

# Geochemical Interpretation Of Groundwater In Torehalla Watershed, Karnataka, India – A Pca Approach

Dr. Manjunath. H. N<sup>1</sup>, Prof. Suresh T.S<sup>2</sup>

<sup>1</sup> Associate professor, SJB Institute of Technology, Bangalore

<sup>2</sup> Professor, Dept. of Geology, Bangalore University, Bangalore, India.

**Abstract-** Principal Component Analysis (PCA) is a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables. It gives a better approach to understand the inter dependence among the variants. When hydrochemical data is subjected to PCA hidden complex chemical relationship is highlighted, thus the different chemical environment in the data set is extracted using loadings and scores. In the study area 71 groundwater samples collected were analyzed for pH, EC, major cations and anions. The analysed hydro chemical data has been subjected to PCA using statistical package PCA.BAS. The principal component analysis transforms the chemical variable into orthogonal principal components. The first four principal components accounts for 76.20% and the rest constitutes 23.80% of the total variance. These four principal components have used to explain hydro chemical process. The percent of variance of these components are 48.5%, 10.9%, 9.1% and 7.7% respectively. Principal component scores reflect the chemical environment of the groundwater in the study area Viz., Total Hardness, Fluoride and Alkalinity.

**Keywords-** Groundwater, Principal Component Analysis, Variables, Eigen values, Hydrogeochemistry

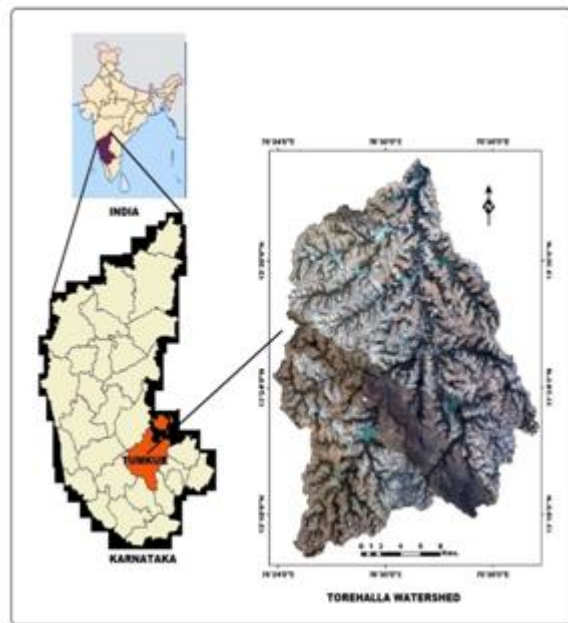
## I. INTRODUCTION

The statistical tools Viz., Cluster, Factor or Principal Component Analyses are used to study the inter- relationships among various ions and for grouping the samples of similar hydrochemical characteristics based on its degree of similarity. Optimal results in multivariate statistical modeling require normal distribution and homoscedasticity. Principal component analysis (PCA) is a mathematical reduction of data without any elaborate assumptions (Anderson, 1958, Morrison, 1964; Wilks, 1963). It is used to emphasize variation and to bring out strong patterns in a dataset. The main purposes of a principal component analysis are the analysis of data to identify patterns and finding patterns to reduce the dimensions of the dataset with minimal loss of information (Mardia et al., 1979 and Mrklas et al., 2006). PCA uses an orthogonal transformation to convert a set of

observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components are less than or equal to the number of original variables. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. The variances of the parameters need to be compatible so that the interpretation of the final results makes logical sense. Where there are vast differences in the variances of the data of the parameters, the results will be influenced more by the parameters with the highest variances in the distribution of their dataset (Davis, 2002). Kaiser (1960) developed a methodology for cutting the number of factors in the final model to size based on merit. The Kaiser criterion requires that a worthy factor in the final model has an eigen value of at least 1. The principal components can be used as predictor or criterion variables in subsequent analyses. In this context an attempt has been made to understand the variation of ionic concentrations of groundwater through particular patterns using PCA for the study area.

## II. STUDY AREA

The Torehalla watershed lies between latitude 13°15' to 13°35'N and longitude 76°20' to 76°40' E, covering an area of 569 Sq.kms. (Map-1). The watershed covers parts of Tiptur and Chikkanayakanahalli taluks of Tumkur district. The mean rainfall in the watershed is 556 mms. The area is of semi-arid type of climate. The topography in general is rolling hills and drainage pattern is Dendritic to semi – Dendritic. The major soil types are red sandy soil, loamy sand and sandy clay loam. The major litho units in the area are granites, gneisses and schists, which are also the aquifers.



Map-1: Study Area

### III. METHODOLOGY

Groundwater samples were collected from 71 locations distributed all over the study area were analyzed for pH, EC, major cations and anions by adopting standard analytical procedures and the results are tabulated in Table-1. The analysed hydrochemical data has been subjected to PCA using statistical package PCA.BAS. Investigation of relationship among a set of variates is conventionally carried out by regression techniques. But the investigation of the dependence among variates is better approached by the use of multivariate techniques like Cluster, Factor and Principal Component analyses. Factor analysis was first introduced by Spearman (1904). Later hydrologic studies by Snyder (1962), Wong (1963) and Wallies (1965) have generated considerable interest on this subject. The term Principal component analysis was coined by Hotelling (1933).

A Principal component analysis involves the extraction of uncorrelated linear combinations of the original variables in such a manner that each successively selected linear combination (termed Principal component) has a smaller variance. A set of observed interdependent variates is transformed into an orthogonal set of variants called Principal components. The first Principal component accounts for as much as possible of the variance of the 'n' observed variates. The second Principal component accounts for as much as possible of the residual variance not accounted by the first Principal component. Likewise, each succeeding Principal component accounts for as much as possible of the residual variance not accounted for by all the previous Principal components.

In Principal component analysis, the data is standardized by subtracting from each observation the mean of the data set and dividing by the standard deviation. The correlation matrix is determined to know the interdependence of variables, the values are tabulated in Table- 2. Further the correlation matrix is used in determination of eigen values and eigen vectors. Kaiser (1960) has suggested that eigen values less than 1.0 are generally not significant. The rationale of choosing 1 is that at least the variance is as large as that of single standardised original variable to be accepted. Later Spencers (1966a) said that in a geochemical data eigen value > 0.50 can often be given physical meaning. There are two techniques called R-mode and Q-mode methods. In the present analyses as interest was centered around the interrelation among variable, R-mode methods has been employed. The PC analysis has been carried out on water chemistry data of the watershed, with constituents Viz., Ca, Mg, Na, K, Fe, Cl, SO<sub>4</sub>, NO<sub>3</sub>, HCO<sub>3</sub>, F, EC, pH, TDS and TH.

### IV. RESULTS AND DISCUSSION

The Principal component results are tabulated in Table -3. In the present analysis, components with eigen value less than 1.0 are not generally significant. However Spencers(1966a) and Grande et al., (1996) have suggested that in a geochemical data eigen value > 0.50 can often have physical meaning. In PCA, each component captures a certain amount of the overall variance in the observed variables, and the components are always listed in order of how much variation they explain. The eigen values measure how much of the variance of the observed variables a factor explains.

Based on the loadings of the different constituents in these factors, the following chemical grouping has been derived:

- Component 1: EC, Cl, TH, SO<sub>4</sub>, Ca
- Component 2: F, Fe, Na, HCO<sub>3</sub>
- Component 3: K, pH, NO<sub>3</sub>, Na
- Component 4: K, HCO<sub>3</sub>, NO<sub>3</sub>, Mg

The order of constituents in each component is in the order of decrease in loadings. In interpretation, all the values of correlation coefficients with a value greater than 0.7 are taken into consideration. Examination of the correlation matrix (Table - 2) reveals that Ca has a strong positive correlation with Cl, EC, TDS and TH. But moderate positive correlation with SO<sub>4</sub> and Mg and low positive correlation with other parameters except Fluoride which has weak correlation with calcium. Mg has strong correlation with TH and moderate

positive correlation with Cl, EC, TDS and  $\text{SO}_4$ . However, negative correlation is observed against Fe and F. Sodium has strong correlation with Cl, EC and TDS and moderate positive correlation with  $\text{HCO}_3$ ,  $\text{SO}_4$  and  $\text{NO}_3$  and weak positive correlation with rest of the parameters.  $\text{K}^+$  is negatively correlated with F and moderately positive correlation with remaining parameters.

Chloride has strong positive correlation with EC, TDS, TH and  $\text{SO}_4$  moderate with  $\text{NO}_3$ ,  $\text{HCO}_3$  weak with rest of the parameters. Sulphate has a good positive correlation with EC and TDS and moderate correlation with TH and  $\text{NO}_3$  and negatively correlated with F. Nitrate has negative correlation with F and pH but shows moderate positive correlation with other variables.  $\text{HCO}_3$  has strong correlation with EC and TDS and negative relation with pH. Fluoride is negatively correlated with EC, pH, TDS and TH. EC has strong correlation with TDS and TH. pH has weak positive relation with TDS and TH. TDS has strong positive relation with TH.

Initially the correlation matrix of the chemical parameters are computed. This is followed by the determination of Eigen values, eigen vectors, trace and cumulative percent of trace and the same is presented in the Table - 3. The first four eigen values accounts for 76.20% and the rest constitutes 23.80% of the total variance. Thus the first three values are used to explain hydrochemical process. The percent of variance of components are 48.5%, 10.9%, 9.1% and 7.7% respectively.

The First Component which is 48.5% has good loading over EC & TDS, Cl, TH,  $\text{SO}_4$ , Ca in descending order and moderate loading over rest of the parameters. Since most of the cations have good correlation with TH, the first factor can be called as 'TH' factor. There is equal loading of TDS and EC which explains the direct inter relationship between them. The Second principal component which accounts for 10.9% of the total variance is loaded over F, Fe, Na,  $\text{HCO}_3$  and poorly loaded over other parameters indicating negative correlation. In this case, parameters seem to be more independent. However based on the highest loading among themselves, 'F' has the highest loading. Hence it can be accounted as 'F' factor. Also since the main aquifers in the study area are granites and gneisses, which are the source for F further substantiates the same. The third Component has good loading over K, pH,  $\text{NO}_3$  and Na, and moderate to poor loading over the rest of the parameters. This factor accounts for 9.1% of total variance. Thus can be accounted as alkalinity factor. The fourth component accounts for 7.7% of the total variance. This component has good loading over K,  $\text{HCO}_3$ ,  $\text{NO}_3$  and Mg. Thus even this also accounts for alkalinity

factor. Thus the PCA analysis points out that major components in waters of the study area are Hardness, Fluoride and Alkalinity. This also has been substantiated in the plot of Hill piper diagram also.

Thus the PCA analysis is useful in grouping and regrouping of chemical parameters of least significance. Hence the results of the PCA enable to know the water - rock interaction, ion exchange process and influence of anthropogenic activity over the groundwater.

## V. CONCLUSION

PCA is appropriate when obtained measures on a number of observed variables to develop a smaller number of artificial variables called principal components that will account for most of the variance in the observed variables. The principal components may then be used as predictor or criterion variables in subsequent analyses. Initially the correlation matrix of the chemical parameters are computed. This is followed by the determination of Eigen values, eigen vectors, trace and cumulative percent of trace. The first four eigen values accounts for 76.20% and the rest constitutes 23.80% of the total variance. Thus the first three values are used to explain hydrochemical process. The percent of variance of components are 48.5%, 10.9%, 9.1% and 7.7% respectively. However

## VI. ACKNOWLEDGEMENT

The authors acknowledge the financial assistance received from the University Grant commission, New Delhi as a Major Research Project in carrying this research work.

## REFERENCES

- [1] Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York.
- [2] Davis, J.C., 2002. Statistical and Data Analysis in Geology, John Wiley & Sons, Inc. London, 543 pp.
- [3] Hotelling, H., 1933. Analysis of a Complex of Statistical Variable into Principal Component. J. Educ. Psycho., 24:417 – 441, 498-520.
- [4] Kaiser, H.F., 1960. The Application of Electronic Computers to Factor Analysis. Educ. Psycho. Measmt., 20 pp.
- [5] Morrison, D.F., 1964. Multivariate Statistical Methods. McGraw Hill, New York.
- [6] Mardia, K., Kent, J.T., and Bibby, J.M (1979). Multivariate Analysis, Academic press, New York, pp:213-254.
- [7] Mrklas, O., Bently, L. R., Lunn, S. R. D., and Chu, A., 2006. Principal Analysis of Groundwater Chemistry Data

- during Enhanced Bioremediation. *Water, Air and Soil Pollution*, Vol.169, pp: 395-411.
- [8] Spearman, C., 1904. General Intelligence Objectively Determined and Measured, *Am. J. Psycho.*, 15, 201 – 293.
- [9] Spencer, D.W., 1966a. Factor Analysis. *Woods Hole Oceanographic Inst. Ref.*: 66 – 69.
- [10] Snyder, W.M., 1962. Some Possibilities for Multivariate Analysis in Hydrologic Studies. *J. Geophys. Res.*, 67(13): 721 – 730.
- [11] Wilks, S.S., 1963. *Mathematical Statistics*. John Wiley and Sons, New York.