Assessment of Ground Water Quality and Source of Contamination by Multivariate Factor Analysis

***Deena Prasad Prajapati[1] , Sandhya Shrivastava[2]**
Instrumental and Analytical Laboratory ,
**Department of Chemistry ,Government P.G. College Seoni (M.P.) 480661, India*

Abstract: **In this study, the factor analysis techniques are applied to ground water quality data sets obtained from the Satna, city, Madhya Pradesh (India). The data obtained were standardized and subjected to Factor extraction to simplifying its interpretation and to define the parameters responsible for the main variability in water quality. The objective is to evaluate the mutual correlations among the various water quality parameters to reveal the primary factors that affect reservoir water quality, and the differences among the various water quality parameters. The factor analysis resulted in three factors explaining more than 77.4% of the total variation in ground water quality data set. The first factor indicates the variation in water quality is due to anthropogenic sources and second factor shows variation in water quality due to organic sources that are taking place in the system. Finally the results of factor analysis reflect a good look on the water quality monitoring and interpretation of the ground water.**

Keywords: **Factor Analysis, ground water, water quality, sources of pollution, Anthropogenic.**

Introduction

Water quality monitoring has one of the highest priorities in environmental protection policy Due to pressure of human activity, urbanization and industrialization, the groundwater sources are degraded gradually; therefore pure, safe, healthy and odourless drinking water is a matter of deep concern. There are many pollutants in groundwater due to seepages viz. organic and inorganic pollutants, heavy metals, pesticides, fluorides etc. The quality of water is identified in terms of its physical, chemical and biological parameters .The quality of water is identified in terms of its physical, chemical and biological parameters (Sargaonkar and DeshPande, 2003). The particular problem in the case of water quality monitoring is the complexity associated with analysing the large number of measured variables (Saffran, 2001). The data sets contain rich information about the behaviour of the water resources. The classification, modelling and interpretation of monitoring data are the Multivariate statistical methods including factor analysis have been used successfully in hydrochemistry for many years. Surface water, groundwater quality assessment and environ mental research employing multi-component techniques are well described in the literature (Liu et al., 2003; Boyacioglu, 2006; Praus, 2005). Multivariate statistical approaches allow deriving hidden information from the data set about the possible influences of the environment on water quality (Sheela et al. 2012; Spanos et al., 2003; Sharma et al., 2005). In the present study, a water quality data matrix, obtained from 8 stations of Satna City Madhya Pradesh, has been subjected to Factor analysis. From this, the study determines the information needed to make effective decisions about managing water resources.

*Address for Communication

Materials and Methods

Study area

The Satna city is situated between Longitude (80”21’and 81”23’East and Latitude is 23”58’ and 25”12’North)satna Geographical area 742432 Hectare/7502 Km^2 and total area 2.4% of MadhayaPradeshis. the height level from sea 317 Mtr.and the Average rain fall 617.6MM(Year 2010-2011) Satna is the northern part of Rewa Commissioner’s Division in Madhaya Pradesh state of India .selected for the proposed studies due to the largest transport system and running of famous Satna Birla Cement Plant,Prism Cement Plant (Mankahari) ,Wire Cable Plant (Satna), which is the largest Cement Factory in India the central region. The city is located 500kilometers west of the capital Bhopal, on the main Howrah–Mumbai rail line, and National Highway 7. There are different sources of drinking water viz., hand pumps, tube wells, open wells, etc. Samples of ground water were collected in clean polyethylene bottles from different sources located in map (No. 1-8) from study area. Using Multivariate Factor Analysis we analyzed the data for different parameters as pH, EC, TDS, Ca, Mg, TH, Cl, Na+, K+, SO42-, NO3-, etc.

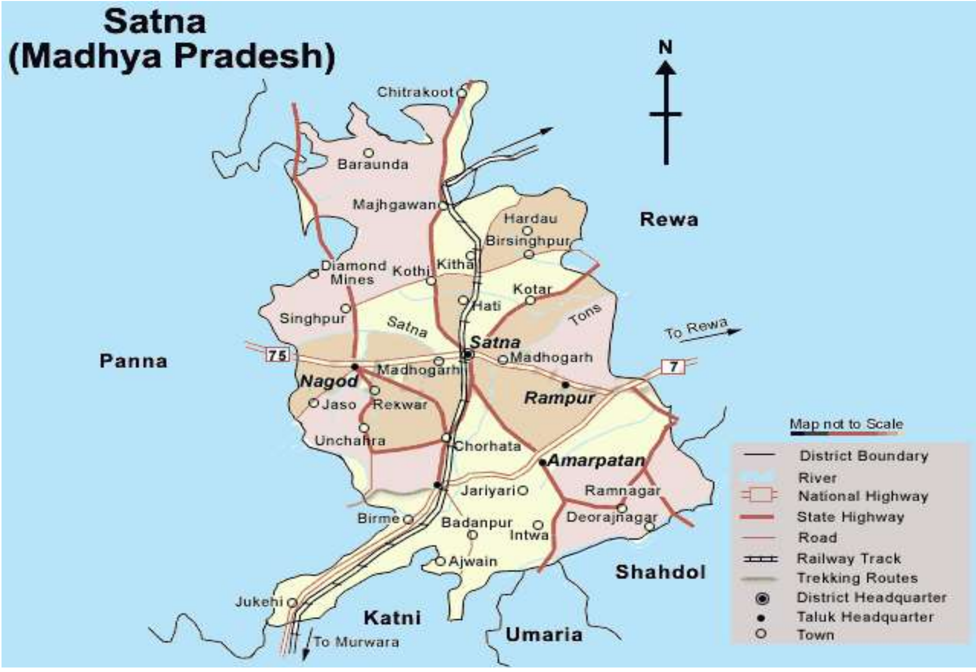

Figure1: Location of sampling sites (No. 1-8)

2 Multivariate Factor Analysis The Factor analysis (FA) is a multivariate statistical technique employed for the purpose of data reduction with a view to determining the sources of elements and their controlling factors. It enables the geographical distribution of the resulting factors to be determined. The geological interpretation of factors yields insight into the main processes, which may govern the distribution of hydrochemical variables. The hydrochemical data were statistically analyzed. The first step was to standardize the raw data to form Table 2. Standardization tends to increase the influence of variables whose variance is small, and reduce the influence of variables whose variance is large. Furthermore, the standardization procedure eliminates the influence of different units of measurement, and makes the data dimensionless. Factor analysis takes data contained in a correlation matrix and rearranges them in a manner that better explains the structure of the underlying system that produced the data. Therefore, the correlation coefficient matrix measures how well the variance of each constituent can be explained by relationships with each of the others. Then, the variances/co-variances and correlation coefficients of the variables are computed, the correlation coefficient is

$$\frac{rxy = [\sum(x - xm)]2[\sum(y - ym)]2}{\sqrt{[\sum(x - xm)]2[\sum(y - ym)]2}}$$

In this expression the correlation coefficients r x,yis simply the sum (over all samples) of the products of the deviations of the x-measurements and they measurements on each sample, from the mean values of x and y , respectively, for the complete set of samples. Table 3 presents the matrix of correlation coefficient for hydrochemical data. Eigen values and eigenvectors were calculated for the covariancematrix. Then, the data were transformed into factors. Table 4 presents the eigen value and the percentages of variance associated with each factor. These values are summed to express as a cumulative eigen value and percentage of variance, respectively. Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable (Panda et la., 2006; Yu et al., 2003). There are three stages in factor analysis (Gupta et al., 2005):

* For all the variables a correlation matrix is generated .
* Factors are extracted from the correlation matrix based on the correlation coefficients of the variables .
* To maximise the relationship between some of the factors and variables, the factors are rotated.

A first step is the determination of the parameter correlation matrix. It is used to account for the degree of mutually shared variability between individual pairs of water quality variables. Then, eigenvalues and factor loadings for the correlation matrix are determined. Eigenvalues correspond to an eigenfactor which identifies the groups of variables that are highly correlated among them. Lower eigen values may contribute little to the explanatory ability of the data. Only the first few factors are needed to account for much of the parameter variability. Once the correlation matrix and eigenvalues are obtained, factor loadings are used to measure the correlation between the variables and factors. Factor rotation is used to facilitate interpretation by providing a simpler factor structure. After rotation of the factor loading matrix (e.g., by varimax rotation, which involves scaling the loadings by dividing them by the corresponding communality), the factors can often be interpreted as origins or common sources (Kuppusamy and Giridhar, 2006, Zeng and Rasmussen, 2005).

Results and Discussion

Factor analysis was applied to data sets and statistical description of the measured water quality parameters (mean, standard deviation, minimum and maximum values are listed in Table 1. The correlation matrix of variables was generated and factors extracted by the Varimax rotation calculated eigenvalues, percent total variance, factor loading and cumulative variance are given in Table 1. The factor analysis generated three significant factors which explained >77% of the variance in data sets

Figure 2.

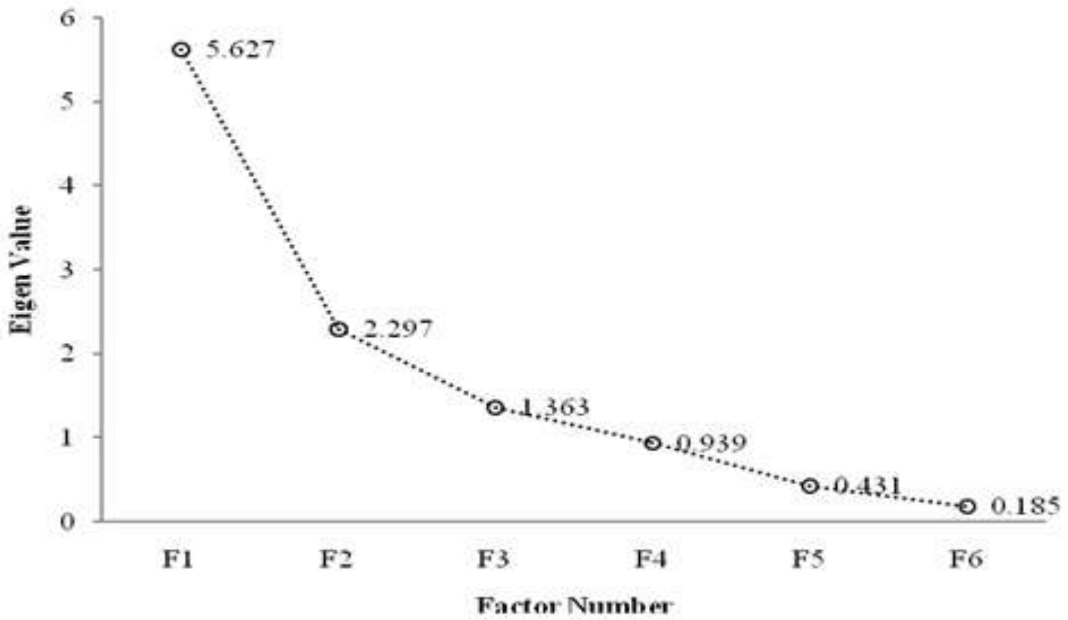

Table No. 1

Parameter	F1	F2	F3
pH	0.034	**-0.516**	0.238
EC	**0.947**	0.150	0.109
TDS	0.636	**-0.748**	-0.109
Mg-H	-0.423	-0.274	**-0.710**
Ca-H	**0.964**	0.190	0.091
TH	-0.133	**-0.807**	0.531
NO3	**0.830**	0.336	-0.184
PO4	**0.864**	-0.241	-0.314
SO4	**0.821**	0.227	0.098
Cl	**0.838**	0.129	0.083
Na	0.438	**-0.735**	-0.330
K	0.385	-0.319	0.127
Eigen Value	5.627	2.297	1.363
Variability (%)	46.890	19.140	11.362
Cumulative %	46.890	66.030	77.392

Table 2 :Correlation matrix of parameters

Var	pH	EC	TDS	Mg-H	Ca-H	TH	NO_3^-	$PO4^{-3}$	SO_4^{-2}	Cl^-	Na^+	K^+
pH	1											
EC	-[illegible]	1										
TDS	0.194	0.463	1									
Mg-H	-[illegible]	-[illegible]	0.040	1								
Ca-H	-[illegible]	0.969	0.419	-0.563	1							
TH	0.538	-[illegible]	0.448	-0.108	-0.246	1						
NO_3^-	-[illegible]	0.733	0.307	-0.231	0.826	-0.389	1					
$PO4^{-3}$	-[illegible]	0.723	0.724	-0.006	0.794	-0.096	0.741	1				
SO_4^{-2}	0.224	0.727	0.380	-0.389	0.804	-0.151	0.950	0.660	1			
Cl^-	0.005	0.887	0.503	-0.502	0.817	-0.280	0.643	0.609	0.662	1		
Na^+	-[illegible]	0.321	0.893	0.192	0.255	0.441	0.155	0.625	0.158	0.245	1	
K^+	0.198	0.420	0.440	-0.191	0.427	0.120	-0.027	0.507	-0.047	0.385	0.295	**1**

For factor loadings the values which are greater than the radius of the equilibrium circle of contribution (Radius = (No of extracted factors/ No of variables) 0.5 = (4/12) 0.5 are considered (Al-Rawi and Shihab, 2005).

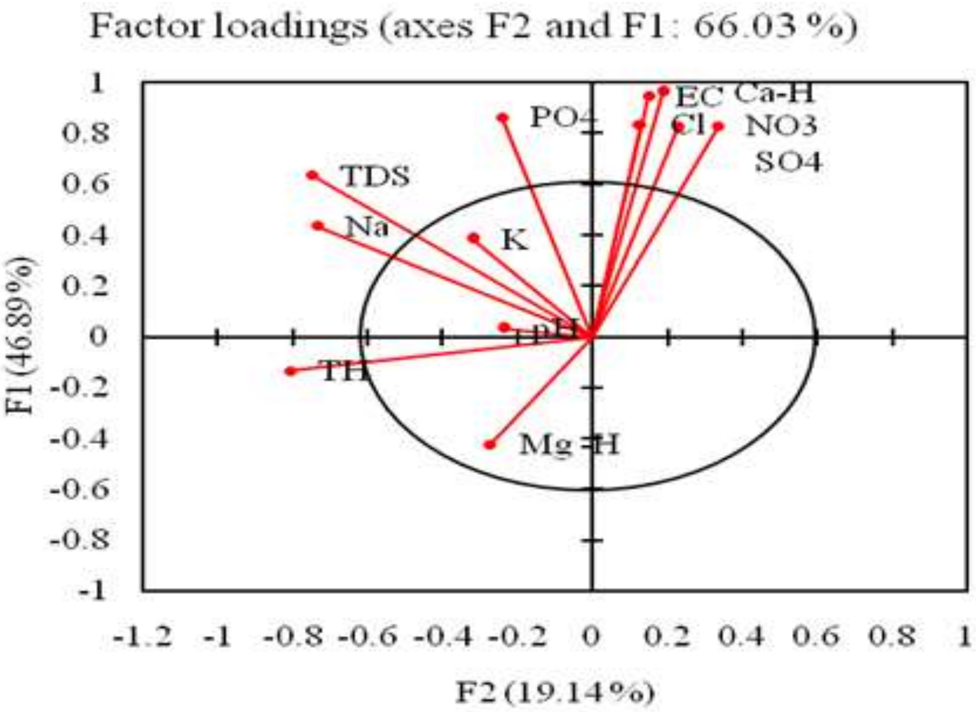

Figure 3:
Correlations of water quality parameters with Factor I and II The following factors were indicated considering the hydro chemical aspects of the water:- Factor I: It accounts for 46.89% of the variance in water quality of ground water within the study area **Figure 3**. It represents the strong loadings of EC, Ca-H, $NO3^-$, $PO4^{3-}$, $SO4^{2-}$ &Cl^- ,which originates from anthropogenic causes like industrial and agricultural pollution in surface water. So we named this factor as „anthropogenic origin‟. The electrical conductivity (EC) is positively correlated with the concentration of ions, which can thus be indirectly calculated from EC. Therefore, EC can be regarded as a water salinization index.

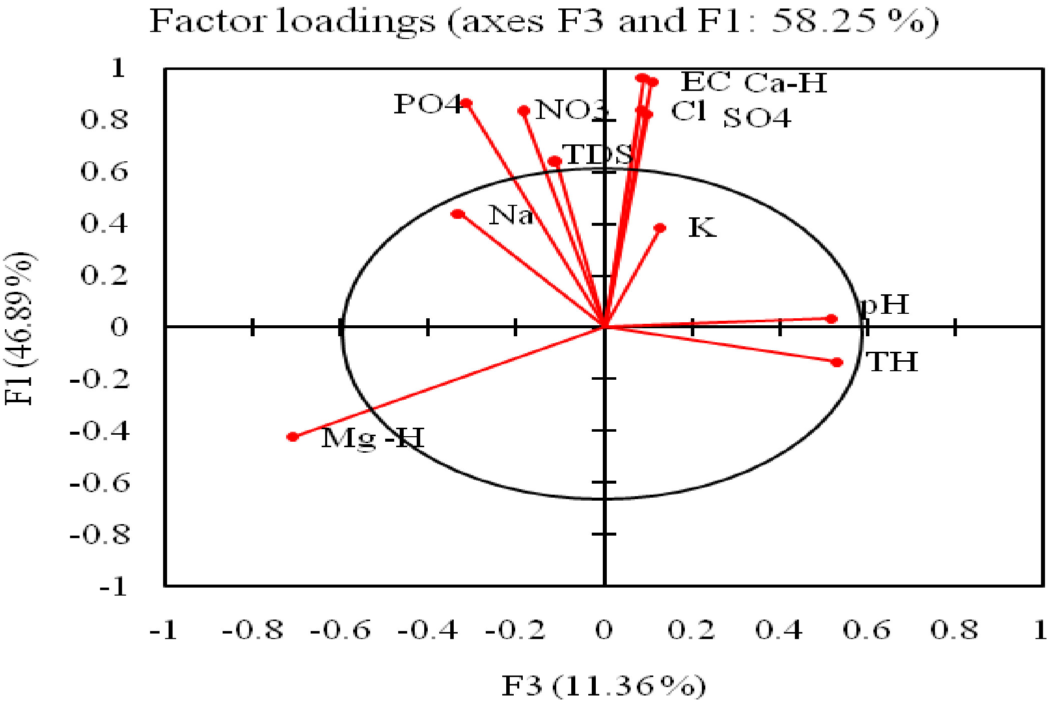

Figure 4:

Correlations of water quality parameters with Factor I and IIIFactor II: It account for > 19% of the variance in water quality **Figure 4.** It is mainly dominated by TDS and TH. TDS is strongly correlated Table 2 with all major ions calculated. This emphasizes that the colloidal fraction (as dissolved) of water constituting water turbidity. The runoff from catchment area also contributes in the variation in total solids and total hardness. These parameters are also significantly correlated to each other Table 2. Factor III: It account for 11.36% of the variance in water quality. It includes mainly pH which comes from natural sources **Figure 4**.

Conclusions

Factor analysis is an effective means of manipulating, interpreting, and representing data concerning water pollutants. Factor analysis was applied to ground water samples collected from an area in the vicinity of industrial site. The results showed that a three factor which explains >77% of the variation in quality of ground water. The data indicate the ground water of studied locations is highly deteriorated as it is polluted with high amount of anthropogenic causes like industrial and agricultural. This study shows that Factor analysis is a useful method that it could assist decision makers in determining extent of pollution via practical pollution indicators. This could also provide crude guidelines for selecting the priorities of possible preventative measures in the proper management of surface water.

References

1. 01. H. Boyacioglu, Surface water quality assessment using factor analysis. Water SA, 32 (2006) 389-394.
2. 02. X. Zeng, T.C. Rasmussen, (2005) Multivariate statistical characterization of water quality in lake lanier in Georgia, USA, J. Environ. Qual. 34, 1980-1991
3. 03. P. Praus, (2005) Water quality assessment using SVD- based principal component analysis of hydrological data, Water SA, 417-422.
4. 04. S. Yu, J. Shang, J. Zhao, (2003) Factor analysis and dynamics of water quality of the Songhua River Northeast China. Water, Air Soil Pollut. 144, 159-169.
5. 05. A. K. Gupta, S.K. Gupta, R.S. Patil, (2005) Statistic al analyses of coastal water quality for a port and harbour region in India. Environ. Monit. Assess. 179-200.
6. 06. J. D. Sharma, P. Sharma, P. Jain, D. Sohu, (2005) Chemical Analysis of Drinking Water of Sanganer Tehsil, Jaipur District. Int. J. Environ. Sci. Technol. 2, 373-379.
7. 07. O. Arslan, (2009) A GIS-Based Spatial-Multivariate Statistical Analysis of Water Quality Data in the Porsuk River, Turkey. Water Qual. Res. J. Can. 44, 279-293.
8. 08. K.P. Singh, A, Malik, S. Sinha, (2005) Water Quality Assessment and Apportionment of Pollution Sources of Gomti River (India) Using Multivariate Statistical Techniques: A Case Study.Anal.Chim.Acta, 538, 355-374.
9. 09. M.R. Kuppusamy, V.V. Giridhar, (2006) Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore, Environ. Int. 32, 174-179.
10. A.Mustapha, Ado Abdu, (2012), Application of principal component analysis and multiple

regression models in surface water quality assessment. J. Environ. Earth Sci. 2(2), 16-24.
11. Sati Al-Rawi, A.S Shihab, (2006) Application of factor analysis as a tool for water quality management of Tigris river within Mosul City. Raf J Sci. Geol. 16, 56-64.
12. A.M. Sheela, J. Letha, S. Joseph, M. Chacko, S.P. Sanalkumar, J. Thomas, (2012) Water quality assessment of a tropical coastal lake system using multivariate cluster, principal component and factor analysis, Lakes & Reservoirs: Research & Management, 17, 143-159.
13. U. C. Panda, S. K. Sundaray, P. R. Rath, B. B. Nayak, D. B. Bhatta, (2006) Application of factor and cluster analysis for characterization of river and estuarine water systems- A case study: Mahanadi River (India), J. Hydrol. 331, 434-445.
14. RamchanderMerugu, SwethaSandilyaGarimella, KarunakarRaoKudle, D.Ramesh, M.P. PratapRudra (2012). Optimization studies for defluoridation of water using Aspergillusnigerfungal biosorbent. Int J chem. Tech Research. 4 (3): 1089-1093
15. SandeepShukla, “WATER QUALITY ASSESSMENT OF GORAMA DAM OF HANUMANA REWA (M.P.) INDIA, WITH SPECIAL REFERENCE TO ZOOPLANKTON” International Journal of Innovative Researchin Science, Engineering and TechnologyVol. 2, Issue 6, 2013,ISSN: 2319-8753.
16. SandeepShukla, “WATER QUALITY STATUS OF BENISAGAR DAM CHHATARPUR DISTRICT (M.P.), INDIA” International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 11, November 2013 ISSN: 2319-8753
17. T.P. Sagar,“WATER QUALITY ASSESSMENT OF THE DALSAGAR LAKE AT SEONI (M.P.) International Multidisciplinary Research Indian Stream Research Journal Volume3 | Issue10,2014, ISSN No: 2230-7850.

