ADVANCES IN BIORESEARCH
Volume 3, Issue 1, March 2012: 64 - 68
ISSN 0976-4585

Website: www.soeagra.com/abr.htm

Research Article

Categorization of Pollution Load in surface Water System using Multivariate Techniques

Sukhdev Kundu
Shoolini University, Solan-Himachal Pradesh, India
Corresponding Author email: drkundu24@yahoo.co.in

ABSTRACT
The pollution loadings in the Ghaggar River Basin surface waters arrive from various point and non-point sources. Multivariate statistical techniques, including cluster analysis (CA), principal component analysis (PCA) and factor analysis (FA), were used to characterize most significant variables and to check their spatial variations. The data sets which contained 15 parameters and 31 monitoring sites were generated during 2006. The significant concentrations of different parameters (highly, moderately and weakly) were determined by using PCA/FA methods. CA results classified the surface water stations based on their similarity and dissimilarity among the parameters. Different clusters of sampling sites pointed out that each cluster had a water quality of its own which was dissimilar from the other clusters. PCA and FA concluded that anthropogenic activities and natural sources were the major reasons of river water pollution.
***Keywords**: Multivariate statistical analysis, Principal component analysis, Factor analysis, Cluster analysis*

INTRODUCTION

Rivers are an important source of surface water and play an important role in daily life. On the other hand, the quality of river water is matter of serious concern. They are among the most vulnerable water bodies to pollution because of their role in transporting municipal and industrial wastes and runoffs from agricultural lands in their vast drainage basins. The various types of pollution such as domestic and industrial wastewaters may potentially affect the river water quality as the common ways of releasing these types of pollutants via river. In our study area, river is getting wastewaters from various point and non-point sources viz. industrial, municipal, domestic and agricultural run-offs.

In this study, PCA/FA and CA techniques were used to characterize the most significant variables and to identify the locations that are affected by the specific variables. Earlier multivariate statistical analysis has also been successfully applied in a number of hydrogeo-chemical studies (1-3). All the studies show that multivariate statistical analysis can help to interpret the complex data sets and assess the water quality. It is also useful in verifying spatial variations caused by natural and anthropogenic factors. Lomniczi *et al.* (2007) (4) has also characterized the pollution sites by using principal component analysis (PCA), while working on Arenales River (Salta) Argentina. The objective of our study was to investigate the main pollutant contributor in surface water during the year 2006 and to determine the specific sampling locations that affected by the pollutants by using factor loading from the principal component. The spatial characterization was done using cluster analysis (CA) to evaluate the most significant parameters that affect the aquatic health of the river (5).

MATERIALS AND METHODS

Study area and Monitoring sites

The study area is situated in Haryana and Punjab states of India (Figure 1). The Ghaggar, a major river of Haryana originates from the lower Siwalik Hills of Himachal Pradesh and Haryana. It runs along the foot of the lower Siwaliks Hills and flows through Haryana and Punjab to Rajasthan and then disappears in the sands of the Thar Desert. The Ghaggar River receives several small tributaries and drains that flow through urban areas with domestic, industrial and agricultural wastes.

Data Preparation

The data sets of the 31 monitoring sites, which comprise 15 water quality parameters monitored for 2006 were subjected for spatial characterization using principle component analysis (PCA)/factor analysis (FA) and cluster analysis (CA). The monitored physico-chemical variables are temperature, pH, electrical conductivity (EC), total dissolved solids (TDS), carbonates (CO_3^{2-}), bicarbonates (HCO_3^-), chlorides (Cl^-), sulphates (SO_4^{2-}), phosphates (PO_4^{3-}), fluoride (F^-), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+) and total hardness (TH).

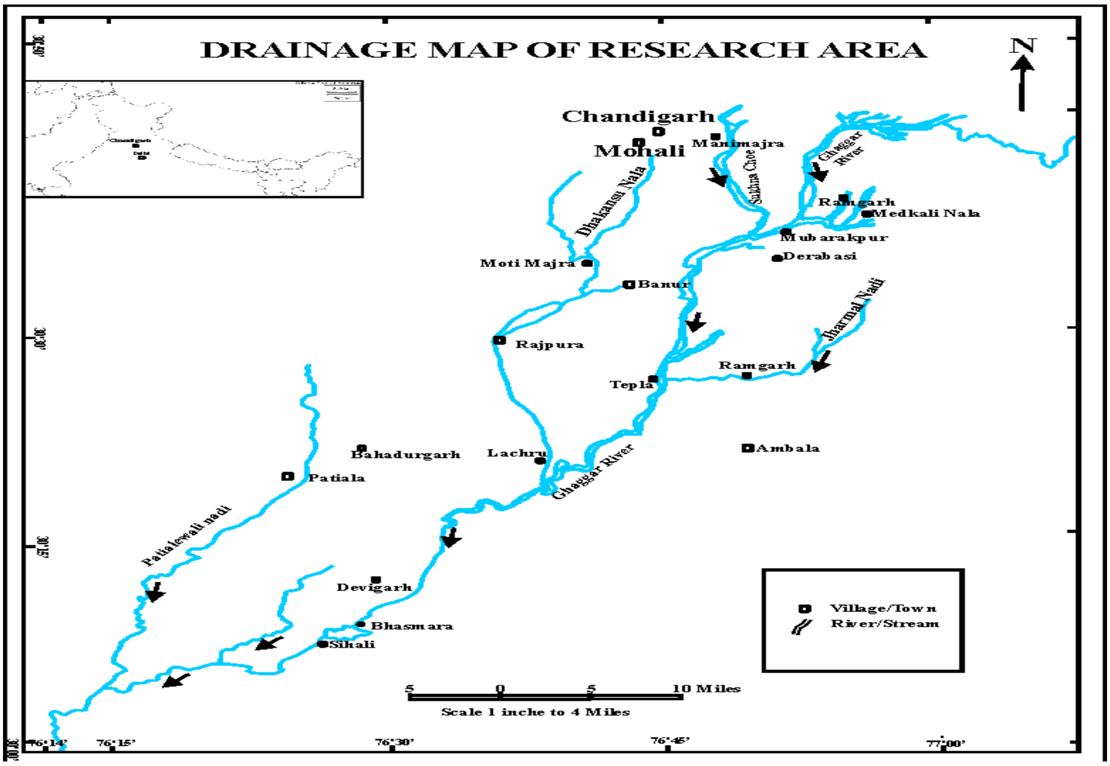

Figure 1: Selected study area of Ghaggar River System

Multivariate Statistical Methods

The multivariate analyses of the river water quality data sets were performed through cluster analysis (CA), principal component analysis (PCA)/factor analysis (FA) and descriptive statistics. In our study, cluster analysis (CA) was performed on the normalized data sets by means of the Ward's method, using squared Euclidian distances as a measure of similarity. The spatial variability of water quality in the whole river basin was determined by cluster analysis using the linkage distance, which represented the quotient between the linkage distances for a particular case divided by the maximal linkage distance. In other words, the relationships among the sampling stations were obtained through CA using Word's method (linkage between groups) and were amalgamated into dendrogram plot as shown in Figure 2. Principal component analysis **(**PCA)/Factor analysis (FA) was employed to investigate the compositional pattern between the examined water quality parameters at the particular monitoring sites. All mathematical and statistical computations were carried out using Microsoft Office Excel 2003 and SPSS 11.5.

RESULTS AND DISCUSSION

Site Similarity and Dissimilarity

The results of CA are shown in Figure 2. The physico-chemical parameters were used as variables to show the spatial heterogeneity among the sites as a result of sequence in their relationship and the degree of contamination. All the 31 sampling sites of the river systems were grouped into five statistically significant clusters: Cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5. The number of stations clustered to form a group is in the sequence of I > II > III > IV > V. From this, it was cleared that clusters I and II were highly significant and variables viz., EC, TDS, and Cl^- had more concentrations in these clusters. These different clusters of sampling sites pointed out that each cluster had a water quality of its own which was dissimilar from the other clusters. Hence, CA results revealed that this approach is useful in offering reliable classifications of surface waters in the selected regions and optimizing the design of a future spatial sampling strategy. Thus, it can be said that for quick spatial assessments of water quality, one site sampled in each cluster is sufficient to determine the water quality of the entire network.

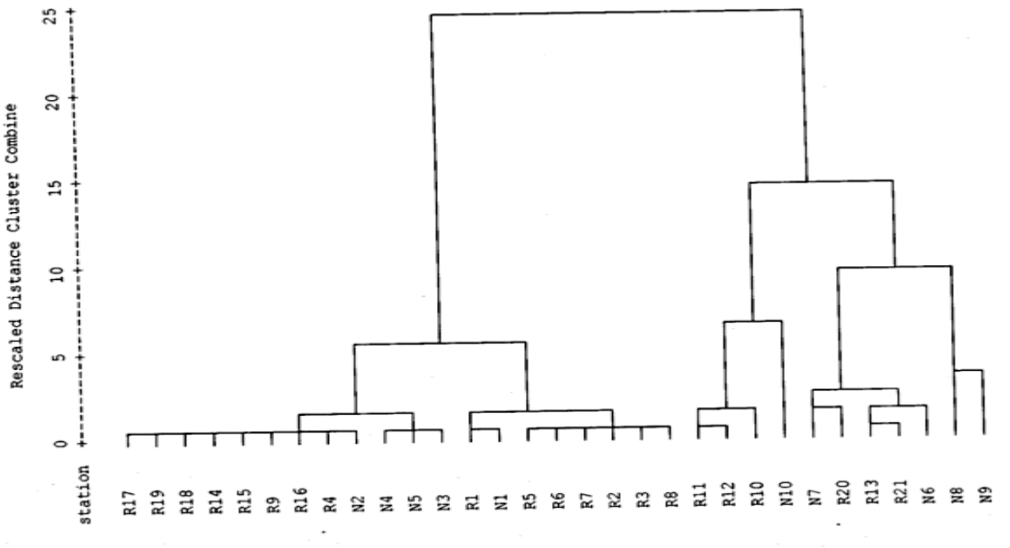

Figure 2: dendrogram showing relationship among the stations

Table 1: Basic situation of the 31 monitoring sites

Code	Station	Source
R1	Badisher-Koti	Ghaggar
R2	Bijdoli-Ki-Doli	Ghaggar
R3	Thapali-Narda	Ghaggar
R4	Burjkotian	Ghaggar
N1	Kambali	Kaushalya Nadi
N2	Kalka Dobighat	Jhajra Nadi
N3	Surajpur	Jhajra Nadi
N4	Surajpur	Kaushalya Nadi
N5	Amravati	Jhajra + Kaushalya
R5	Chandimandir	Jhajra + Kaushalya + Ghaggar
R6	Panchkula S-3	Ghaggar
R7	Dafarpur	Ghaggar
N6	Mubarkpur	Sukhna Choe
R8	Mubarkpur-Camp	Ghaggar
R9	Bhankarpur	Ghaggar
N7	Ibrahimpur	Dhabi Nallah
N8	Toana	Jharmal Nadi
R10	Tepla	Ghaggar
R11	Devinagar	Ghaggar
R12	Nanheri	Ghaggar
R13	Utsar	Ghaggar
N9	Surala	Dhakansu Nallah
R14	Surala-D/S	Ghaggar
R15	Maru	Ghaggar
R16	Devigarh-D/S	Ghaggar
R17	Mohamdpur	Ghaggar
R18	Tatiana	Ghaggar
R19	Rattakhera	Ghaggar
N10	Ratanheri	Patiala Nadi
R20	Ratanheri-D/S	Ghaggar
R21	Bhadshapur	Ghaggar

Table 2: Factor loading values and explained variance of parameters

Parameter	PC	
	Factor 1	Factor 2
Temp	.231	**.695**
pH	-.418	**.732**
EC	**.762**	.370
TDS	**.767**	.351
CO_3^{2-}	.260	-.467
HCO_3	**.682**	.054
Cl^-	**.799**	.255
SO_4^{2-}	.175	**.717**
PO_4^{3-}	.271	**.738**
F^-	**.523**	**.750**
Na^+	**.650**	.431
K^+	.385	**.724**
Ca^{2+}	**.624**	.188
Mg^{2+}	**.586**	-.353
TH	**.744**	-.155
Variance (%)	40.54	18.69
Cumulative (%)	40.54	59.24

Data structure determination and source identification

The PCA and FA were performed on the normalized data sets (15 variables) (Table 2). The value gives a measure of the significance of the factor and factor with highest Eigen value is the most significant. Eigen value of 1.0 or greater is considered significant and is kept back in table and rest less than 1.0 is omitted. Factor loadings were categorized as strong, moderate and week corresponding to the absolute loading values of >0.75, 0.75-0.50, 0.50-0.30, respectively (6). Two principal components (PC) were obtained with Eigen values >1 summing more than 59% of the total variance in the water data sets, which is adequate to give good prior information regarding data structure. The first component (PC1), accounting for 40.5% of the total variance with strong positive loading on EC, TDS and Cl^- and moderated positive loading on HCO_3^-, F^-, Na^+, Ca^{2+}, Ma^{2+} and TH. This factor contains variables that are associated to natural and anthropogenic sources. The second component (PC2) that accounted for 18.69% of the total variance represented the inorganic parameters, which included pH, temperature, PO_4^{3-}, SO_4^{2-}, F^- and K^+. SO_4^{2-}, PO_4^{3-} and K^+ pollution could be associated with anthropogenic pollution sources. These anthropogenic activities may associate with agricultural based product industry and municipal wastes. Therefore, as the inorganic factors (PC1 & PC2) have the largest proportion of the total variance, then it can be concluded that anthropogenic and natural sources were the major causes of river water pollution.

CONCLUSIONS

In conclusions, CA grouped the sampling stations into five clusters of similar water quality characteristics. A group clustered with a large number of stations construed the spatial similarity in their physico-chemical composition amongst them. PCA and FA revealed the major pollution contributor into the river basin using factor loadings values. PCA and FA have helped to identify that the parameters responsible for water quality parameter were mainly related to inorganic and physico-chemical. Overall monitoring stations receive various types of anthropogenic and natural pollutants. Some urgent actions for river water conservation must be taken to protect the living organisms in aquatic system.

ACKNOWLEDGEMENTS

Author is grateful to University Grant Commission (UGC), India for providing financial grant. The author is also thankful to Chairman, Department of Geology (CAS), Panjab University, Chandigarh, for providing laboratory facilities.

REFERENCES

1. Parashar, C., Verma, N., Dixit, S. and Shrivastava, R. (2008). Multivariate analysis of drinking water quality parameters in Bhopal, India. *Envir. Monit. Ass.* **140**: 119-122.
2. Yang, Li, Linyu, XU and Shun, Li. (2009). Water quality analysis of the Songhua River Basin using multivariate techniques. *J. Wat. Res. Prot.* **2**: 110-121.
3. Zali, M.A., Retnam, A. and Juahir, H. (2011). Spatial characterization of water quality using principal component analysis approach at Juru River Basin, Malaysia. *World App. Sci. J.* **14**: 55-59.
4. Lomniczi, I., Boemo, A. and Musso, H. (2007). Location and characterization of pollution sites by principal component analysis of trace contaminants in a slightly polluted seasonal river: A case study of the Arenales River (Salta, Argentina). *Water SA*. **33**(4): 479-485.
5. Juahir, H., Zain, S.M., Yusoff, M.K., Hanidza, T.I.T, Samah, M.A.A., Toriman, M.E. and Mokhtar. (2010). Spatial water quality assessment of Langat River Basin (Malaysia) using envirometric techniques. *Env. Monit. Ass.* **173**(1-4): 625-641.
6. Liu, C.W., Lin, K.H. and Kuo, Y.M. (2003a). Application of factor analysis in assessment ogf groundwater quality in a Blackfoot disease area in Taiwan. *Sci. Tot. Env*. **313**(1-3): 77-89.