INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ENGINEERING, SCIENCE & MANAGEMENT

ASSESSMENT OF SURFACE WATER QUALITY USING MULTIVARIATE STATISTICAL TECHNIQUES – A REVIEW

Jinal Patel[1] and Minakshi Vaghani[2]
M.E, Environmental Engg., Sarvajanik College Of Engg. And Tech., Surat, Gujarat, India [1]
Assistant Prof., Civil Engg., Sarvajanik College Of Engg. And Tech., Surat, Gujarat, India[2]

Abstract*: Multivariate statistical techniques, such as cluster analysis (CA), Principal component analysis (PCA), Factor analysis (FA) and Discriminant analysis (DA), use for the evaluation of temporal/spatial variations and the interpretation of a large complex variable surface water quality data set. Hierarchical cluster analysis grouped sampling sites into three clusters, i.e., relatively less polluted (LP), medium polluted (MP) and highly polluted (HP) sites, based on the similarity of water quality characteristics. Factor analysis/principal component analysis, applied to the data sets of the three different groups obtained from cluster analysis, which indicate that the parameters responsible for water quality variations are mainly related to discharge and temperature (natural), organic pollution (point source: domestic wastewater) in relatively less polluted areas; organic pollution (point source: domestic wastewater) and nutrients (non-point sources: agriculture runoff) in medium polluted areas; and organic pollution and nutrients (point sources: domestic wastewater, wastewater treatment plants and industries) in highly polluted areas in the basin. Discriminant analysis gave the best results for both spatial and temporal analysis. DA allowed a reduction in the dimensionality of the large data set, delineating a few indicator parameters responsible for large variations in water quality. Thus, this study illustrates the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, and in water quality assessment, identification of pollution sources/factors and understanding temporal/spatial variations in water quality for effective River water quality management.

Keywords*: Water quality; Cluster analysis; Principal component analysis; Factor analysis; Discriminant analysis

INTRODUCTION

Water quality has become one of the major environmental concerns worldwide and is influenced by natural and anthropogenic disturbance, such as wastewater, runoff effluents, land reclamation, atmospheric deposition and climate change. In recent years, more and more attention has been paid to surface water quality because of its strong linkage with human well being. The quality of river at any point reflects several major influences, including the lithology of the basin, atmospheric inputs, and climatic conditions and governed by both natural process and anthropogenic effects. So, wastewater from agricultural, industrial and urban activities and often natural

 ISSN : 2394-1766

processes such as erosion and weathering degrades water quality and impair their use for drinking, industrial, agriculture, recreation or other purposes. Clean river water is a vital commodity for the well-being of human societies, and damage of inland aquatic system was one of the most serious environmental problems of the last century. Surface water, due to their role in carrying-off the domestic and industrial wastewater and run-off from agricultural land in their vast drainage basin are among the most vulnerable water bodies to pollution. Discharges from municipals and industries are considered as a point source while surface runoff is as a non-point source due to its characteristics that are highly influenced by spatial and seasonal changes.

Seasonal variations in precipitation, surface runoff, interflow, groundwater flow and pumped in and outflows have a strong effect on river discharge and, subsequently, on the concentration of pollutants in surface water. Therefore, the effective, long-term management of surface water requires a fundamental understanding of hydro-morphological, chemical and biological characteristics. However, due to spatial and temporal variations in water quality, a monitoring program, providing a representative and reliable estimation of the quality of surface waters, is necessary.

The application of different multivariate statistical techniques such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors/sources that influence water systems and offers a valuable tool for reliable management of water resources as well as rapid solution to pollution problems. Multivariate statistical techniques has been applied to characterize and evaluate surface and freshwater quality, and it is useful in verifying temporal and spatial variations caused by natural and anthropogenic factors link to seasonality.

STATISTICAL METHODS

All mathematical and statistical calculations can be implemented using STATISTICA 8 and Microsoft Office Excel 2007.

A. *Cluster Analysis (CA)*

Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between Cluster) heterogeneity. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set, and is typically illustrated by a dendrogram. The objectives of this study is to identify similarities or dissimilarities among monitoring sites and to distinguish each cluster analysis applied to find out the similarity groups between monitoring sites. Cluster analysis has to perform on the mean values of the parameters for each of the stations. It yielded a dendrogram, grouping all the sampling sites into three statistically significant clusters. The results separates three different groups (clusters), which consisting all the measured values.

1) *Group A:* This group (cluster) could be regarded as relatively less polluted (LP) sites. In this area, there should the urbanization and industrialization level relatively low. Hence, the impact of human activities on the surface water ecosystem is relatively low.
2) *Group B:* This group (cluster) could be regarded as relatively highly polluted (HP) sites. These sites are located industrial and city area; therefore, these sampling stations received pollutants mostly from domestic wastewater, wastewater treatment plants and industrial effluents. Beside these, untreated sewage of agriculture activities and surface runoff increased pollutants at sites.
3) *Group C*: This group (cluster) could be regarded as relatively moderate polluted (MP) sites. These sites may receive pollution from non-point sources.

The results indicate that the CA technique is useful in offering reliable classification of surface water in the selected region and make it possible to design a future spatial sampling strategy in an optimal method, which can reduce the number of sampling stations and associated costs of sites.

B. *Principal Component Analysis (PCA) / Factor Analysis (FA)*

PCA is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA applied to the normalized data to compare the compositional pattern between the analyzed water samples and to identify the factors that influenced each samples. Rotation of the axis defined by PCA produced a new set of factors, involving primarily subset of the original variables is divided into groups. PCA was applied to the data set to compare the compositional pattern between the analyzed water samples and to identify the factor that reflecting each other. PCA was performed on the raw data set comprising all the water quality parameters with observation to identify the factors that contribute to pollution sources.

PCA provides information on the most meaningful parameters which describe the whole data set interpretation, data reduction and to summarize the statistical correlation among constituent in the water with minimum loss of original information. PCA are sensitive to outliers, missing data and poor linear correlation between variables due to inadequate assigned variables. Therefore, a detailed pretreatment data set needs to perform in order to get clearer image in complex data. It is a prominent technique for pattern recognition in attempts to explain the variance of a large set of inter-correlated variables and transforming into a smaller set of independent (uncorrelated) variables (principal components). The PC's can be expressed using the equation below:

$$Zij = ai1x1j + ai2x2j + \ldots + aimxmj$$

Where Z is the component score, a is the component loading, x is the measured value of a variable, i is the component number, j is the sample number, and m is the total number of variables.

FA follows PCA. The main applications of factor analytic techniques are, to lessen the number of variables and to discover structure in the relationships between variable, that is to classify variables.

C. Discriminant analysis (DA)

Discriminant analysis (DA) is used to classify cases into categorical-dependent values, usually a dichotomy. If discriminant analysis is effective for a set of data, the classification table of correct and incorrect estimates will yield a high correct percentage. In DA, multiple quantitative attributes are used to discriminate between two or more naturally occurring groups. Temporal variations in water quality evaluate through DA. In DA, multiple quantitative attributes are used to discriminate between two or more naturally occurring groups. DA provides statistical classification of samples and it is performed with prior knowledge of membership of objects to a particular group or cluster. DA helps in grouping samples sharing common properties. The DA technique builds up a discriminant function for each group, which operates on raw data and this technique constructs a discriminant function for each group.

SOME STUDY OF MULTIVARIATE STATISTICAL TECHNIQUES

The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), assists in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors that influence water environment systems and offers a valuable tool for reliable management of water resources. In recent years, many studies related to these methods have been carried out.

A. H. Pejman; G. R. Nabi Bidhendi et al.(2009), studied spatial and seasonal variations of water quality in Haraz River Basin were evaluated using multivariate statistical techniques, such as cluster analysis, principal component analysis and factor analysis. Water quality data collected from 8 sampling stations in river during 4 seasons (Summer and Autumn of 2007, Winter and Spring of 2008) were analyzed for 10 parameters. Result of PCA/FA evinced that, a parameter that can be significant in contribution to water quality variations in river for one season, may less or not be significant for another one.

Adamu Mustapha and Ado Abdu (2012), Principal component analysis (PCA) and multiple linear regressions were applied on the surface water quality data with the aim of identifying the pollution sources and their contribution toward water quality variation along Jakara River. PCA was used to investigate the origin of each water quality parameters which identified latent pollution sources.

Bhattacharyya Rama, Kumar Manoj et al. (2013), appled multivariate statistical approaches on water quality data of Damodar river ,India for better interpretation of water chemistry. Principal component analysis (PCA)/Factor analysis (FA) and cluster analysis (CA) along with correlation analysis was performed on the data matrix for apportionment of sources of chemicals found in the river. PCA/FA and CA interpreted both geogenic and anthropogenic factors responsible for influencing the water chemistry.

Shrestha and Kazama(2006) were used multivariate statistical methods, such as CA , DA and PCA/FA to analysis and interpretation of complex data sets , and in water quality assessment, including 12 parameters at 13 sites of the Fuji river basin from 1995-2002 to identification of pollution sources/factors and understanding temporal/spatial variations in water quality for effective river water quality management.

Kazi et al. (2009) using CA and PCA methods (with monitoring at five different sites for 36 parameters in Manchar Lake, generated during 2005-06) evaluated and interpreted complex water quality data sets and apportioned of pollution sources to get better information about water quality and to design a monitoring network.

Saroj Gyawali1, KuaananTechato et al.(2012), studied Multivariate statistical techniques, such as cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA), and factor analysis (FA)application to evaluate temporal and spatial variations of water quality and to identify potential pollution sources of U-tapao River Basin (URB) from 21 monitoring stations of river during five years (2007-2011) and analyzed for 12 parameters. The study illustrates the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, and in water quality assessment, identification of pollution sources/factors and understanding temporal/spatial variations in water quality for effective River water quality management.

Xiaoyun Fan, Baoshan Cui et al.(2012), PCA and CA multivariate techniques have been used for water quality spatial assessment and pollution sources identification in Pearl River Delta, China region. They conclude that the method used here can offer a useful tool for assessment of water quality and management of water resources in some regions with a large number complex water quality datasets involved.

REFERENCES

[01] A. H. Pejman; G. R. Nabi Bidhendi; A. R. Karbassi; N. Mehrdadi; M. Esmaeili Bidhendi, “Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques”, Int. J. Environ. Sci. Tech., 6 (3), 467-476, 2009,ISSN: 1735-1472

[02] Adamu Mustapha and Ado Abdu, “Application of Principal Component Analysis & Multiple Regression Models in Surface Water Quality Assessment”, Journal of Environment and Earth Science ISSN 2224-3216 , 2(2), 2012

[03] Bhattacharyya Rama, Kumar Manoj and Padhy Pratap Kumar, “Index Analysis, Graphical and Multivariate Statistical Approaches for Hydrochemical Characterisation of Damodar River and its Canal System,Durgapur, West Bengal, India”, International Research Journal of Environment Sciences, ISSN 2319–1414 Vol. 2(2), 53-62, February (2013)

[04] Mohd Fahmi Mohd Nasir, Mohd, Saiful Samsudin, Isahak Mohamad, Mohammad Roshide Amir Awaluddin, Muhd Ariffin Mansor, Hafizan Juahir, Norlafifah Ramli, “River Water Quality Modeling Using Combined Principle Component Analysis (PCA) and Multiple Linear Regressions (MLR): A Case Study at Klang River, Malaysia”, World Applied Sciences Journal, ISSN 1818-4952,pp:73-82, 2011

[05] Reghunath, R., Murthy, T.R.S., Raghavan, B.R., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research 36, 2437- 2442.

[06] Saroj Gyawali, KuaananTechato, Chumpol Yuangyai and Sathaporn Monprapusson, “Evaluation of Surface Water Quality using Multivariate Statistical Techniques: A Case Study of U-tapao River Basin, Thailand”, KMITL Sci. Tech. J. Vol. 12 No. 1 Jan. - Jun. 2012

[07] Shrestha, S.; Kazama. F., (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environ. Model. Software, 22 (4), 464-475

[08] Singh, K.P., Malik, A.D. Mohan, S. 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 resources,38(18): 3980-3992

[09] 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. Analytica Chimica Acta 538, 355-374.

[10] Varol, M. and Sen, B. 2009. Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. Environmental Monitoring and Assessment, 159, 543-553.

[11] Vikram Bhardwaj,, Dhruv Sen Singh and A K Singh, "Water quality of the Chhoti Gandak River using principal component analysis, Ganga Plain, India", J. Earth Syst. Sci. 119, No. 1, February 2010, pp. 117–127

[12] Xiaoyun Fan, Baoshan Cui, Hui Zhao, Zhiming Zhang, Honggang Zhang, "Assessment of river water quality in Pearl River Delta using multivariate statistical techniques", science direct, Procedia Environmental Sciences 2 (2010) 1220–1234

 ISSN : 2394-1766