ISSN 2229-5518

Assessment of Sediment Quality in Hussainsagar Lake and Its Inlet Channels Using Multivariate Statistical Techniques

[1*]Dr. A.Sridhar Kumar,[2]Y.Seeta,[3]M.Satyanarayana,[4]Prof.P.Manikya Reddy

Abstract— Concerns about the sediment quality in lake Hussainsagar located in Hyderabad, India have been rapidly increasing recently due to urbanization and industrialization pollution. This study analyzes twelve chemical and heavy metal parameters at the four sampling stations during the year 2013 by using multivariate statistical techniques like Hierarchical clustering, principle component and factor analysis. FA identified five factors responsible for data structure explaining 71.05% total variance and allowed to group selected parameters according to common features. Cd, Cr, Zn, Ag, Ni, & As were associated with similar contributions from anthropogenic sources, whereas SO_4^{-2} and Cl- are derived from natural sources.

Keywords: Cluster analysis, Factor analysis, Heavy metals, Hussainsagar lake, Multivariate statistical techniques, Principle component, Sediment quality.

1 Introduction

Primary industrial processes that release a variety of metals into waterways include mining, smelting and refining [1]. In fact, almost all industrial processes that produce wastewater discharges are potential sources of heavy metals to the aquatic environment [2], [3]. Domestic wastewater, sewage sludge, and urban runoff are also major heavy metal sources to rivers, estuaries and coastal waters [4]. Accumulation of heavy metals in the sediment bed of Lake and re-suspension of heavy metals in to lake water [5] identified, sediment bed as sink for heavy metals in lakes. Therefore, the effective long-term management of lakes requires a fundamental understanding of hydro morphological, chemical and biological characteristics.

The application of different multivariate statistical techniques, such as principal component analysis (PCA), factor analysis (FA), cluster analysis (CA), and discriminate analysis (DA), assists in the interpretation of complex data matrices for a better understanding of sediment quality and ecological characteristics of a study area. These techniques provide the identification of possible sources that affect water environmental systems and offer a valuable tool for reliable management of water resources as well as rapid solution for pollution issues [6], [7], [8]. Multivariate statistical techniques have been widely adopted to analyze and evaluate surface and freshwater water quality, and are useful to verify temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality [9], [10].

The objective of the present study was to analyze 14 chemical and heavy metal parameters at the channels confluence points (where channels joins in to the lake) during the year 2013 in Hussainsagar basin in India. The data matrix obtained from field measurement was subjected to the CA, PCA, and FA techniques, as well as to evaluate information about the similarities between sampling stations and to ascertain the important contributions of heavy metal sources among sediment quality parameters in the Hussainsagar basin.

2 Materials and Methods

The study area and sample locations (Fig 1) in Hussainsagar basin is located in Hyderabad in India and is situated at 17° 22′ of northern Latitude and 78° 29′ of the eastern longitude. The changes in land use of the lake catchments have a direct impact on the water body. Silting of lake has commenced during the last twenty years due to high industrial and domestic waste water in the catchments increased the rate of sedimentation with high concentration of heavy metals, this resulted in reduction of water holding capacity of the lake, the water quality as well as sediment quality has deteriorated.

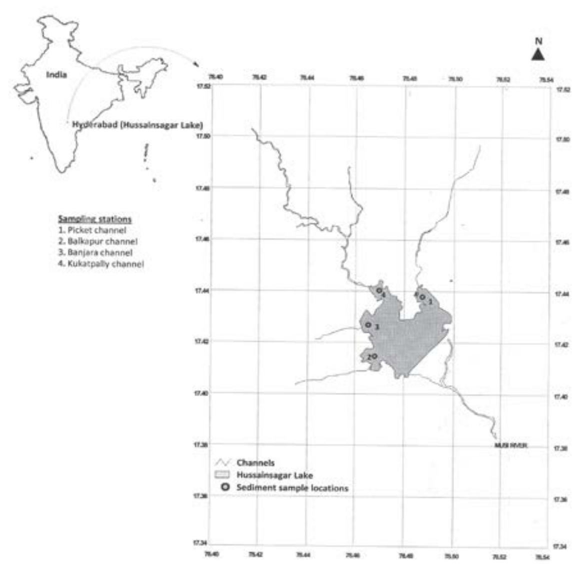


Fig. 1. Map of study area and sediment quality monitoring stations (listed 1-4) in the Hussainsagar lake basin.

For this study, all four major inlet channels at lake mouth

*Author: Dr. A. Sridhar Kumar, Department of Environpnmental Science, Osmania University Hyderabad, Telangana State, India.
E-mail: meetsreedhar1@gmail.com
Co-Author: Y.Seeta, Research Scholer, Department of Environpnmental Science,
Co-Author: M. Satyanarayana, MD, Clinicapro Data Research, Hyderabad
Co-Author: Prof. P. Manikya Reddy, Dept of Botony, Osmanaia University

are consider, where the flow is coming last 20 – 30 years. The details of the sample collection points, name of the inlet channel and locations are given in fig-1. The method of sample collection is through drilling and underwater drilling by using bore logs and tubes for getting undisturbed (core) samples.

The sediment samples were collected from all four major channels at confluence of the lake during the year 2013. Samples were collected from different layers, 1st layer 0-0.45 m. and 2nd layer 1.0-1.45 m. depth below the bed surface at four major inlet channels joining to the lake up to 750 meter radius towards the lake. The sediment samples were air dried, powdered and sieved and duplicate samples were used for chemical analysis and analyzed as per IS: 2720 (1980). The muck is an organic material with considerable amount of soil with high content of water and black in appearance. The chemical characteristics like nitrates, phosphates, sulphates, chlorides and heavy metal were determined.

3 Data Treatment and Multivariate Statistical Methods

The data sets were subjected to four multivariate techniques: Cluster analysis (CA), principle component analysis (PCA) factor analysis (FA) and discriminate analysis [11], [12], [13]. Descriptive statistics was applied to raw data, whereas PCA, FA and CA were applied to experimental data, standardized through z-scale transformation in order to avoid misclassification arising from the different orders of magnitude of both numerical values and variance of the parameters analyzed[14], [15].

3.1. Cluster Analysis

CA is an unsupervised pattern recognition method that divides a large amount of cases into smaller groups or clusters based on the characteristics of the process. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between the analytical values form the samples [16]. In the present study, we used Euclidian distance average linkage method (within the group) of cluster analysis. The number of clusters was also decided by practicality of the results as there is ample information available on the study sites. The spatial variability of water quality in the lake was determined from hierarchical CA using the linkage distance [17], [18].

3.2. Factor Analysis/Principal Component Analysis

Factor analysis technique extracts the eigen values and eigen vectors from co-variance matrix of original variables. The principle components (PC) are the uncorrelated (orthogonal) variables obtain by multiplying original correlated variables with eigen vector, which is a list of coefficients (loading or weightings). Thus principal components are weighted linear combinations of original variables. PC provides information on the most meaningful parameters, which describe whole data set affording data reduction with minimum loss of original information [19], [20]. It is a powerful technique for pattern recognition that attempts to explain the variance of large set of inter-correlated variables and transforming in to a smaller set of independent (uncorrelated) variables (principle component). Factor analysis further reduce the contribution of less significant variables obtained from PCA and the new group of variables known as varifactors, are extracted through rotating the axis defined by PCA. A varifactors can include unobservable, hypothetical, latent variables, while a PC is a linear combination of observable water quality variables [21], [22]. PCA of the normalized variables was performed to extract significant PC's and to further reduce the contribution of variables with minor significance. These PC's were subjected to varimax rotation (raw) generating varifactors.

4. Results and Discussion

The measured results of 14 chemical and heavy metal sediment quality parameters at four sampling stations during the year 2013 at Hussainsagar lake are presented in table 1. In picket channel (C1) the parameters like Nitrates, Phosphates, Mercury, Silver and Zinc concentration were high during the study period, and showing high mean values compare to other locations. In Kukatpally channel (C4) the concentrations of Arsenic, Cadmium, Copper, Lead and Nickel are showing high mean values compare to other locations. Whereas in Balkapur and Banjara channels (C2&C3), the concentrations of Chloride, Total Chromium, Iron, and the Sulphates showing high average values compared with other locations.

From the descriptive analyses (table 1), it is also noticed that, the sediment samples collected from the surface of the lake bed are showing higher values than the samples collected from 1 meter below the bed surface. This shows that the contamination is not percolating down and is getting accumulated on the surface. From the results, Sediment analysis of Balkapur, Banjara, and Picket channel confluence points, the average concentrations of heavy metals were well within the permissible limits, where as in Kukatpally channel showing higher average concentration for cadmium.

Table 1. Mean and Standard Deviation of sediment quality parameters at different locations In Hussainsagar during the year 2013.

Parameter	Layers	Station C1	Station C2	Station C3	Station C4
Cl-	1st Layer	0.17 ± 0.11	0.19 ± 0.10	0.24 ± 0.18	0.15 ± 0.04
	IInd Layer	0.16 ± 0.06	0.41 ±0.76	0.20±0.13	0.13±0.03
NO_3	1st Layer	32.00±5.10	23.90±7.78	30.92±7.57	28.53±4.74
	IInd Layer	28.40±4.43	24.90±4.43	28.00±6.30	24.24±4.52
PO_4	1st Layer	0.70±0.24	0.44±0.12	0.56±0.33	0.61±0.30
	IInd Layer	0.75±0.42	0.45±0.24	0.58±0.34	0.48±0.29
SO_4^{-2}	1st Layer	0.12±0.04	0.05±0.02	0.14±0.05	0.10±0.04
	IInd Layer	0.11±0.04	0.05±0.02	0.12±0.03	0.09±0.02
As	1st Layer	1.70±0.84	2.89±0.86	2.09±1.00	4.60±3.84
	IInd Layer	1.02±0.61	1.65±1.07	1.86±0.70	2.93±1.72
Cd	1st Layer	0.98±0.38	0.57±0.32	1.24±0.79	62.12±11.40
	IInd Layer	0.74±0.39	0.78±0.43	0.98±0.67	57.88±10.75
Cr	1st Layer	3.60±0.45	7.13±1.14	1.50±0.41	4.72±1.64
	IInd Layer	3.41±1.03	5.19±1.40	1.70±1.36	3.92±1.10
Cu	1st Layer	4.58±0.86	2.51±1.46	2.54±0.63	4.93±1.67
	IInd Layer	4.11±0.73	3.35±1.92	2.55±0.70	4.21±1.36
Fe	1st Layer	26.70±7.51	41.60±7.08	21.48±5.25	41.54±10.95
	IInd Layer	28.68±7.43	34.64±12.37	24.41±10.67	38.84±9.48
Pb	1st Layer	12.12±1.91	7.90±1.92	6.20±1.32	12.78±4.64
	IInd Layer	10.04±2.98	8.59±2.09	5.88±1.08	10.06±3.10
Hg	1st Layer	1.96±1.88	0.98±0.66	1.20±0.70	1.66±1.35
	IInd Layer	1.45±1.54	0.49±0.30	0.78±0.34	0.86±0.55
Ni	1st Layer	1.83±0.54	1.90±0.53	1.49±0.25	2.27±0.75
	IInd Layer	1.90±0.52	1.88±0.45	1.49±0.25	2.36±1.09
Ag	1st Layer	0.04±0.02	0.02±0.01	0.01±0.00	0.03±0.03
	IInd Layer	0.03±0.02	0.02±0.01	0.01±0.00	0.02±0.02
Zn	1st Layer	20.85±9.49	17.37±4.82	14.12±4.64	17.75±8.39
	IInd Layer	18.89±8.73	15.65±2.72	13.60±3.74	13.44±6.14

ISSN 2229-5518

Note: Values represent mean ± standard deviation
Units: Except chlorides and sulphates (in %) all are in mg/kg
Stations: 1. Picket (C1) 2. Balkapur (C2) 3. Banjara (C3) 4. Kukatpally (C4)

Overall, the discharge of municipal sewage, industrial effluents, the storm water discharge containing diluted sewage and other impurities on the land surface from over 240 square kilometers area of watershed have resulted in dumping of high amounts of organic matter, nitrogen and phosphorous in to the water and indicating the decreased the sediment quality. This situation suggests a strong variability due to presence of anthropogenic sources from the catchment affecting the sediment quality.

4.1. Spatial Similarity with Cluster Analysis

Cluster analysis was applied to find out the similarity groups between the parameters. It produced a dendogram (Fig 2), grouping all fourteen parameters in to three meaningful clusters.

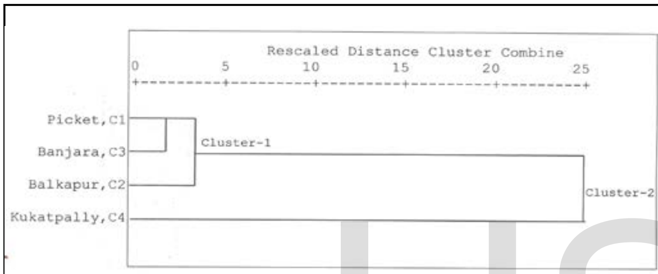


Fig 2. Dendrogram showing clustering of sampling sites according to sediment quality characteristics of the Hussainsagar

Cluster 1 formed by Picket (C1), Balkapur (C2) and Banjara (C3) stations because of similar or low distances based on the water quality parameters average concentrations and corresponds to relatively moderately polluted sites within the group and these stations are had a secondary level sewerage treatment plant at upstream side of the lake and carrying less dry whether flows in to the lake. Cluster 2 formed by Kukatpally station (C4) because of similar or dissimilar and showing moderate distances based on the sediment quality parameters average concentrations and correspond to highly polluted sites within the group. These stations receive pollution either sides of the channels from domestic and industrial areas. The results indicate that the clusters are showing similarities and dissimilarities between the stations. There are other reports [23], [24] with similar approach has successfully been applied to water quality programs.

4.2. Principal Component Analysis and Pollution Identification

Pattern recognition of correlations among 14 parameters was best summarized by PCA/FA. In this study, the covariance matrix coincided with the correlation matrix which was presented in table 2, because FA/PCA was applied to normalized data. Overall, the correlations between variables were relatively weak. There are some positive correlations between some variables such as sulphate, lead, copper, silver and arsenic. The negative correlations were revealed between some variables such as chromium and ferrous. Correlation coefficients of two elements were very useful, because they numerically represented the similarity between two elements of the two sediment quality variables. This also indicated that PCA could successfully reduce the dimensionality of the original data set. Therefore factor analysis of the present data set further reduced the contribution of less significant variables obtained from PCA.

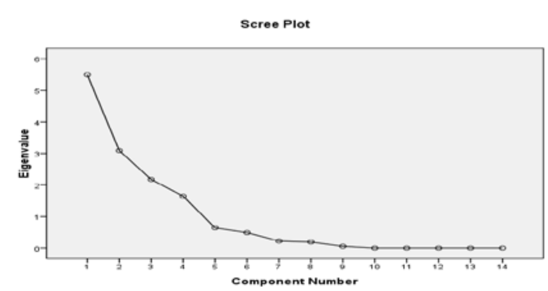
TABLE 2. CORRELATION MATRIX OF SEDIMENT QUALITY PARAMETERS OF HUSSAINSAGAR BASIN.

	NO_3	PO_4	SO_4^{-2}	Cl-	Pb	Hg	Cr	As	Cu	Ni	Fe	Ag	Zn	Cd
NO_3	1													
PO_4	0.23	1												
SO_4^{-2}	0.55**	0.31*	1											
Cl-	0.09	0.22	0.05	1										
Pb	0.08	0.14	0.02	-0.17	1									
Hg	0.17	0.1	0.27*	-0.04	0.34**	1								
Cr	-0.32*	-0.14	-0.56**	0.08	0.37**	-0.01	1							
As	-0.14	0.18	0.02	-0.06	0.21	0.06	0.30*	1						
Cu	0.07	0.13	0.09	-0.17	0.54**	0.26*	0.18	0.14	1					
Ni	0.09	0.05	-0.06	0.08	0.21	0.18	0.27*	0.05	0.31*	1				
Fe	-0.23	-0.18	-0.31*	-0.07	0.41**	-0.01	0.62**	0.22	0.46**	0.27*	1			
Ag	0.09	0.17	0.17	-0.09	0.48**	0.28*	0.37**	0.43**	0.25*	0.01	0.14	1		
Zn	0.19	0.18	0.31*	0.06	0.24*	0.34**	0.26*	0.07	0.26*	0.17	0.07	0.59**	1	
Cd	-0.14	-0.13	-0.08	-0.22	0.42**	0.08	0.16	0.42**	0.44**	0.45**	0.47**	0.1	-0.09	1

**Values are statistically significant at $p<0.01$
*Values are statistically significant at $p<0.05$

This also indicated that, PCA could successfully reduce the dimensionality of the original data set. Therefore factor analysis of the present data set further reductions the contribution of highly influenced variables obtained from PCA. Elements belonging to a given factor were defined by factor matrix after varimax rotation, with those having strong correlations grouped in to factors. The identification of factors is based on dominant influence.

The Scree plot (Fig 3) was applied to identify the number of PCs to be retained to understand the underlying data structure. Based on the Scree plot and the eigenvalues >1 criterion, seven factors were chosen as principal factors, explaining 71.05% of the total variance in the data set. The corresponding varifactors (VFs), variables loadings, eigenvalues, and ex-



plained variance are presented in table 3.

ISSN 2229-5518

Fig 3. Scree plot of the characteristic roots (eigen values) of principle component analysis.

As per the classification Liu et al., [25] classified the factor loadings as "strong", "moderate", and "weak", corresponding to absolute loading values of >0.75, 0.75–0.50, and 0.50–0.30, respectively. The first factor (VF1), explaining 17.88% of total variance, had strong loadings on Cadmium and Nickel and represented anthropogenic sources. VF2, which explained 17.34% of total variance, had a moderate loading on Sulphates and Chromium and represents the natural and anthropogenic sources.

TABLE 3. LOADING OF 14 PARAMETERS ON SIGNIFICANT VFS FOR SEDIMENT QUALITY DATA SET.

Parameters	Four significant PCs				
	VF1	VF2	VF3	VF4	VF5
NO_3	.084	.688	.210	-.161	.169
PO_4	.002	.434	.126	.457	.475
SO_4	-.023	.841	.226	.101	.004
Cl	-.099	-.011	-.027	-.040	.865
Pb	.558	-.050	.491	.195	-.178
Hg	.256	.300	.526	-.117	-.104
Cr	.246	-.794	.398	.116	.195
As	.164	-.130	.091	.881	-.031
Cu	.715	.077	.296	.067	-.141
Ni	.735	-.024	.015	-.148	.352
Fe	.598	-.543	.139	.063	-.038
Ag	.019	-.055	.790	.444	-.081
Zn	.032	.087	.841	-.031	.173
Cd	.782	-.090	-.174	.333	-.240
Eigen value	2.504	2.428	2.265	1.427	1.323
% of total variance	17.883	17.346	16.180	10.190	9.452
Cumulative % of variance	17.883	35.229	51.409	61.598	71.050

VF3, explaining 16.18% of total variance, had a positive loading on Zinc and Silver and represented anthropogenic sources. VF4, explaining 10.19% of total variance, had a strong loading on Arsenic and represented anthropogenic sources. VF5, explaining 9.45% of total variance, had a strong loading on Chloride and represented natural sources.

5. CONCLUSION

Euclidian distance average method analysis grouped 4 sampling stations into two clusters of mostly similar sediment quality characteristics and confirmed the existence of two types of sediment quality (less and highly polluted stations). The PCA and FA assisted to extract and recognize the factors or origins responsible for sediment quality variations. PCA/FA identified five latent factors that explained 71.05% of total variance, broadly natural and anthropogenic pollution factors controlling their variability in sediments of Hussainsagar basin. Thus, this study illustrates the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, and in sediment quality assessment, identification of pollution sources/factors and understanding spatial variations.

ACKNOWLEDGMENTS

The authors are thankful to Prof P. Manikya reddy, Dept of Botany, Osmanaia University-Hyderabad, for his continuous supervision throughout this study.

REFERENCES

[1] Bryan, G.W., (1976). *Heavy Metal Contamination in the Sea*. In: Marine Pollution (R. Johnston (ed.)). Academic Press: London: New York: San Francisco. Pp. 185-302.

[2] Klein, L.A., Lang, M., Nash, N., and Kirschner, S.L., (1974). Sources of Metals in New York City Waste - Water. *Journal of the Water Pollution Control Federation, 46: 265.*

[3] Barnhart, B. J., (1978). *The Disposal of Hazardous Wastes*. Environmental Science & Technology, 12: 1132-1136.

[4] Connell, D.W., and G. J., Miller. (1984). *Chemistry and Ecotoxicology of Pollution*. John Wiley & Sons, New York Chichester: Brisbane: Toronto: Singapore. 444 pp.

[5] Forstner, V., and Wittman, G.T.M., (1981). *Metal Pollution in the Aquatic Environment*. Springer Verlag. 486 P.

[6] Singh, K.P., Malik, A., Mohan, D., Sinha, S., (2004). *Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River* (India): A case study. Water Res. 38:3980–3992.

[7] Li, R., Dong, M., Zhao, Y., Zhang, L., Cui, Q., He, W., (2007). *Assessment of water quality and identification of pollution sources of plateau lakes in Yunnan* (China) J. Environ. Qual.; 36:291–297.

[8] Kazi, TG., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, HI., Sarfraz, R.A., Baig, J.A, Shah, A.Q., (2009). *Assessment of water quality of polluted lake using multivariate statistical techniques*: A case study. Ecotox. Environ. Safe. 72:301–309.

[9] Singh, K.P., Malik, A., Sinha, S., (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.

[10] Kim, J.H., Choi, CM., Kim, S.B., Kwun, S.K., (2009). *Water quality monitoring and multivariate statistical analysis for rural streams in South Korea*. Paddy Water Environ. 7:197–208.

[11] Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Hued, A.C., Bistini, M.A., (2001). *Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality*. A case study: Suquia river basin (Cordoba-Argentiana), Water Research 35: 2881-2894.

[12] Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Vousta, D., Anthemidis, A., Sofoniou, M., Kouimtzis, T., (2003). *Assessment of the surface water quality in Northern Greece*. Water Research 37: 4119-4124.

[13] Singh, K. P., Malik, A., Mohan, D., Sinha, S., (2004). *Multivariate statistical techniques for the evalution of spatial and temporal variations in water quality of Gomti River* (India): a case study. Water Research 38: 3980-3992.

[14] Liu, CW., Lin, KH., Kuo, Y.M., (2003). *Application of factor analysis in the assessment of groundwater quality in*

a Blackfoot disease area in Taiwan. Sci. Total Environ. 313:77–89.

[15] Simeonov V, Stratis JA, Samara C, Zachariadis G, Vousta D, Anthemidis A, Sofoniou M, Kouimtzis Th., (2003). *Assessment of the surface water quality in Northern Greece*. Water Res.2003; 37:4119–4124.

[16] Otto, M., (1998). *Multivariate methods*. In: Kellner, R., Mermet, J.M., Otto, M., Widmer, H.M. (Eds), Analytical Chemistry. Wiley-VCH, Weinheim.

[17] Simeonov V, Stratis JA, Samara C, Zachariadis G, Vousta D, Anthemidis A, Sofoniou M, Kouimtzis Th., (2003). *Assessment of the surface water quality in Northern Greece*. Water Res.2003; 37:4119–4124.

[18] Zaharescu DG, Hooda PS, Soler AP, Fernandez J, Burghelea CI., (2009). *Trace metals and their source in the catchment of the high altitude Lake Respomuso, Central Pyrenees*. Sci. Total Environ. 407:3546–3553.

[19] Vega, M., Pardo, R., Vega, M., Barrado, 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.

[20 Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M., Fernandez, I., (2000). *Temporal evaluation of ground water composition in an alluvial aquifer (Pisuerga river, Spain) by principle component analysis*, Water Research 34: 807-816.

[21] Panda, U.C., Sundaray, S.K., Rath, P., Nayak, B.B., Bhatta, D., (2006). *Application of factor and cluster analsis for characterization of river and esturine water systems*-a case study: Mahanadhi River (India), Journal of Hydrology 331 (3-4): 434-445.

[22] Davis, J.C., (1986). *Statistics and Data Analysis in Geology*, 2nd edn., Wiley, New York.

[23] Kim, J.H., Choi, CM., Kim, S.B., Kwun, S.K., (2009). *Water quality monitoring and multivariate statistical analysis for rural streams in South Korea*. Paddy Water Environ. 7:197–208.

[24] Shrestha, S., Kazama, F., (2007). Assesment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan, Environmental Modelling and Software 22 (4): 464-475.

[25] Liu, CW., Lin, KH., Kuo, Y.M., (2003). *Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan*. Sci. Total Environ. 313:77–89.