Mysore J. Agric. Sci., 50 (1) : 33-38, 2016

Factors Contributing to the Pollution of Selected Lakes in and around Bengaluru District

S. M. Lekhashree, K. N. Krishnamurthy and C. A. Srinivasamurthy
Department of Agricultural Statistics, Applied Mathematics and Computer Science,
College of Agriculture, UAS, GKVK, Bengaluru - 560 065

Abstract

In the present study, data on water quality parameters for 16 lakes located in and around Bengaluru Urban and Rural Districts over a period of three years *viz.,* 2008 to 2010 was considered to know the factors responsible for pollution. The study reveals that the first three factors explain 46.37 per cent, 22.71 per cent and 16.62 per cent of variances, respectively. The potential factors identified were lead, zinc, copper, iron, dissolved oxygen, electrical conductivity, total dissolved solids, pH and suspended solids.

The lakes in Bengaluru occupy 4.8 per cent of the city's geographical area (640 sq km covering both rural and urban areas). These Lakes were created basically for hydrological reasons; and for impounding the monsoon run off in the valley for use in lean seasons. They also helped in checking floods, in recharging and maintaining the ground water table. Lakes and tanks also act as sediment traps and prevent clogging up of natural valleys and reduce erosion by regulating run off.

Lakes serve as an important life supporting system by helping in recharging of aquifers and regulating hydrological regimes, beside, providing habitats and breeding grounds for variety of fish, birds and other aquatic life. In urban areas, lakes assume special importance in providing drinking water, recreation, irrigation and fishing. However, these lakes are often seen as main reason for development of urban area due to pressure of human activities like urbanization and industrialization.

In both urban and rural districts of Bengaluru we can observe the production activities like Agriculture, Horticulture, Dairy, Poultry, Sericulture, Fishery, Animal husbandry, etc. In Bengaluru rural district, there is 65 per cent cultivable area and 77 per cent population is engaged in agriculture. Amongst that, 65 per cent marginal land holding and 22 per cent small land holding farmers are there (Source: District profile-Bengaluru Rural). In Bengaluru urban district there is 36.7 per cent net cultivable area. Among that 82 per cent of cultivable area is under rain fed condition (Source: District profile-Bengaluru Urban). These activities are entirely depending on lake water. But after 1950's, the rate of urbanization increased rapidly. To create accommodation for the increasing population, tank beds were converted into layouts, bus depots, playgrounds and so on. As such several tanks / lakes have been erased from the map.

In the last few decades, Bengaluru has observed rapid industrialization and urbanization. Information technology sector in Bengaluru over the last decade resulted in migration and settlement of large population. The uncontrolled industrialization and urbanization has led to excessive pollution of water bodies in Bengaluru city. Rapid urbanization resulted in scarcity of potable water and this made the situation worst. The agencies are not able to provide sanitation to all parts of the city.

The comprehensive monitoring has revealed that most of the lakes in and around Bengaluru metropolitan area are polluted because of untreated municipal waste water joining the lakes. The causes are due to improper sewage maintenance, missing links which are not connected to trunk sewers of Bangalore Water Supply and Sewerage Board (BWSSB), *i.e.* not reaching the sewage treatment plant. The sewage treatment plants established by BWSSB are under loaded and as such bypasses of sewage effluents are taking place without any treatment.

The quality of water depends on several physico-chemical parameters. It is difficult to describe a simple straight forward reason for the deterioration of water quality in a particular lake. The water quality varies both spatially and temporally. The lakes nearer to the industrial area get more pollutants. Because of the lake water being stagnated, the day by day accumulation of pollutants will increase the concentration.

There are several studies reported by researchers identifying the factors responsible for pollution level of lakes and rivers. The prominent among them are Nazire *et al.* 1999; Hulya *et al.* 2005; Papatheodorou *et al.* 2007; PetrPraus, 2007; Abbas *et al.* 2008 ; Hulya and Hayal, 2008; Sangam *et al.* 2008; Smeti *et al.* 2009; Santhosh Kumar and Bhasker, 2005; Venkatesharaju *et al.* 2010 ; Yang-Hui *et al.* 2010; Mahmood *et al.* 2011; and Adamu, 2012. In the present study, factor analysis was used to identify the parameters responsible for pollution of lakes in Bengaluru District.

Material and Methods

The 16 lakes of Bengaluru District which were selected for the present study are Begur Lake, Bellandur Lake, Gottigere Lake, Haralakunte Lake, Jakkur Lake, Kalkere Lake, Karihobanahalli Lake, Lalbagh Lake, Madavara Lake, Puttenahalli Lake, Sanky tank, Ulsoor Lake, Varthur Lake, Yadiyur Lake, Yelahanka Lake, and Yellemallappa Lake.

The water quality parameters measured at 6 sample sites each over three years from 2008 - 2010 are pH, Dissolved oxygen (DO), Bio - Chemical Oxygen Demand (BOD), Electrical conductivity (EC), Total dissolved solids (TDS), Suspended solids (SS), Lead (Pb), Copper (Cu), Zinc (Zn) and Iron (Fe).

Factor analysis (FA) is a multivariate statistical technique, which gives the general relationship between measured variables by showing multivariate patterns that may help to classify the original data. FA is designed to transform the original variables into new uncorrelated variables called factors, which are linear combinations of the original variables. It is a data reduction technique and suggests how many variables are important to explain the observed variances in the data. This treatment provides a small number of factors that usually account for approximately the same amount of information as the original set of observations. The observed variables are modeled as linear combinations of the potential factors including the error terms as follows :

$$Z_{ij} = a_1 f_{1j} + a_2 f_{2j} + ... + a_m f_{mj} + e_{ij}$$

Where, Z_{ij}= Measured variable, a_i= i^{th} Factor loading, f_{ij}= Factor score

e_{ij}= Error term, i=1, 2,…m, j=1, 2, 3….p

The first thing to do before conducting factor analysis is to look into the inter-correlation between variables. If there is any variable that does not correlate with any other variables or the variables which are highly correlated with other variables are excluded before the analysis. The correlation between the variables can be checked using correlation matrix of all variables. This matrix will be created as part of the main factor analysis.

Although mild multi - collinearity exists, it is not a problem for factor analysis, it is important to avoid extreme multi-collinearity and variables that are perfectly correlated. Because, it becomes impossible to determine the unique contribution to a factor of the variables that are highly correlated as was the case in multiple regression. Therefore, at this early stage we look to correlation matrix and eliminate any variables that do not correlate with any other variable ($R<0.05$) or that correlate very highly with other variables ($R>0.9$). Hence, R-matrix can be used. If the determinant of the R-matrix is more than 0.00001, then it can be concluded that there is a problem of multi-collinearity and hence, no need of elimination of variables. If it is less than this value then look through the correlation matrix for variables that correlated very highly (i.e. $R>0.9$) and consider eliminating one of the variables before proceeding.

Kaiser-Meyer-Olkin test : This test is conducted to test the sampling adequacy, used to examine the appropriateness of factor analysis.The test statistics is given by

$$r_{xy} - \frac{\sum_{j=1}^{m} \sum_{j\neq k}^{m} r_{jk}^2}{\sum_{j=1}^{m} \sum_{j\neq 1}^{m} r_{jk}^2 + \sum_{j=1}^{m} \sum_{j\neq k}^{m} P_{jk}^2}$$

Where, r_{jk} = Correlation between j^{th} and k^{th} variables.

p_{jk} = Partial correlation between j^{th} and k^{th} variables.

m = Number of pairs of observations.

r_{xy} = Karl Pearson's coefficient of correlation between x and y

It varies between 0 and 1. A value of 0 indicates that the sum of partial correlation is large relative to the sum of the correlation, indicating diffusion in the pattern of correlation. Hence, factor analysis is likely to be inappropriate. A value close to 1 indicates that pattern of correlations are relatively compact and so factor analysis should yield distinct and reliable factors.

Kaiser (1974) recommended accepting values greater than 0.5 as acceptable. Furthermore, values between 0.5 and 0.7 are mediocre, value between 0.7 and 0.8 are good, values between 0.8 and 0.9 are greater and values above 0.9 are superb.

Bartlett's test of Sphericity : Bartlett's test was used to test the null hypothesis that the original correlation matrix is an identity matrix. For factor analysis one should have some relationship between variables and if the R-matrix were an identity matrix then all correlation coefficients would be zero. Therefore, this test provides a minimum standard, which should be passed before a factor analysis is conducted. The test statistics is given by

$$\chi^2 = \frac{1 - n + (2p + 5)}{6} \log|r| \sim \chi^2_{a,} \frac{P(p\text{-}1)}{2} df$$

Where, n = Number of observations,

p = number of parameters (Variables)

In order to determine the number of factors to be used, the variance and co-variances of the variables are computed. Then, the eigen vectors are evaluated for the covariance matrix and data can be transformed into factors. An eigen value gives a measure of significance of the factor. The factors with highest eigen values are the most significant factors. Eigen values of 1.0 or greater were considered significant (Shreshta and Kazama, 2007). Factor loading is classified as strong, moderate and weak corresponding to absolute loading values of greater than 0.75, between 0.50 and 0.75 and less than 0.50, respectively (Liu *et al.*, 2003).

Then Kaiser Varimax rotation scheme was implemented to evaluate factor loadings that correlate the factors and variables. The next step of the analysis is to compute the contribution of each factor based on the factor scores. The factor scores are projection of data onto corresponding eigen vectors based on which factor loadings are evaluated that correlate the factors and variables.

RESULTS AND DISCUSSION

Table I Shows the Karl Pearson's correlation coefficient between all the pairs of water quality parameters. It was observed that majority of variables had correlation coefficients more than 0.05 and there was some relationship between the parameters while none was more than 0.9. Hence, no correlation coefficient was particularly large and there is no problem of high correlation. Therefore, there was no need of eliminating any variables at this stage.

Table II shows Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of Sphericity. The Kaiser value was found to be 0.543. According to Kaiser (1974) it is acceptable, which means pattern of correlations are relatively compact and so factor analysis should yield distinct and reliable factor. Bartlett's test statistic was 179.228 ($p<0.00001$) which is highly significant. So, there was significant relationship between parameters which clearly implies that the correlation matrix is not an identity matrix. Therefore, Factor analysis was found to be highly appropriate.

Table III shows the eigen values associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction 10 linear components were identified. The eigen values associated with each factor represents the variance explained by that particular linear component.

Factor 1 explained 46.375 per cent of the total variance. It clearly shows first few factors explaining relatively larger amount of variance. Whereas, the

TABLE I

Correlation between water quality parameters

	pH	DO	BOD	Pb	Cu	Zn	Fe	EC	TDS	SS
pH	1.000									
DO	0.588 **	1.000								
BOD	-0.050	-0.372 *	1.000							
Pb	0.324 *	0.035	-0.065	1.000						
Cu	-0.300 *	0.005	0.067	0.883 **	1.000					
Zn	-0.237 *	0.011	0.211	0.825 **	0.862 **	1.000				
Fe	0.336 *	-0.126	0.353 *	0.755 **	0.803 **	0.823 **	1.000			
EC	0.488 **	0.356 *	0.794 **	0.275 *	0.371 *	0.443 **	0.574 **	1.000		
TDS	-0.476 **	-0.305 *	0.782 **	0.275 *	0.369 *	0.430 **	0.542 **	0.894 **	1.000	
SS	0.525 **	0.198	0.198	-0.075	-0.007	0.069	0.014	0.047	-0.023	1.000

* Significant at 5% level ** Significant at 1% level

TABLE II

Kaiser-Meyer-Olkin and Bartlett's test values

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.543
Bartlett's Test of Sphericity	Approximate Chi-Square	179.228
	Degree of freedom	45
	p value	0.000

TABLE III

Extracted values of various factor analysis for water quality parameters

Component	Initial Eigen Values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cum %	Total	% of Variance	Cum %	Total	% of Variance	Cum %
1	4.638	46.375	46.375	4.638	46.375	46.375	3.675	36.753	36.753
2	2.272	22.715	69.090	2.272	22.715	69.090	2.986	29.86	66.612
3	1.662	16.621	85.711	1.662	16.621	85.711	1.910	19.099	85.711
4	0.675	6.749	92.460						
5	0.364	3.641	96.101						
6	0.194	1.939	98.040						
7	0.150	1.495	99.535						
8	0.037	0.369	99.905						
9	0.008	0.077	99.982						
10	0.002	0.018	100.000						

subsequent factors explains the small amount of variance. There are three factors with eigen values >1, the eigen value associated with these factors are again displayed in the columns labeled extraction sum of squared loadings. The values in this part are same as the values before

Extraction, except that the values for the discarded factors are ignored. In the final part of the table labeled rotation sums of squared loading, the eigen values that the factors after rotation are displayed. Rotation has the effect of optimizing the factors structure and one consequence for these data is that the relative importance of the four factors is equalized, before rotation. Factor one account for only 36.756 per cent of variance compared to 29.860 per cent and 19.099 per cent, respectively. High positive loading indicates strong linear correlation between the factor and parameter. Similarly, Factor 2 and 3 explained 22.715 and 16.621 per cent of the total variability, respectively.

Table IV shows that, Factor 1 was positively and strongly (>0.75) loaded with eigen values of Pb (0.969), Cu (0.971), Zn (0.835), Fe (0.910) and weakly loaded with EC (0.307) and TDS (0.303). Factor 2 is positively and highly loaded with DO (0.952), EC (0.897), TDS (0.888) and moderately loaded with Zn (0.371). Factor 3 was positively and strongly loaded with pH (0.866), SS (0.813) and moderately loaded with BOD (0.621).

TABLE IV

The factor loading values and the explained variance of water quality parameters

Para meters	Factor 1	Factor 2	Factor 3
pH	-0.254	-0.242	0.866
BOD	0.130	-0.442	0.621
DO	-0.014	0.952	0.118
Pb	0.969	-0.045	-0.111
Cu	0.971	0.080	-0.055
Fe	0.910	0.217	0.046
Zn	0.835	0.371	-0.071
EC	0.307	0.897	-0.209
TDS	0.303	0.888	-0.180
SS	-0.005	0.199	0.813

Figure 1 shows the variables in the rotated factor space. It is the alternative method of representation of loadings, to achieve a better interpretation. This will help to see how the variables are organized in the common factor space, while, the total amount of variance explained remains equal but the variance explained per component changes.

Figure 2 shows the Scree plot of the Eigen values. By using Kaiser's criteria extracted 3 factors, this criteria is accurate when the number of variables are less than 30 and number of communalities after extraction are greater than 0.7. Table IV shows the communalities before and after extraction. In this case, both the conditions are satisfied. So, Kaiser's criterion

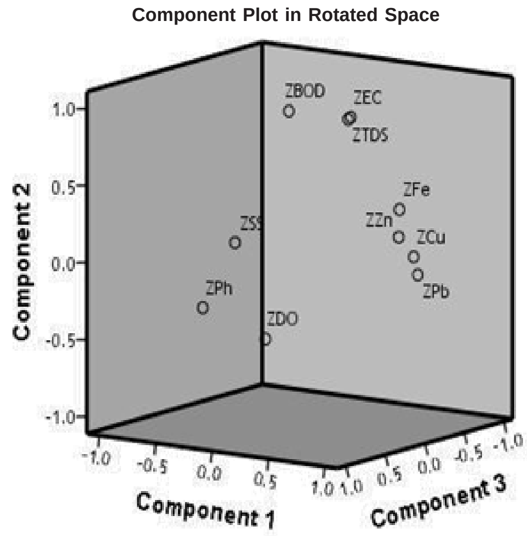

Fig. 1 : Showing component plot in rotated space

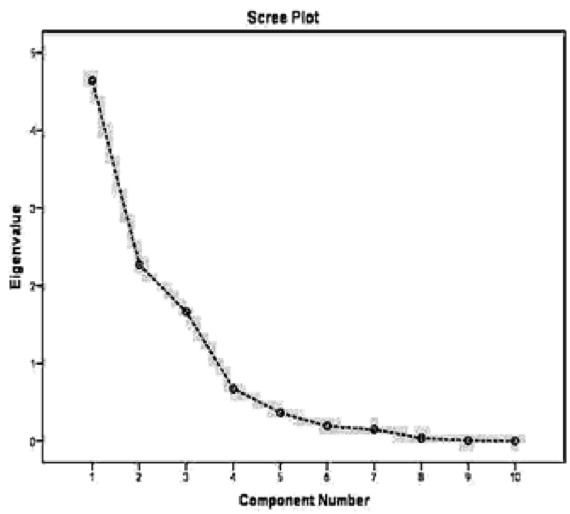

Fig. 2 : Showing the Scree plot of the Eigen values

is accurate. It is once again proved in the Scree plot of Eigen values. Here, we can observe the inflexion on the curve. The curve is tail off after three factors, reaching stable plateau. Therefore, we can justify that retaining of three factors is correct.

Based on the statistical analysis of data on factors responsible for pollution of selected lakes around Bengaluru, it can be concluded that the parameters such as Lead (Pb), Copper (Cu), Zinc (Zn) and Iron (Fe) were strongly loaded with a total variability of 46.37 per cent. On the other hand, DO, EC and TDS had explained variations to an extent of 22.71 per cent. However, the parameters pH and SS explained variation to an extent of 16.62 per cent. These were the potential factors for pollution of the lakes.

REFERENCES

ABBAS, F. M., ANEES AHMAD, NORLI ISMAIL, AZHAR MAT EASA AND KHALID OMAR., 2008, Assessment of Surface Water through Multivariate Analysis. *Journal of Sustainable Development*, **1** (3) : 166 - 169.

ADAMU MUSTAPHA, 2012, Identification of Anthropogenic Influences on water quality of Jakara River, North-western Nigeria. *Journal of Applied Sciences in Environmental Sanitation*, **7** (1) : 11 - 20.

HULYA BOYACIOGLU, HAYAL BOYACIOGLU AND ORHAN GUNDUZ., 2005, Application of Factor analysis in the Assessment of Surface Water Quality in Buyuk Menderes River Basin. *European water*, **9** (10) : 43 - 49.

HULYA BOYACIOGLU AND HAYAL BOYACIOGLU, 2008, Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environ Geol*., **54** : 275 – 282.

KAISER, H. F., 1974, An index of factor simplicity, *Psychometrica*, **39** : 31 - 36.

LIU, C. W., LIN, K. H. AND KUO, Y. M., 2003, Application of factor analysis in the assessment of ground water quality in a back foot disease area in Taiwan. Sci. *Total Environ*., **313** (1 - 3) : 467 - 476.

MAHMOOD ASIF, WAQAS MUQBOOL, MUHAMMAD WASEEM MUMTAZ AND FAROOQ AHMAD, 2005, Application of Multivariate Statistical Techniques for the Characterization of Ground Water Quality of Lahore, Gujranwala and Sialkot (Pakistan). *Pak J. Anal. Environ. Chem*., **12** (1&2) : 102 - 112.

NAZIRE MAZLUM., ADEM OZER. AND MAZLUM, 1999, Interpretation of Water Quality Data by Principal Components Analysis, *Tr. J. of Engineering and Environmental and Environmental Science*, **23** : 19 - 26

PAPATHEODOROU, G., LAMBRAKIS, N. AND PANAGOPOULOS, G., 2007, Application of Multivariate statistical procedures to the hydrochemical study of a coastal aquifier : an example from Crete, Greece. *Hydrological processes*, **21** : 1482 - 1495.

PETR PRAUS., 2007, Urban water quality evaluation using multivariate analysis. *Acta Montanistica Slovaca*, **12** (2) : 150 - 158.

SANGAM SHRESTA, FUTABA KAZAMA AND TAKASHI NAKAMURA, 2008, Use of Principal component analysis, Factor analysis and Discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong river. *Journal of Hydro-informatics*, **10** (1) : 43 - 56.

SANTOSHKUMAR SARKAR AND BHASKAR DEB BHATTACHARYA., 2010, Water Quality Analysis of the Coastal Regions of Sundarban Mangrove Wetland, India Using Multivariate Statistical Techniques. Environmental Management, book pp . 258.

SHRESTHA, S. AND KAZAMA, F., 2007, Assessment of water quality using multivariate statistical techniques: A case study of Fuji river basin, Japan. *Environmental Modelling and Software*, **22** : 464 - 475.

SMETI, E. M., LYTRAS, E. S., TZOUMERKAS, P. C. AND GOLFINOPOULOS, S. K., 2009, Multivariate Statistical Analysis of Surface Water Intended for Human Consumption. Proceeding of the 11th International Conference on Environmental Science and Technology.

VENKATESHARAJU, K., SOMASHEKAR, K. AND PRAKASH, K., 2010, Study of seasonal and Spatial Variation in Surface Water quality of Cauvery River Stretch in Karnataka. *Journal of Ecology and the Natural Environment*, **2** (1) : 001 - 009.

YONG - HUI YANG, FENG ZHOU, HUAI CHENG GUO, HU SHENG, HUI LIU AND CHENG - JIE HE, 2010, Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environ Monitoring and Assess*, **170** : 407 - 416.

(***Received :*** August, 2015 ***Accepted :*** December, 2015)