ASSESSMENT OF GROUNDWATER QUALITY USING MULTIVARIATE STATISTICAL ANALYSIS IN PARTS OF PERAMBALUR TALUK, PERAMBALUR, TAMILNADU

[1]S.Vinodhini, [2]G.Pavithira, [3]J.Theboral, [4]A.Ayyappan

Assistant Professor
[1,2,4]Department of Civil Engineering, [3]Department of Agriculture Engineering
Sri Ramakrishna College Engineering
Perambalur-621 113.

Abstract- **This paper evaluate the groundwater quality in parts of Perambalur Taluk, Peramblur, Tamilnadu by using multivariate statistical methods such as Descriptive analysis (DA), Hierarchical Cluster analysis (HCA), Principle component analysis (PCA),Factor analysis (FA). Totally 90 Samples were collected by spatial wise (Bore well, Hand pump and Open well) respectively the period of August, September, October 2016 from the study area in 10 locations. The descriptive analysis shows the cations order for august month Na > Mg >K >Ca and for September and October month Na > K> Mg> Ca the anions order for august month Cl> HCO_3 >So_4> No_3> F and for September and October month Cl> So_4 > HCO_3> No_3 > F. Factor analysis result shows the types of pollution present in the ground water. The percentage of variance of pollution is 24.77% of domestic wastewater pollution, 20.70% of agricultural pollution, 12.79% of biological pollution and 11.65% of mineral pollution present in the ground water samples. Cluster analysis defines the group of samples that have similar chemical and physical characteristics. Cluster 1 establish F, No_3, Mg, K, pH, So_4, Ca, Na, cluster 2 establish HCO_3,Cl and cluster 3 establish EC, TDS. This kind of study help to determine the status of groundwater quality.**

INTRODUCTION

Ground water is one of the prime sources of fresh water. It is an important source of drinking water for the world's population. The ground water is free from pollution and is very useful for domestic use in small towns and isolated farms. In 21st century our natural water resources have been used unconsciously which leads to its over exploitation. Ground water is the part of precipitation that infiltrates through the soil to the water table. The water table is the surface below which all openings in the rock or unconsolidated materials are filled with water. Water quality is extremely important because constant access to good quality water is necessary for life as well as the economy. In recent times, there has been a tremendous increase in demand for freshwater and water shortage in arid and semiarid regions due to population increase, urbanization, industrialization, and intense agricultural activities in many parts of world. Due to inadequate supply of surface waters, most of the people in India are depending mainly on groundwater resources for drinking and domestic, industrial, and irrigation uses. Water is extremely essential for the survival of all living organisms.

Multivariate statistical Analysis

Multivariate hydro statistical analysis were performed on water quality data using hierarchical cluster analysis (HCA) and principle component analysis (PCA) / factor analysis(FA).

Principle component analysis / Factor analysis

The variables were grouped using factor analysis and the result are represented by the factor loading matrix, Eigen value, total and cumulative variance. PCA basically is a dimension reduction intelligent data analysis technique. It reduces a large number of variables to a small number of variables which are the principle components.

The principle component generated through PCA can be expressed as:

$$y_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{j3} + \ldots\ldots\ldots\ldots + a_{im}x_{mj}$$

Where y = component score, a = component loading, x = measured value of the variable, i = component number, j = sample number, and m = total number of variables.

Cluster analysis

The assumptions of cluster analysis techniques include homoscedasticity (equal variance) and normal distribution of the variables. However, an equal weighing of all the variables requires long transformation and standardization (z-scores) of data. Comparison based on multiple parameters from different samples are made and the samples are grouped according to their similarity to each other. The classification of samples according to their parameters is termed Q-mode classification. This approach is commonly applied to water-chemistry investigations in order to define groups of samples that have similar chemical and physical characteristics. This is because rarely is a single parameter sufficient to distinguish between different water types. Individual samples are compared with specified similarity/dissimilarity and linkage methods are then grouped into clusters. The linkage rule used here is ward's method (Ward 1963). Linkage rules iteratively link nearby points (samples) by using the similarity matrix. The initial cluster is formed by linkage of the two samples with the greatest similarity. Ward's method is distinct from all the other method because it uses an analysis of variance (ANOVA) approach to evaluate the distance between clusters. Ward's method is used to calculate the error sum of squares, which is the sum of the distances from each individual to the center of the parent group. This form smaller distinct cluster than those formed by other method.

Cluster analysis has been carried out to substitute the geo-interpretation of hydro-geochemical data. Cluster analysis has been useful in studying the similar pair of group of chemical constituents of water. The similarity/dissimilarity measurements and linkage method used for clustering greatly affect the outcome of the Hierarchical Cluster Analysis (HCA) results. After a careful examination of the available combination of similarity/dissimilarity measurements, it was found that using Euclidean distance (straight line distance between two points in c-dimensional space defined by c variable) as similarity measurement, together with Ward's method for linkage, produced the most distinctive groups. In these groups each member within the group is more similar to its fellow member than to any other member from outside the group. The HCA technique does not provide a statistical test of group differences; however, there are tests that can be applied externally for this purpose. It is also possible in HCA result that one single sample does not belong to any of the groups is placed in a group by itself. This unusual sample is considered as residue. The value of chemical constituents have been linked. Then the next most similar pairs of groups and so on, until all the chemical constituents have been clustered in a dendrogram by an average method (Davis 1973; 1986).

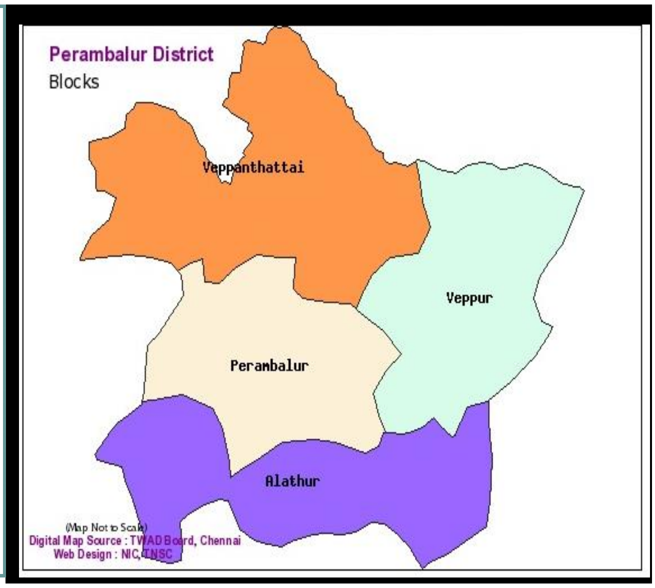
STUDY AREA

Perambalur District is located in the centre part of Tamil Nadu in India. The district occupies an area of 1,750 sq.km and geographically situated between the North latitudes 11°02´ to 11°32´ and East longitudes 78 °38´ to 79° 10´ Perambalur district is bounded on the North by Cuddalore and Salem, South by Trichy, East by Ariyalur, West by Trichy and Salem .

Perambalur district have 3 Taluks (Perambalur, Kunnam, Veppanthattai), 4 Blocks (Perambalur, Kunnam, Veppanthattai, Alathur),and 121 Villages. Perambalur Taluk is taken as the study area for this assessment . Perambalur block is a revenue block in the Perambalur district. It has a total of 20 panchayat villages. Samples are taken from 10 villages in perambalur block, perambalur district,these villages are shown in table (1). In this district highly rich in mineral deposits.Celeste,Lime Stone,Shale,Sand Stone,Canker and Phosphate presents in various parts of district.

TABLE 1 : SAMPLE STATIONS

S.NO	SAMPLE STATIONS
1	Bommanappadi
2	Chattramanai
3	Kalarampatti
4	Ammapalayam
5	Ladapuram
6	Melapuliyur
7	Kurumbalur
8	Aranarai
9	Pudunaduvallur
10	Velur



PERAMBALUR BLOCK VILLAGES

METHODOLOGY

A total of 90 groundwater samples were collected from the study area during August, September &October 2016. Each 30 samples are collected during August, September and October by spatial wise (Bore well, Hand pump, Open well). The sample bottles were labelled, sealed and transported to the laboratory and preserved by adopting standard preservation methods. The major anionic and cationic concentrations were determined in the laboratory using the standard analytical procedures as recommended by the APHA. Sampling and analysis were carried out using standard procedures shown in table (2). These parameters were analysed by multivariate statistical methods are Descriptive analysis (DA), Hierarchical Cluster analysis (HCA), Principle component analysis (PCA), Factor analysis (FA). Finally the study area groundwater quality status has been identified.

Table 2: Instrumental and Volumetric Methods Used for Chemical Analysis of Groundwater in the Study Area

Chemical Parameters	Methods And Instruments
pH, Alkalinity	Field water testing kit
EC	EC Meter
Ca^{2+}, Mg^{2+}	Titration Method (EDTA)
Na^{+}, K^{+}	Flame photometer

TDS, NO3, Cl,	Field water testing kit
HCO3,	Titration Method
SO4, H4SiO4,	UV Spectrophotometer

RESULT AND DISCUSSION

1. DESCRIPTIVE ANALYSIS

The average, mean, and standard deviation values of the water samples collected during three months in different formations are given in table 3, 4, 5. The total cation and anion balance (Freeze and Cherry 1979) is considered to show the charge balance error percentage. The error percentage in the samples of the present study ranges between ±1% and ±10%. Occurrence of errors in chemical analysis of ground water is also due to reagents employed, limitations of the methods and the instruments used, presence of impurities in distilled water, etc.

During the august month Cl is the dominant anion followed by HCO_3, No_3, So_4, F. (Cl> HCO_3 >So_4> No_3> F). The contribution of Cl may be due to chemical weathering of chloride. Na is dominant cation followed by Ca, K, Mg (Na > Mg >K >Ca).

During September and October month Cl is the dominant anion followed by HCO_3, No_3, So_4, F (Cl> So_4 > HCO_3> No_3 > F). The contribution of Cl may be due to chemical weathering of chloride. Na is dominant cation followed by Ca, Mg, K (Na > K> Mg> Ca).

Table 3 Descriptive Statistics for august month

Parameters	Minimum	Maximum	Mean	Std. Deviation
Ca	12.00	92.00	42.3333	19.82046
Mg	28.00	148.00	69.7000	31.79910
Na	6.00	398.00	129.2667	114.50732
K	.10	117.00	31.2933	31.46192
Cl	50.00	560.00	196.0000	118.16412
HCO3	140.00	460.00	295.6667	72.47750
SO4	11.00	325.00	80.9000	79.99369
NO3	.00	45.00	13.6667	12.92774
F	.00	1.50	.5667	.58329
pH	6.50	7.50	7.0500	.33088
EC	390.00	2620.00	1171.7000	633.63103
TDS	636.00	1296.00	980.2667	197.34216

Table 4 Descriptive Statistics for September month

Parameters	Minimum	Maximum	Mean	Std. Deviation
Ca	18.00	96.00	46.2000	21.37578
Mg	32.00	132.00	70.4667	28.80461
Na	12.00	380.00	127.8000	107.26744
K	.80	120.00	35.0000	32.06678
Cl	60.00	500.00	198.0000	101.58740
HCO3	180.00	450.00	297.0000	60.63742
SO4	12.00	320.00	81.6000	78.74341
NO3	.00	45.00	9.6667	13.45063
F	.00	1.50	.4833	.49971
pH	6.50	7.50	7.0500	.33088
EC	400.00	2500.00	1161.9333	616.70956
TDS	684.00	1296.00	974.8000	163.46642

Table 5 Descriptive Statistics for October month

Parameters	Minimum	Maximum	Mean	Std. Deviation
Ca	20.00	98.00	47.8333	20.18036
Mg	28.00	140.00	71.1667	30.16287
Na	10.00	386.00	127.0000	105.56351
K	1.20	116.00	33.8800	30.57624
Cl	66.00	530.00	202.6667	105.71965
HCO3	174.00	440.00	293.5333	61.39231
SO4	10.00	332.00	84.2000	80.30863
NO3	.00	45.00	8.1667	11.77983
F	.00	1.50	.5167	.49971
pH	6.50	7.50	7.0333	.31984
EC	380.00	2560.00	1160.4000	620.16380
TDS	670.00	1320.00	981.1333	169.32049

2. FACTOR ANALYSIS (or) PRINCIPLE COMPONENT ANALYSIS

In this study, water quality variables were grouped using factor analysis. The correlation matrix of variables was generated and factors extracted by the Principal Component Analysis method, rotated by Varimax
The factor analysis generated four significant factors. Parameters were grouped based on the factor loadings. Types of pollution generation based on factor analysis for three month data as shown in table 9.

Table 6 Principal Component Analysis of water quality data for august month

Parameters	Principle Component				Communalities
	1	2	3	4	
Ca	.452	.581	.362	-.315	.772
Mg	.569	.386	.345	-.049	.595
Na	.718	-.227	-.010	.156	.592
K	.638	-.114	-.356	-.289	.630
Cl	-.094	.701	-.345	.479	.848
HCO3	.124	.408	-.102	-.713	.701
So4	.723	-.438	-.025	.118	.729
No3	.029	-.071	.772	.064	.606
F	.340	.377	.079	.586	.608
pH	.262	.155	-.645	-.047	.511
EC	.832	-.400	-.012	.103	.863
TDS	.312	.807	.007	.053	.751
Eigen values	2.973	2.383	1.525	1.325	
% Variance explained by component	24.776	19.857	12.707	11.040	
Cumulative % variance	24.776	44.633	57.341	68.381	

The Factor Analysis (FA) generated four significant factors for the august month, which are explained as 68.381 % of the variance in data sets. Table 6 gives the rotated factor loadings, communalities, Eigen value and the percentage of variance explained by these factors. In order to reduce the number of factors and enhance the interpretability, the factors are rotated. The rotation usually increases the quality of interpretation of the factors. There are several methods of the initial factors matrix to attain simple structure of the data. In this regard, Principal Components Analysis (PCA) is widely used. After PCA 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. The parameters are grouped based on the factor loadings and the following factors are identified:

Factor 1 (F1): Mg, Na, K, So_4, EC
Factor 2 (F2): Ca, Cl, TDS
Factor 3 (F3): **No3**
Factor 3 (F4): F

F1, F2, F3 and F4 have been explained as 24.776%, 19.857%, 12.707 % and 11.040%of the variance respectively. The F1 has a high positive loading in Mg, Na, K, So4, and EC. The High positive loading indicated strong linear correlation between the actor and the parameters.

Table 7 Principal Component Analysis of water quality data for September month

Parameters	Principle Component				Communalities
	1	2	3	4	
Ca	.386	.502	.345	.502	.824
Mg	.569	.349	.554	.105	.785
Na	.727	-.209	-.090	-.138	.611
K	.665	.110	-.360	-.187	.786
Cl	-.076	.673	.145	-.619	.916
HCO3	-.018	.535	-.428	.404	.723
So4	.735	-.352	-.105	-.077	.696
No3	.123	-.356	.694	.330	.793
F	.193	.469	-.181	.107	.727
pH	.195	.322	-.422	.485	.701
EC	.858	-.331	-.098	-.112	.874
TDS	.228	.794	.232	-.313	.837
Eigen values	2.867	2.484	1.535	1.336	
% Variance explained by component	23.895	20.703	12.792	11.137	
Cumulative % variance	23.895	44.597	57.389	68.526	

The Factor Analysis (FA) generated four significant factors for the September month, which are explained as 68.526 % of the variance in data sets. Table 7 gives the rotated factor loadings, communalities, Eigen value and the percentage of variance explained by these factors. In order to reduce the number of factors and enhance the interpretability, the factors are rotated. The rotation usually increases the quality of interpretation of the factors. There are several methods of the initial factors matrix to attain simple structure of the data. In this regard, Principal Components Analysis (PCA) is widely used. After PCA 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. The parameters are grouped based on the factor loadings and the following factors are identified:
Factor 1 (F1): Na, K, So_4, EC
Factor 2 (F2): Cl, TDS
Factor 3 (F3): NO3
Factor 4 (F4): Ca

F1, F2, F3 and F4 have been explained as 23.895%, 20.703%, 12.792 and 11.137% of the variance respectively. The F1 has a high positive loading in Na, K, EC and So_4.The High positive loading indicated strong linear correlation between the actor and the parameters.

Table 8 Principal Component Analysis of water quality data for October month

Parameters	Principle Component				Communalities
	1	2	3	4	
Ca	.309	.586	-.062	.550	.745
Mg	.471	.581	-.255	.154	.648
Na	.758	-.142	-.025	-.098	.605

K	.670	.087	.399	-.189	.652
Cl	-.149	.709	-.081	-.500	.781
HCO3	.035	.266	.780	.207	.723
So4	.759	-.319	-.034	-.115	.692
No3	.064	.075	-.624	.566	.712
F	.056	.291	-.303	-.507	.437
pH	.178	.166	.380	.336	.316
EC	.882	-.277	-.154	-.094	.887
TDS	.147	.859	.016	-.144	.781
Eigen values	2.779	2.315	1.494	1.399	
% Variance explained by component	23.157	19.296	12.448	11.656	
Cumulative % variance	23.157	42.452	54.900	66.556	

The Factor Analysis (FA) generated four significant factors for the October month, which are explained as 66.556 % of the variance in data sets. Table 8 gives the rotated factor loadings, communalities, Eigen value and the percentage of variance explained by these factors. In order to reduce the number of factors and enhance the interpretability, the factors are rotated. The rotation usually increases the quality of interpretation of the factors. There are several methods of the initial factors matrix to attain simple structure of the data. In this regard, Principal Components Analysis (PCA) is widely used. After PCA 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. The parameters are grouped based on the factor loadings and the following factors are identified:

Factor 1 (F1): Na, K, So_4, EC

Factor 2 (F2): Cl, TDS,

Factor 3 (F3): Alkalinity

Factor 4 (F4): Ca, NO3

F1, F2, F3 and F4 have been explained as 23.895%, 19.296%, 12.448% and 11.656 % of the variance respectively. The F1 has a high positive loading in Na, K, TDS, So_4. The High positive loading indicated strong linear correlation between the actor and the parameters.

Table 9 Types of pollution generation based on Factor Analysis

Month	**Factor**	**Parameter**	**Types of pollution**
August	factor 1	Magnesium, Sodium, Potassium, Sulphate, EC	Domestic waste water pollution
	factor 2	Chloride, TDS	Mineral pollution
	factor 3	Nitrate	Biological pollution
	factor 4	Chloride, Fluoride	Agricultural pollution
September	factor 1	Sodium, Potassium, Sulphate, EC	Domestic waste water pollution
	factor 2	Chloride, TDS	Mineral pollution
	factor 3	Nitrate	Biological pollution
	factor 4	Calcium	Agricultural pollution
October	factor 1	Sodium, Potassium, Sulphate, EC	Domestic waste water pollution
	factor 2	Calcium, Chloride, TDS	Mineral pollution
	factor 3	Alkalinity	Agricultural pollution
	factor 4	Calcium, Nitrate	Biological pollution

In summary, four factors representing,

- Agricultural pollution
- Domestic waste pollution
- Mineral pollution
- Biological pollution

Based on this factor and results of the factor analysis it can be concluded that water quality of the study area was mainly uncontrolled by agricultural, domestic, biological and mineral caused by rapid population growth were also threat for the ground water quality.

3. POLLUTION PRESENT IN WATER

During the august month the percentage of pollution present in the water according to factor analysis 24.776% of domestic waste water pollution, 19.857% of mineral pollution, 12.707% of biological pollution and 11.040% of agricultural pollution present in the water sample. In September 23.895% of domestic waste water pollution, 20.703% of mineral pollution, 12.792% of biological pollution and 11.137% of agriculture pollution present in the water sample and October month samples indicates 23.157% of domestic waste water pollution, 19.296% of mineral pollution, 12.448% of agricultural pollution and 11.656% of biological pollution.

This analysis indicates major pollution occurs due to domestic waste water caused by rapid population growth.

4. CLUSTER ANALYSIS OF WATER SAMPLE

Hierarchical dendrogram for the clustering, (figures 3, 4, 5) for the August, September, October month of the determined physical and chemical parameters for all the studied sites were plotted. Dendrogram in CA provided a useful graphical tool for determining the number of clusters that describe the underlying process leading to spatial variation (Papaioannai.et al 2010). The CA results established that the parameters were principally separated into two big cluster.

- ➢ Cluster 1(8 parameters are included) fluoride, magnesium, potassium, pH, nitrate, sulphate, calcium, sodium
- ➢ Cluster 2 alkalinity, chloride, and hardness.
- ➢ Cluster 3 EC, TDS

A careful consideration of the content of clusters revels that during the august and October month the first clusters included dominant chemical parameters (fluoride, RC, magnesium, potassium, nitrate, sulphate, calcium, sodium) and one physical parameter (pH). The second cluster consist of two chemical parameter (alkalinity, chloride). The third Cluster included the physical parameters EC and TDS. During the September month, the first cluster included dominant chemical parameters (fluoride, potassium, nitrate, sulphate, calcium, sodium, nitrite) and one physical parameter (pH). The second cluster consist of one chemical parameter (alkalinity, Chloride, magnesium). The The third Cluster included the physical parameters EC and TDS.

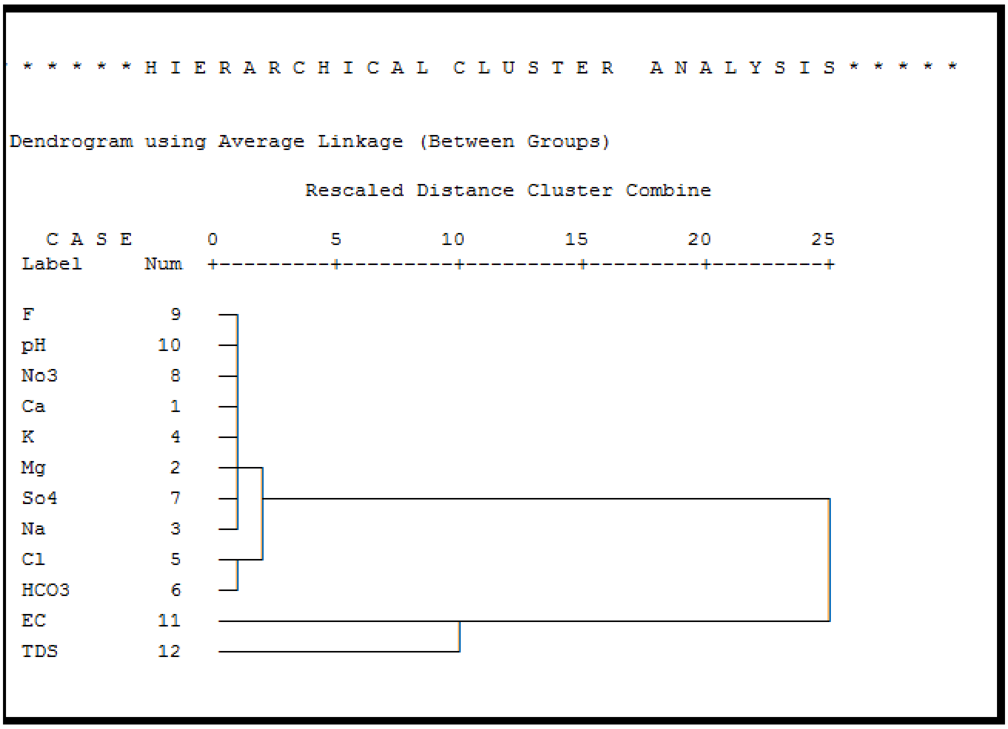


Figure 3 Dendrogram for cluster analysis of water for august month

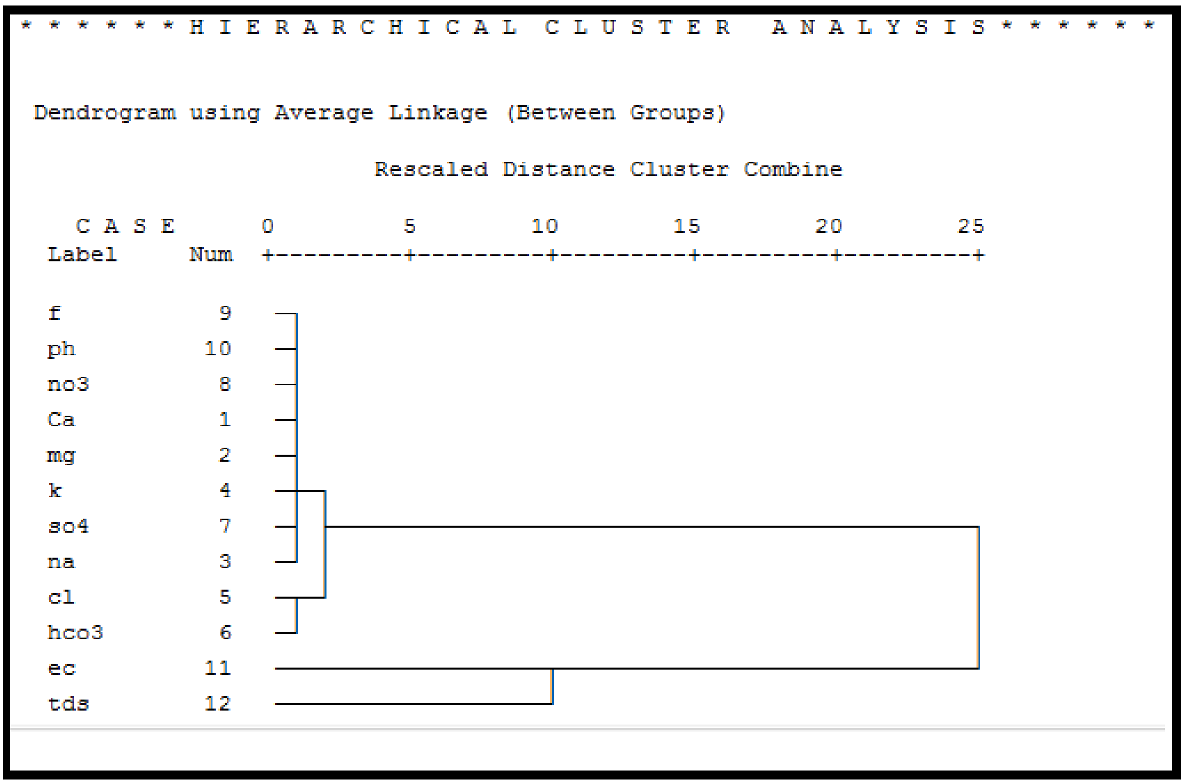


Figure 4 Dendrogram for cluster analysis of water for September month

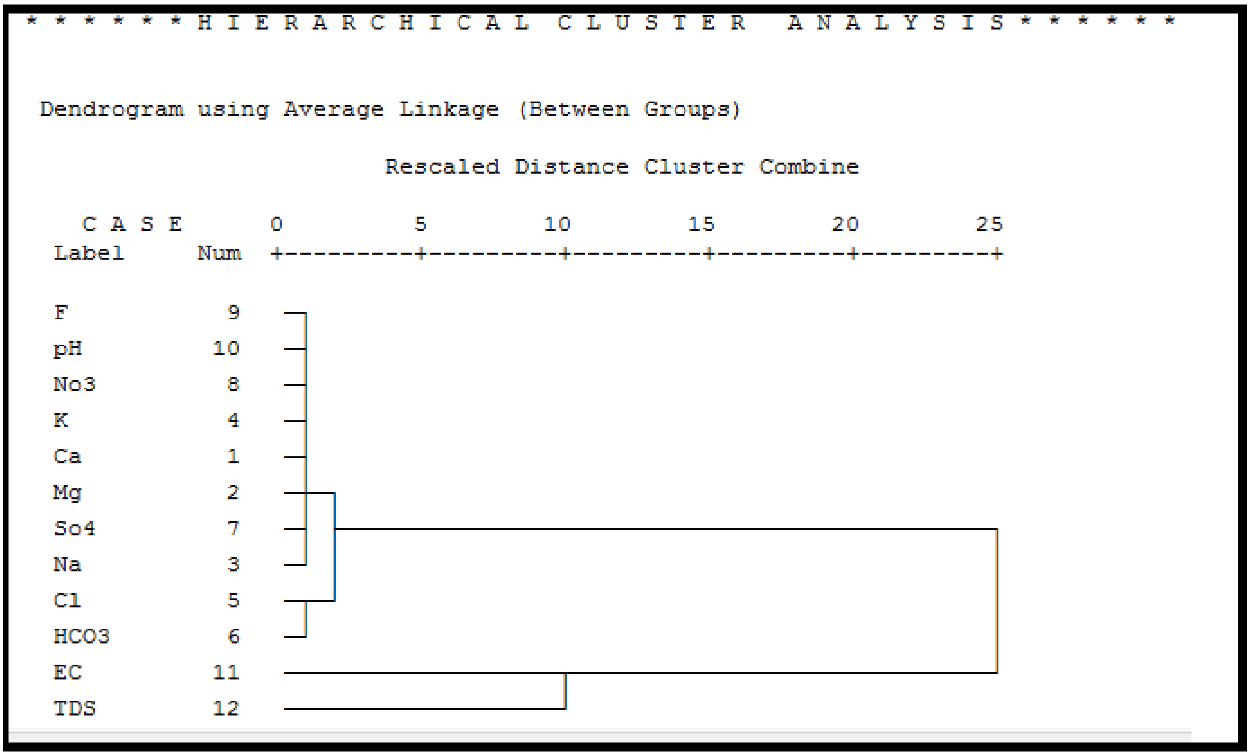


Figure 5 Dendrogram for cluster analysis of water for October month

The data analysis gave an idea of how the single physicochemical parameter should be compared and relater with all the physicochemical values simultaneously, not individually. For instance, within a group of water samples (figure 3, 4, 5) like chloride, alkalinity there is a stronger relation between the group of chemical parameters (chloride, alkalinity) and the physical parameter (EC, TDS) or with parameters (sulphate, calcium, sodium) to the chemical parameters (fluoride, RC, phosphate, ammonia, magnesium, potassium, nitrate) and physical parameter (pH). The study revealed that in all the samples the clustering parameters were more or less same type.

CONCLUSION

Monitoring and assessment of water quality is a continuous process which is must for the proper conservation and management of the water resources. In this study we have analyzed three month water samples from Perambalur Taluk. Water quality monitoring and there indigenous technologies should be adopted to make water fit for domestic and drinking purpose after treatment. From the results of factor analysis identified the pollution type in the study area. Such type of pollution cannot be identified easily. Factor analysis explained the variation in water quality in three factors. It is shown that agricultural pollution, domestic waste water pollution, mineral pollution, biological pollution. Based on this fact and results of the factor analysis it can be concluded that water quality of the study area was mainly uncontrolled by agricultural, domestic and mineral caused by rapid population growth were also threat for the ground water quality.

REFERENCES:

1. Bhattacharyya Rama et.al., " Index Analysis, Graphical And Multivariate Statistical Approaches For Hydro Chemical Characterisation Of Damodar River And Its Canal System, Durgapur, West Bengal, India" in the International Research Journal Of Environmental Sciences Vol.2.(2),53-62, 2013.
2. Florence Lilly P, Paulraj A & Ramachandramoorthy T (2012),*"Assessment of Ground Water Quality of Salem Taluk in Salem District, Tamil Nadu, India"*, in the International Journal of Research in Chemistry and Environment, Vol. 2, pp 323-337.
3. Hulya Boyacioglu & Hayal Boyacioglu (2007),*"Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey"*, in the Journal of Environmental Science and Geology, Vol. 5, pp 275–282.
4. Asif Mahmood, Waqas Muqbool, Muhammad WaseemMumtaz & Farooq Ahmad (2011),*"Application of Multivariate Statistical Techniques for the characterization of Ground Water Quality of Lahore, Gujranwala and Sialkot (Pakistan)"*, in the Pak. J. Anal. Environ. Chem., Vol. 12, pp 102-112.

5. Mohammed Mehdi Heydari & Ali Abasi et al (2013),*"Correlation study and Regression Analysis of Drinking Water Quality in Kashan City, Iran"* in the Journal of Scientific Research at pp 1238-1244.
6. Sati M. Al-Rawin& Abdulmuhsin S.Shihab (2005),*"Application of Factor Analysis as a Tool for Water Quality Management of Tigries River within Mosul city" "* in the Journal of Geology Research at pp 56-64.
7. R.R.Sangpal, U.D.Kulkarani & Y.M.Nandurkar (2011),*" An Assessment Of The Physico-Chemical Properties To Study The Pollution Potential of Ujjani Reservoir, Solapur District, India"* in the journal of Agricultural and biological science at pp 34-38.
8. S.D.Jadhav, D.G.Kanase, R.W.Jawale,, M.S.Jadhav (2007),*" Physico Chemical Assessment of Krishna River at Karad, Satara District (Maharashtra) "* in the Journal of Geology Research at pp 1-6.