Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Advances in Applied Science Research, 2011, 2 (5):84-91

Pelagia Research Library

ISSN: 0976-8610
CODEN (USA): AASRFC

Assessment of Water Quality of Ganga River in Kanpur by Using Principal Components Analysis

Sukarma Thareja*, Siddhartha Choudhuryand Priyanka Trivedi***

**Department of Chemistry, Ch. Ch. College, CSJM Kanpur University, Kanpur, UP, India*
***Department Of Physics, IIT-K, UP, India*

ABSTRACT

Water samples were collected from Jalsansthan Benajhawar Kanpur sampling station on the Ganga River with in Kanpur city in the year 2008(April) -2009(March) and analyzed for 14 water quality variables (physico- chemical) parameters, The data obtained were standardized and subjected to principal components analysis (PCA) to define the parameters responsible for the main variability in water quality variance for Ganga River with in Kanpur city. The PCA produced two significant main components and explain more than 99.316% of the variance Namely, anthropogenic effect and industrial effect; that represent 64.470% and 34.846% respectively of the total variance of water quality in Ganga River. Results reveal that Total dissolved Solids,Total Alkalinity,Total Hardness were the parameters that are most important in assessing variations of water quality in October, November, December, January, February, March, April(post monsoon season) in the river. Results also reveal Turbidity, Suspended Solid were the parameters that are most important in assessing variations of water quality in June, July , August and September in the river(monsoon season). This study suggests that PCA technique is useful tool for identification of important river water quality monitoring months and parameters. Ca^{+2}, Cl^-,SO_4^{-2},*Temperature, Flouride, pH, Fe, Oxygen Consumption(O.C),* Cl^- , Mg^{+2} *are found to be non principal water quality parameters.*

Key Words: Variance, Principal Component Analysis, non principal parameters, Physico-Chemical parameters, anthropogenic effect.

INTRODUCTION

Kanpur is an industrial town with many tanneries and other polluting mills,is alone responsible for about 20% of the total water pollution of Ganga. Water quality monitoring has one of the highest priorities in environmental protection policy [1]. The main objective is to control and minimise the incidence of pollutant-oriented problems, and to provide water of appropriate quality to serve various purposes such as drinking water supply, irrigation water, etc. The quality of water is identified in terms of its physico-chemical parameters [2,3,4,5,6]. The particular problem in the case of water quality monitoring is the complexity associated with analysing the large number of measured variables [7]. The data sets contain rich information about the behaviour of the water resources. The classification, modelling and interpretation of monitoring data are the most important steps in the assessment of water quality. Surface water, groundwater quality assessment and environmental research employing multi-component techniques are well described in the literature [8]. Multivariate statistical approaches allow deriving hidden information from the data set about the possible influences of the environment on water quality [9,10]. Principal component analysis(PCA) is the method that provides a unique solution, so that the original data can be reconstructed from the results. Principal Components(PCs) actually takes the cloud of data points and rotates it such a way that maximum variability is visible. In other words, it identifies the most important gradients. In recent years many studies have been done using principal components analysis in the interpretation of water quality parameters, Lohani [11] utilized principal components technique to provide a quick analytical method for the water quality of Chao Phraya river in Thailand. Shihab,[12] also used this technique in order to describe the variation in water quality in Saddam dam reservoir .Principal component analysis has been successfully applied to sort out hydrogeological and hydrogeochemical processes from commonly collected ground water quality data [13, 14,15, 8,16].To interpret and describe the variation in water quality of Ganga River with in Kanpur city, principal components analysis (PCA) is being used in this study

MATERIALS AND METHODS

Table 1: Descriptive statistics of the data

Variables	Months	Min	Max	Mean	Std. Dev	Variance
Temp	12	16	30	24.5833	5.5834	31.174
pH	12	7.9	8.9	8.4625	0.2327	0.5415
Tur	12	15	825	223.75	296.72	88041.48
T.A	12	108	260	198.83	53.35	2846.152
T.H	12	106	246	182	49.923	2492.364
Ca	12	28	80	44.667	14.67	215.3697
Mg	12	6.318	26.244	17.089	7.48	55.98
Cl	12	7	26	17.25	8.28	68.5682
TDS	12	255	540	407	114.54	13119.45
Fe	12	0.2	0.8	0.616	0.2208	0.04879
O.C	12	2.4	7.8	4.766	1.494	2.2315
SO_4	12	47	126	69.833	23.15	535.969
S.S	12	50	340	138.33	98.79	9760.6
Fluoride	12	0	0.8	0.2916	0.22747	0.051742

All water samples were analyzed following standard method specified by the American Public Health Association[17]. Descriptive statistics of the data set are presented in (Table 1).

Table 2: Showing components, percentage of variance and cumulative %

Components monitoring months as variables	% of Variance	Cumulative%
1	64.470	64.470
2	34.846	99.316

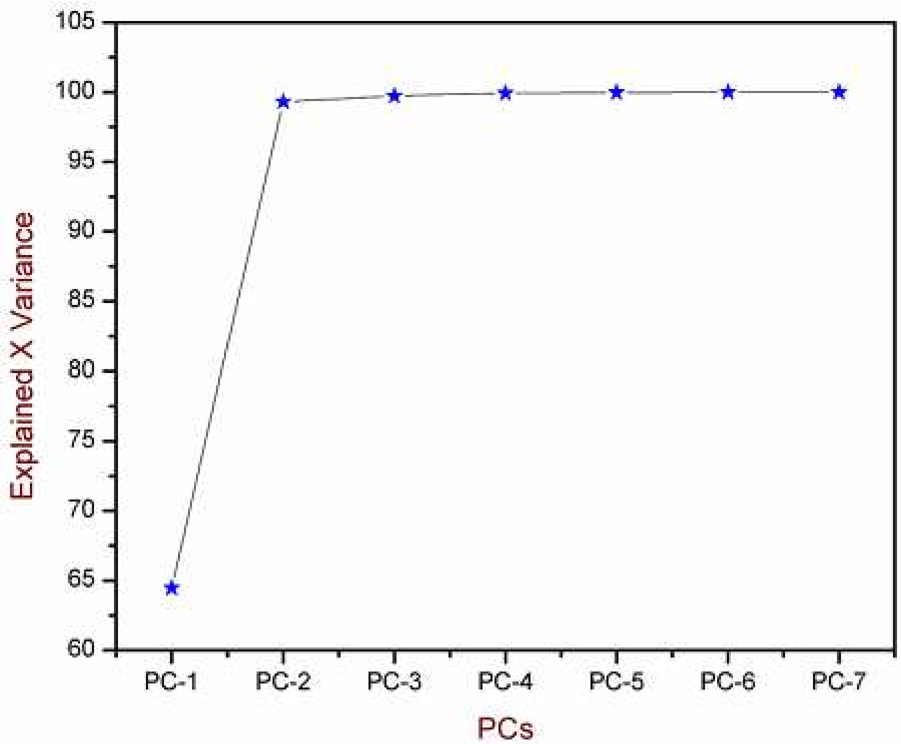

Figure 1: Scree plot showing X-Variance vs PCs

Data Processing

The 168 parameters obtained from the laboratory analysis were used as variable inputs for Principal Components Analysis (PCA). PCA was performed using the SPSS package. Prior to the analysis, the data were standardized to produce a normally distribution of all variables. Since water quality parameters had different magnitudes and scales of measurements, which if not taken into account would have given more weight to certain variables due to their respective variance [18].This PCA technique aims to transform the observed variables to a new set of variables. Principal components(PCs) which are uncorrelated and arranged in decreasing order of importance to simplify the problem. In PCA, data cluster is rotated by subtracting the mean of the data and dividing by the standard deviation. So the centroid of whole data set is zero and relative location of all the points remains the same. This type of ordination reduces the dimensionality of the data set and minimizes the loss of information caused by reduction.From the standardized covariance or correlation matrix of the data, the eigen values and corresponding eigen vectors of covariance matrix were calculated. Then a number of PCs were selected from the initial PC according to their eigen values and scree diagram. The eigen value associated with each principal component tells us how much variation in the data set it explains. They are usually expressed as a percentage of the total variation in the data set. Table 2 represents the determined initial PC and its eigenvalues and percent of variance contributed in each PC. Figure. 1 shows the scree plot of X- Variance (also represents eigenvalue) for each component. Figure 1. a scree plot also showed that the first two PCs are the most significant components which represent more than 99.316% of the variance in water quality of Ganga River. 64.470% by PC1, 34.846% by PC2; showing that first two eigen values are significant as compared to others.

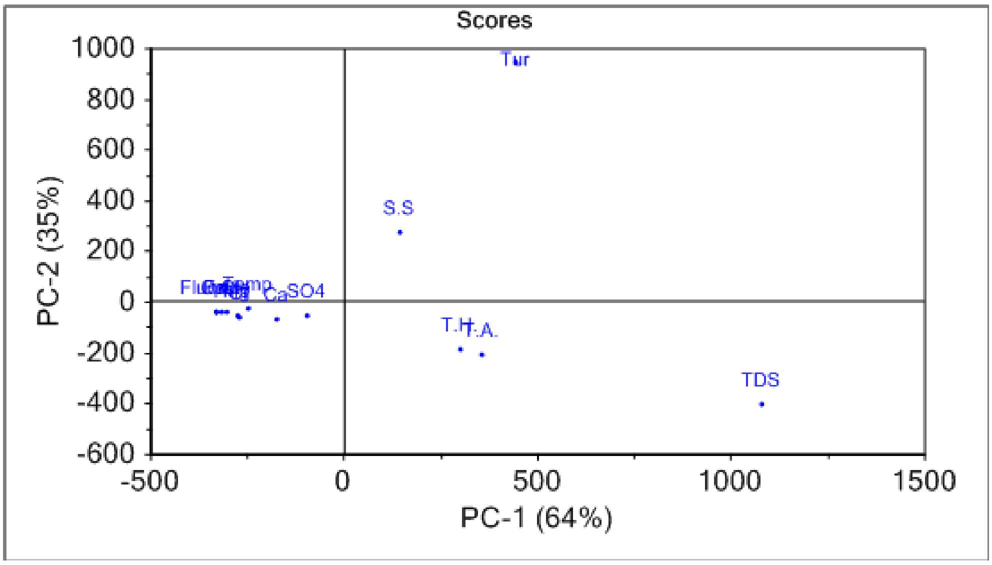

Figure 2a :Score Plot PC1 vs PC2

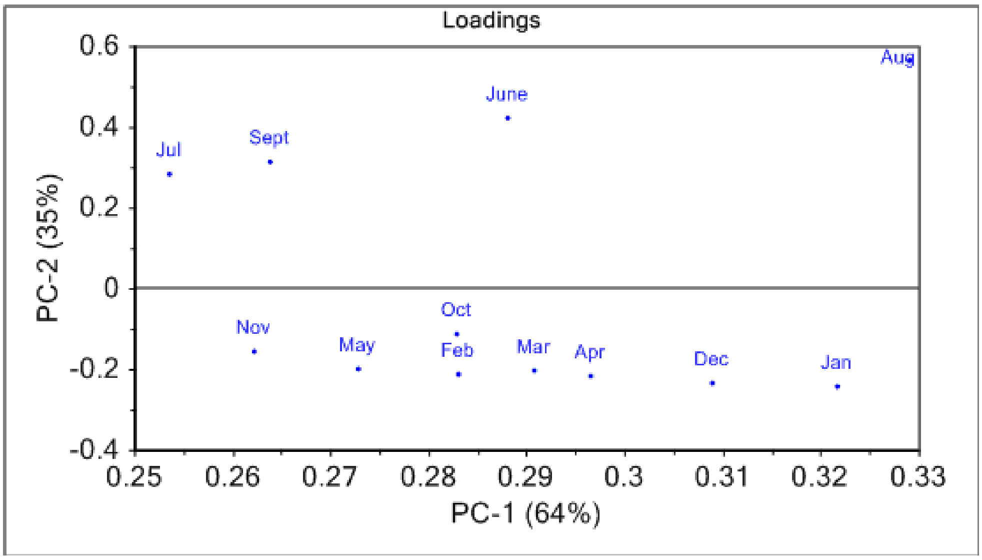

Figure 2 b : Loading plot PC1 vs PC2

RESULTS AND DISCUSSION

The principal components, PC-1 and PC-2 contribute about 99% of the total variance in the data. So these two comonents are sufficient to interpret our analysis. As Fig ure 2a. shows 64% variance along PC-1 axis where TDS lies at the right most postion. It natually tells that TDS contributes for the maximum variance along PC-1 axis. While relating it with the loadings plot(Figure. 2b) we can see that the response of TDS in water is strong for the months of November(Nov), October(Oct), May, February(Feb), March, April, December(Dec) and January(Jan)(Post monsoon season) as these months near to PC-1 axis where as the weak coorelation for the months of July, Septmber(Sep),June and August(Aug)(pre monsoon season) is due their relative position from PC-1 axis is very far resulting in their weak effective contribution on this parameter. As Turb is very far from PC-1 axis and is at maximum along PC-2 axis, it's contribution is significant only in the months of July, Sep, June and Aug . It's correlation with the months of Nov, Oct, May, Feb, Mar, Apr, Dec and Jan is very negligible as seen from the loadings plot.

Positon near to centroid of PC axes for TH and TA shows a significant contribution of these parameters for the months of Nov, Oct, May, Feb, March, April, Dec and Jan as shown by loadings plot. The parameter SS has a considerable percentage of variance along PC-2 axis resulting its effective contribtution in water durings the months of July, Sep , June and Aug. The position of Ca^{+2},Mg^{+2}, SO_4^{-2}, Temperature, flouride, pH, Fe, O.C, Cl^- towards the left most corner along PC-1 axis representing their least contribution with season change during months , and their position near to zero along PC-2 axis shows almost their no contribution at all.It is interesting to note that scores plots for physico- chemical parameters are distributed in particular quadrant of graph of PC1 vs PC2,which depend both on the water quality of the Ganga River and monitoring months. This remark could be taken into account for good management of water bodies.

Table 3: Components score covariance matrix

Component	1	2
1	1	0
2	0	1

Principal Components Interpretations

Components loading (correlation coefficients), Which measure the degree of closeness between the variables and the PC, the largest loading either positive or negative, suggest the meaning of the dimensions; positive loading indicates that the contribution of the variables increases with the increasing loading in dimension; and negative loading indicates a decrease. Component score covariance matrix shown in Table 3 shows that there are no correlation between components, this is, each components represent an discrete unit from others.

Principal Component one; has a high loading of Oct, Nov, March, May, April, Dec, Feb and Jan explain 64.470% of the total variance (Table 2) This component can be ascribed as post Monsoon season effect on water quality of Ganga River. Principal Components two; explain 34.846% of the total variance(Table 2), has high loading of June, July, Aug and Sep and this component reflect the monsoon effect on Ganga River with in Kanpur city.PCA study on physico- chemical parameters found that TDS,TH and TA are significant parameters in post monsoon season Therefore monitoring of these three parameters may be pointed during post monsoon season responsible for main variability in water quality. Turbidity and Suspended Solids are significant parameters in monsoon season as they are responsible for the main variability in water quality in monsoon season, therefore monitoring may be pointed for Turbidity and Suspended Solids for monsoon season. The frequency of sampling for insignificant parameters(pH,Ca^{+2},Mg^{+2},Cl^{-1},Fe,O.C,SO_4,F^{-1},Temp) may be reduced; or special studies may be conducted on them when needed. Monitoring may be pointed for Turbidity and Suspended Solids for monsoon season responsible for variability in water quality.TA,TH and TDS contribute to the construction of component one . The greater alkalinity values may be due to large scale use of river bank as open latrine and consequent washing of excreta in and near by the water body[19]. Variation in hardness of river water is probably due to regular addition of sewage detergents due to huge human activities called Anthropogenic Factor. Anthropogenic

Factor is identified as responsible for explaining physical nature of component one which explains 64.47% of the total variance.Turb and SS contribute to the construction of components Two explained 34.846% of total variance. Urbanized areas contribute large amounts of turbidity to nearby waters, through storm water pollution from paved surfaces such as roads, bridges and parking lots.

Certain industries such as leather tanneries near bank of Ganga has generated very high levels of turbidity from colloidal particles from its waste. Turbidity is commonly linked to total suspended solids (SS) because water with high SS levels typically look murkier and have higher turbidity measurements. Common suspended solids are clay, silt, and sand from soils, bits of decaying vegetation, industrial wastes and sewage. Therefore the Industrial factors is identified as responsible for explaining physical nature of component two.

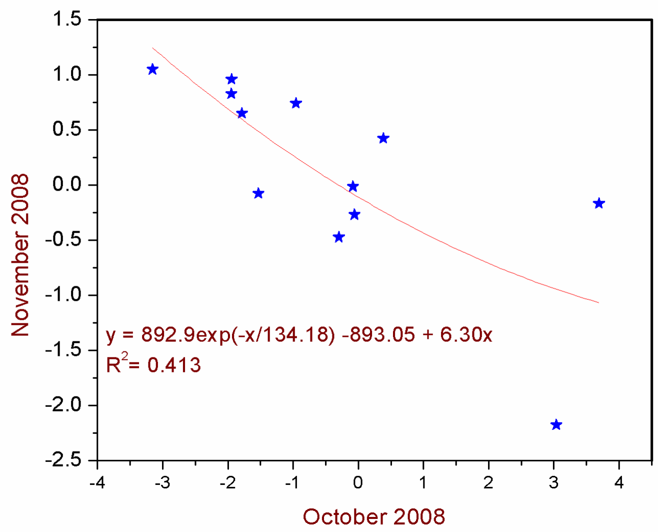

Figure. 3 Relationship between October 2008 and November for data of all the physico-chemical parameters

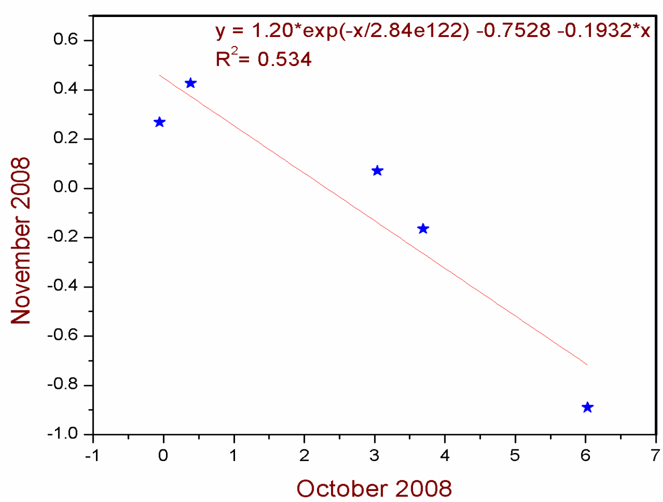

Figure. 4 Relatioship between October 2008 and November for 5 principal parameters

Validation of PCA Results

Before applying the above finding, its scientific reliability must be validated using other independent methods. One way to achieve this goal is to compare the water quality data with 9 non principal physico –chemical parameters .In this study, we developed the comparison in between two cases.In the first case, data from the principal physico- chemical parameters were used to formulate relationship by regression analysis for month of October 2008 and November. Comparison of the relationship for October 2008-November, using data of all 14 physico-chemical parameters(Figure. 3) with that obtained using data of rest of five physico-chemical(principal physico-chemical parameters) (Figure. 4) showed that addition of non principal parameters did not improve the curve fitting between October (X) and November2008 (Y), as indicated by correlation coefficient(R^2). The R^2 value for regression equation(y = 892.9 exp(-x/134.18)+6.30x – 893.05)for data of all 14 physico- chemical parameters was 0.413, whereas the R^2 value for the regression equation(y =1.20exp(x/2.84e122)+0.1932x-0.7528) for data of 5 principal parameters was 0.534.The latter is better than former.Therefore the 9 physico-chemical are considered non-principal, since the addition of data of these 9 physico-chemical parameters did not improve the curve-fitting.

CONCLUSION

1. It is interesting to note that scores plots for physico - chemical parameters and months are distributed in particular quadrant of graph of PC1 vs PC2 which depends on the water quality of the Ganga River. This remark could be taken into account for good management of water bodies.
2. From the 12 components the first two components are sufficient to explain the monitoring area as well physico-chemical water quality parameters. These components explain more than 99.0 of the total variance of the original data set in Ganga River.
3. PCA result also show that TA,TH and TDS are significant parameters in post monsoon season while Turbidity and SS are significant parameters in monsoon season. For this reason, it is worthwhile to stress that these significant parameters should be observed more systematically.
4. Thus two factors were identified as responsible for the data structure, explaining more than 99% of total variance in the dataset. The first factor called the anthropogenic factor explained 64.470% of the total variance. The second factor named the industrial factor explained 34.846% respectively.
5. The outcome shows there is a potential for improving the efficiency and economy of the monitoring network in Ganga river by reducing the number of physico- chemical monitoring parameters of water quality from 14 to 5. This reduction may result in significant cost saving in monitoring program without sacrificing important river water quality
6. PCA results show that 9 physico-chemical parameters(pH,Ca^{+2},Mg^{+2},Cl^{-1},Fe,O.C,SO_4^{-2},F^{-1},Temp)identified as less important in explaining the annual variance of data set, and therefore could be non-principal parameters. Identification of less important water quality parameters can be seen in fig.2a , which show component loading for PC1 and PC2. Observation of these insignificant variables may not be important for this analysis, may be eliminated from monitoring program.
7. However it should be noted that only one year annual mean values of water quality parameters were used in this study. Prior to making any critical decision in eliminating water quality physico- chemical parameters in Ganga river in Kanpur, the PCA with longer time scale(ie more than 3 years) should be performed.
8. To create increasing awareness among people that to maintain the Kanpur Ganga river water to its highest quality and purity ,the present study may prove to be useful in achieving this goal.

Acknowledgment
ST is grateful to Dr. Pervez E. Deen, Principal Christ Church College, Kanpur for his encouragement throughout the work.

REFERENCES

[1] V. Simeonov, J.W. Einax, I. Stanimirova, , & J.Kraft, 2002. *Analytical and Bioanalytical Chemistry*,**2002**,374,898.
[2] A. Sargaonkar, &V. Deshpande, .*Environmental Monitoring and Assessment*, **2003**, 89,43.
[3] P. Pavendan et al Pelagia Research Library *Eur. J. Exp. Bio*.,2011,1(1):183-189
[4] Shashikant R. K uchekat et al., 2011. *Der chemica Sinica*, **2011**,2(4),229
[5] Yadav S.S and Kumar Rajesh, 2011. *Adv. Appl. Sci. Res.* 2011,2(2):197
[6] Mane P.C et al. *Der chemica Sinica*, **2010**,*1(3),53*
[7] K. Saffran, **2001**. Canadian water quality guidelines for the protection of aquatic life, CCME water quality Index . Users manual. Excerpt from Publication No.1299:ISBN1-896997-34-1.
[8] Praus, *Water SA*.,**2005** Vol**.** 31, No. 4, October: 417-422.
[9] Spanos et al.,. *Theoretical and Applied Climatology*,**2003**, 75, 167.
[10] Animesh Agarwal et al. *Adv. Appl. Sci. Res.* **2011**,2(2),185
[11] B.N Lohani, and G. Todino. *Journal of Environmental Engineering* **984**,110(6),1163.
[12] A.S.Shihab, **1993**. Application of Multivariate Method in the Interpretation of water Quality Monitoring Data of Saddam Dam Reservoir, Confidential 13.
[13] R.Jayakumar, and L.Siraz, *Environmental Geology***, 1997,**31(3/4)June,174.
[14] Jothivenkatachalam K. et al, *Der chemica Sinica*, **2011**,2(2),199
[15] S.R. Salman, and Y.H AbuRuka'h, *Environmental Geology*, **1998**,38(3) September ,265
[16] S.B. Olobaniyi, and F.B.Owoyemi, F.B., 2006., *AJST*, **2006**,Vol. **7** (1), 73.
[17] APHA, **1985** APHA,AWWA and WPCF, Standard methods for the examination of water and wastewater,16th Ed.:1268.
[18] Davis, Statistics and data analysis in geology, John Wiley and Sons, inc.,Newyork,**1973**.
[19] Narasimha Rao and Jaya Raju, 2001. *J.aqua. Biol*. **2001**.16(1 and 2),11