distance between variables and the origin indicates the quality of the variables on factor map. From the above figure, almost all the variables are positively correlated rather than SAR which is negatively correlated.
Quality of representation of the variables in factor map is called as **cos2.**A high cos2 indicates a good representation of the variable on the principal component. But low cos2 indicates that the variable is not perfectly represented by the PCs.

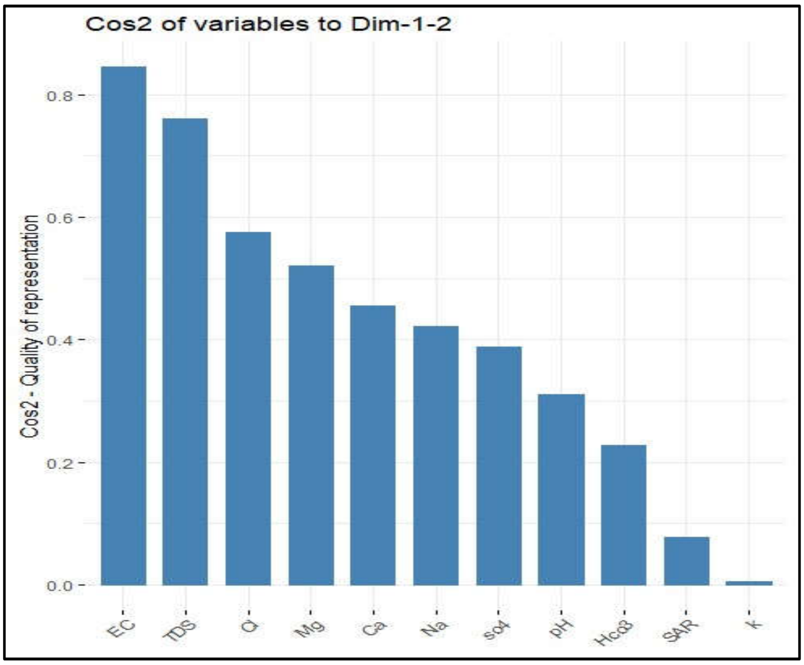

Figure3: Representation of variables in PC 1 and PC 2

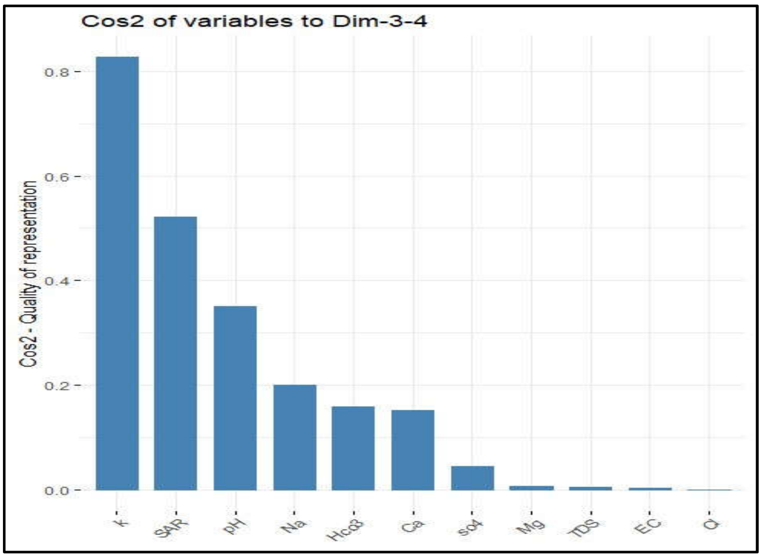

Figure 4: Representation of variables in PC 3 and PC 4

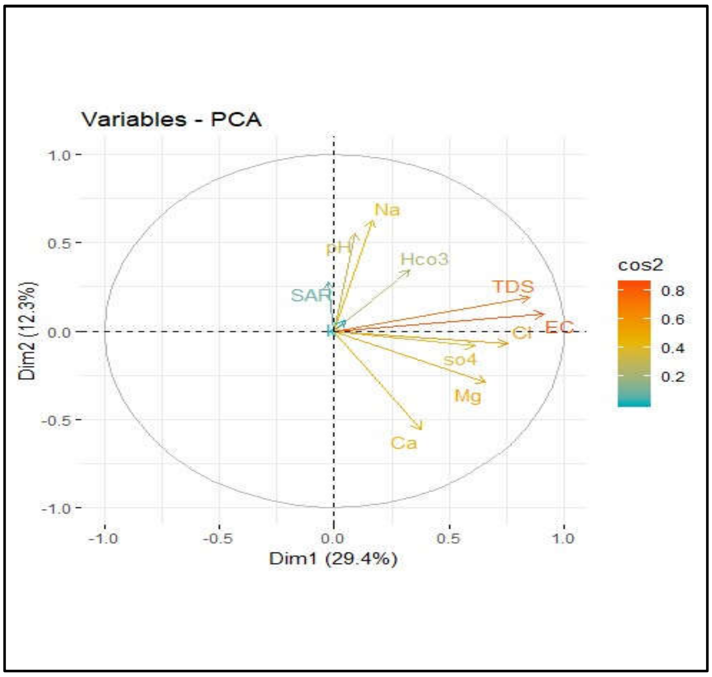

Figure5: cos2 Variance Plot of Physico-chemical Parameters

The variables that are close to the centre of the plot (Figure 3) are less important for the Study. From the factor plot it represents that high cos2 (TDS, EC) represents that the variables is perfectly represented on PCs and low cos2 (pH, HCO_3, SAR) shows that the variables is not perfectly represented by the PCs.

RESULTS AND DISCUSSION

Based on PCA method, the PC's showing eigen value greater that one are retained which explains about 62.32% of total variation. For further analysis, one can also choose up to eight PC's which explains about 90% of variation from the original observations.

From the Figure 3 and Figure 4, the variables represented from the principal components are graphed in which EC, TDS, Cl, Mg, Ca, Na variables highly contributed in first and second PC's. Whereas in third and fourth PC's,potassium and sodium absorption ratio are highly contributed. From the Table2, it is observed that there is a highly significant correlation of EC and TDS with other parameters like Na, Ca, Mg, Cl, SO_4 and HCO_3. Hence

EC and TDS components (highly represented in 1st PC) explain a large amount of variation in the study area.
The EC that showed a significant correlation with parameters such as TDS, Na, Ca, Cl which can be related to water salinity. EC and high TDS limit the absorption of water by crops because of the salt that stays in the roots. Due to the difficult access to water, the growth rate of plants is reduced, which limits agricultural production [11][13].The EC and the TDS in groundwater samples are significantly correlated with cations and anions (Ca, K, Na, Cl, Ca, HCO_3 and SO_4) which can be the result of ionic changes in the aquifer [11][3].

CONCLUSION

The variables like EC, TDS, Cl, Mg, Ca and Na are contributed much in the groundwater samples taken from Andipalayam and Neruperichel blocks of Tiruppur district. From correlation and PC's result, it's concluded that EC and TDS are the highly sensitive parameter from the rest of the parameters taken in consideration. Multivariate Statistical Methods are useful to evaluate the groundwater quality. The multivariate approach showed that the relationships between EC, TDS, pH, Na, Ca, Mg, Cl, SO_4, HCO_3, K and SAR, proving to be convenient for the confirmation and refinement of interpretations through the statistical results.

ACKNOWLEGEMENT

Authors are thankuful toDr.S.Mahimairaja, Department of Environmental science for providing the dataset.

References:

1. Abdi, Hervé, and Lynne J. Williams. 2010. "Principal Component Analysis." *John Wiley and Sons, Inc. WIREs Comp Stat* 2: 433.
2. 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).*doi:10.1007/s12665-018-7937-x.
3. Brahman, K.D.; Kazi, T.G.; Afridi, H.I.; Naseem, S.; Arain, S.S.; Ullah, N. Evaluation of high levels of fluoride, arsenic species and other physicochemical parameters in underground water of two sub districts of Tharparkar, Pakistan: A multivariate study. Water. Res. **2013**, 47, 1005–1020.
4. Delbari, M., Amiri, M., &Motlagh, M. B. (2016). Assessing groundwater quality for irrigation using indicator kriging method. *Applied Water Science, 6*(4), 371-381.
5. Department of Environmental Science- Report on Integrated Remediation for improving and Managing polluted soils and waters in Tiruppur. (Vol-I) Tiruppurdistrict , Part-1 Tiruppur Taluk.
6. Dutta, S., Dwivedi, A.,& Suresh Kumar, M. (2018). Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. Environmental Monitoring and Assessment, 190(12). doi:10.1007/s10661-018-7100-x
7. Kaiser, Henry F. 1961. "A Note on Guttman'sLower Bound for the Number of Common Factors." *British Journal of Statistical Psychology* 14: 1–2.
8. Kengnal, P.; Megeri, M.N.; Giriyappanavar, B.S.; Patil, R.R. Multivariate Analysis for the Water Quality Assessment in Rural and Urban Vicinity of Krishna River (India). Asian J. Water Environ. Pollut. **2015**,12, 73–80.
9. Mahapatra, S. S., Sahu, M., Patel, R. K., & Panda, B. N. (2012). *Prediction of Water Quality Using Principal Component Analysis. Water Quality, Exposure and Health, 4(2), 93–104.*doi:10.1007/s12403-012-0068-9.
10. PranabSabhapandit, PrasantaSaikia and Abani Kr. Mishra .2010 Interpretation of ground and surface water quality using principal component analysis in Gohpur subdivision of Sontipur district, Assam, India. *Int. J. Chem. Sci.: 8(3), 2010, 1523-1536.*
11. Prieto-Amparán, J., Rocha-Gutiérrez, B., Ballinas-Casarrubias, M., Valles-Aragón, M., Peralta-Pérez, M., &Pinedo-Alvarez, A. (2018). Multivariate and Spatial Analysis of Physicochemical Parameters in an Irrigation District, Chihuahua, Mexico. *Water, 10*(8), 1037.
12. Razack, M., &Dazy, J. (1990). Hydrochemical characterization of groundwater mixing in sedimentary and metamorphic reservoirs with combined use of Piper's principle and factor analysis. *Journal of Hydrology, 114*(3-4), 371-393.
13. Ward, J.H., Jr. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 1963, 58, 236–244.

