International Journal of Science, Environment and Technology, Vol. 4, No 1, 2015, 193 – 204
ISSN 2278-3687 (O)

STATISTICAL TECHNIQUES FOR ASSESSMENT OF INLAND WATER QUALITY

V.S. Bharti[*1], C.S. Purusothaman[2], V. K. Yadav[3], A.B. Inamdar[4] and S.P. Shukla[5]
[1,3&5]ICAR-Central Institute of Fisheries Education, Versova, Mumbai-400061, India
[2]ICAR-Central Marine Fisheries Research Institute, Post Box No. 1603, Ernakulam North. P.O. Kochi – 082018, Kochi
[4]Centre of Studies in Resource Engineering (CSRE), IIT Bombay, Powai, Mumbai - 400076, India
E-mails: [2]cspuru@hotmail.com, [3]vinod_iasri@yahoo.co.in, [4]abi@iitb.ac.in, [5]spshukla @cife.edu.in and vsbharti@cife.edu.in (*Corresponding Author*)

Abstract: Complex water quality of Lake was assessed with Multivariate statistical techniques during the 2-year (2011-13) monitoring program. Cluster analysis grouped whole lake into three zones. Discriminant Analysis (DA) calculated three parameters Dissolved Oxygen (DO), Chlorophyll a (chl a) and Total Suspended Solids (TSS) to 100% correct assignment in the spatial analysis of the three clusters in stepwise mode and 91.7% in standard mode, whereas in temporal analysis, DA calculated six parameters (DO, Chl-a, TSS, transparency, air temperature and water temperature) to 100% correct assignment in both standard and stepwise modes. The Principal component analysis (PCA) assisted to extract and recognize the factors responsible for water quality variations in four seasons of the year. Principal components (PCs) captured more than 60% of the total variance in each season. Strong loadings included DO, TSS, Available phosphorous (avail. P), total phosphorus (TP), temperature, ammonia-N and nitrate-N. The natural parameters (temperature and DO), the inorganic parameter (TSS) and the organic nutrients (Avail P, TP, nitrate-N and ammonia-N) were the most significant parameters contributing to water quality variations in all seasons, and indicate organic and sediment pollution in the lake.
Keywords: Cluster analysis; Discriminant Analysis; Principal component analysis; pollution, inland water quality

Introduction

The natural environment is a multivariate complex system. The world scientific community is presently unanimous about intangible benefits provided by lakes and wetlands. Efforts were made by Costanza *et al.* (1997) to quantify and value the benefits provided by natural capital like oceans, forests, grasslands, wetlands and lakes. It was found that non-market services in the form of water regulation, water supply, waste treatment and recreational aspects constitute the major portion of overall benefits/services provided by Inland Water bodies. Lake in watersheds with substantial eroded soil, effluent discharge and urban land use

*Received Jan 21, 2015 * Published Feb 2, 2015 * www.ijset.net*

experience increased inputs and varying compositions of organic matter (Sickman *et al.*, 2007), and excessive concentrations of phosphorus and other nutrients from detergent and chemical application and watershed releases (Easton *et al.*, 2007). Reliable information on the quality of freshwater needs regular monitoring programs. Analyzing the large number of measured physico-chemical, biological and optical variables is difficult, and it is also necessary to draw meaningful conclusions from the data-set, without losing useful information for quality assessment with respect to sustainability. This also helps to optimize the monitoring network by recognizing the representative parameters delineating the sampling station and identifying the pollution source. A variety of methods are being used to display the information which is concealed in the water quality variables. The application of different multivariate statistical techniques such as hierarchical cluster analysis (HCA), discriminant analysis (DA) and principal component analysis (PCA) helps in the interpretation of complex data matrices to better understand the water quality.

In the present study, different multivariate statistical techniques were used to alienate information on the similarities or dissimilarities between sampling stations, identifying water quality parameters responsible for spatial and temporal variations, and to identify the source of pollution.

Materials and Methods

Site Description

Lake (Powai) is located in the north-eastern suburb of Mumbai in the western region of India (Fig. 1). It lies between the longitudes of 72°53′45”and 72°54′30” E, and latitudes of 19°7′ 15” and 19°8′15” N, 55 m above the mean sea level. The lake has a water-spread area of about 2.1 km^2 and the depth varies from about 3 m (at the periphery) to 12 m at its deepest part. The catchment area of the lake is about 600 ha with the average annual rainfall of 1917.3 mm. About 94% of the annual rainfall of the Mumbai Suburban district is received during the south-west monsoon months of June-September. Powai Lake is located downstream of Vihar Lake on Mithi River, which brings substantial discharge of exogenous nutrients and suspended solids into the lake. The lake water is not used for drinking purposes though used for non-potable purposes like gardening, angling and industrial use. This lake was included in "National Lake Conservation Plan" by the Ministry of Environment and Forests (Government of India) in the year 1991.

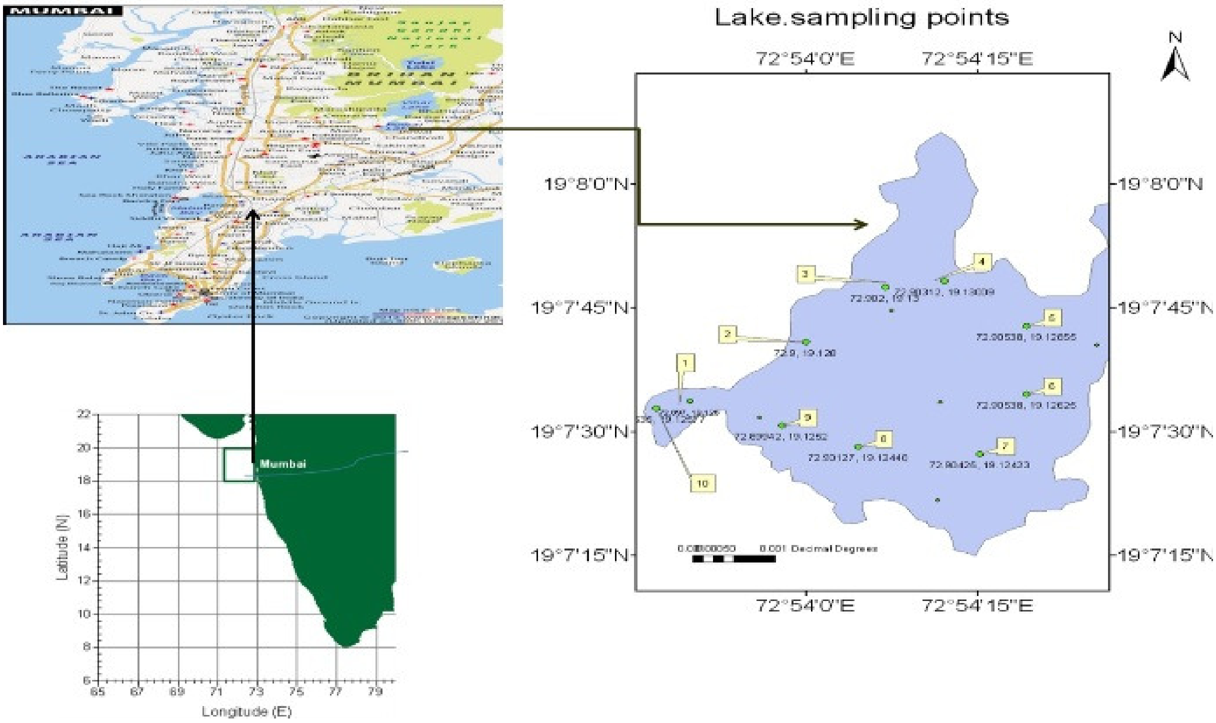

Fig. 1: Map of the study area and surface water quality sampling stations in Powai Lake

Sample collection

The geographic locations of the sampling points were recorded using the global positioning system (GPS). Water samples were collected in 2-L polyethylene bottles cleaned with metal-free soap, rinsed many times with distilled water. All water samples were stored in insulated cooler containing ice and transported the same day to the laboratory. All the samples were kept at 4°C until processing and analysis. For CDOM, about 100 ml water were filtered under dim light and gentle vacuum through 0.22 µm Millipore polycarbonate membrane filter, which had been briefly (5 min) soaked in purified water, into precombusted amber colored bottle. For chl a about 300–500 ml were filtered through Whatman GF/F filter (0.7 µm, 25 mm) and immediately put in an organic solvent for pigment extraction.

Analytical methods

In the present investigation, water samples were collected at ten stations in four different seasons (autumn and winter of 2011-12; spring and summer of 2013). Sampling site selection criteria included natural conditions as well as human activities: stations 1 and 10 represent natural conditions near the dam; stations 2 and 3 are affected by sewage discharge from the guesthouse and hostel of the Indian Institute of Technology - Bombay (IITB); stations 5, 6 and 7 are affected by almost all types of pollutants from residential and to a lesser extent industrial activities; Station 9 is an idol immersion site. The selected water quality parameters consisted of dissolved oxygen (DO), pH, water temperature (Water T), air temperature (Air T), nitrate-N, ammonia-N, total phosphorus (TP), transparency, total suspended solids (TSS), available phosphorus (avail P), Chlorophyll - a (Chl-a) and $a_{CDOM}(440)$. The temperature, pH and transparency of the water were measured on site by a thermometer, pH meter and Sechhi disc respectively (APHA, 1998). DO was determined by the Winkler azide method and TSS was determined gravimetrically at 100-105°C. Ammonia-N and TP were analyzed by cadmium reduction and ascorbic acid method (Biochrom, Libra 32 pc spectrophotometer), respectively.

Absorbance measurements of CDOM were made in a Biochrom, Libra 32 pc spectrophotometer in 1-cm path length quartz glass cells using purified water as reference. Absorbance spectra were obtained for the 200–800 nm range at 1-nm intervals. With all samples, the optical density at 700 nm was low (0.0000±0.0001), and for uniformity, it was null-point corrected at this wavelength. The absorption coefficient a(λ) was calculated from the corrected absorbance using the equation (1):

$$a(\lambda) = 2.303A(\lambda)/0.1(m^{-1}) \qquad \text{..........................(1)}$$

Where 2.303 is ln (10) and 0.1 is the cuvette path length in meters. The magnitude of aCDOM(440) was used as a proxy to the concentration of CDOM (Menon et al., 2011).

Statistical methods

All mathematical and statistical calculations were implemented using SPSS 16.0 and Microsoft Office Excel 2010. Hierarchical agglomerative cluster analysis was performed on the data-sets by means of the Ward's method, using squared Euclidean distances as a measure of similarity. The spatial variability of water quality in the whole lake was determined from CA, using the dendrogram. Discriminant Analysis operates on original data in standard and stepwise modes to confirm the groups found by cluster analysis, and evaluates spatio-temporal variations in terms of the discriminant variables. Here, the monitoring stations (spatial) and periods (temporal) were the grouping variables, and the measured parameters were the independent variables. Principal component analysis is designed to transmute the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. These principal components were subjected to varimax rotation (raw) generating factors (Love *et al.,* 2004; Abdul-Wahab *et al.,* 2005).

Results

Mean and standard deviation of water quality parameters in four seasons are presented in Tables 1. Chlorophyll - a concentration was higher during summer and spring as compared to autumn and winter, The variation in the trend of transparency is approximately similar in autumn and summer seasons, and transparency was higher in autumn and summer as compared to winter and spring because of monsoon and sandy loam texture which helps in better settlement of suspended particle from water column.

Spatial similarity with cluster analysis

CA was applied to find out the similarity groups between the sampling stations. It resulted in a dendogram (Fig. 2) grouping all the ten sampling stations into three statistically meaningful clusters. Four stations (1, 4, 8 and 10) formed Cluster 1 which comprises relatively less polluted sites: stations 1 and 10 are in the dam area having higher water depth, Station 8 is near the garden area and no sewage outlet is present at stations 4 and 8 indicating that these stations are relatively less disturbed by sewage entry. Cluster 2 included Station 2 and corresponds to moderately polluted sites; this station receives pollution from Raymond guesthouse. Five stations (3, 5, 6, 7 and 9), which are highly polluted, formed Cluster 3; wastewater and untreated sewage are discharged from Rennaisance Hotel at Station 3; wastewater from IITB campus is discharged at Station 5; stations 6 and 7 have discharge points dumping sewage water into the lake; Station 9 is an established idol immersion site . There are other reports (Kumar, 2011) with similar results in water quality programmes.

Table 1. Water quality parameters mean and standard deviation in four seasons

Season		DO (mg/L)	Water T (^{0}C)	TP (mg/L)	Nitrate-N (mg/L)	Ammonia-N (mg/L)	TSS (mg/L)	Chl-a (mg/m^3)	a_{CDOM}(440) (m^{-1})	Avail P (mg/L)	SDD (cm)	Air T (^{0}C)
Spring	Mean	4.20	26.2	0.277	0.594	0.006	6.95	17.37	0.2159	0.195	49.6	33.57
	SD	0.77	0.48	0.071	0.439	0.0048	1.49	8.98	0.0229	0.0279	3.94	1.72
Summer	Mean	7.45	31.65	0.253	0.5100	0.013	15.19	39.48	0.1860	0.293	40.3	34.1
	SD	0.810	0.818	0.097	0.497	0.015	10.22	12.64	0.055	0.138	1.63	1.52
Autumn	Mean	3.80	30.2	0.216	3.9700	0.0690	136.8	14.85	0.220	0.118	69.8	30.8
	SD	0.82	0.447	0.035	1.7200	0.0440	24.37	5.85	0.099	0.0141	8.40	1.09
Winter	Mean	5.860	25.56	0.318	0.1370	0.008	23.60	14.16	0.193	0.1719	53.6	29.02
	SD	1.090	0.291	0.153	0.0440	0.006	10.00	4.09	0.031	0.0220	3.97	1.24

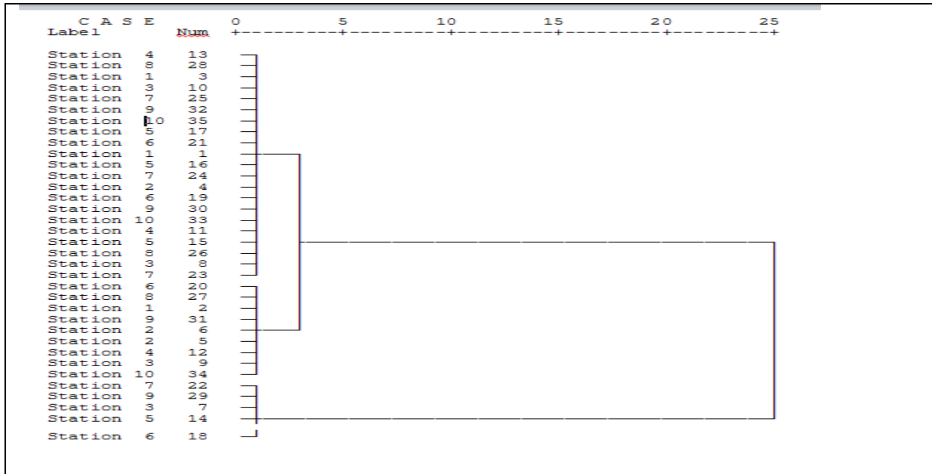

Fig. 2: Dendrogram of cluster analysis for sampling stations according to surface water quality parameters of Powai Lake

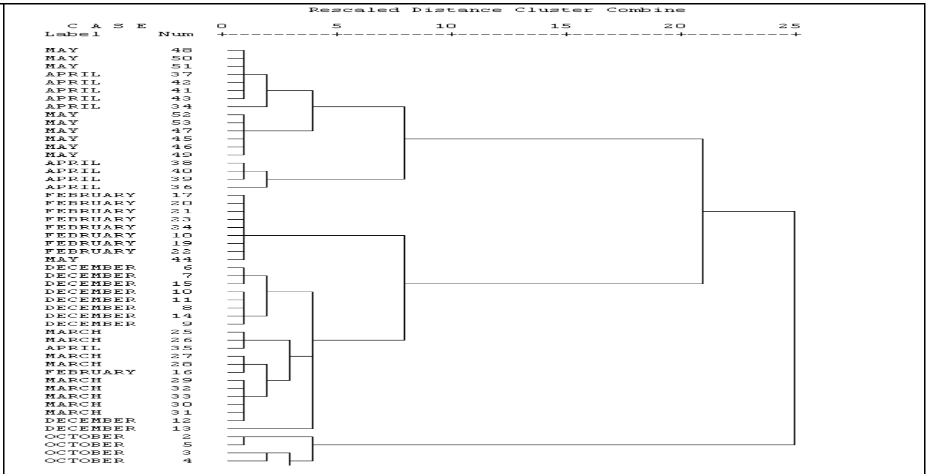

Fig. 3: Dendrogram of cluster analysis for seasons according to surface water quality parameters of Powai Lake

Table 2. Results of temporal DA **Table 3.** Results of spatial DA

	Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.	Wilks' Lambda	Chi-square	df	Sig.
Standard	1 through 2	0.005	236.101	24	0.000	0.004	148.024	24	0.000
	2	0.107	99.452	11	0.000	0.146	50.921	11	0.000
Stepwise	1 through 2	0.009	225.645	12	0.000	0.016	127.426	6	0.000
	2	0.139	93.775	5	0.000	0.305	36.789	2	0.000

	Table 4. Standardized canonical discriminant unction coefficients for DA of temporal variation				**Table 5.** Standardized canonical discriminant function coefficients for DA of spatial variation			
	Standard		Stepwise		Standard		Stepwise	
	1	2	1	2	1	2	1	2
SDD	0.740	0.098	0.480	0.083	1.032	-0.078		
Air T	-0.028	0.914	0.024	0.733	0.163	0.708		
Water T	0.204	0.420	0.079	0.565	1.460	-1.564		
pH	-0.118	0.440			0.137	0.106		
DO	-0.530	0.512	-0.570	0.579	-1.018	0.385	-0.641	0.039
Avail P	-0.290	0.254			-0.894	0.945		
Total P	-0.158	-0.254			-0.233	-0.066		
Ammonia-N	0.344	0.323			0.358	0.309		
Nitrate-N	0.428	-0.105			0.055	-0.145		
Chl-a	-0.240	0.833	-0.069	0.654	0.126	-0.907	-0.030	0.968
TSS	0.680	0.234	0.919	0.300	0.771	0.349	1.102	0.195
$A_{CDOM}(440)$	-0.388	-0.148			-0.528	-0.032		

Temporal similarity with cluster analysis

An initial exploratory approach involved the use of hierchical CA on data sorted by season. Temporal CA generated a dendrogram as shown in Fig. 3 grouping the months into three clusters. Cluster 1 comprised October representing autumn. December, February and March formed the second cluster representing winter. April and May formed the third cluster representing summer. Temporal variation in lake water quality is not determined absolutely by seasonal effects; the nature of discharge also plays a prominent role.

Temporal variation in lake water quality

The temporal variation in lake water quality was evaluated using DA with clusters based on temporal CA. The objective of DA was to test the significance of discriminant functions and determine the most significant variables associated with the differences among the clusters. Wilks' lambda and the chi square for each discriminant function ranged between 0.005 and 0.139, and 93.77 and 236.10, respectively, at $P < 0.011$ (Table 2) showing that the temporal DA is effective and meaningful. The discriminant functions obtained from the standard, stepwise modes of DA are shown in Table 4. In the stepwise mode, variables are included step by step beginning with the most significant one until no significant changes are obtained. The standard DA mode constructs DFs containing all parameters. The stepwise mode suggests that SDD, air temperature, water temperature, DO, Chl-a and TSS are significant parameters for discriminating among the seasons for the temporal variation of water quality. Both the standard and stepwise DFs rendered the assigning 100% of the cases in the three groups.

Spatial variation in lake water quality

Spatial variation in lake water quality was evaluated using DA, with cluster based on spatial CA. The objective of DA was to test the significance of discriminant functions and determine the most significant variables associated with the differences among the clusters. Stations were the dependent variables and the measured parameters were the independent variables. Wilks' lambda and the chi square for the function were 0.004 to 0.3 and 36.78 to 148.02, respectively, at $P < 0.011$ showing that spatial DA is effective and meaningful (Table 3). The discriminant function obtained from the standard and stepwise modes of DA are shown in Table 5. The stepwise mode suggests that DO, chl-a and TSS are the significant parameters for discriminating among the three groups for the spatial variation of water quality in Powai Lake (Table 5). Both the standard and stepwise mode DFs rendered the corresponding group memberships correctly assigning 100% and 92% of the cases in the three groups, respectively.

Principal component analysis

PCA was executed on variables for the ten different sampling stations in four seasons in order to identify important seasonal water quality parameters. The factors with the highest eigenvalues are the most significant ones. The classification of factor loading is thus 'strong', 'moderate' and 'weak', corresponding to the absolute loading values of > 0.75, 0.75-0.50 and 0.50-0.30, respectively (Liu *et al.,* 2003). Corresponding variable loadings and explained variance are presented in Table 7, and the strong loading values have been highlighted. The two factors of PCA include totally more than 60% of the total variance in each season respecting water quality data-sets. The most significant water quality parameters that can contribute to evaluate seasonal variations of water quality are signalized in Table 6. Available phosphorus is the most significant parameter contributing to water quality variations in all the four seasons. DO and aCDOM(440) with positive strong loading values as the most significant parameters contributed to water quality variations in three seasons. It has this meaning that macrophytes photosynthesise and liberate oxygen throughout the year and the addition of dissolved organic matter from the household areas through sewage is a continuous process in Lake.

Table 6. The most significant parameters that contribute to water quality variations in each season

Season	Parameter with strong positive factor loading	Parameter with strong negative factor loading
Spring	Avail P, TP, pH, Nitrate-N	$a_{CDOM}(440)$
Summer	Avail P, DO, Nitrate-N , $a_{CDOM}(440)$, Water T, Air T, TSS	Total P, pH
Autumn	Avail P, TP, DO , Nitrate-N, $a_{CDOM}(440)$, pH, Water T, Air T, TSS	
Winter	Avail P , DO, Chl-a, Ammonia-N	

TSS with a strong positive loading is another of the most significant parameters in water quality variations during summer and autumn, and it can be a demonstrator of erosion effect due to construction activity and sediment drainage during rainy season. The temperature and pH with strong loadings are as the most significant parameters contributing to water quality variations in the four seasons, and these represent the seasonal impacts of temperature and decomposition. The most significant parameters that contribute to water quality variation in each season represent eutrophication (phosphorus), seasonal influence (temperature), sediment pollution (natural runoff) and organic pollution (domestic sewage), respectively.

Discussion

Normally, DO is higher during winter as at low temperature, the DO holding capacity is higher but in the study area during summer, DO was higher due to phytoplankton bloom. Transparency was higher in autumn and summer as compared to winter and spring because of monsoon and sandy loam texture of sediment which helps in better settlement of suspended particle from water column. DA allowed a reduction in the dimensionality of the large data-set, delineating a few indicator parameters responsible for the variations in water quality. PCA resulted in four factors explaining 80% of the total variance in different seasons. The spring and winter factor obtained represents eutrophication and organic pollution (phosphorus). Summer and autumn factor represents natural pollution which includes the surface runoff from the catchment area and the seasonal effects of temperature.

Conclusion

Temporal variation in lake water quality is not determined absolutely by seasonal effects; the nature of discharge also plays a prominent role. Thus based on statistical analysis, it is possible to plan for optimum sampling strategy that can lessen the number of sampling stations and affiliated recurring costs. Concretely, the number of monitoring stations and month could be reduced, and could be only chosen from groups 1, 2 and 3. Furthermore, pollution at stations 3, 5, 6, 7 and 9 indicates direct discharge of sewage from 17 culverts from road side, which can be treated well before discharge or strict measures can be taken to stop sewage discharge through these discharge points.

Table 7. The factor loadings value and explained variance of water quality parameters in four seasons

Spring			Summer			Autumn			Winter		
Parameter	Factor 1	Factor 2	Parameter	Factor 1	Factor 2	Parameter	Factor 1	Factor 2	Parameter	Factor 1	Factor 2
$a_{CDOM}(440)$	-0.906	0.254	Water T	0.903	0.223	DO	0.979	0.170	DO	0.819	-0.305
Avail P	0.895	-0.033	$a_{CDOM}(440)$	0.855	0.022	Nitrate-N	0.965	-0.033	Chl-a	0.803	-0.069
pH	0.813	-0.119	DO	0.851	-0.352	Water T	0.952	-0.140	SDD	0.743	0.533
Total P	0.790	0.280	Air T	0.807	0.418	pH	0.926	-0.261	Air T	0.732	0.061
Water T	0.652	0.614	Total P	-0.796	0.022	TSS	0.898	-0.318	pH	-0.647	-0.238
DO	-0.150	0.030	Nitrate-N	-0.760	0.330	Air T	0.757	0.016	Water T	0.643	0.590
Nitrate-N	-0.251	0.849	Chl-a	-0.073	-0.017	Chl a	-0.657	0.325	Nitrate-N	-0.562	0.557
Chl-a	-0.139	-0.702	pH	0.147	-0.774	Avail P	-0.356	0.932	Ammonia-N	-0.067	0.926
Ammonia-N	-0.244	0.696	TSS	0.315	0.757	$a_{CDOM}(440)$	0.282	0.892	Avail P	-0.332	0.782
Air T	-0.327	0.613	Avail P	0.068	0.753	Total P	-0.068	0.757	TSS	0.158	0.601
TSS	-0.227	-0.605	Ammonia-N	-0.233	0.725	SDD	-0.077	0.704	Total P	0.071	0.594
			SDD	0.136	0.704	Ammonia-N	-0.217	0.535	$a_{CDOM}(440)$	-0.077	-0.311
Total variance	33.34	26.96	Total variance	36.20	26.74	Total variance	50.918	24.672	Total variance	31.621	27.404

Extraction Method: Principal component analysis rotation: Varimax with Kaiser normalization rotation converged in 3 iterations

Acknowledgements

The authors would like to thank the Director, ICAR-Central Institute of Fisheries Education (CIFE) for financial and scientific support and to the crew members of the boat of the Maharastra State Angling Association for providing assistance during sampling.

References

[1] Abdul-Wahab, S.A., Bakheit, C.S., Al-Alawi, S.M., 2005. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. - Environ. Model. Software, 20 (10): 1263-1271.

[2] APHA, 1998. Standard methods for the examination of water and wastewater. - American Public Health Association, 18th Ed., Washington DC, USA.

[3] Costanza, R.R., d'Arge, R., de Groot, R., Farbe, S., Grasso, M., Limburg, K., Hannon, B., Naeem, S., O'Neill, R.V., Raskin, R.G., Sutton, P. and Vanden, B. M., 1997 The value of the world's ecosystem services and natural capital.- Nature, 387: 253-260.

[4] Easton, Z.M., Gerard-Marchant, P., Walter, M.T., Petrovic, A.M. and Steenhuis, T.S. 2007. Identifying dissolved phosphorus source areas and predicting transport from an urban watershed using distributed hydrologic modeling. - Water Resour. Res., 43 (11): 1-16.

[5] KumarRatheesh, R. 2011. A comparative study of biodiversity between Powai and Vihar lakes in Mumbai. M.F.Sc. Dissertation, Central Institute of Fisheries Education (Deemed University), Versova, Mumbai. 128pp.

[6] Liu C W, Lin K H, Kuo Y M, 2003. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. - Sci. Total Environ. , 313 (1-3): 77-89.

[7] Love, D., Hallbauer, D., Amos, A. and Hranova, R. 2004. Factor analysis as a tool in groundwater quality management: Two southern African case studies. Phys. Chem. Earth, 29 (15-18): 1135-1143.

[8] Menon, H.B., Sangekar, N.P., Lotliker, A.A. and Vethamony, P. 2011. Dynamics of chromophoric dissolved organic matter in Mandovi and Zuari estuaries - A study through in situ and satellite data. - ISPRS Journal of Photogrammetry and Remote Sensing, 66: 545-552.

[9] Sickman, J.O., Zanoli, M. J. and Mann, H.L. 2007. Effects of urbanization on organic carbon loads in the Sacramento River, California. Water Resour. Res., 43, (22): 1-15.