World Applied Sciences Journal
14 (Exploring Pathways to Sustainable Living in Malaysia: Solving the Current Environmental Issues): 66-72, 2011
ISSN 1818-4952; © IDOSI Publications, 2011

Pattern Recognition of Kedah River Water Quality Data by Implementation of Principal Component Analysis

[1]Isahak Mohd, [1]Muhd Ariffin Mansor, [1]Mohammad Roshide Amir Awaluddin,
[1]Mohd Fahmi Mohd Nasir, [1]Mohd Saiful Samsudin, [1]Hafizan Juahir and [2]Norlafifah Ramli

[1]Department of Environmental Sciences, Faculty of Environmental Studies,
43400 UPM Serdang, Selangor, Malaysia
[2] Department of Environment Malaysia, Surface Water Monitoring Unit,
Water and Marine Division, Federal Government Administrative Centre, 62574 Putrajaya, Malaysia

Abstract: This study examines Kedah River Basin, Kedah, Malaysia, to achieve the objective of identifying and recognizing pollutant sources contributing to the water quality using a large dataset extending over a period of eight years, from the year 1997 to 2006. Principal Component Analysis was applied to simplify and provide a better understanding for the complex relationships among water quality parameters such as DO, BOD, COD, SS, pH, NH3-NL, temperature, conductivity, turbidity, salinity, dissolved solids, total solids, NO3, Cl, Ca, PO4, As, Hg, Cd, Cr, Pb, Zn, Ca, Fe, K, Mg, Na, Oil and Grease, MBAS, *E.coli* and Coliform. Graphical presentation of the data also helps a better view of the overall analysis to appoint sources of pollutant in accordance to their effect. Similar pattern of water quality data reveals nine Principal Components responsible for the data structure and explained 73% of the total variance of the data set. PC score model provided apportionment of various sources contributing to the water quality. Consequently the nine causes of pollutants involved are natural causes in terms of strong river current and geological location of this river, industrial and factories effluent discharge, construction, coal and metal mining, agricultural and sewage plant, human waste and illegal oil dumping.

Key words: Principal Component Analysis · Water quality · Sources apportionment and source of variations

INTRODUCTION

Pollution modeling is generally pointed toward the identification of man-made sources of chemicals. By understanding the spatial and temporal variation of these pollutants, control measures can be applied to bring levels into compliance with environmental standards [1,2].

A river is defined as any natural stream of water that flows in a channel with defined banks. The source of a river may be a lake, a spring, or a collection of small streams, known as headwaters. From their source, all rivers flow downhill, typically terminating in the sea/ocean as sketched in. Sometimes a river flows into the ground or dries up completely before reaching another body of water. A river is a component of the water cycle, thus its quality is directly affected by anthropogenic activities happening throughout the water cycle. Fish survival, diversity and growth; recreational activities, urban, industrial and private water supplies, irrigation and livestock watering, waste disposal and aesthetics value are all affected by the chemical, biological, physical and microbiological conditions in watercourses and in subsurface aquifers [3]. Environmental data may be highly complex and depend on unpredictable factors that are usually characterized by their high variability. The main origins of this variability are geogenic (relating to the formation of the earth), hydrological, meteorological and also anthropogenic (such as different emitters and dischargers) [4].

In this research, Principle Component Score was implemented to recognize pattern in the data as well as reducing the number of possible pollutant sources deteriorating the water quality. Principle Component Analysis (PCA) is recommended as an exploratory tool to uncover unknown trends in the data that involved large and complex dataset which required appropriate method

Corresponding Author: Hafizan Juahir, Department of Environmental Sciences,
Faculty of Environmental Studies, 3400 UPM Serdang, Selangor, Malaysia.

in explaining its meaning. Exploratory data analysis has been used to evaluate the water quality of rivers and seasonal, spatial and anthropogenic influences have been evidenced [5-23]. When applied on conditions, PCA will explore correlations between samples or conditions. The nature of PCA is to 'summarize' the data, therefore it is not considered as a clustering tool. PCA does not attempt to group data by user-specified criteria as does the clustering methods. According to Shlens (2009), Principal Component Analysis is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics - for example prediction of bacteria *Chlorophyll a* in reservoirs and distribution of metal contaminants in soil because it is simple, non-parametric method for extracting relevant information from confusing data sets. This is because the reliability of PCA in reducing the dimensions and complexions of a data matrix in order to pinpoint a clear picture of the particular dataset. With minimal effort PCA provides a roadmap to reduce a complex data set to a lower dimension. The problem of data reduction and interpretation of multi-constituent chemical and physical measurements can be determined through the application of multivariate statistical methods and exploratory data analysis [2, 10]. The usefulness of multivariate statistical tools in the treatment of analytical and environmental data is reflected by the increasing number of data-driven model that uses this software achieved greater and more appropriate interpretation in a simplified manner [8, 9]. The objectives of this research are to determine source apportionment of river water quality by implementing PCA, in order to the lower management costs and increase the efficiency of law. By pointing out which parameters are significantly becoming worse year after year, the pollutant agent can also be identified, thus can be tracked from its sources.

MATERIALS AND METHODS

In this study, a total of 410 samples concluding 30 water quality parameters ranging from the year 1997 until 2006 were used. The location of the sampling stations in grid reference is between east 102° 44.426' north 02° 33.565 and east 101° 23.075' and north 03° 52.792'. Large number of data covers all possible conditions and reduces any flaw that may occur during the construction of this model. The parameters involved and examined in chemometrics analysis, or principle components are; dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, ammoniacal nitrogen (NH3-NL), temperature (T), conductivity (Cond), turbidity (Tur), salinity (Sal), dissolved solids (DS), total solids TS), nitrate (NO3), chloride (Cl), calcium (Ca), phosphate (PO4), Arsenic (As), mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb), zinc (Zn), ferum (Fe), potassium (K), magnesium (Mg), sodium (Na), oil and grease(OG), surfactant (MBAS), *Eschericia Coli* (*E.-Coli*) and Coliform.

The first thing to do when handling a secondary data is a pre-treatment to find any void or unreadable numerical in cells and improve the data obtained and thus increasing the effectiveness of multivariate analysis. Then the following procedures were performed to the data: (1) missing data were estimated using average values and (2) values below the limit were replaced by limits of detection (Mudge, 2007). The first step in assessing this confusing data set is analyzed it using PCA. Each of the parameter can now be known as principle component and the result of the statistical evaluation is principle component scores. Additional factors provide marginally less explanatory capability and were not examined further. The eigenvalue, also called the characteristic root of each discriminant function, reflects the ratio of importance of the dimensions which classify cases of the dependent variable.

Study Area: Sungai Kedah is more than 100 km long, originating from the mountainous areas bordering Perlis and Thailand in the north and northeast. From here the river flows through hilly terrain and finally through a wide coastal plain. As a major river system in the state, Sungai Kedah flows through the districts of Kubang Pasu, Kota Setar, Padang Terap and Pendang. At present there are 54 water quality stations, 26 water discharge (flow) stations and one groundwater stations in the river basin. The natural basin is approximately 60 km wide and 80 km long and covers an area of 2,920 km2. The basin ranges from 400 meters high to the coastal plains. The coastal plain is the centre of rice cultivation. The state capital was founded over 250 years ago at the confluence of Sungai Anak Bukit and Sungai Kedah, which was the centre for the rice trade. Land use of the area along Kedah River can be divided into four; agriculture 62%, forest coverage 28%, urban areas 6.6% and water bodies 3%.

Data: The 9-year data obtained from nine different stations sums up to 12 741 reading taken from this river basin. These water quality parameters were measured within duration of 1997-2006 at nine different water quality monitoring stations, namely 2KD01, 2KD02, 2KD03, 2KD04, 2KD05, 2KD06, 2KD07, 2KD08 and 2KD09 as can

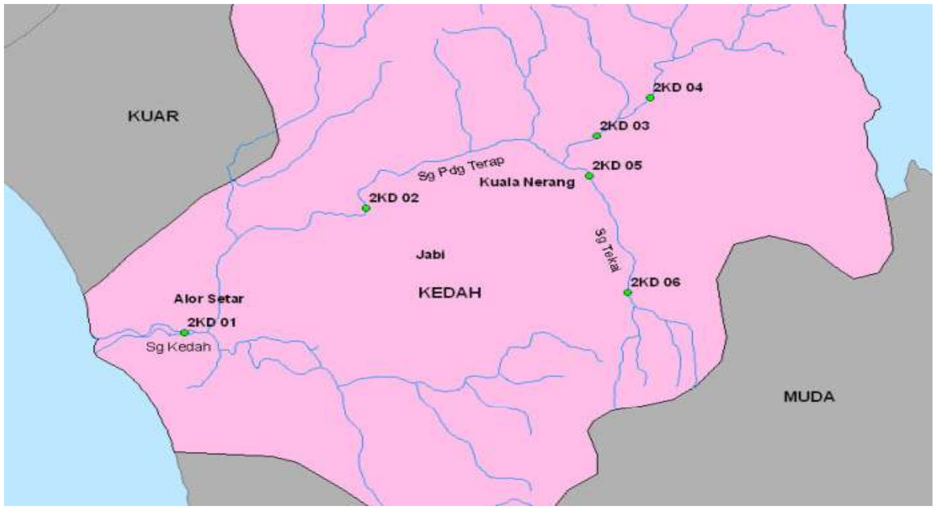

Fig. 1: Map showing the monitoring stations along the river

be seen in Figure 1. Although measured using automatic probes, these data were not perfectly complete, with a number of missing data in each parameter. Missing data or below detection limit value is unreadable when using XLSTAT add-ins in Excel and to overcome this situation, approximation using average of nearest neighboring data were performed. Some data are closed or under the limitation of measurement. This kind of flaw can be identified with mathematical figure (<) in front of the parameters value. What can be done here is to multiply it by two to make it legitimate for XLSTAT analyzing process.

Statistical Method: Principle component analysis and artificial neural network were performed on these 30 parameters to rank their relative significance and to describe their interrelation. XLSTAT software is used to analyze the data with Principal Component Analysis.

Principal Component Analysis: The first step taken is to perform PCA on all 30 water quality parameters in order to exclude insignificant data. After that, eigenanalysis were performed to extract data with eigenvalue higher than 1 and a new group of variables was created as a resemblance of the whole data set using one selection criteria, i.e. the corrected average eigenvalue [4]. The eigenvalue, also called the characteristic root of each discriminant function, reflects the ratio of importance of the dimensions which classify cases of the dependent variable. Varimax rotation was also performed to obtain varimax factors of the principle components. In this sudy, only varimax factors with values more than 0.70 (positive or negative) will be discussed. In previous study [28], variables with loadings greater than 0.7 are considered strong, variables with loadings from 0.7 to 0.5 are moderate and variables with loadings lower than 0.5 are considered a weak variable. Rotations were done in order to maximize the number of factors in common to the number of variables used in this study, which in this case is 30 because the number of parameters involved are 30.

Score value (s^{kj}) for j^{th} observation in k^{th} PC was obtained from the weight of variables in PCs and standardized variables by using the following equation;

$$S^{kj} = t1^{k}\, z1^{j} + t2^{k}\, z2^{j} + ... + t^{pk}\, z^{pj} \qquad (1)$$

Where j = 1, 2, . . . n is the number of observation; k = 1, 2, . . . q, the number of selected PC number; p the number of independent variables; *Skj* the standardized score value of jth observation in k^{th} PCs; z^{pj} the standardized value of p^{th} variable of jth observation, calculated from $z = yp - \square/sx$, where *yp* is the original value of pth variable; and t^{pk} the standardized weight of the p^{th} variable in k^{th} PCs. One approach was employed in using PC scores that is, utilizes only 9 PCs with eigen values greater than 1 out of 29 principles components. Eigenvalues can be thought of as quantitative assessment of how much a component represents the data. The higher the eigenvalues of a component, the more representative it is of the data.

RESULTS AND DISCUSSION

Value of $x2$, calculated as 484.63 by Barttlett's sphericity test (d.f 435 and p-value 0.05), implies that the principal component analysis is applicable to this dataset.

Table 1: Factor loadings after varimax rotation

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
DO	-0.022	-0.672	-0.080	0.025	0.025	-0.197	-0.176	0.021	0.470
BOD	0.057	0.174	0.076	0.066	-0.122	0.794	0.101	0.107	0.024
COD	0.217	0.552	-0.137	0.540	-0.038	0.189	0.079	-0.150	0.090
SS	0.005	-0.018	-0.017	0.937	0.018	0.022	-0.012	-0.016	-0.044
PH	-0.043	-0.075	0.144	-0.130	-0.005	0.042	0.057	0.049	0.812
NH3-NL	0.069	0.428	0.059	-0.134	-0.013	0.343	0.231	-0.055	-0.351
TEMP	0.012	0.707	-0.001	-0.095	0.103	-0.094	0.220	0.051	0.058
COND	0.581	-0.106	0.326	-0.069	0.504	0.081	0.092	0.287	-0.034
SAL	0.599	-0.127	0.195	-0.072	0.552	0.017	0.071	0.273	-0.055
TUR	0.032	-0.030	0.092	0.904	-0.017	-0.019	0.034	0.018	-0.054
DS	0.969	0.017	0.094	-0.047	0.173	0.038	0.020	0.045	-0.019
TS	0.926	0.009	0.084	0.278	0.171	0.044	0.014	0.037	-0.033
NO3	0.199	-0.148	0.202	-0.006	0.459	0.648	-0.137	-0.046	-0.098
CL	0.978	0.045	0.038	-0.018	0.023	-0.008	0.013	0.029	-0.020
PO4	0.001	-0.062	0.045	0.043	0.168	0.444	0.099	-0.064	0.060
AS	0.280	0.020	0.017	0.019	0.802	0.056	-0.050	-0.021	-0.043
HG	0.109	-0.393	0.056	-0.115	-0.020	0.332	0.086	0.324	-0.363
CD	0.020	-0.052	0.843	-0.004	-0.010	0.014	0.054	-0.203	0.017
CR	0.144	0.035	0.909	0.044	0.036	0.094	-0.019	0.215	0.037
PB	0.137	-0.010	0.921	0.021	0.067	0.044	-0.041	-0.003	0.037
ZN	0.104	0.160	-0.040	0.072	0.633	-0.118	0.059	-0.004	0.220
CA	0.501	0.057	0.360	-0.084	0.168	0.396	-0.068	0.038	0.070
FE	-0.122	0.217	0.103	0.402	0.176	-0.255	0.086	0.200	0.236
K	0.682	0.048	0.021	0.015	0.498	0.113	-0.032	0.015	0.003
MG	0.903	0.057	0.143	-0.027	-0.219	0.072	0.021	0.018	0.034
NA	0.938	0.074	-0.055	-0.028	0.138	0.030	-0.023	-0.040	-0.019
OG	0.120	0.027	0.031	-0.006	0.049	0.041	-0.043	0.888	0.043
MBAS	0.497	0.586	-0.078	0.054	-0.071	-0.052	-0.083	0.202	-0.003
E. coli	-0.008	0.098	0.015	0.029	0.064	-0.048	0.806	-0.014	-0.095
COLIFORM	0.064	0.134	-0.049	0.036	-0.077	0.104	0.786	-0.036	0.115
Eigenvalue	7.587	2.930	2.494	2.213	1.706	1.665	1.166	1.161	1.044
Var %	25.290	9.766	8.312	7.375	5.687	5.551	3.886	3.871	3.479
CV %	25.290	35.056	43.369	50.744	56.431	61.982	65.868	69.739	73.218

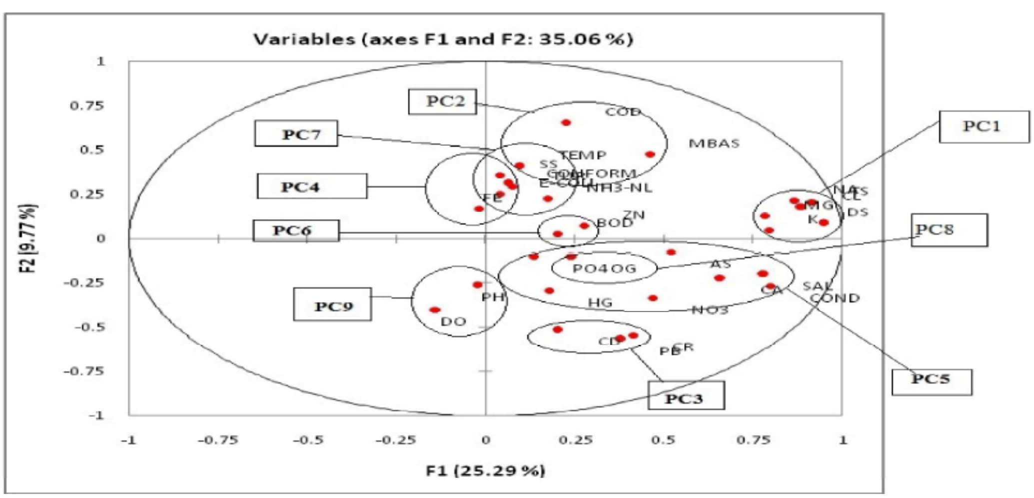

Fig. 2: Scores of river water samples and loadings of variables on the plane

Table 2: Eigenvalues, variance and cumulative percentage of PCs

Principle Components	Eigenvalues	Percentage of total variance	Cumulative percentage
PC1	7.587	21.271	21.271
PC2	2.930	7.164	28.435
PC3	2.494	9.437	37.872
PC4	2.213	7.744	45.616
PC5	1.706	7.757	53.373
PC6	1.665	6.192	59.566
PC7	1.166	5.021	64.586
PC8	1.161	4.260	68.846
PC9	1.044	4.372	73.218

The first step in assessing this dataset is analyzing it using PCA. Each of the parameter can now be known as principle component (PC) and the result of the statistical evaluation is principle component scores (PCS). PCA on 30 parameters yielded nine principle components explaining approximately 73% of sample variances as can be seen in Table 1. Additional factors provide marginally less explanatory capability and were not examined further. There is one eigenvalue for each discriminant function. Only nine out of 30 variables are further analyzed using varimax rotation because the value needed for this rotation is more than 1 [25]. The interpretation of this rotation can bring out unobservable and latent data [6]. Similar approach based on PCA has been used to identify the main components in water quality by [6, 8, 9, 13 and 25].

In Figure 2, both the scores of samples and loadings of variables are presented, corresponding to PC1 and PC2. Samples sites can be classified in nine different groups. The overlapping circle shows that the parameter involved in each PC weakly or moderately influence the classification of each PC.

The first principle component that (PC) explains total variances of 21.27% (Table 2) has significant positive loadings on dissolved and total solids, chloride, magnesium and sodium. High loadings on dissolved and total solids were due to natural effects of strong river current that prohibited tiny physical solids from settling. Inert solids are produced in all montane rivers as the energy of the water helps grind away rocks into gravel, sand and finer material.

Second PC, as in Table 2, shows strong positive loading (0.707) for temperature. Fast moving waters in the river make it colder water whereas slow flowing river is likely to be warmer. Other than the river current, fluctuations of water temperature can also associated with cloud coverage [28]. Cloudy days and nights would affect the efficiency of a river to obtain temperature increase from insulation and temperature decrease from radiation cooling. PC2 charted 7.2% of the total variance and 28.4% of the cumulative percentage of variance.

The third PC can be explained with 9.44% of variance from factor loadings after varimax rotation. Presence of excess Cd, Cr and Pb in the third PC accounted for strong positive loadings for all of the components. Cadmium and chloride can be found in most of fertilizers used for agricultures. Documented liquid fertilizers are often used in agricultural activities and this heavy metals can absorb into soil and infiltrate into groundwater. The presence of lead in river systems is largely due to anthropogenic inputs such as fossil fuel burning, mining activities and metal smelters [29].

The fourth PC that explains 7.74% of the total variance shows strong positive loading both on suspended solids and turbidity. These two variables (water quality parameter) are connected and influenced each other. High loadings suggested that construction work and sediment erosion along the river may cause small debris from land carried into the river by surface runoff [30]. Particular section of the river with the monitoring station would pick up high reading of suspended solids and increasing turbidity.

Zn and As shows strong readings in factor loadings 5 that explains a total variance of 7.76%. Arsenic in nature can be found at low level, meaning that anthropogenic activities might have triggered the concentration of this metal in this river basin. Arsenic doesn't evaporate but it do dissolve in water and can be found in many antifouling paints [31] used for fishermen to coat their boats, weed killers which might be used by farmers to kill off weeds surrounding their crops and also insecticides for protection against enormous group of insects. For PC6, the total variance after varimax rotation is 6.19% and account for 59.57% for the cumulative percentage.

In most countries, rivers are heavily contaminated by the natural and soil bacteria, but also large number of organisms produced from sewage [32]. PC6 shows significant loadings of BOD, where the cause of these significant loadings may be due to organic source and sewage treatment produced into the river with very high concentrations of bacteria and [33].

Table 4.1.1 shows PC 7, 8 and 9 have 5.02%, 4.26% and 4.37% of total variance respectively. PC 7 accounted for significant loadings of *E. coli* and Coliform, which is suspected to originate from animal faeces, surface runoffs and discharges from sewage treatment plant [32]. A previous study [32] also associated *E. Coli* level during the dry season exceeds the level in wet season.

PC 8 shows a strong positive loading on Oil and Grease. Oil and grease not only originate from petroleum oil but also natural and vegetable oils [34]. Sediments and aquatic or non-aquatic decaying biota are often rich in natural oils, which make up for significance in measurements [34]. Surface runoffs from road drainage may also contain fossil fuel from vehicles as a result of leakage.

The last principal component shows significant positive loadings for pH. The level of pH can vary as a result of the bedrock and soil composition in which a river flows through, the amount of organic material and plant in the river, chemical dumping and acid rain precipitation. However in this case the pH level shows strong positive loadings with moderate positive loading of DO, which indicates the level of organic material and plant growth such as algae is sufficiently excessive to change the pH of this river. Organic materials and plants produced carbon dioxide, which change the pH level in the water as in previous study [35].

CONCLUSION

In this study PCA serves as useful effective tools for assessing a total of 30 water quality parameters obtained from various sampling station along Kedah River Basin. The multivariate tool shows that out of 30 parameters, nine principal components or variables were extracted from the data. Origin of pollutants were interpreted from these nine principal components, either it's a point source or a nonpoint source. It shows that management of a river basin as important as Kedah River can be done using secondary data collected by DOE without further sampling which would consume more cost. Management in the river should be controlled in order to maintain good quality of water that can be used for any purpose.

ACKNOWLEDGEMENTS

I would like to express my most appreciations to all of the individuals who had contributed their precious time and energy to this project paper especially my supervisor, Dr. Hafizan Bin Juahir for his continous guidance, great concern, valuable comments, fruitful advices and constructive suggestion from the beginning until the completion of this research. Thanks also to all my project partners Mohd Fahmi Mohd Nasir, Mohd Saiful bin Samsudin, Muhd Roshide Amir Awalluddin and Mohd Ariffin Mansor for their help and guidance to write this paper.

REFERENCES

1. Juahir, H., S.M. Zain, A.Z. Aris, M.K. Yusof, M.A.A. Samah and M. Mokhtar, 2010. Hydrological Trend Analysis Due to Land Use Changes at Langat River Basin. Environment Asia 3(special issue) pp: 20-31.
2. Infometrix Inc., 1995. Chemometrics in Environmental Science. Chemometrics Application Overview.
3. Najah, A., A. Elshafie, O. Karim and O. Jaffar, 2009. Prediction of Johor River Water Quality Parameters Using Artificial Neural Networks. European Journal of Scientific Res., 28(3): 422A435.
4. Juahir, H., S.M. Zain, R.A. Khan, M.K. Yusof, M. Mokhtar and M.E. Toriman, 2009. Using Chemometric in Assessing Langat River Water Quality and Designing a Cost-effective Water Sampling Strategy. Maejo International J. Science and Technol., 3(01): 26-42.
5. Brown, S.D., B. Blank, S.T. Sum and L.G. Weyer, 1994. Chemometrics. Anal. Chem., 66: 315R-359R.
6. Simeonov, V., J.A. Stratis, C. Samara, G. Zacharidis, D. Voutsa and A. Anthemidis, 2003. Assessment of the Surface Water Quality in Northern Greece. Water Res., 37: 4119-4124.
7. Einax, J.W., D. Truckenbrodt and O. Ampe, 1998. River Pollution Data Interpreted by Means of Methods. Microchem. J., 58: 315-324.
8. Vega M., R. Pardo, E. Barrado and L. Deban, 1998. Assessment of Seasonal and Polluting Effects on the Quality of River Water by Exploratory Data Analysis. Water Res., 32: 3581-3592.
9. Helena, B., R. Pardo, M. Vega, E. Barrado, J.M. Fernandez and L. Fernandez, 2000. Temporal Evaluation of Groundwater Composition in An Alluvial Aquifer (Pisuerga River, Spain) by Principal Component Analysis. Water Res., 34: 807-816.

10. Brown, S.D., S.T. Sum and F. Despagne, 1996. Chemometrics. Anal. Chem., 68: 21R-61R.
11. Massart, D.L. and L. Kaufman, 1983. The Interpretation of Analytical Data by the Use of Cluster Analysis. Wiley, New York.
12. Shlens J., 2009. A Tutorial on Principal Component Analysis.Center for Neural Science (New York University), NY. 10003-6603.
13. Singh, K.P., A. Malik, D. Mohan and S. Sinha, 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.
14. Bartels, J.H.M., T.A.H.M. Janse and F.W. Pijpers, 1985. Classification of The Quality of Surface Waters by Means of Pattern Recognition. Analytica Chimica Acta, 177: 35-45.
15. Grimalt, J.O., J. Olive and J.I. Gomez-Belinchon, 1990. Assessment of Organic Source Contributions in Coastal Waters by Principal Component and Factor Analysis of The Dissolved and Particulate Hydrocarbon and Fatty Acid Contents. International J. Environment Analytical Chemistry, 38: 305-320.
16. Librando, V., 1991. Chemometric Evaluation of Surface Water Quality at Regional Level. Fresenius J. Anal. Chem., 339: 613-619.
17. Aruga, R., G. Negro and G. Ostacoli, 1993. Multivariate Data Analysis Applied to The Investigation of River Pollution. Fresenius J. Anal. Chem., 346: 968-975.
18. Elosegui, A. and J. Pozo, 1994. Spatial vs Temporal Variability in The Physical and Chemical Characteristics of The Aguera Stream (Northern Spain). Acta Ecologica Đ Int. J. Ecol., 15: 543-559.
19. Pardo, R., E. Barrado, M. Vega, L. Deban and M.L. Tascon, 1994. Voltammetric complexation capacity of waters from the Pisuerga River. Wat. Res., 28: 2139-2146.
20. Battegazzore, M. and M. Renoldi, 1995. Integrated Chemical and Biological Evaluation of The Quality of The River Lambro (Italy). Wat. Air Soil Poll., 83: 375-390.
21. Voutsa, D., G. Zachariadis, C. Samara and T. Kouimtzis, 1995. Evaluation of Chemical Parameters in Aliakmon River in Northers Greece. 2. Dissolved and Particulate Heavy Metals. J. Environ. Sci. Hlth. Part A: Environ. Sci. Engng, 30: 1-13.
22. Mudge, S.M., 2007. Multivariate Statistical Methods in Environmental Forensics. Environmental Forensics, 8: 155.
23. Wenning, R.J. and G.A. Erikson, 1994. Interpretation and Analysis of Complex Environmental Data Using Chemometric Methods. Trends Anal. Chem., 12: 446-447.
24. Liu, C.W., K.H. Lin and Y.M. Kuo, 2003. Application of Factor Analysis in the Assessment of Groundwater Quality in A Blackfoot Disease Area in Taiwan. Sci. Total Environ., 313: 77-89.
25. Wunderlin, D.A., M.D.P. Diaz, M.V. Ame, S.F. Pesce, A.C. Hued and M.D. Bistoni, 2001. Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality, A Case Study of Suquia River Basin (Cordoba, Argentina). Water Res., 35: 2881-2894.
26. Kim, J.O. and C.W. Mueller, 1987. Introduction to Factor Analysis: What It Is and How to Do It. Sage University Press, Newbury Park.
27. Dalal, S.G., P.V. Shirodkar, X.N. Verlekar, T.G. Jagtap and G.S. Rao, 2009. Apportionment of sources affecting water quality: Case study of Kandla Creek, Gulf of Katchchh. Environmental Forensics, 10(2): 101-106.
28. Maybeck, H.F., 2004. Rapid River Water Temperature Fluctuations.
29. Véron A.J., T.M. Church and R. Flegal, 1998. Lead Isotopes in the Western North Atlantic Transient Tracers of Pollutant Lead Inputs. Environ. Res. Section. A., 78: 04-11.
30. Kannel P.R., S. Lee, S.R. Kanel and S.P. Khan, 2007. Chemometric Application in Classification and Assessment of Monitoring Locations of an Urban River System. Analytica Chimica. Acta, 582: 390-399.
31. Kopylov, N.I., Y.M. Kaplin, V.P. Litvinov and Y.D. Kaminskii, 2007. Large-Scale Use of Arsenic in the Production of Antifouling Coatings. Theoretical Foundations of Chemical Engineering, 41(5): 780-785.
32. Chigor V.N., V.J. Umoh and S.I. Smith, 2010. Occurrence of *Escherichia Coli* O157 in a River Used for Fresh Produce Irrigation in Nigeria.
33. Wilson, G. and E.B. Pike, 1990. The bacteriology of water.In: Edward Arnold(ed) Principles of Bacteriology, Virology and Immunity. Vol. I. General Bacteriology and Immunity. pp: 243- 264.
34. Irwin, R.J., M. VanMouwerik, L. Stevens, M.D. Seese and W. Basham. 1997. Environmental Contaminants Encyclopedia. National Park Service, Water Resources Division, Fort Collins, Colorado.
35. Umamaheswari, S. and N. Anbusaravanan, 2009. Water Quality of Cauvery River Basin in Tricchirippalli, India. International J. Lakes and Rivers, 2(1): 1-20.