International Journal of Hybrid Information Technology
Vol.8, No.4 (2015), pp.89-96
http://dx.doi.org/10.14257/ijhit.2015.8.4.11

Application of Principal Components Analysis for Interpretation and Grouping of Water Quality Parameters

Sarita Gajbhiye[1*], S.K. Sharma[2] and M.K. Awasthi[3]

[1]*Department of Water Resources Development and Management*
Indian Institute of Technology, Roorkee (Uttrakhand), India
**gajbhiyesarita@gmail.com*
[2,3]*Department of Soil and Water Engineering*
J.N.K.V.V. Jabalpur (M.P.)

Abstract

*Water samples were collected from two sites (Urdana Nala and Moti Nala) in Jabalpur city, were analyzed for their physio- chemical characteristics. This analysis result was compared with the WHO & United State Sanility labolatory standards of irrigation/drinking water quality parameters with the following water quality parameters namely pH***,** *Electrical conductivity, Cu, Cr, SO_4, Fe, NO_3, Chloride, TH, TA, Na. The statistical analysis like Factor Analysis (FA) and Principle Component Analysis (PCA) of obtained data were carried out. The PCA extract and to define the parameters responsible for the main variability in water quality variance for Jabalpur city. Moti Nala and Urdana Nala water constrains high Na ions. But Cu, Cr, SO_4, Fe, NO_3, Cl, TH are within limits. TA is higher in most of the samples. The PCA produced more important parameter NO_3, Cr, Fe, TH, Na, EC, SO_4, Na in Urdana Nala site. $NO_{3,}$ Cu, Na, Fe, TA, pH, $SO_{4,}$ TH, Cr in Moti Nala site. Finally the results of PCA reflect a good look on the water quality monitoring and interpretation of the surface water.*

Keywords: *Principle Component Analysis, Water quality, Factor analysis*

1. Introduction

Water is one of the essential inputs for sustenance of all living being. The need of water for various purposes is increasing at a fast rate due to the continual increase in population, rapid urbanization with change in life style and growing industrialization (Gajbhiye *et al.*, 2014b: Sharma *et al.*, 2014b). The World's total water resources are estimated at 1.37 X 10^8 million ha-m.Water quality plays an important role in groundwater protection and quality conservation. Hence it is very important to assess the water quality not only for its present use but also from the viewpoint of a potential source of water for future consumption. The water resources are being utilized for drinking, irrigation and industrial purpose. There is growing concern on deterioration of water quality due to geogenic and anthropogenic activities.

The quality of water is identified in terms of its physical, chemical and biological parameters. Polluted surface waters cannot achieve a balanced ecosystem. A balanced ecosystem is one in which living things and the environment interacts beneficially with one another. Water quality obviously plays a critical role in this relationship (Ntengwe 2006), as it is key to the maintenance of a well-balanced environment.

Factor analysis technique is very useful in the analysis of data corresponding to large number of variables, analysis via this technique produce easily interpretable results, and this method have been used successfully in hydrochemistry for many years, surface water, ground water quality assessment and environmental research employing multi component techniques are will described in the literature (Praus, 2005).

ISSN: 1738-9968 IJHIT

In recent years many studies have been done using principal components analysis in the interpretation of water quality parameters, (Lohani, 1984) utilized principal components technique to provide a quick analytical method for the water quality of Chao Phraya river in Thailand. Also (Shihab, 1995) used this technique in order to describe the variation in water quality in Saddam dam reservoir and the regulating lake and to reduce the number of water quality parameters needed for monitoring the lake water, he noted that (90%) of the seasonal variation in the water quality of the reservoir are exhibited in only three parameters (Algae, EC and TS) and he referred that the monitoring may be pointed to these parameters responsible for the main variability in water quality, While the frequency of sampling for insignificant parameters may be reduce; or special studies may be conducted on them when needed. (Mazlum *et al.*, 1996) determine factors that caused variations in water quality at the monitoring station on the Porsuk Tributary in the Sakarya river basin by using principal component analysis (PCA); researchers referred that PCA is more reliable than factor analysis and it a pure mathematical technique without any assumption.

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 (Ouyang et al. 2006; Shrestha and Kazama 2007). 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. (Mazlum *et al.*, 1996) determine factors that caused variations in water quality at the monitoring station on the Porsuk Tributary in the Sakarya river basin by using principal component analysis (PCA); researchers referred that PCA is more reliable than factor analysis and it a pure mathematical technique without any assumption. principal components analysis has been successfully applied to sort out hydrogeological and hydrogeochemical processes from commonly collected ground water quality data (Jayakumar and Siraz, 1997), (Salman and Abu Ruka'h, 1998), (Praus, 2005), (Olobaniyi and Owoyeni, 2006). (Iyer *et al.*, 2003) constructed a statistical, model which based on the PCA for coastal water quality data from the Cochin coast in south west India, which explain the relationships between the various physicochemical variables that have been monitored and environmental conditions effect on the coastal water quality. In this study principal components analysis (PCA) are being used in order to interpret and grouping the water quality parameter.

2. Material and Methods

2.1 Study Area and Data Source

The locations Jabalpur was chosen for the present analysis (Figure-1). The experiment is conducted on two point source are- Municipal waste (Urdana Nala) and Municipal waste (Moti Nala). Jabalpur with sub-tropical, sub-humid climate is situated on the Kymore plateau and Satpura hills, agro-climatic region at $23^0 9”$ latitude and $79^0 58”$ E longitude at an altitude of 411.78 meter above the mean seas level at has hot dry summer and cold winter with occasions showers. In some years, during the winter, freezing point is touched and during the summer the temperature rises as high as 46^0C making the climate of this place quite extreme (Gajbhiye *et al.* 2010, 2014a).

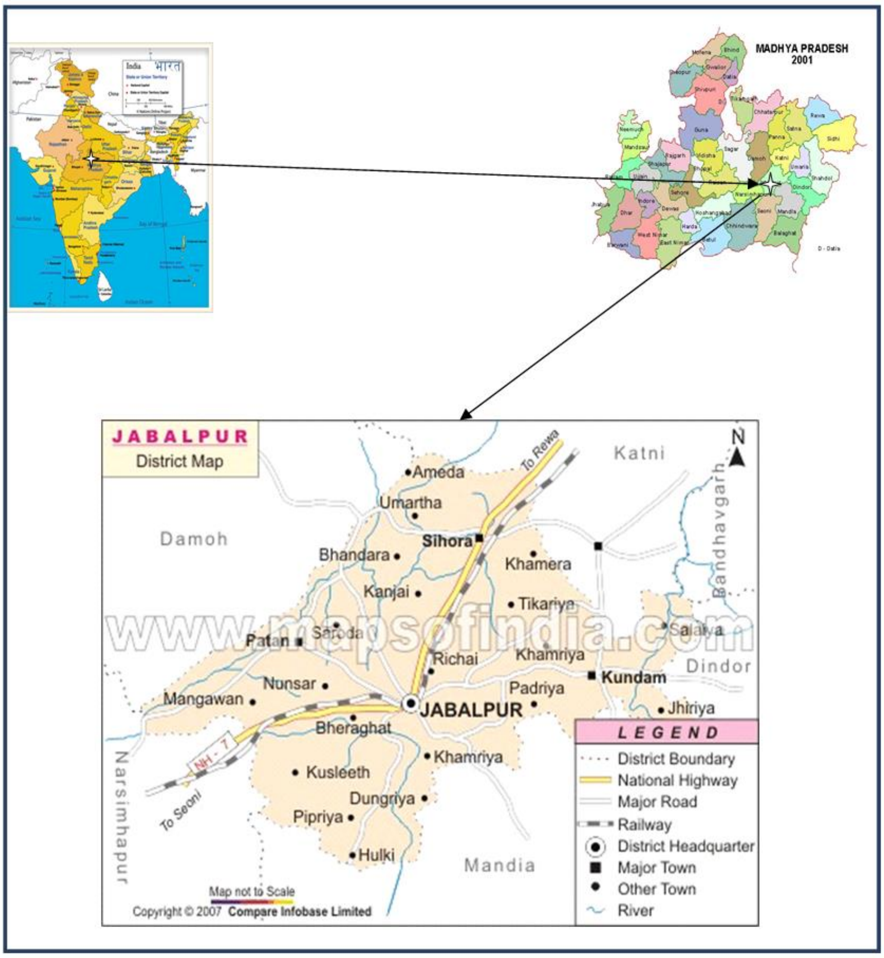

Figure 1. Location Map of the Study Area

2.2 Water Sampling

Water samples from nalas *i.e.* sources of waste water were collected and either collected from open well or bore well or from tube well during the pre-monsoon period of year 2008. The water sample collected in plastic container. All samples were collected, preserved and stored for analysis (APHA 1992) as outlined in Standard Methods for the Examination of Water and Wastewater.

2.3 Determination of Properties of Water

For water sample 11 parameters (variable) had been done these parameters including: pH, Electrical conductivity (EC),Copper (Cu), Chromium (Cr), Sulphate (So_4), Iron (Fe), Nitrate (No_3), Chloride (Cl^-), Total Hardness (TH), Total Alkanity (TA) and Sodium(Na). All examinations were done according to standard specifications presented by the Water quality Kit (La Motte, smart water analysis, code- 1951- 01 EX2).

2.4 Principle Component Analysis/Factor Analysis

Statistical computations were made using SPSS 14.0 software. Principal component analysis (PCA) is a powerful tool that attempts to explain the variance of a large dataset of intercorrelated variables with a smaller set of independent variables (Simeonov *et al.* 2003). PCA technique extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The Principal Components (PC) are 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 information on the most meaningful parameters, which describe the whole data set while affording data reduction with a minimum loss of original information (Hair *et al.* 1995; Sharma 1996; Vega *et al.* 1998). The first principal component is that linear combination of the original variables which contributes a maximum to their total variance; the second principal component, uncorrelated with the first, contributes a maximum to the residual variance, and so on until the total variance is analyzed. Since the method is so dependent on the total variance of the original variables, it is most suitable when all the variables are measured in the same units (Sharma *et al.*, 2013, 2014a).

3. Results and Discussions

3.1 Water Quality

Analysis of the water sample shown in table 1.This table indicate that the concentration of pH water samples of the Urdana Nala and Moti Nala is ranging from 7.42 to 7.61. These values are within the prescribed limit of WHO.The highest average value of EC (1.66 ds/m) for Urdana Nala which is under High category as per USSSL. Thus all the samples fall under high category.The copper in Urdana Nala, Moti Nala are 0.13 mg/l, 0.90 mg/l respectively. The copper content of all water samples are within the permissible limit as per WHO. The highest value of Cr (0.23 mg/l) for Moti Nala. It is followed by Urdana Nala (0.07 mg/l). Sulphate content in drinking water exceeding 400 mg/l impart bitter, medicinal taste and may cause gasto-intestine irritation and catharsis. The SO4 value of water samples of the Urdana Nala, Moti Nala are 12 mg/l, 60 mg/l respectively. These all values of SO4 are within the prescribed limit of WHO. The highest value of Fe (0.92 mg/l) was recorded for the Urdana Nala. The iron concentration of all water samples is within the permissible limit as per WHO. Nitrate indicates the pollution in ground water due to sewage percolation beneath the surface. Nitrate in natural water come from organic sources or from industrial and agricultural chemicals. The highest value of NO3 (4.8 mg/l) for Urdana Nala. The Nitrate concentration of all water samples are within permissible limits as per WHO. The Chloride content of water samples of Urdana Nala, Moti Nala, are 36 mg/l, 50 mg/l, respectively. The chloride content all water samples are within the permissible limit as per WHO. The minimum value of TH is 235 mg/l for Moti Nala. Total hardness of all water samples are within the permissible limit as per WHO. The major portion of alkalinity in natural water is caused by hydroxide, carbonate and bicarbonate. Alkalinity in itself is not harmful to human being. The total alkalinity values of water samples are 790 mg/l, 900 mg/l, for Urdana Nala, Moti Nala, respectively. Water samples from Urdana Nala, Moti Nala have shown higher TA. According to National Academy of Science, the higher concentration of sodium can be related to cardiovascular disease and women toxemia associated with pregnancy. The highest value of Na (520 mg/l) for the Moti Nala. It is followed by Urdana Nala (380 mg/l). Water samples showed higher value of sodium for Moti Nala and Urdana Nala as per WHO limitation.

Table 1. Physio-chemical Property of Water Samples

Parameter	Location	
	Urdana Nala	Moti Nala
pH	7.58	7.61
EC (ds/m)	1.66	1.16

Cu (mg/l)	0.13	0.90
Cr (mg/l)	0.07	0.23
SO_4 (mg/l)	12	60
Fe (mg/l)	0.39	0.48
NO_3 (mg/l)	4.8	0.71
Chloride (mg/l)	36	50
TH (mg/l)	240	235
TA (mg/l)	790	900
Na (mg/l)	380	520

4. Factor Analysis

4.1 Compositional Relation

Urdana Nala - Compositional relations in water show (Table 2) that strong correlations (> 0.85) exist between pH and NO3, Fe and Na. Also, good correlations (> 0.70) exist between, pH and Cr, EC and Sulphate. Some more moderate correlated parameters (> 0.60) are EC with TA, Cr with Fe, Cr with NO3 and SO4 with Na.

Moti Nala - Compositional relations in water show (Table 3) that strong correlations exist between Cu with Cr, Fe, NO3, TA and Cr with TA, NO3, Fe and Fe with TA, NO3 and NO3 with TA. Also, good correlations exist between pH and Sulphate, SO4 and Fe, TH and Na. Some more moderately correlated parameters are SO4 with TA and Cu with SO4.

It is very difficult at this stage to group the parameters into components and attach any physical significance. Hence, in the next step, the principal component analysis has been applied. The correlation matrix is subjected to the principal component analysis.

Table 2. Correlation Matrix of Urdana Nala

	pH	**EC**	**Cu**	**Cr**	**SO_4**	**Fe**	**NO_3**	**Chloride**	**TH**	**TA**	**Na**
pH	1.0										
EC	-0.06	1.0									
Cu	-0.57	-0.05	1.0								
Cr	0.71	-0.42	-0.25	1.0							
SO_4	0.49	0.75	-0.61	0.00	1.00						
Fe	0.31	-0.26	-0.52	0.61	0.21	1.0					
NO_3	0.90	-0.27	-0.70	0.67	0.30	0.32	1.0				
Chloride	-0.01	0.43	0.50	-0.30	0.04	-0.87	-0.13	1.0			
TH	-0.59	-0.10	0.17	-0.01	-0.33	0.49	-0.54	-0.66	1.0		
TA	-0.79	0.61	0.43	-0.68	0.0	-0.45	-0.80	0.29	0.37	1.0	

					1						
Na	0.46	0.20	-0.71	0.31	0.68	0.86	0.32	-0.65	0.20	-0.31	1.0

Table 3. Correlation Matrix of Moti Nala

	pH	**EC**	**Cu**	**Cr**	**SO_4**	**Fe**	**NO_3**	**Chloride**	**TH**	**TA**	**Na**
pH	1.0										
EC	-0.58	1.0									
Cu	0.41	-0.08	1.0								
Cr	0.27	-0.07	0.97	1.0							
SO_4	0.75	0.01	0.64	0.49	1.0						
Fe	0.37	0.04	0.98	0.95	0.70	1.0					
NO_3	-0.00	0.09	0.89	0.96	0.29	0.88	1.0				
Chloride	-0.53	0.45	-0.56	-0.52	-0.55	-0.59	-0.35	1.0			
TH	0.14	0.18	0.08	-0.03	0.24	0.01	-0.04	0.47	1.0		
TA	0.43	-0.05	0.98	0.95	0.65	0.95	0.87	-0.43	0.23	1.0	
Na	-0.28	0.27	-0.39	-0.47	-0.14	-0.41	-0.40	0.49	0.73	-0.31	1.0

5. Principal Component Analysis

Urdana Nala - The principal component analyses for the water sample of dairy waste shown in Table 4. It includes loading for the rotated component matrix, eigenvalues for each component, per cent and cumulative per cent of variance explained by each component. It indicates that the first three principal components together account for 91.44% of the total variance in the dataset, in which the first principal component is 44.72%, second principal component is 69.80%, and the third principal component is 91.44% of the total variance. The eigenvalues of the first three principal components (>1) can be used to assess the dominant hydro geochemical processes. The concentrations of pH, Cr and NO_3 show high positive loadings (0.75-0.95) whereas concentrations of SO_4, Fe and Na have low positive loadings (0.16-0.34) for the first principal component. In the second principal component Fe, Na and TH have high positive loadings (0.70-0.91) and the concentration of Cr, SO_4 and NO_3 shows low positive loading (0.01–0.32). For the third principal component, the concentrations of EC and SO_4 show high positive loadings (0.83-0.94), the concentrations of Na show moderate positive loadings (0.61), the concentrations of pH, Fe, NO_3, Chloride and TA show low positive loadings (0.11-0.33).

Moti Nala - The principal component analyses for the water sample of dairy waste shown in Table 5. It includes loading for the rotated component matrix, eigenvalues for each component, per cent and cumulative per cent of variance explained by each component. It indicates that the first three principal components together account for 89.28% of the total variance in the dataset, in which the first principal component is 53.99%, second principal component is 73.39%, and the third principal component is 89.28% of the total variance. The eigenvalues of the first three principal components (>1) can be used to assess the dominant hydrogeochemical processes. The concentrations of Cu, Cr, Fe, NO_3 and TA show high positive loadings (0.94-0.97) the concentrations of SO4 show moderate positive loadings (0.57) whereas concentrations of pH, EC and TH have low positive loadings (0.10-0.20) for the first principal component. In the second principal component pH have high positive loadings (0.95) the concentrations of SO4 show moderate positive loadings (0.62), and the concentration of Cu, Cr, Fe, TH and TA shows low positive loading (0.04-0.24). For the third principal component, the concentrations of TH and Na show high positive loadings (0.81-0.96), the concentrations of pH, EC, SO_4, Chloride and TA show low positive loadings (0.07-0.49).

Table 4. Rotated Component Matrix of Physio Chemical Data (Urdana Nala)

Parameter	Component		
	1	2	3
pH	0.92	-0.02	0.33
EC	-0.41	-0.28	0.83
Cu	-0.52	-0.40	-0.56
Cr	0.75	0.32	-0.17
SO_4	0.16	0.03	0.98
Fe	0.34	0.91	0.11
NO_3	0.95	0.01	0.16
Chloride	-0.10	-0.96	0.08
TH	-0.54	0.76	-0.28
TA	-0.92	-0.16	0.18
Na	0.27	0.70	0.61
Eigen values	4.91	2.76	2.38
% age of variance by component	44.72	25.08	21.63
Cumulative %age of variance	44.72	69.80	91.44

Table 5. Rotated Component Matrix of Physio Chemical Data (Moti Nala)

Parameter	Component		
	1	2	3
pH	0.20	0.95	0.07
EC	0.15	-0.70	0.33
Cu	0.96	0.24	-0.05
Cr	0.96	0.10	-0.18
SO_4	0.57	0.62	0.26
Fe	0.97	0.19	-0.07
NO_3	0.94	-0.17	-0.18
Chloride	-0.41	-0.60	0.49
TH	0.10	0.04	0.96
TA	0.95	0.22	0.09
Na	-0.33	-0.19	0.81
Eigen values	5.94	2.13	1.74
% age of variance by component	53.99	19.39	15.89
Cumulative %age of variance	53.99	73.39	89.28

6. Conclusions

1. Water samples from Moti Nala Urdana Nala (domestic areas) are severely polluted in terms of sodium, however other elements were within permissible limits.
2. The intercorrelationship among the eleven selected water quality parameters revealed that strong correlations (correlation coefficient more than 0.9) exist in pH, NO_3 and SO_4 in Urdana Nala site. Cu, TA, pH and TH in Moti Nala site.
3. The important parameter found $NO_{3,}$ Cr, Fe, TH, Na, EC, $SO_{4,}$ in Urdana Nala site. $NO_{3,}$ Cu, Na, Fe, TA, pH, $SO_{4,}$ TH, Cr in Moti Nala site.

References

[1] APHA-AWWA-WPCF, Standard methods for the examination of water and wastewater. 18th Ed. Washington DC. American Public Health Association, American Water Works Association and Water Pollution Controll Federation, pp. 10–137, (**1992**).
[2] J. C. Davis, "Statistics and data analysis in Geology", John Willey & Sons., New York. (**1973**), pp. 646.
[3] S. Gajbhiye, M. K. Awasthi, S. K. Sharma, "Assessment of Ground Water Quality", LAMBERT Academic Publishing, Germany, ISBN 978-3-659-53985-5, (**2014a**).
[4] S. Gajbhiye, S. K. Mishra and A. Pandey, "Simplified sediment yield index model incorporating parameter", Arabian Journal of Geosciences, (**2014b**)
[5] S. Gajbhiye, S. K. Sharma and M. Jha, "Application of Principal Component Analysis in the assessment of water quality parameters", Sci-fronts A Journal of Multiple Science, vol 4, (**2010**), pp. 67-72.
[6] J. F. Hair, R. E. Anderson, R. L. Tatham and W. C. Black, "Multivariate data analysis with readings (4th ed.)", London: Prentice-Hall, (**1995**).
[7] C. S. Iyer, M. Sindhu, S. G. Kulkarni, S. S. Tambe and B. D. Kulkarni, "Statistical analysis of the physico-chemical data on the coastal water of Cochin", J. Environ. Monit., vol. 5, (**2003**), pp. 324-327.
[8] R. Jayakumar and L. Siraz, "Factor analysis in hydrogeochemistry of Coastal aquifers- preliminary study", Environmental Geology, vol. 31, nos. ¾ , (**1997**), pp. 174- 177.
[9] B. N. Lohani and G. Todino, "Water Quality Index of Chao Phraya River", Journal of Environmental Engineering, vol. 110, no. 6, (**1984**), pp. 1163-1176.
[10] N. Mazlum, A. Ozer and S. Mazlum, "Interpretation of Water Quality Data by Principal Components Analysis", J. of Engineering and Environmental Science, vol. 23, (**1996**), pp. 19-26.
[11] F. W. Ntengwe, "Pollutant loads and water quality in streams of heavily populated and industrialized towns", Physics and Chemistry of the Earth, vol. 31, (**2006**), pp. 832–839.
[12] S. B. Olobaniyi and F. B. Owoyemi, "Chemical Facies of groundwater in the Deltaic Plain Sands Aquifer of Warri", Western Niger Delta, Nigeria, AJST, vol. 7, no. 1, (**2006**), pp. 73-81.
[13] P. Praus, "Water Quality Assessment Using SVD-Based Principal Component Analysis of Hydrological Data", Water SA., vol. 31, no. 4, (**2005**), pp. 417- 422.
[14] S. R. Salman and Y. H. AbuRuka'h, "Multivariate and principle component Statistical Analysis of contamination in urban and agricultural Soils from north Jordan", Environmental Soils from north Jordan, Environmental Geology, vol. 38, no. 3(**1998**), pp. 265-270.
[15] S. Sharma, "Applied multivariate techniques", New York: Wiley, (**1996**).
[16] S. K. Sharma, S. Gajbhiye and S. Tignath, "Application of Principal Component Analysis in Grouping Geomorphic Parameters of Uttela Watershed for Hydrological Modeling", International Journal of Remote Sensing & Geoscience (IJRSG), (**2013**).
[17] S. K. Sharma, S. Gajbhiye and S. Tignath, "Application of Principal Component Analysis in Grouping Geomorphic Parameters of a Watershed for Hydrological Modeling", Applied Water Science, (**2014a**).
[18] S. K. Sharma, S. Gajbhiye, R. K. Nema and S. Tignath, "Assessing vulnerability to soil erosion of a watershed of tons River basin in Madhya Pradesh using remote sensing and GIS", InternationalJournal of Environmental Research and Development, vol. 4, no. 2, (**2014b**), pp. 153-164.
[19] A. S. Shihab, "Application of Multivariate Method in the Interpretation of Water Quality Monitoring Data of Saddam Dam Reservoir", Confidential (**1993**).
[20] V. Simeonov, J. A. Stratis, C. Samara, G. Zachariadis, D. Voutsa and A. Anthemidis, "Assessment of the surface water quality in Northern Greece", Water Research, vol. 37, (**2003**), pp. 4119–4124.
[21] M. Vega, R. Pardo, E. Barrato and L. Deban, "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis", Water Research, vol. 32, (**1998**), pp. 3581–3592.
[22] D. A. Wunderlin, M. P. Diaz, M. V. Ame, S. F. Pesce, A. C. Hued and M. A. Bistoni, "Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality", A case study: Suquira river basin (Cordoba— Argentina). Water Research, vol. 35, (**2001**), pp. 1894–2881.
[23] S. Yerel, "Water Quality Assessment of Porsuk River", Turkey, E-Journal of Chemistry, vol. 7, no. 2, (**2010**), pp. 593-599.

