Assessment of Water Quality of Bennithora River in Karnataka through Multivariate Analysis

Sukarma Thareja[1*], Priyanka Trivedi[1]
[1]Department of Chemistry, Christ Church College, CSJM Kanpur University, UP, India
*E-mail address: sukarmathareja@hotmail.com

Abstract: The evaluation of river water quality is a critical element in the assessment of water resources. The objective was to identify physico-chemical parameters that are less important in assessing annual variations of river water quality. Eight physico-chemical parameters were used for monitoring river water quality from June (2005) to May (2006) for Bennithora River (Krishna Basin) near Gulbarga city of Karnataka state in India and they were selected for the purpose of the study. Significant variations among the parameters and interesting correlations were observed throughout the period of study. Multivariate technique, Principal component analysis (PCA) was applied to evaluate the annual correlation of water quality parameters. Results show that 5 physico-chemical parameters are identified as less important in explaining the annual variance of the data set, and therefore could be the non-principal parameters (Water temperature, pH, total alkalinity, Cl_2, NH_4–N). This study suggests that PCA technique is useful tool for identification of non-principal water quality physico-chemical parameters. The outcome of this study also shows that there is a potential for improving the efficiency and economy of the monitoring network in the Bennithora River by reducing the number of physico-chemical parameters from 8 to 3. This reduction may result in significant cost saving without sacrificing important water quality data. [Nature and Science 2010;8(6):51-56]. (ISSN: 1545-0740).

Keywords: pollution, Monitoring River water quality, Principal Component analysis

1. Introduction

Water is a common chemical substance that is essential for the survival of all known forms of life. The major proportion of all water quality degradation worldwide is due to anthropogenic causes[1]. Our country India is a developing democratic country. To maintain the democratic system the foremost points that we must look into to preserve our resources e.g. water resources. Water resources are the main areas which are very closely associated with the daily life of masses. Impact of the science and technology has made our water resources polluted. The municipal and industrial wastewater discharge constitutes the constant polluting source, on the concentration of pollutants in river water, example Gomti River [2] at Lucknow, U.P in India.

As a responsible citizen and scientist of India we must make sure that our water resources remain minimum polluted. So there is a need to develop a systematic program to clean the above set pollution present almost in every river water. It is imperative to prevent and control the rivers pollution and to have reliable information on the quality of water for effective management. In view of the variations in the hydrochemistry of rivers, regular monitoring programs are required for reliable estimates of the water quality[3a,3b].The first step is to collect the data base of polluted water in terms of physico- chemical parameters and then analyze the data so that with minimum money we could clean the polluted river water.

We are presenting a case study of Bennithora River [4] near Gulbarga city Karnataka state, in India in which the physico-chemical parameters like: Atmospheric temperature, water temperature, pH, DO (dissolved oxygen), TA (total alkalinity), Cl_2, PO_4-P, NH_4-N were analyzed by principal component analysis (PCA).

2. Materials and Method:

2.1 Study Area:

The Bennithora dam is on Bennithora River (Krishna Basin) near Gulbarga city (76^0-04' to77^0- 42' N and 16^0-12' to 17^0 - 46' E). This reservoir spreads over area of 45sq. miles. The maximum depth is 25 feet during monsoon and minimum is 15 feet during the dry period.

2.2 Data:

In this study, eight 8 physico-chemical parameters obtained from Bennithora River[5] were used for analysis. The investigations of physico-chemical parameters were carried out during June 2005 to May 2006. The water samples were collected on monthly basis during 9 AM to 11 AM and brought to laboratory for the further analysis and analyzed for physico-chemical parameters, following the standard method (APHA, 1985)[6]. All the samples were analyzed for Atmospheric Temperature (Atm. Temp) (^{0}C) Water Temperature (Water Temp), pH, DO (mg/l) (dissolved oxygen), total alkalinity (TA) (mg/l), Cl_2 (mg/l) PO_4-P (mg/l) NH_4-N (mg/l).

PCA was performed in this study to identify the potential for reducing the number of physico chemical parameters of river water. Eight parameters water quality parameters from 12 months were examined in this study. The procedures used for PCA is described below.

2.3 Principal component analysis (PCA):

The PCA is performed on MATLAB 7.0.1. Software, version 7, The Math Works, Inc. Software.

Mathematically PCA normally involve the following five major steps [7]:

1. Coding the variables x_1, x_2, -----, x_n (which are 8 physico- parameters in our present study) standardization of the measurements to ensure that all have equal weights in analysis 2. Calculate the covariance matrix C, 3. Calculate the Eigen values and corresponding Eigen vectors of the covariance matrix, 4. Discarding any component that account for small proportions of variation in data set, 5. Choosing components and forming a feature vector.

3. Results and discussion:

3.1. Correlation of water quality parameters

Data in Table 1 provide the correlation matrix of the water quality parameters obtained from Vijaykumar et al[5]

Table 1: Correlation Matrix

	Atm. Temp	Water Temp	pH	DO	TA	Cl_2	PO_4-P	NH_4-N
Atm. T	1.0000							
W.T	0.7797	1.0000						
pH,	-0.3330	-0.3537	1.0000					
DO	0.3509	0.2286	-0.4156	1.0000				
TA	0.0124	-0.1972	0.0217	0.2277	1.0000			
Cl_2,	-0.4423	-0.2204	0.3119	-0.6644	0.0992	1.0000		
PO_4-P	-0.3758	-0.4543	0.4795	-0.2516	0.5731	0.3001	1.0000	
NH_4-N	-0.0585	0.1371	0.1006	0.1216	0.4449	0.2781	0.0847	1.0000

In general, the river water temperature has relatively weak to fair correlations, i.e., most of the correlation coefficients are less than 0.7 (absolute value) with other parameters for the entire months.

The correlation coefficients between Atm. Temp and other parameters were less than or equal to 0.55 except with Water Temp (0.7797). There is a negative correlation between Atm. Temp with pH,Cl_2, PO_4 - P ; Water Temp with pH, TA, Cl_2, PO_4 -P and pH with DO respectively. This is shown in Table 1. A negative correlation is also shown by DO with Cl_2, PO_4 -P There is hardly any correlation exist between Cl_2 and TA; NH_4-N and PO4 -P.

It is interesting to observe that a high positive correlation (0.7797) exist between Atm. Temp and Water Temp. Different types of correlations are shown in Fig.1, 2 and 3 respectively.

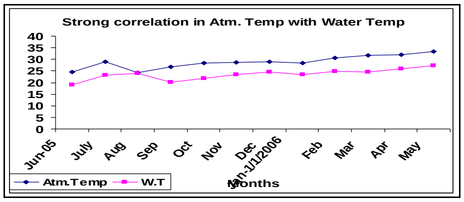


Fig.1- Shows Strong correlation in Atm. Temp with Water Temp

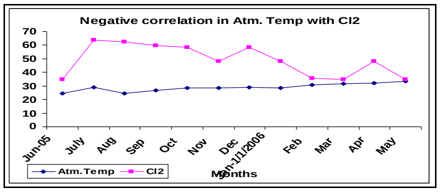


Fig.2- Shows Negative correlation in Atm. Temp with Cl_2

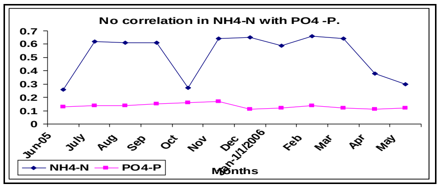


Fig.3- Shows No correlation in NH_4-N with PO4 -P.

Principal component analysis:

In PCA, the number of components is equal to the number of variables. However, a component is not only comprised of a single variable but rather all of the variables used in a study. For example there are eight variables (physico – chemical parameters) used in this study, which produce 8 components. In each component, there are 8 variables as shown in Eq. (1) below. The PCA results showed that of the 8 components, the first component accounted for about 38.06% the second component accounted for about 20.00% the third component accounted for about 15.46% and the fourth component accounted for about 10.59% of the total variance in the data set. These four components together accounted for about 84.11% of the total variance and the rest of the components only accounted for about 15.89 %. Therefore, we focus our discussions only on the first four components.

From the eigenvectors obtained in the PCA, the first component, Z_1 for June 2005 to May 2006 months can be given as follows:

$$\mathbf{Z_1 = +0.4445\, x_1 + 0.4179\, x_2 - 0.3795\, x_3 + 0.3562\, x_4 - 0.1553\, x_5 - 0.3943\, x_6 - 0.4120\, x_7 - 0.0866\, x_8 \ldots\ldots\ldots\ldots (1)}$$

Where x is the monitoring 8 physiochemical parameters, the subscripts denote the parameter numbers, and the coefficients are the eigenvectors. This component had almost different loadings (i.e., different coefficient values in Eq. (1)) on all variables and therefore is a measure of overall performance of the parameters.

It is apparent that Z_1 has an extremely high correlation with the measured data as it accounts for 38.06% of the data variance. This equation shows that the first component, Z_1 will be high if x_1 to x_2 and x_4 are high but x_3, x_5 x_6, x_7, and x_8 are low. Hence, Z_1 represents a difference among parameters. The low coefficients of x variables such as those associated with x_8 mean that the values of these variables have little effect on Z_1

Similarly, the second, third and fourth components for June 2005 to May 2006 months can be given as:

$$\mathbf{Z_2 = +0.1366x_1 + 0.0461x_2 - 0.0585x_3 + 0.3976\, x_4 + 0.6825\, x_5 - 0.0783\, x_6 + 0.3035\, x_7 + 0.5036\, x_8 \ldots\ldots\ldots (2)}$$

$$\mathbf{Z_3 = -0.2356x_1 - 0.5140\, x_2 - 0.0703\, x_3 + 0.3703\, x_4 + 0.0761\, x_5 - 0.5190\, x_6 + 0.1785\, x_7 - 0.4809\, x_8 \ldots\ldots\ldots (3)}$$

$$\mathbf{Z_4 = +0.5169x_1 + 0.2678x_2 + 0.5033x_3 - 0.1872x_4 + 0.0891x_5 - 0.1948\, x_6 + 0.4347x_7 - 0.3715x_8 \ldots\ldots\ldots (4)}$$

This equation (2)) shows that the second Z_2, will be high if x_1, x_2, x_4, x_5, x_7 and x_8 are high but x_3 and x_6 are low. Hence, Z_2 represents a difference among the parameters. The low coefficients of x variables such as those associated with x_2 mean that the values of these variables have little effect on Z_2.

Equation (3)) shows that the third Z_3, will be high if x_4, x_5 and x_7 are high but x_1, x_2, x_3, x_6 and x_8 are low. Hence, Z_3 represents a difference among the parameters. The low coefficients of x variables such as those associated with x_5 mean that the values of these variables have little effect on Z_3.

Equation (4)) shows that the third Z_4, will be high if x_1, x_2, x_3, x_5 and x_7 are high but x_4, x_6 and x_8 are low. Hence, Z_4 represents a difference among the parameters. The low coefficients of x variables such as those associated with x_5 mean that the values of these variables have little effect on Z_4.

A graphical representation of the first four components for loadings is given in Fig.4, 5,6 and 7 These Figures were constructed using the eigenvectors from the first four components. The components measured the difference among the parameters. Fig. 1 (First component) shows that parameter NH_4-N has the lowest (eigenvector) value. Furthermore, the second component shows that three monitoring parameters Water Temp, pH, Cl_2 had the lowest absolute loading (eigenvector) values, which could indicate that parameters Water Temp, pH, Cl_2 are less important in monitoring water quality variations. However, any conclusion based upon Z_2 would be in appropriate since Z_2 only accounted for 20.00% of the total variance.

The third component shows that two monitoring parameters pH and TA have the lowest absolute loading (eigenvector) values, which could indicate that these parameters are less important in monitoring water quality variations. Any conclusion based upon Z_3 would be in appropriate since Z_3 only accounted for 15.46% of the total variance.

The fourth component shows that only one monitoring parameter TA has the lowest absolute loading (eigenvector) value, which could indicate that parameter TA, is less important in monitoring water quality variations. Any conclusion based upon Z_4 only accounted for 10.59% of the total variance. It should be pointed out that a loading reflects only the relative importance of a variable within a component, and does not reflect itself.

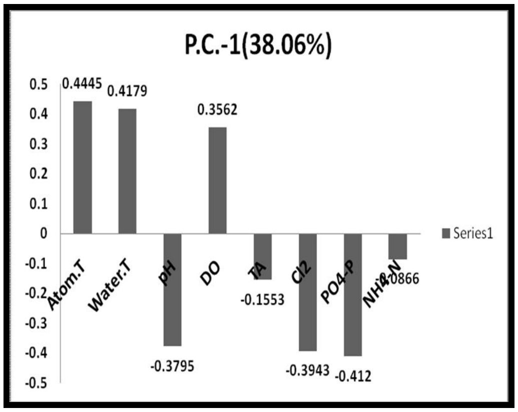


Fig.4 Component loading for the first component(PC1)

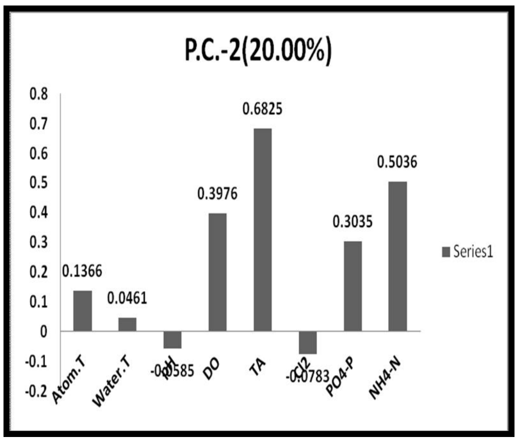


Fig.5 Component loading for the second component (PC2)

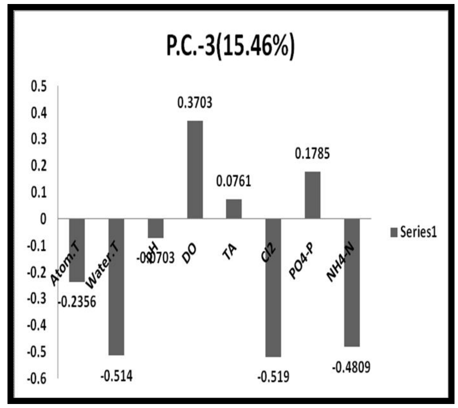


Fig.6 Component loading for the third component (PC3)

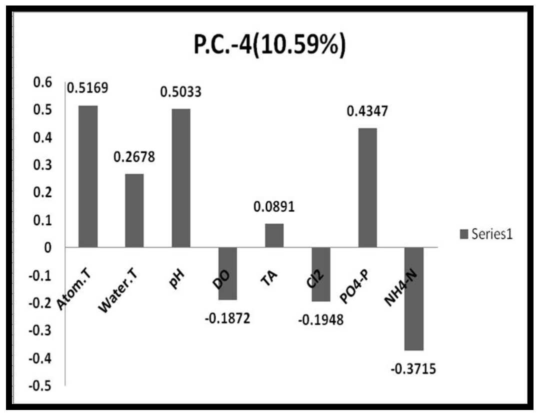
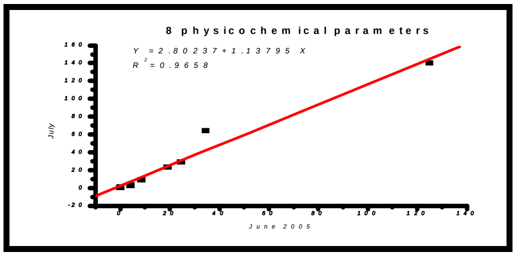


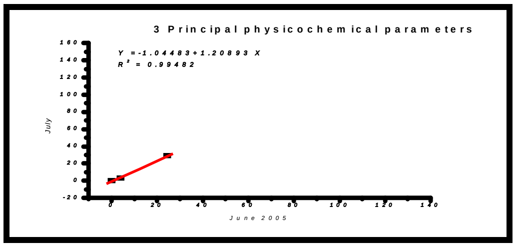
Fig.7 Component loading for the fourth component (PC4)

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 and without the 5 nonprincipal physico-chemical parameters. In this study, we developed the comparisons in between two cases. In the first case, data from the principal physico-chemical parameters were used to formulate the following four relationships by regression analysis: (1) June2005 versus July; (2) August versus September; (3) October versus November; (4) December versus January 2006; In the second case, data from all of the stations (i.e., principal and non-principal stations) were used to formulate the aforementioned four relationships by regression analysis. These two cases were then compared to determine if the addition of data from the five non principal parameters improved the regression relationships. Comparison of the relationship between June 2005 and July obtained using data of all the 8 physicochemical parameters with that obtained using data of rest of three physicochemical parameters (Fig. 8) showed that the addition of the five non-principal parameters did not improve the curve fitting between July (Y) and June2005(X), as indicated by correlation coefficients (i.e., R^2 values).The R^2 value for the regression equation (Y =2.80237+1.13795 X)for data of all the 8 physicochemical parameters was 0.9658, whereas the R^2 value for the regression equation (Y= -1.04483+1.20893 X)for data of the 3 principal parameters was 0.99482. The latter is slightly better than the former. Similar results were obtained for the relationships between August versus September (Fig. 9), October versus November (Fig. 10) December versus January 2006 (Fig. 11). That is, the R^2 values obtained for data of 3 principal parameters were slightly better than those of all the 8 physicochemical parameters. Therefore, the 5 Physico-chemical parameters are considered to be less important parameters since the addition of data of these 5 physico-chemical parameters did not improve the curve-fitting.

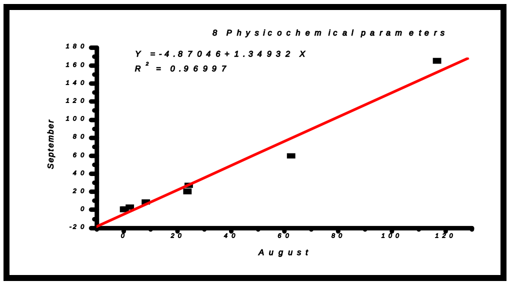


A

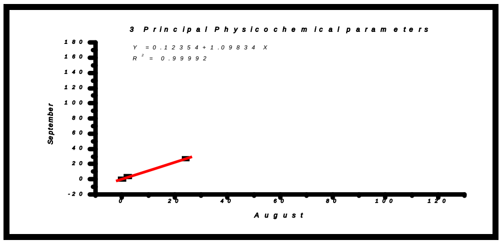


B

Fig. 8 Relationship between June 2005 and July for data of all the parameters (A),
Relationship between June 2005 and July for data of 3 principal parameters (B)

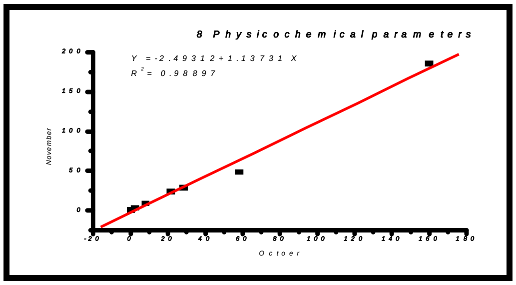


A

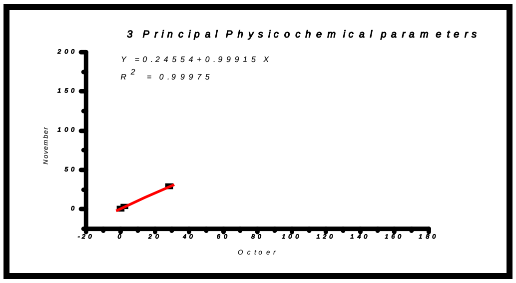


B

Fig. 9 Relationship between August and September for data of all the parameters (A), Relationship between August and September for data of 3 principal parameters (B)

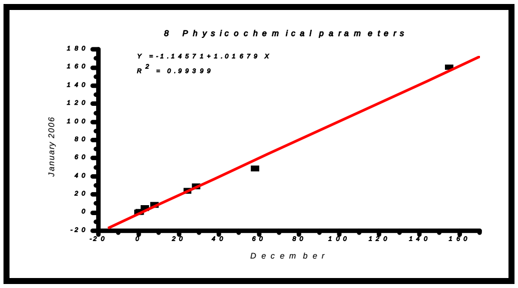


A

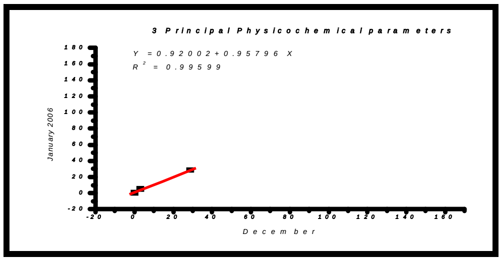


B

Fig. 10 Relationship between October and November for data of all the parameters (A), Relationship between October and November for data of 3 principal parameters (B)



A



B

Fig. 11 Relationship between December and January 2006 for data of all the parameters (A), Relationship between December and January 2006 for data of 3 principal parameters (B)

4. Summary and Conclusions:

1) The outcome showed that there was a potential for improving the efficiency and economy of the monitoring network in the Bennithora River Karnataka by reducing the number of monitoring Parameters from 8 to 3. This reduction may result in significant cost savings without sacrificing important river water quality data.
2) In this study, water quality data for 8 physical and chemical parameters collected were collected on monthly basis along the Bennithora River (Krishna Basin) near Gulbarga city Karnataka in India from months June (2005) to May (2006) were analyzed, using the PCA technique. The correlation matrix of the water quality parameters was obtained from [7].
3) The correlation coefficients between Atm. Temp and other parameters were less than or equal to 0.55 except for Water Temp (0.7797). There is a negative correlation between Atm. Temp.and pH, Cl_2, PO_4 – P; Water Temp and pH, TA, Cl_2, PO_4 – P; pH and DO respectively. A negative correlation is also shown by DO with Cl_2, PO_4 $^-$ P.
4) It is interesting to observe that a high positive correlation (0.7797) exist between Atm Temp and Water Temp and there is hardly any correlation between PO_4-P and NH_4-N; Cl_2 and TA respectively.
5) PCA results show that 5 physico-chemical parameters (Water Temp, pH, TA, Cl_2, NH_4-N). identified as less important in explaining the annual variance of the data set, and therefore could be the non-principal parameters .Identification of less important water quality parameter as can be seen in Figs. 1, 2, 3 and 4 respectively which show component loading(eigenvector) for PC1, PC2, PC3 and PC4 respectively.
6) Results from this analysis could prove valuable in evaluating the potential for reducing the water quality measured parameters
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 Bennithora River Karnataka, the PCA with longer time scale (i.e. more than 3 years) should be performed. Further study i.e. Principal Factor Analysis (PFA) is also needed to identify the principal physical, chemical, and biological parameters that are important in predicting monthly/seasonal variations in surface water quality for the entire Bennithora River monitoring network. The work is in progress in this direction.

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

Correspondence to:
Sukarma Thareja
Department of Chemistry, Christ Church College
CSJM Kanpur University
Kanpur 208001, UP, India
Telephone: +91- 0512- 2598306
Emails:sukarmathareja@hotmail.com;
sukarmathareja@yahoo.com

5. References:

[1]. O.K. Adeyemo O.A. Adedokun, R.K. Yusuf, E.A. Adeleye Global NEST Journal, 10, No 3, 326(2008)

[2]. Kunwar P. Singh, Amrita Malik, Sarita Sinha, Analytica Chimica Acta, 538. 355 (2005)

[3]. a) Y. Ouyang, P. Nkedi-Kizza, Q.T. Wuc, D. Shinde, C.H. Huang water research 40 3800(2006).

[3]. b) Ying Ouyang, Water Research 39, 2621 (2005).

[4]. A.Pappa Preliminary Studies on luminology of gulbarga tank M.Phil dissertation ,Gulbarga University,Gulbarga.(1995)

[5]. K.Vijaykumar,M.Rajashekar, Shashikanth. M And. B. Vasanthkumar. Water Quality Of Bennithora River In Karnataka wgbis.ces.iisc.ernet.in/energy/lake2006/programme/programme/.../Rajasekhar

[6]. APHA, AWWA and WPCF. Standard methods for the examination of water and wastewater, 16 th Ed. 1268(1985).

[7]. I.Lindsay and A. Smith: A tutorial on Principal Components Analysis. http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components

Date of Submission: 14.3.2010