Study of Variations in Water Quality of Mumbai Coast Through Multivariate Analysis Techniques

Article *in* Indian Journal of Geo-Marine Sciences · July 2009

3 authors:

Indrani Gupta
National Environmental Engineering Research Institute

Shivani S Dhage
National Environmental Engineering Research Institute

Rakesh Kumar

Indian Journal of Marine Sciences
Vol. 38(2), June 2009, pp. 170-177

Study of variations in water quality of Mumbai coast through multivariate analysis techniques

Indrani Gupta, Shivani Dhage & Rakesh Kumar
National Environmental Engineering Research Institute, 89/B, Dr. A. B. Road, Worli, Mumbai –400 018, India
[E.mail: indranig1@rediffmail.com]

Received 16 October 2007; revised 10 July 2008

Multivariate statistical techniques, such as Cluster Analysis (CA), Discriminant Analysis (DA), and Principal component analysis (PCA) were applied to evaluate the temporal/spatial variations in marine water quality of Mumbai and to identify pollution sources. Hierarchical CA grouped 12 sampling sites into three clusters of similar water quality characteristics. DA gave the best results both spatially and temporally. It provided an important data reduction as it used only four parameters (DO, Total coliform, Ammonical nitrogen and pH) affording 100% correct assignment in temporal analysis. For spatial DA, DO and temperature; Feacal strptococii, DO and Total coliform; temperature and phosphate were used for summer, monsoon and winter seasons respectively. DA gave 100% correct assignment in spatial analysis except for summer season, step wise mode DA rendered 91.6% correct assignment. PCA resulted in four factors explaining 81.4% of the total variance. The first factor obtained represents organic pollution from domestic waste water. The second factor represents natural pollution which includes the surface run off. The third factor represents nutrient pollution whereas the fourth factor represents seasonal effects of temperature.

[Keywords: Marine water quality, cluster analysis, discriminant analysis, principal component analysis.]

Introduction

Marine water quality has become a matter of serious concern because of its effects on human health and aquatic ecosystems including a rich array of marine life. With the growth of population and industrialization, marine water receives a large amount of pollution from municipal and industrial sources, as also from surface run-off[1]. Reliable information on quality of water needs regular monitoring programs. This results in a huge and complex data matrix comprised of a large number of physico-chemical and microbiological parameters. This data matrix is often difficult to interpret for drawing meaningful conclusions[2]. It is necessary to draw meaningful conclusions from this data set, without losing useful information. This also helps to optimize the monitoring network by recognizing the representative parameters delineating the sampling sites and identifying pollution sources[3].

The application of different multivariate statistical techniques such as Cluster Analysis (CA), Discriminant Analysis (DA) and Principal Component Analysis (PCA) helps in the interpretation of complex data matrices to better understand the water quality. These techniques have been applied by many researchers to characterize and evaluate freshwater, marine water and sediment quality[3-12]. It is useful to examine temporal and spatial variations caused by natural and anthropogenic factors[13,14]. In the present study, a large data matrix, obtained during one whole year monitoring program, has been subjected to different multivariate statistical techniques such as Cluster Analysis (CA), Discriminant Analysis (DA) and Principal Component Analysis (PCA). This is to alienate information on the similarities or dissimilarities between sampling sites, identify water quality parameters responsible for spatial and temporal variations in marine water quality and to also identify the pollution sources.

Materials and Methods

The sampling locations are shown in Fig. 1. Field survey were conducted to obtain data on marine water quality from 12 monitoring sites that cover the complete range of the west coast of Mumbai. The software package SPSS 13 was used to carry out the analysis.

The beaches and sea fronts along the coast were monitored fortnightly for 12 months. The water samples were collected from three locations at each beach for three hours i.e. one hour before tide, during the tide and one hour after the tide over three seasons

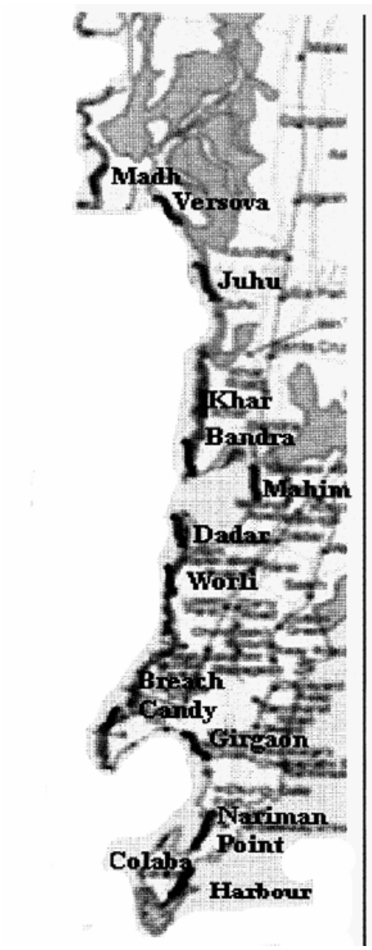

Fig. 1—Sampling Locations of the Beaches and Seafronts of Mumbai

viz. summer, post monsoon and winter. Ten parameters viz. Temperature, Turbidity, pH, Dissolved oxygen (DO), Biochemal Oxygen Demand (BOD), Ammonical Nitrogen, Phosphate Phosphorus, Total Coliform (TC), Faecal Coliform (FC) and Faecal Strptococii (FS) were monitored. Analysis for these parameters followed standard method[15]. Mean and standard deviations of water quality parameters are given in Table 1.

Cluster Analysis

The objective of cluster analysis is to identify relatively similar, that is, homogeneous groups of objects. In the present case it is beaches and sea fronts of Mumbai with respect to marine water quality parameters. Different measures for the similarity with respect to distance between parameters and different algorithms for finding a cluster are applied[16] in the present study.

The agglomerative hierarchical cluster analysis according to Ward[17] with squared Euclidean distances was applied to detect multivariate similarities in marine water quality. In agglomerative hierarchical clustering, clusters are formed by grouping cases into bigger and bigger clusters until all cases (here beaches and sea fronts) are member of a single cluster. According to Liu (2003)[8], the first step in cluster analysis should be data standardization (mean = 0; variance =1). Standardization tends to increase the influence of variables whose variance is small and reduce the influence of those whose variance is large. This will also eliminate the effect of scale of measurement of data.

The squared Euclidean distance (D^2) between location I and location II is calculated from the standardized values as follows

$$D^2 = (Z_{BOD\ I} - Z_{BOD\ II})^2 + (Z_{FC\ I} - Z_{FCII})^2 + \ldots \quad (1)$$

where Z_{BODI} & Z_{BODII} are the standardized values of BOD at locations I and II. Similarly, $Z_{FC\ I}$ and Z_{FCII} are similar values of FC. The squared Euclidean distance between these two locations or cases is labeled as coefficient. The first two cases combined are those with the smallest distance or greatest similarity.

The Wards method[17] is distinct from all other methods because it uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the within cluster variability of clusters that are formed at each step. The Wards method yields clearly structured and relatively stable clusters or groups over a wide range of similarities. In general, this method is regarded as very efficient . However, it tends to create clusters of small size. Among the various methods available in the literature, Ward's method has been used in the present study. Temporal cluster analysis was carried out for the whole data. Spatial cluster Analysis were conducted for each season viz. summer, winter and monsoon.

Discriminant Anlysis (DA)

DA is a method of analyzing dependence that is a special case of canonical correlation used to analyze dependence. Discriminant Analysis (DA) was performed on the original data without affecting the results and comparability with other chemometrics methods and constructed a discriminant function for each group[18] as follows:

Table 1—Statistical Descriptives of Water Quality Parameters

Site		temperature	pH	turbidity	DO	BOD	Ammonical Nitrogen	Phos- phate	TC*	FC**	FS***
Bandra	Mean	28.8	7.7	35.0	6.6	1.7	0.4	0.1	1.0	7.4	4.6
	Stdev	2.1	0.3	28.9	0.6	0.6	0.3	0.1	2.0	14.0	9.3
Breach Candy	Mean	28.6	7.7	35.2	6.3	2.3	1.3	0.1	0.8	2.0	4.2
	Stdev	2.0	0.3	20.8	0.6	1.6	4.8	0.0	1.8	1.1	5.6
Colaba	Mean	28.6	7.6	36.0	6.0	2.4	0.4	0.1	0.8	3.3	2.8
	Stdev	2.2	0.5	17.5	0.7	2.1	0.5	0.1	1.2	4.3	2.8
Dadar	Mean	29.1	7.7	38.8	5.0	6.1	1.8	0.3	2.1	8.6	8.0
	Stdev	2.0	0.4	17.8	1.1	2.7	1.4	0.3	3.0	15.3	15.2
Girgaon	Mean	29.0	7.7	50.6	5.7	3.7	0.3	0.1	1.5	7.4	7.5
	Stdev	2.0	0.3	42.2	0.6	1.6	0.3	0.1	2.5	12.7	12.0
Juhu	Mean	29.4	7.8	43.6	6.2	2.5	0.6	0.1	7.5	5.5	3.2
	Stdev	1.7	0.2	28.4	0.6	1.0	0.5	0.1	1.9	15.7	4.6
Khar	Mean	29.1	7.8	40.3	6.5	3.0	0.4	0.1	1.5	11.3	6.1
	Stdev	2.3	0.2	30.4	0.8	1.9	0.3	0.1	2.4	21.0	11.1
Madh	Mean	28.9	7.8	33.4	6.5	2.2	0.2	0.1	0.4	2.6	5.5
	Stdev	1.8	0.1	21.7	0.5	1.7	0.2	0.1	0.7	4.1	12.9
Mahim	Mean	29.0	7.7	35.0	3.7	10.4	3.2	0.5	13.8	89.3	15.2
	Stdev	2.1	0.3	21.3	1.8	5.3	2.1	0.5	29.5	218.0	22.6
Nariman Point	Mean	28.9	7.6	31.0	5.8	1.9	0.3	0.1	0.7	3.3	53.9
	Stdev	2.0	0.4	16.5	0.8	1.0	0.2	0.0	1.2	6.9	7.0
Versova	Mean	29.0	7.8	42.8	6.1	2.7	0.6	0.2	1.6	5.5	4.4
	Stdev	1.9	0.2	25.0	0.6	2.0	0.3	0.5	2.0	7.2	6.8
Worli	Mean	28.9	7.7	30.3	5.8	2.7	0.6	0.2	1.2	6.4	4.8
	Stdev	1.9	0.2	21.2	0.9	1.6	0.5	0.3	2.1	9.6	5.8

*-10^6 times the value **-10^5 times the value *** -10^4 times the value

$$F(G_i) = k_i + \sum_{j=1}^{n} w_{ij}.p_{ij}$$

where i is the number of groups (G), k_i is a constant inherent to each group, n is no of parameters used to classify a set of data into a given group, and w_{ij} is the weight coefficient, assigned by DA to a given parameter (p_{ij}). There are two most commonly used algorithms for variables selection. All parameters, and step wise method. In a stepwise method the first variable included in this analysis has the largest acceptable value for the selection criterion. The selection criterion is the minimization of Wilks lamda. Wilks lambda is = $SS_{within\text{-}groups}/SS_{total}$ After the first variable is entered the value of the criterion is reevaluated for all variables not in the model and the variable with the largest acceptable criterion is entered next. At this point the variable entered first is reevaluated to determine whether it meets the removal criterion. If it does, it is removed from the model. Variables are removed until none remain that meet the removal criterion. Variable selection terminates when no more variables meet entry or removal criterion. DA was performed on the original data based on two different modes i.e. all parameters and step wise to construct the best discriminant functions to confirm the clusters determined by means of CA.

Principal Component Analysis (PCA)

PCA starts by building the correlation matrix for the data. Diagonalization of this matrix provides its eigen values and eigen vectors. Since the variance explained by each eigen vector is proportional to its eigen value, only those eigen vectors with eigen values greater than 1 are selected as significant independent variables (components). Sum of eigen values is equal to the total no of variables. Correlation of Principal components and original variables is called loadings. The eigen vectors or components are more easily interpretable if a VARIMAX rotation, which transfers the eigen vectors to make each of them representative of individual sources of variation, is applied[19-21]. Then each component may be identified as a source of pollution by determining its most inter related parameters.

Data Treatment

Cluster analysis requires variables to conform to a normal distribution. The normality of the distribution of each variable was checked by analyzing kurtosis and skewness before multivariate statistical analysis[18]. Skewness and kurtosis values were high for the original data and were in the range -1.6 to 7.7 and 0.25 to 65.3 respectively. After log transformation of all parameters all skewness and kurtosis values were significantly reduced to ranges from -0.409 to 0.58 and -0.666 to 1.404 respectively, which were less than the critical values. For CA all parameters were also z-scale standardized (mean = 0 and variance = 1) to minimize the effects of differences in measurement units and variance to render the data dimensionless[21].

Results and Discussion

Temporal Grouping

An initial exploratory approach involved the use of hierchical CA on standardized log- transformed data sorted by season. Temporal CA generated a dendrogram as shown in Fig. 2 grouping the 12 months into three clusters. Cluster 1 comprised July, August and September representing the rainy season. April and May formed the second cluster representing the Summer season. The rest of the months October, November, December, January, February and March clustered together representing winter season. The temporal pattern of marine water quality was to some extent consistent with the three seasons with only exception of October and March getting clustered with the winter months which are transition months. Temporal variation in marine water quality is not absolutely determined by seasonal effects but also the nature and frequency of discharge.

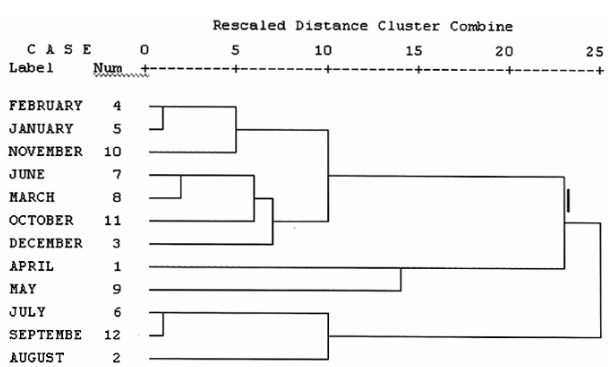

Fig. 2—Dendrogram of temporal clustering of monitoring periods

Spatial Grouping

Winter Season: The dendrograms showing the spatial clustering for Winter is given in Fig. 3. The dendrogram shows the clusters being combined and the values of the coefficients at each step. The dendrogram plots do not show the actual distances but rescales them to numbers between 0 and 25. Thus the ratio of the distances between steps is preserved. The results clearly show that in winter, Cluster I comprised of only Mahim as it is distinctly more polluted than the rest. Versova is the site, which is next heavily polluted location. All other ten sites have similar features and were affected by similar sources and corresponds to relatively cleaner areas.

Summer Season: Fig. 4 gives the dendogram of cluster analysis for summer season. The results show that in summer, Cluster I comprised of Mahim and Dadar as they are distinctly more polluted than the rest. Colaba, Madh, Girgaun and Breach Candy form the second cluster and have similar features and were affected by similar sources and corresponds to relatively cleaner areas. Other six sites form the third group which are moderately polluted.

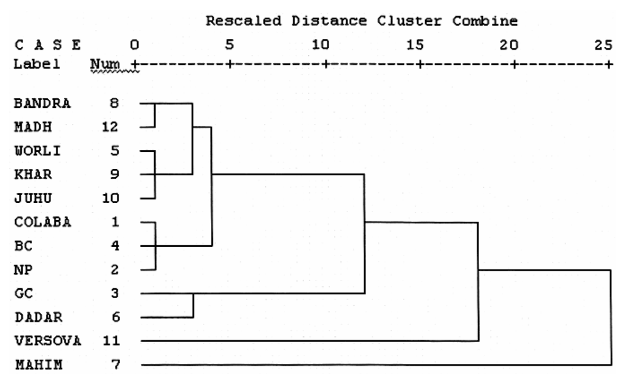

Fig. 3—Dendrogram of spatial clustering for winter season

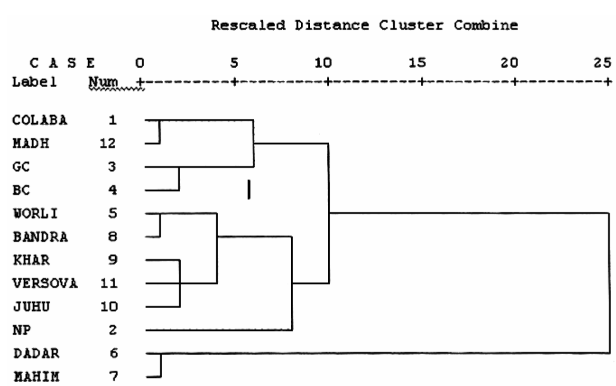

Fig. 4—Dendrogram of spatial clustering for summer season

Monsoon Season: The dendrogram for monsoon is given in Fig. 5. The results show that in monsoon Cluster I comprised of only Mahim as it is distinctly much more polluted than the rest. Cluster II comprises of Girgaon, Worli and Dadar. All other eight sites have similar features.

The results clearly show that through out the year, Cluster I comprised of only Mahim as it is distinctly much more polluted than the rest. Dadar and Versova are the sites which are next heavily polluted. All other nine sites have similar features and were affected by similar sources and correspond to relatively cleaner areas.

Temporal/Spatial variation in marine water quality

Temporal variation in marine water quality was evaluated using DA, with clusters based on temporal CA. The objective of the DA was to test the significance of discriminant functions and determine the most significant variables associated with the differences among clusters. Wilks lambda and the Chi square for each discriminant function ranged between 0.001 to 0.12 and from 11.68 to 36.85 respectively at $p < 0.011$ showing that the temporal DA was effective and meaningful (Table 2).

The discriminant functions obtained from the standard, stepwise modes of DA are shown in Table 3. In the step wise mode variables are included step by step beginning with the most significant, until no significant changes are obtained. The standard DA mode constructs DFs containing all parameters. The stepwise mode suggests that DO, TC, NH_3 and pH are significant parameters for discriminating among the three periods for the temporal variation of water quality. Both the standard and stepwise mode DFs rendered the corresponding group memberships correctly assigning 100% of the cases in the three groups.

The spatial DA was performed using the original set of 12 sites after classification into three major groups obtained from the spatial CA. Sites were the dependent variables and the measured parameters were the independent variables. Wilks lambda and the Chi square for each discriminant function ranged between 0.00 to 0.45 and from 6.73 to 58.1 respectively at $p < 0.009$ showing that the spatial DA was effective and meaningful except for monsoon season when $p = 0.143$ for standard mode (Table 4).

The stepwise mode suggests that DO and temperature are significant parameters for

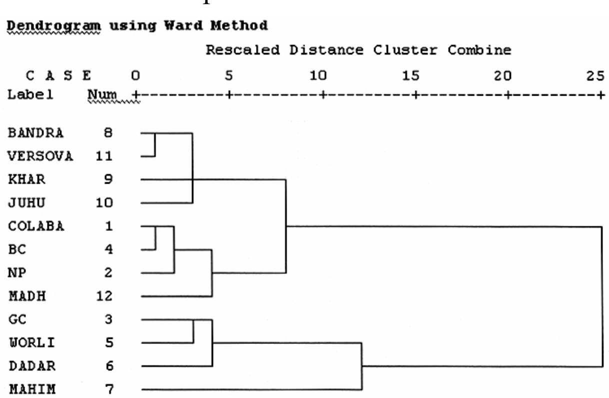

Fig. 5—Dendrogram of spatial clustering for monsoon season

Table 2—Results of Temporal DA

	Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
Standard	1 through 2	0.001	36.856	8	0.000
	2	0.016	16.621	6	0.011
Stepwise	1 through 2	0.001	36.856	8	0.000
	2	0.120	11.682	3	0.009

Table 3—Standardized Canonical Discriminant Function Coefficients for DA of Temporal variation

	Standard mode		Stepwise	
	1	2	1	2
pH	8.217	2.568	6.643	0.209
Turbidity	1.605	0.871		
DO	1.844	3.736	1.272	1.619
BOD	0.153	-1.565		
NH_3	-9.033	-2.533	-7.481	-0.446
PO_4	-1.883	-2.011		
TC	4.704	4.158	3.114	1.419

Table 4—Results of Spatial DA

	Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
		Summer			
Standard	1 through 2	0.000	58.054	18	0.000
	2	0.017	20.448	8	0.009
Step wise	1 through 2	0.022	32.392	4	0.000
	2	0.453	6.730	1	0.009
		Monsoon			
Standard	1 through 2	.000	43.199	18	0.001
	2	.087	12.191	8	0.143
Step wise	1 through 2	.001	52.862	6	0.000
	2	.153	15.020	2	0.001
		Winter			
Standard	1 through 2	.000	54.523	16	0.000
	2	.009	25.940	7	0.001
Step wise	1 through 2	.000	68.874	4	0.000
	2	.025	31.315	1	0.000

Table 5—Standardized Canonical Discriminant Function Coefficients for DA of spatial variation for three seasons

	Summer				Monsoon				Winter			
	Standard		Step wise		Standard		Step wise		Standard		Step wise	
	1	2	1	2	1	2	1	2	1	2	1	2
Temperature	18.42	-1.77	0.03	1.02	3.97	-1.05			1.57	-0.15	0.97	-0.30
Turbidity	6.49	2.21			5.56	-2.82			0.96	-0.44		
BOD	-19.30	2.94			5.81	-0.94			6.85	6.50		
FC	25.99	-4.95			3.34	-0.71			-0.49	0.25		
FS	28.26	4.57			2.31	0.37	-0.03	1.19	-0.63	1.11		
DO	-10.82	-0.42	0.99	-0.23	-4.78	1.48	-0.76	1.25	-8.75	-6.47		
NH_3	-8.87	0.48			-3.48	0.81			0.55	0.38		
PO_4	-12.74	5.94			-1.98	1.20			0.16	0.96	0.15	1.00
TC	-21.84	-7.25					1.25	-0.05				

discriminating among the three groups for the spatial variation of water quality for summer season (Table 5). The standard and stepwise modes rendered the corresponding group memberships, correctly assigning 100% and 91.6% of the cases of the three groups. The parameters significant are FS, DO and TC for monsoon season and temperature and PO_4 for winter season. Both the standard and stepwise modes rendered the corresponding group memberships correctly assigning 100% of the cases in the three groups for both monsoon and winter seasons

Based on the above results, hierarchal CA provided a classification of marine water quality of Mumbai beaches and sea fronts that aids in designing an optimal spatial monitoring plan with a sharply reduced number of monitoring sites. Concretely, the number of monitoring sites could be reduced and could only be chosen from group A , B and C. Further more the pollution at Mahim, and Dadar is extremely high indicating direct sources, which can be treated well before discharge. As identified by DA, box and whisker plots of the selected parameters showing seasonal trends are given in Fig. 6.

Source identification

Prinicipal component analysis was performed on the standardized log transformed normalized data set. Only the PCs with eigenvalues greater than 1 were considered. PCA of the whole data set yielded 4 data sets explaining 81.4% of the total variance. Table 6 summarizes the PCA results comprising the loadings, eigenvalues and percentage of variance. First factor which explained 30.6% of the total variance is correlated with DO, BOD, TC, FC and FS. Thus factor 1 represented organic pollution from domestic waste water and sewage treatment works. The second factor which explained 20.4% of the total variance is due to turbidity and Phosphate. Phosphate is an important component of detergents. This could be natural pollution which included the run off from agricultural fields. The third factor is related to pH and Ammonical Nitrogen and it represents nitrogenous nutrient pollution. The fourth factor is related to temperature and it represents seasonal effects of temperature.

Table 6—Rotated Component Matrix[a]

	Component			
	1	2	3	4
Temperature	.149	-.230	.074	.641
pH	.052	.021	.993	-.014
Turbidity	.094	.985	.023	-.066
DO	.950	.053	.079	.101
BOD	.553	.176	-.166	-.323
NH_3	.052	.021	.993	-.014
Phosphate	.094	.985	.023	-.066
TC	.800	.096	-.110	.018
FC	.832	.018	.082	.164
FS	.950	.053	.079	.101
Eigenvalue	3.06	2.04	1.95	1.08
% of variance	30.6	20.4	19.6	10.8

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a]Rotation converged in 5 iterations.

Conclusions

The temporal pattern of marine water quality was to some extent consistent with the three seasons with only exception of October and March getting clustered with the winter months which are transition months.

The results of the spatial clustering for Winter shows Cluster I comprises of only Mahim as it is distinctly more polluted than the rest. Versova is the site, which is next heavily polluted location. All other ten sites have similar features and were affected by similar sources and corresponds to relatively cleaner areas. The results for summer show that Cluster I

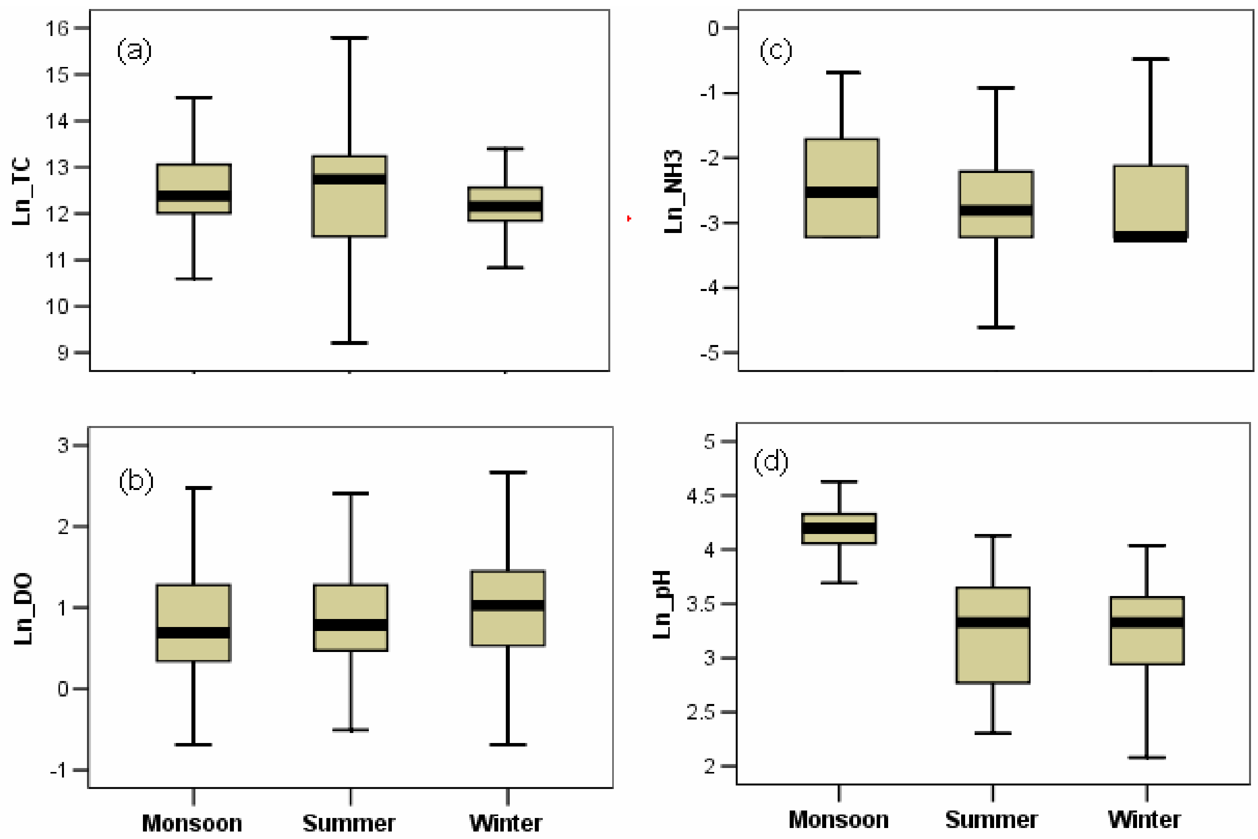

Fig. 6—Temporal variation of (a) In(pH) (b) In(DO) (c) In(NH_3) and (d) In(pH)

comprised of Mahim and Dadar. Colaba, Madh, Girgaun and Breach Candy form the second cluster or group. The rest of the sites form third cluster. Results show that in monsoon Cluster I comprised of only Mahim. Cluster II comprises of Girgaon, Worli and Dadar. All other sites have similar features.

DA gave the best results both spatially and temporally. It provided an important data reduction as it used only four parameters (DO, Total coliform, Ammonical nitrogen and TC) affording 100% correct assignment in temporal analysis. For spatial DA, DO and temperature; Feacal strptococii, DO and Total coliform; temperature and phosphate were used for summer, monsoon and winter seasons respectively. DA gave 100% correct assignment in spatial analysis except for summer season, step wise mode DA rendered 91.6% correct assignment. Therefore, DA allowed a reduction in the dimensionality of the large data set, delineating a few indicator parameters responsible for variations in water quality. PCA resulted in four factors explaining 81.4% of the total variance. The first factor obtained represents organic pollution from domestic waste water. The second factor represents natural pollution which includes the surface run off from agricultural fields. The third factor represents nutrient pollution whereas the fourth factor represents seasonal effects of temperature.

References

1 Singh K.P., Malik, A. & Sinha, S., Water Quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. *Analytica Chimica Acta*, 538, (2004) 355-374

2 Dixon, W. & Chiswell B., Review of aquatic monitoring program design. *Water Research,* 30(1996) 1935-1948

3 Shrestha, S. & Kazama, F., Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan, *Environmental Modeling & Software,* 22 (2007) 464-475

4 Shin P.K.S. & Fong K.Y. S., Multiple Discriminant Analysis of Marine Sediment Data *Marine Pollution Bulletin,* Vol. 39, Nos. 12,(1999) pp. 285-294

5 Carrasco, M., Lopez-Ramirez J.A., Benavente, J., Lopez-Aguayo, F. & Sales D., Assessment of urban and industrial contamination levels in the bay of Cadiz, SW Spain, *Marine Pollution Bulletin,* 46, (2003) 335-345

6 Bengraine, K. & Marhaba, T.F., Using principal component analysis to monitor spatial and temporal changes in water quality. *J. Hazard. Mater*. *B*, 100, (2003) 179-195

7 Voncina, D.B., Dobcnik, D., Novic, M. & Zupan, J., Chemometric characterization of the quality of river water, *Anal. Chim. Acta* 462 (2002) 87–100

8 Liu, C.W., Lin, K.H. & Kuo, Y.M., Application of factor analysis in the assessment of groundwater quality in a blackfoot disease are in Taiwan, *Sci. Total Environ*. 313 (2003) 77-89.

9 Reghunath, R., Murthy, T.R.S. & Raghavan, B.R., 'The utility of multivariate statistical techniques in hydro-geochemical studies: an example from Karnataka, India', *Water Research* 36 (2002) 2437-2442.

10 Astel, A., Biziuk, M.,Pryjazny, A. & Namiesnik, J., 'Chemometrics in monitoring satial and temporal variation in drinking water quality', *Water Research* 8 (2006) 1706-1716

11 Zhou Feng, Guo Huaicheng & Hao Zejia., Spatial distribution of heavy metals in Hong Kong's marine sediments and their human impacts: A GIS-based chemometric approach, *Marine Pollution Bulletin* (2007) in press.

12 Kowalkowskia, T., Zbytniewskia, R., Szpejnab J. & Buszewski, B., Application of chemometrics in river water classification, *Water Research* 40 (2006) 744-752.

13 Helena, B., Pardo, R., Vega, M., Barrado, E., Ferna´ ndez,
1. J.M. & Ferna´ ndez, L., Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis, *Water Research* 34 (2000) 807-816

14 Singh K.P., Malik, A. Singh V.K., Madan, D. & Sinha, S., Multivariate statistical techniques for the evaluation of spatial and temporal variation in water quality of Gomti river (India) a case study, *Water Research* 38 (2005) 3980-3992

15 Standard Methods, for the Examination of Water and Waste water', prepared and published by American public health Association, 17 th Edition

16 Henrion, R. & Henrion, G., *Multivariate Daten analyse*. (Springer, Berlin), 1995

17 Ward, J.H., Ward's method, *J. Am. Stat. Assoc.* 58 (1963) 236-246

18 Johnson, R.A. & Wichern, D.W., *Applied Multivariate Statistical Analysis,* 5th edition. (Prentice-Hall, New Jersey), 1992

19 Hopke P.K., *Receptor Modelling in Environ Metal Chemistry*, (Wiley-Interscience, New York), 1985

20 Thurston G.D. & Spengler J.D., A quantitative Assessment of Source Contributions to Inhalable Particulate Matter Pollution in Metropolitan Boston, *Atmospheric Environment*, 19 No. 1 (1985) 9-25

21 Zhou Feng, Guo Huaicheng, Liu Yong & Jiang Yumei., Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong, *Marine Pollution Bulletin* (2007) in press