GIAP Journals

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

EVALUATION AND MODELLING OF GROUND WATER QUALITY DATA OF ALLAHABAD CITY BY ENVIRONMETRIC METHODS

Girdhari Lal Chaurasia, Mahesh Kumar Gupta, Praveen Kumar Tandon#
#Department of Chemistry, University of Allahabad, Allahabad, UP, India-211002,
E-mail: pktandon1@gmail.com

Article History: Received on 05th March 2016, Revised on 17th September 2016, Published on 10th December 2016

ABSTRACT

water is an essential resource for all the organisms, plants and animals including the human beings. It is the backbone for agricultural and industrial sectors and all the small business units. Increase in human population and economic activities have tremendously increased the demand for large-scale suppliers of fresh water for various competing end users.The quality evaluation of water is represented in terms of physical, chemical and Biological parameters. A particular problem in the case of water quality monitoring is the complexity associated with analyzing the large number of measured variables. The data sets contain rich information about the behavior of the water resources. Multivariate statistical approaches allow deriving hidden information from the data sets about the possible influences of the environment on water quality. Classification, modeling and interpretation of monitored data are the most important steps in the assessment of water quality. The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA) and factor analysis (FA) help to identify important components or factors accounting for most of the variances of a system. In the present study water samples were analyzed for various physicochemical analyses by different methods following the standards of APHA, BIS and WHO and were subjected to further statistical analysis viz. the cluster analysis to understand the similarity and differences among the various sampling stations. Three clusters were found. Cluster 1 was marked with 3 sampling locations 1, 3 & 5; Cluster-2 was marked with sampling location-2 and cluster-3 was marked with sampling location-4. Principal component analysis/factor analysis is a pattern reorganization technique which is used to assess the correlation between the observations in terms of different factors which are not observable. Observations correlated either positively or negatively, are likely to be affected by the same factors while the observations which are not correlated are influenced by different factors. In our study three factors explained 99.827% of variances. F1 marked 51.619% of total variances, high positive strong loading with TSS, TS, Temp, TDS, phosphate and moderate with electrical conductivity with loading values of 0.986, 0.970, 0.792, 0.744, 0.695, 0.701, respectively. Factor 2 marked 27.236% of the total variance with moderate positive loading with total alkalinity & temp. with loading values 0.723 & 0.606 respectively. It also explained the moderate negative loading with conductivity, TDS, and chloride with loading values -0.698, -0.690, -0.582. Factor F3 marked 20.972 % of the variances with positive loading with PH, chloride, and phosphate with strong loading of pH 0.872 and moderate positive loading with chloride and phosphate with loading values 0.721, and 0.569 respectively.

KEYWORDS: Water, Physico-chemical, Environ-metric methods, Cluster Analysis, FA/PCA.

1. INTRODUCTION

Water is essential to sustain life. Improving access to safe drinking-water can result in tangible benefits to health [1, 2, 3, and 4]. Water is the most precious resource of nature and it is the basic need for the life for all organisms, plants, animals and human beings. Water is the fundamental resource which is essential for different sectors, such as agricultural sector, industrial sector and all the small business units. Increase in human population and economic activities have tremendously increased the demand for large-scale suppliers of fresh water for various competing end users. The decline in the quality and quantity of surface water resources can be attributed to water pollution and the improper management of the resources [5]. Many regions around the world are simultaneously impacted by urbanization processes and industrial and agricultural activities, and many cities in developing countries have been developed without adequate and proper planning. This has led to indiscriminate actions, including dumping of wastes into the water and washing and bathing in open surface water bodies [6]. The deteriorating water quality affects men, animals, and also the plant life with far-reaching consequences. From the environmental, economical, and/or social point of view, it is important to identify these sources and their contribution to the total contamination of an area [7].

The quality evaluation of water is represented in terms of physical, chemical and biological parameters. Polluted surface &groundwater cannot achieve the balanced ecosystem. The balanced ecosystem is the system of mutual interacting and sharing the benefits among organisms and environmental interactions. Water is a fundamental resource of almost of ecosystems hence the quality of water is prime factors for the maintenance of ecosystem homeostasis and balanced Ecosystem[8]

GIAP Journals
Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

A particular problem in the case of water quality monitoring is the complexity associated with analyzing the large number of measured variables.The data sets contain rich information about the behavior of the water resources. Multivariate statistical approaches allow deriving hidden information from the data set about the possible influences of the environment on water quality. Classification, modeling and interpretation of monitored data are the most important steps in the assessment of water quality [9].

The application of environ-metric methods (principal component analysis, cluster analysis and factor analysis etc.) is the best classification and modeling tools for the environmental raw data generated by the different sampling sites and it is a reliable method in which no chance of misinterpretation of the monitoring data. Environ-metric methods have advantage of visualization of the huge amount of raw analytical measurements and also can extract the informations about the possible pollutions point and nonpoint and other possible pollution sources [10].Environmetric methods including cluster analysis &factor analysis have been successfully used by various researchers for many years. Environ-metric methods have been used for the assessment of surface water, groundwater and other environmental research by various workers. These methods extract the hidden information in the single sample and whole data and also the impacts of the environmental factors on the water quality and assist the environmental managers in terms of aiding the decision making process [11, 12].

The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA) and factor analysis (FA) help identify important components or factors accounting for most of the variances of a system [13,14].They are designed to reduce the number of variables to a small number of indices while attempting to preserve the relationships present in the original data.

Factor analysis is used to know the correlations among the observations in reference to factors which are not easily observable. Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable [15]. Observations that are highly correlated (either positively or negatively) are likely to affected by the same factor and those are uncorrelated are likely to affected by different factors.

1.1The main objective of the study

- Collection of groundwater samples in Allahabad city from the handpumps.
- Analysis of physicochemical parameters of groundwater samples collected from 5 different locations
- To apply the CA, PCA/ FA techniques in the evaluation of correlations of water quality parameters

2. MATERIALS AND METHODS:

2.1 The study area

Allahabad district forms the part of central Ganga alluvial Plain of quaternary age in the north. Vindhyan formations are exposed in the southern part of the district generally known as Yamuna par area. The district has an area of 7261 sq.km. and lies between north latitudes 24º47´ and 25º47´ and east longitudes 81º09´ and 82º21´ with total population of 62,36,447 as per 2001 census (density:859 persons/sq.km.). Two important rivers namely Ganga and Yamuna divide the district into three physiographic Units (1) Trans Ganga Region (2) Doab Region & (3) Trans Yamuna Region. The district is drained by river Ganga & Yamuna & its tributaries.The groundwater samples were collected from different locations in Allahabad city in the month of April 2014 which is shown below in location map figure 1:

2.1.1 Hydrogeology:

Ground water in the district occurs both in alluvium (unconsolidation) and in the weathered & jointed sandstones areas which are underlain by hard rock's (consolidated). In the unconsolidated or alluvial formation ground water occurs under unconfined to confined conditions in the shallow and deeper aquifers respectively and depth to water ranges between 2 to 20 metres during pre-monsoon period, while in the post monsoon period it stands between 1 to 18.00 meters. Ground water exploration reveals 3 tier aquifer systems in the alluvialarea[16].The geohydrological map shown in figure:2

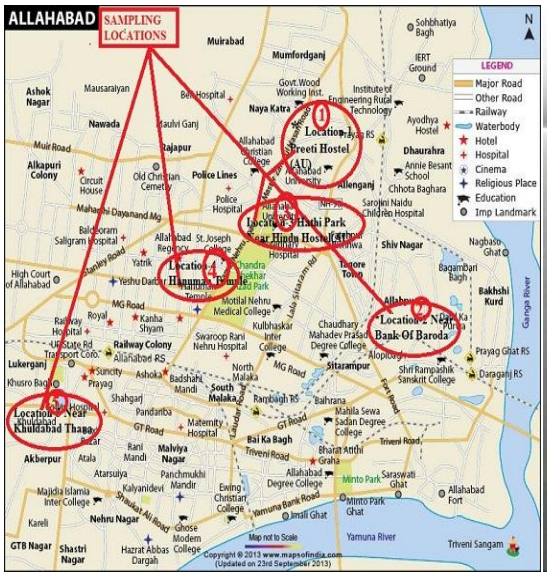
I Aquifer Group 00.00-110 mbgl

II Aquifer Group 120-250 mbgl

III Aquifer Group 260-400 mbgl

GIAP Journals

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248



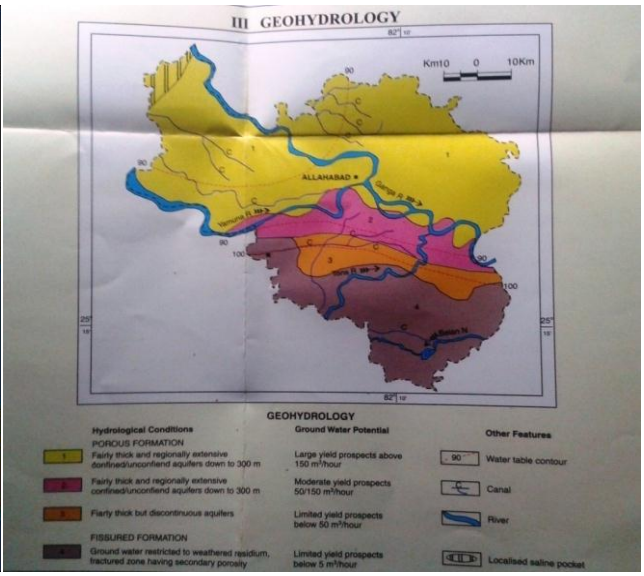


Figure 1: Location map of sampling

(Source: District Allahabad city (web source)

Figure 2: Map geohydrology of Allahabad

(District resource map provided by Northern Region GSI, Lucknow**)**

2.2 Physicochemical Properties of Water:

The water samples were collected in duplicate from the hand pump India mark No.4 in one liter sampling bottles. The temperature was measured at the site by the Hg-thermometer and all other parameters were measured by following the APHA [17] standard procedure and were compared with the standards provided by BIS [18] & WHO [19]. The chemical analysis was done for the following nine parameters viz. temperature, PH, TDS, TSS, TS, total alkalinity, electrical conductivity, chloride and phosphate. The methods used for analysis & Descriptive Statistics is shown below in Table 1.

Table1:Analytical methods used & descriptive statistics of various physicochemical parameters

Parameter	Abbreviation	Unit	Method	Minimum	Maximum	Mean	Std. dev
Temp	^{0}C	centigrade	Hg-thermameter	26.000	28.000	27.000	1.000
PH	pH	pH unit	pH-meter	5.450	6.960	6.048	0.577
Total Solid	TS	mg l^{-1}	Gravimetric	911.000	1464.000	1117.400	212.766
Total Suspended Solid	TSS	mg l^{-1}	Gravimetric	433.000	774.000	581.000	134.718
Total dissolved solids	TDS	mg l^{-1}	Gravimetric	445.000	690.000	536.400	104.816
Electrical-Conductivity	EC	µs cm^{-1}	Electrometric	878.000	1390.000	1068.800	220.098
Total Alkalinity	T-Alk $CaCO_3$	mg l^{-1}	Titrimetric	57.640	361.680	241.208	114.963

GIAP Journals

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

Chloride	Cl	mg l^{-1}	Argentometric Titration	94.300	160.000	118.534	25.888
Phosphate	$PO4^-$	mg l^{-1}	Spectro-photometric	5.000	10.000	7.260	2.022

2.3Multivariate Statistical Methods:

Multivariate statistical methods ground water quality data sets were subjected to three multivariate techniques: cluster analysis (CA),principal component analysis (PCA) and factor Analysis (FA). All mathematical and statistical computations were made using Microsoft Office Excel 2003, Xlstat 2015, and SPSS 16.0

2.3.1 Cluster analysis:

Cluster analysis is multivariate statistical techniques which is applicable for the understanding of the each object which is similar with others in respect to predetermined selected criterion.Hierarchical agglomerative clustering is the most common approach for the study of the similarity relationships among one sample and the whole datasets and is typically illustrated in the form of dendrogram. Dendrogram provide the picture of the clustering of objects groups with dramatic reduction of dimensionality of the original raw data set [20, 21].

2.3.2 Principal component analysis /factor analysis:

Principal component analysis/factor analysis (PCA/FA) is a powerful pattern recognition tool that attempts to explain the variance of a large dataset of intercorrelated variables with a smaller set of independent variables [22]. PCA technique extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The Principal Components (PC) is the uncorrelated (orthogonal) variables obtained by multiplying the original correlated variables with the eigenvector, which is a list of coefficients (loadings or weightings). Thus, the PCs are weighted linear combinations of the original variables. PC provides hidden information on the most significant parameters, which describe the entire data set while attempting data reduction with a minimum loss of original information [23,24, and 25].

FA further reduces the contribution of less significant variables obtained from PCA and the new group of variables known as varifactors (VF) is extracted through rotating the axis defined by PCA. A VF can include unobservable, hypothetical, latent variables; while a PC is a linear combinationof observable water-quality variables [25, 26].

The factor loadings is assigned on the basis of loading value as “strong,” “moderate” and “weak,” corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively[27].

3.RESULT & DISCUSSION:

3.1 Cluster analysis

Cluster analysis was carried out with 5 sampling stations and 9 physicochemical parameters and find out the three clusters with similarities with sampling stations. The Dendrogram of the clusters are shown in figure-3

Cluster -1 marked 3 sampling; location1, 3 & 5; Cluster-2 marked sampling location-2 and cluster-3 mark sampling location-4 on the basis of the sampling locations.

3.2 Factor Analysis: factor Analysis were applied on 5 sampling locations 9 parameters datasets. Descriptive statistics of datasets are presented at **Table- 1**.

The correlation matrix were generated of variables and factors extracted by centroid method and varimax rotated.

The result of factor analysis including Eigen values and total variance and cumulative variances is shown in table-2 & figure-4.0.Three factors F1, F2 & F3 explained 99.827 % of Variances.

Factor 1: Total Suspended Solid [TSS], Total Solid [TS], Temp, Total Dissolved Solid [TDS], Phosphate, Electrical Conductivity [EC]

Factor 2: Total Alkalinity [TA], Conductivity [EC], TDS, Temp, Chloride

Factor 3: PH, Chloride, Total Alkalinity and Phosphate

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

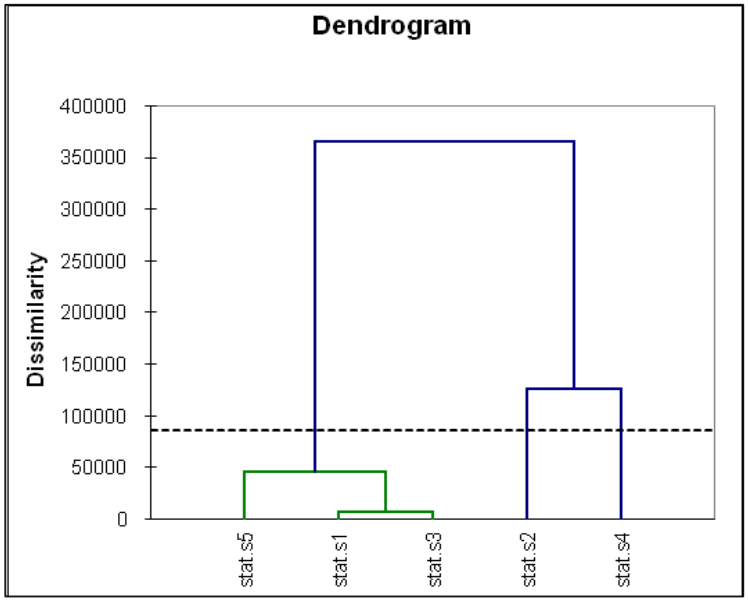


Fig 3.0: Dendrogram showing observations at 5 sampling locations

Factor F1: explained 51.619% of total variances, high positive strong loading with TSS ,TS, Temp, TDS , Phosphate and moderate with Electrical Conductivity with loading value 0.986, 0.970, 0.792,0.744, 0.695 0.701, 0.744, 0.695 respectively. F1 is assign as urban land use factor and biogeochemical dissolution of phosphate rocks due to overexploitation of groundwater the resulting concentration increased. Because water affected by urban land use include Na^+ , K^+ and Cl^- etc. may be differentiated from other land uses through the use of biogeochemical fingerprints [28].

Factor 2: explained the 27.236% of the total variance with moderate positive loading with Total Alkalinity& Temp with loading value 0.723 & 0.606 respectively, it's also explained the moderate negative loading with Conductivity, TDS, and Chloride with loading values -0.698, -0.690, -0.582. The main parameters are Total Alkalinity and conductivity. The electrical conductivity is known as salinization factor hence the Factor F2 is marked as Alkalinity-Salinity Factor. The reason for the increased alkalinity is the released of salts in the groundwater from various underground lithographic rocks.

Factor F3: explained 20.972 % of the variances with positive loading with PH, Chloride, and Phosphate with strong loading of PH 0.872 and moderate positively loading with chloride and phosphate with loading values 0.721, and 0.569 respectively. The factor F3 shows the PH is the main governing factor and the water quality is affected by physicochemical sources which mixed with groundwater and changes the PH either acidic or basic.Hence the Factor analysis can identified the most significant factor from the datasets and sampling stations can reduce and can left the non-significant parameters and sampling locations from huge sampling dataset.

Table 2: Factor loading matrix and total variance explained

	F1	F2	F3
Temp	**0.792**	0.606	
PH			**0.872**
TS	**0.970**		
TSS	**0.986**		

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

TDS	**0.701**	-0.690	
Conductivity	0.695	**-0.698**	
Total Alkalinity		**0.723**	-0.651
Chloride		-0.582	**0.721**
Phosphate	**0.744**		0.569
Eigenvalue	**5.162**	**2.724**	**2.097**
Total Variability (%)	**51.619**	**27.236**	**20.972**
Cumulative %	**51.619**	**78.856**	**99.827**

Values above 0.5 have been only shown

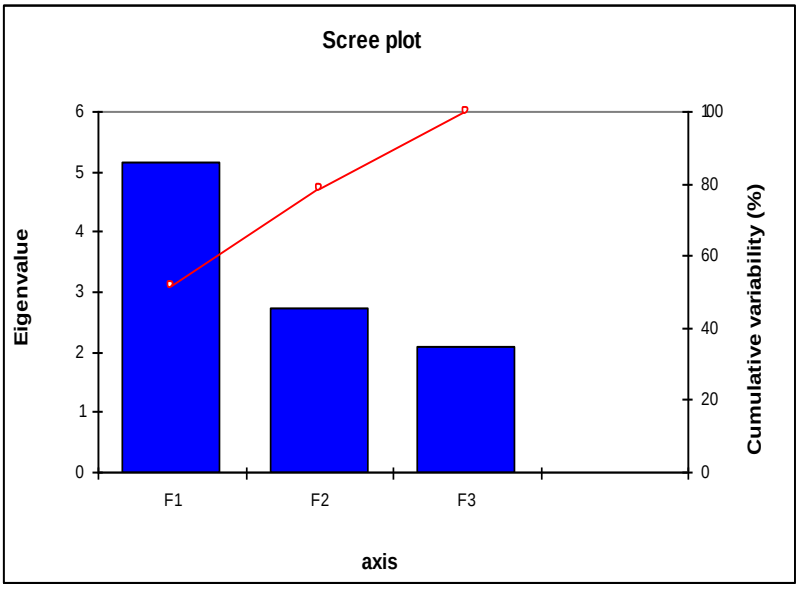


Figure 4.0 Scree plot showing Eigen values and cumulative variability%

4. CONCLUSION:

Today the increased demand of water for different purposes such as drinking, industrial use, and agricultural sectors pressurized the extensive extraction of groundwater resources resulting in the lowering of the ground water level. These results in increased pollutants load of the sewage and solid waste which seepage into the groundwater and alter the water quality for drinking purposes. The effect of different geological formations and the layers that water taped from, which also play a role in the water quality distribution .Further, increased extraction causes the dissolving of the rock materials into drinking water hence the dissolved solid levels increase in water and makes it unfit for drinking. The monitoring of the groundwater's general pollution and following the measured parametersthat are above the permitted concentration level can be used to search for the source of

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

pollution, for planning prevention measures, and to prevent pollution. In our study we have used the physico-chemical, Descriptive statistical analysis were carried out to evaluate the condition of the water quality of the study area and further the study proceed to know the hidden information in datasets and the relationships among the different variables of different sampling stations by the application of Cluster analysis and PCA/Factor analysis. Hierarchical cluster analysis grouped the 5 sampling locations into 3 clusters of similar water quality characteristics. Based on the information achieved by the cluster analysis & PCA/Factor analysis it is possible to reduce the complicated data and sampling stations into smaller, and significant sampling stations.The analysis of variant species of pollutants is more advantageous than a single one, where PCA/Factor Analysis was helpful to reduce and extract the most effective groups of environmental pollutants and also to assign water quality within areas under investigation. Three factors explained three factors explained 99.827 % of Variances. Factor F1 as urban land use factor and biogeochemical dissolution of phosphate rocks due to overexploitation of groundwater the resulting concentration increased. Factor 2 the Total Alkalinity and conductivity is the main parameter electrical conductivity is known as salinization factor hence the Factor F2 is marked as Alkalinity-Salinity Factor. The factor F3 shows the PH is the main governing factor and the water quality is affected by physicochemical sources which mixed with groundwater and changes the PH either acidic or basic. Hence the Factor analysis can identified the most significant factor from the datasets and sampling stations can reduce and can left the non-significant parameters and sampling locations from huge sampling dataset.

Hence there is great need towards the awareness about the conservation and management of the groundwater quality & quantity and policy towards the reducing the overuse of the groundwater extraction. There is need to develop method for sustainable uses of water resource in different sectors. Today's the need of integrated methods monitoring and early alarming systems ,modeling &assessment tools and techniques such as Water footprint Assessment, water resource conservation techniques, quality monitoring tools and techniques for effective water resource management.

5.FUTURE SCOPE OF WORK

The present work will be beneficial for the researchers and other workers who are engaged in water resources conservation, development and management. Environ-metric methods will be useful for the studying as early warning system and also in water & Environmental monitoring programs for monitoring, design a future, optimal studying strategies and also applicable to visualization of large amount of data and their quicker classification of potentially harmful pollutants of unknown water samples and also useful for the evaluation of spatial variations and source identification at the monitoring sites.Hence the study will motivate the researchers, eco-planing managers and other policy maker's person to involve the environ-metric methods in their planning proposals to evaluate and find out the effective, reliable potent, cost effective and less time consuming methods for their study.

ACKNOWLEDGEMENTS

G L Chaurasia acknowledges the University Grant Commission for providing financial assistance (JRF & SRF) and the Geological Survey of India, Northern Region Lucknow, for providing the Allahabad District Resource Map.

References:

[1] EU Directive (1998/83/EC) Official journal of the European communities. No L 31/1

[2] EU Directive (2000/60/EC) Official journal of the European communities. No. L 297/1

[3]EU Directive (2006/118/EU) Official journal of the European communities No. L 372/1

[4] WHO (**2004**) Guidelines for drinking-water quality, vol 1. World Health Organization, Geneva

[5] Mustapha A, Nabegu AB, Surface water pollution source identification using principal component analysis and factor analysis in Getsi River, Kano, Nigeria. Austr J Basic ApplSciVol (5), (**2011**), 1507-1512.

[6] Cukrov N, Tepic N, OmanovićD, Logen S, Bura-NakićS, Vojvodic E etal, Qualitative interpretation of physico-chemical and isotopic parameters in the Krka River (Croatia) assessed by multivariate statistical analysis. Int J Environ Anal ChemVol (92), No (10), (**2012**), 1187–1199.

[7] Tobiszewski M, Tsakovski S, Simeonov V, Namiesnik J, Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. Chemosphere Vol (80), No (7), (**2010**), 740-746.

[8] Ntengwe, F. W. (**2006**). Pollutant loads and water quality in streams of heavily populated and industrialised towns. Physics and Chemistry of the Earth, 31, 832–839.

Green Chemistry & Technology Letters
Vol 2, No 4, December 2016, pg. 211-218
eISSN: 2455-3611, DoI: 10.18510/gctl.2016.248

[9] Boyacioglu, H. (**2006**).Surface water quality assessment usingfactor analysis. Water S.A., 32(3), 389–393.

[10] Simeonov, V., Einax, J. W., Stanimirova, I., & Kraft, J.(**2002**). Environmetric modeling and interpretation of river water monitoring data. Analytical and Bioanalytical Chemistry, 374, 898–905.

[11] Kotti, M. E., Vlessidis, A. G., Thanasoulias, N. C., &Evmiridis, N. P. (2005). Assessment of river water quality in Northwestern Greece. Water Resources Management, 19, 77–94.

[12] Lopez, F. J. S., Garcia, M. D. G., Vidal, J. L. M., Aguilera, P. A., &Frenich, A. G. (**2004**). Assessment of metal contamination in Doñana National Park (Spain) using Crayfish (Procamburusclarkii). Environmental Monitoring and Assesment, 93, 17–29.

[13] Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., & Huang, C.H. (**2006**). Assessment of seasonal variations in surface water quality. Water Research, 40, 3800–3810.

[14] Shrestha, S., &Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin; Japan. Environmental Modelling & Software, 22, 464–475.

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

[16] http://www.cgwb.gov.in/District_Profile/UP/Allahabad.pdf

[17] APHA.Standard methods for the examination of water and waste water. American Public Health Association Washington D.C. (**2005**).

[18]. Bureau of Indian Standards, Indian Standards (IS: 10500) Drinking Water Specification: New Delhi (2004).

[19] World Health Organization; Guidelines for drinking Water Quality, Recommendation 2ndEdition; Geneva, WHO Vol.1, (2008).

[20] Tabachnick, B.G., & Fidell, L.S. (**1996**). Using multivariate statistics (3rd ed.). New York: Harper Collins CollegePublishers.

[21] Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin; Japan. Environmental Modelling & Software, 22, 464–475

[22] Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G.,Voutsa, D., &Anthemidis, A., et al. (**2003**).Assessment of the surface water quality in Northern Greece. Water Research, 37, 4119–4124.

[23] Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (**1995**). Multivariate data analysis with readings (4th ed.).London: Prentice-Hall.

[24] Sharma, S. (1996).Applied multivariate techniques. New York: Wiley.

[25] Vega, M., Pardo, R., Barrato, E., &Deban, L. (**1998**).Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 32, 3581–3592.

[26] Wunderlin, D. A., Diaz, M. P., Ame,M. V., Pesce, S. F., Hued, A. C., &Bistoni, M. A. (**2001**). Pattern recognitiontechniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquira river basin (Cordoba—Argentina). Water Research, 35, 1894–2881.

[27] Liu, C. W., Lin, K. H., &Kuo, Y. M. (**2003**).Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science of the Total Environment, 313, 77–89.

[28] Lindeman, M. A. (2004). Exploring the effects of urban and agricultural land use on surface water quality. 2004 Denver annual meeting. Paper no. 72–9. Geological Society of America abstracts with programs. 36, 184.