Chemometrics data analysis of marine water quality in Maharashtra, west coast of India

Indian Journal of Geo-Marine Sciences
Vol. 42(1), February 2013, pp. 97-105

Chemometrics data analysis of marine water quality in Maharashtra, west coast of India

Indrani Gupta, Abhaysinh Salunkhe, Nanda Rohra & Rakesh Kumar
CSIR- National Environmental Engineering Research Institute, 89/B Dr. A.B. Road, Worli, Mumbai 400 018, India
[E-mail: i_gupta@neeri.res.in, abhay.salunkhe@rediffmail.com, nandarohra@ymail.com,r_kumar@neeri.res.in]

Received 5 July 2011; revised 18 March 2012

Various chemometrics methods were used to analyze data sets of marine water quality for 9 parameters measured at 34 different sites of Maharashtra from 2007 to 2009 to determine spatial variations in marine water quality and identify pollution sources. Hierarchical cluster analysis (CA) grouped the 34 monitoring sites into three groups based on similarities in marine water-quality characteristics. Discriminant analysis (DA) was important in data reduction because it used five parameters (DO, Ammonia, *p*H, FC and temperature) to correctly assign 96.4% of the cases. In addition, principal component analysis (PCA) identified three latent pollution sources for organic pollution, industrial pollution and fecal pollution. Furthermore, water quality index was calculated based on four parameters viz. *p*H, DO, BOD and FC.

[Keywords: Marine water quality, Cluster analysis, Discriminant analysis, Principal component analysis, Maharashtra]

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. However, with the growth of population and commercial industries, marine water has received large amounts of pollution from a variety of sources such as domestic sewage, industrial wastes, storm water drains which collect and channel the water from all non-porous surfaces, such as pavements and tarred roads, within catchments. Pollution sources at sea include accidental and deliberate discharges of oil, operational dumping of garbage (particularly plastics), and controlled dumping. It is therefore essential to prevent and control marine water pollution, and regularly implement monitoring programs which help to understand the temporal and spatial variations in marine water quality. In Maharashtra, to evaluate the state of health of marine water and its long-term changes and to measure compliance with the water quality standards, the Maharashtra Pollution Control Board (MPCB) implements a marine monitoring program. In the last decade, methods such as cluster analysis (CA), discriminant analysis (DA) and principal component analysis (PCA) have become accepted in identifying temporal and spatial variation and sources of pollution in river, and sea water. (Feng *et al*., 2007, Kowalkowski *et al*., 2006, Carrasco *et al*, 2003, Astel *et al*., 2006, Lou *et al*., 2006, Gupta *et al*., 2009). In the present study, a large data matrix, obtained during three years 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 variations in marine water quality and to also identify the pollution sources.

Materials and Methods

Sampling locations are shown in Fig. 1. Beaches, sea fronts and creeks were monitored by MPCB once every month. Data for 2007-2009 was used for analysis. There was very few data for the year 2007. Selected parameters included temperature, *p*H, DO, BOD, Faecal Coliform, Total Coliform, ammonical nitrogen, nitrate and phosphate.

An index is a mean device to reduce a large quantity of data down to a simplest form. Water quality indices help to evaluate the water quality profile of a river in its entire stretch and to identify the reaches where the gap between the desired and the existing water quality is significant enough to warrant urgent pollution control measures. In India the NSF WQI is being used by CPCB, with a slight modification in weights *(Abbasi, 2002, CPCB, 2001)*. The NSF WQI is expressed mathematically as:

$$\text{NSFWQI} = \sum_{i=1}^{p} W_i I_i,$$

where, I_i = sub index for ith water quality parameter
W_i = weight (in terms of importance) associated with water quality parameter
p = number of water quality parameters.

The modified weights (W_i) and the equation for the sub indices (I_i) as per CPCB, are given in Tables 1 and 2 respectively. The range of the NSF WQI corresponding to various designated best use classification is given in Table 3. Marine water quality and source identification were performed using several chemometrics methods, such as Cluster Analysis, Discriminant Analysis and Principle Component Analysis.

Cluster Analysis

CA is an unsupervised pattern detection method that partitions all cases into smaller groups or clusters of relatively similar cases that are dissimilar to other groups. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set, and is typically illustrated by a dendrogram. Dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between analytical values from the samples. In the present case it is beaches and sea fronts of Maharashtra with respect to marine water quality parameters. Hierarchical CA was applied on standardized data using Ward's method with squared Euclidean distances. Spatial variations in marine water quality were determined from hierarchical CA using linkage distance.

Discriminant function analysis

Discriminant analysis is a statistical technique to study the differences between two or more objects with respect to several variables simultaneously. Main purpose of a discriminant function analysis is to predict group membership based on a linear combination of the interval variables. Procedure begins with a set of observations where both group membership and the values of the interval variables are known. DA technique builds up a discriminant function which operates on raw data and this technique constructs a discriminant function (Johnson and Wichern, 1992). The end result of the procedure is a model that allows prediction of group membership when only the interval variables are known. A second purpose of discriminant function

Table 1— Original and Modified Weights for the computation of NSF WQI based on DO, Fecal Coliforms, *p*H and BOD*

Water Quality parameters	Original Weights from NSF WQI	Modified Weights by CPCB
DO	0.17	0.31
FC	0.15	0.28
*p*H	0.12	0.22
BOD	0.1	0.19
Total	0.54	1.00

* CPCB 2001

Table 2—Sub –Index Equations for Water Quality Parameters (NSF WQI)*

Water Quality Parameters	Range Applicable	Equation
DO (Percent saturation)	0-40% saturation	IDO = 0.18+0.66 × (% Saturation DO)
	40-100% saturation	IDO = -13.55+1.17 × (% Saturation DO)
	100-140% saturation	IDO = 163.34-0.62 × (% Saturation DO)
B.O.D.(mg/l)	0-10	IBOD = 96.67-7 (BOD)
	10-30	IBOD = 38.9-1.23 (BOD)
	> 30	IBOD = 2
*p*H	2-5	IpH = 16.1+7.35 × (*p*H)
	5-7.3	IpH = -142.67+33.5 × (*p*H)
	7.3-10	IpH = 316.96-29.85 × (*p*H)
	10-12	IpH = 96.17-8.0 × (*p*H)
	<2 , >12	IpH=0
Fecal Coliform (counts/100Lml)	$1\text{-}10^3$	IFC = 97.2-26.6 × log(FC)
	$10^3\text{-}10^5$	IFC = 42.33 -7.75 × log(FC)
	$>10^5$	IFC = 2

* Abbasi 2002

Table 3— NSF WQI for Various Designated Best Use*

Serial No	NSF WQI	Description of Quality (1978)	Class by CPCB	Remarks
1	63-100	Good to Excellent	A	Non polluted
2	50-63	Medium to Good	B	Non polluted
3	38-50	Bad	C	Polluted
4	38 & less	Bad to Very Bad	D, E	Heavily polluted

*CPCB 2001, Abbasi 2002

analysis is an understanding of the data set, as a careful examination of the prediction model that results from the procedure can give insight into the relationship between group membership and the variables used to predict group membership.

Principal Component Analysis

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 1,3.

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 (Johnson and Wichern 1992). Skewness and kurtosis values were high for the original data and were in the range -1.4 to 5.1 and -0.7 to 26.8 respectively. After log transformation of all parameters all skewness and kurtosis values were significantly reduced to ranges from -1.5 to 1.8 and -1.5 to 6.6 respectively. For Kurtosis all values were less than the critical values except for Ammonical Nitrogen and Nitrate. For skewness the values were slightly more 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 (Feng *et al*., 2006).

Results and Discussions

The sampling locations are shown in Figure 1. Mean and standard deviations of water quality parameters are given in Table 4. Fig. 2 gives the box plots of *p*H, DO, BOD, FC, TC, Ammonia, Nitrate and Temperature.

Water quality Index based four parameters showed that at Karamvane, Mandavi, Bhagawati Bunder and Ganapati phule were good to excellent. Water Quality Indices of 29 other sites varied between Medium to Good and Bad. Water Quality Indices of all the locations is given in Table 5.

Cluster Analysis

Spatial cluster analysis was carried out with average concentrations for three years. Very few data was available for the year 2007. Spatial cluster analysis produced a dendrogram with four groups (Fig. 3). Group 1 comprised of Karamvane, Mandavi, Bhagawati Bunder and Ganapati phule which were very clean beaches. Group 2 comprised of Ulhas creek at Versova, Uttan sea, Ulhas creek at Gaimukh, Vashi creek at vashi, Ulhas creek ar Mumrba, Bhayander, Vashi creek at Airol, Bassein creek and Ulhas creek ar Kolshet which were moderately polluted. Group 3 consists of Arnala Sea, Kalwa creek, Dadar Chowpatty, Malabar Hill, Charni Road, Gateway of India, Worli seaface, Haji Ali, Mahim, Nariman Point, Juhu, Versova, Elephanta caves, Sawta and Dahanu which were heavily polluted.

The results of Water quality index given in Table 5 and those of Cluster Analysis do not match since water quality index considers only 4 parameters viz. *p*H, BOD,DO and FC. The water quality index of Haji Ali is medium to good based on 4 parameters but the cluster analysis shows it to be heavily polluted. The reason being average TC is 2106 at Haji Ali where as the average TC at other locations vary between 50 to 832. Similarly the difference between results of WQI and cluster analysis for other sites can be explained.

Discriminant function analysis

Table 6 indicates whether there is a statistically significant difference among the dependent variable means for each independent variable. Only Ammonia and Nitrate are not statistically significant. Table 7 shows that. Discriminant analysis used five parameters (DO, Ammonia, *p*H, FC and temperature) to correctly assign the cases. Tables 8 and 9 gives the percentage of the variance accounted for by the 3 discriminant functions generated. Significant of the function is also shown. Because there are three groups' two discriminant functions were generated. Table 10 provides an indication of the success rate for prediction of membership of the grouping variable's categories using the discriminant function developed from the analysis. 96.4% of the original cases were correctly classified.

Principal Component Analysis

PCA were applied to standardized log-transformed data. Table 11 gives the trotated component matrix. Only the PCs with eigenvalues >1 were considered

Fig. 1— Location of monitoring sites

Table 4— Mean and Standard Deviation of significant parameters at Monitoring Locations

Serial No	Monitoring Site	Station code		*p*H	DO mg/l	BOD mg/l	FC/100 mL	TC/100 ml.	Ammonia-N mg/l	Nitrate-N mg/l	Phophate	Temp. ºC
1	Arnala Sea	135	Mean	7.9	4.4	10.5	195.8	439.1	11.9	1.0	75.1	28.1
			Stdev	0.2	1.2	3.8	182.9	357.4	45.1	0.5	259.8	2.0
2	Bassein creek	1316	Mean	7.9	5.1	8.3	355.6	894.7	0.3	1.4	--	27.0
			Stdev	0.3	1.1	2.1	316.9	669.3	0.2	0.8	--	1.4
3	Bhagawati Bunder	25	Mean	6.9	5.9	2.5	3.0	49.9	0.6	4.7	2.0	26.6
			Stdev	0.7	0.5	0.3	1.0	28.6	0.2	3.4	1.9	1.3
4	Bhayander	138	Mean	7.8	4.5	10.7	276.4	462.1	0.2	2.0	0.7	28.1
			Stdev	0.3	1.1	3.4	451.7	435.2	0.3	1.0	1.8	2.0
5	Charni Road	2166	Mean	7.7	4.7	13.8	417.6	829.5	0.5	1.4	0.3	32.4
			Stdev	0.3	1.1	12.3	475.0	641.0	1.1	1.0	0.1	1.9
6	Dadar Chowpatty	38	Mean	7.7	4.4	18.7	453.5	790.1	0.5	1.1	0.4	32.9
			Stdev	0.4	0.8	21.6	420.2	521.9	1.4	0.9	0.4	2.4
7	Dahanu	152	Mean	8.0	4.8	11.3	193.2	559.1	0.7	0.1	0.2	29.8
			Stdev	0.2	0.9	3.2	88.3	190.8	0.4	0.1	0.2	2.7
9	Ganapati phule	23	Mean	6.8	6.0	2.4	2.5	52.1	0.5	4.8	1.6	26.8
			Stdev	0.7	0.4	0.3	1.3	31.1	0.2	3.4	1.7	1.1
10	Gateway of India	2165	Mean	7.7	4.6	15.7	467.1	682.5	0.5	1.3	--	32.4
			Stdev	0.4	1.0	17.4	469.4	405.6	1.1	0.6	--	2.2
11	Haji Ali	36	Mean	7.8	4.3	14.1	377.2	2106.4	0.3	1.1	0.3	32.9
			Stdev	0.4	1.2	12.2	395.7	5269.7	0.4	0.7	0.3	2.4
12	Juhu	40	Mean	7.9	4.6	11.0	328.3	493.2	0.2	1.3	0.2	31.9
			Stdev	0.4	0.6	3.0	398.0	503.1	0.3	1.1	0.2	1.8
13	Kalwa creek	129	Mean	7.3	4.3	14.2	364.1	755.0	1.6	2.4	0.8	27.7
			Stdev	0.4	1.2	17.0	274.1	514.3	2.0	1.6	0.4	2.2
14	Karamvane	155	Mean	7.3	6.1	2.4	3.5	73.3	0.5	1.2	0.8	25.8
			Stdev	0.5	1.2	0.7	1.5	36.4	0.2	0.8	0.0	3.3
15	Mahim	1318	Mean	7.8	3.5	18.9	304.7	831.7	1.1	1.6	30.9	3.5
			Stdev	0.7	1.8	10.9	352.8	663.5	2.1	0.9	1.1	1.8
16	Malabar	35	Mean	7.7	4.2	14.5	419.3	814.6	0.5	1.1	0.3	33.2
			Stdev	0.4	1.2	11.3	433.8	490.5	1.2	0.7	0.4	2.4
17	Mandavi	26	Mean	7.3	6.1	2.4	3.5	73.3	0.5	1.2	1.2	25.8
			Stdev	0.5	1.2	0.7	1.5	36.4	0.2	0.8	0.6	3.3
19	Nariman Point	33	Mean	7.7	4.2	47.1	347.5	701.9	0.6	1.2	0.7	32.8
			Stdev	0.5	0.9	153.3	392.6	527.6	1.1	0.8	0.9	2.4
22	Sawta	151	Mean	7.6	5.5	8.5	112.0	362.5	0.2	0.5	0.2	29.9
			Stdev	1.8	1.2	3.7	25.7	135.0	0.2	0.3	0.1	3.1
23	Elephanta caves	1317	Mean	7.7	4.2	10.1	95.4	371.2	0.3	1.5	32.0	32.0
			Stdev	0.3	0.8	1.4	74.7	266.0	0.4	0.7	3.2	3.2
25	Ulhas creek at Versova	137	Mean	7.6	4.4	13.1	248.4	519.2	0.1	2.3	0.2	28.0
			Stdev	0.3	1.4	6.9	288.7	507.3	0.2	1.3	0.2	1.5
26	Ulhas creek at Gaimukh	134	Mean	7.5	4.9	10.3	254.8	562.3	0.2	2.8	0.3	27.5
			Stdev	0.4	1.4	3.6	149.3	304.3	0.2	1.8	0.1	1.9
27	Ulhas creek Mumbra Reti bunder	128	Mean	7.5	5.1	9.0	328.6	673.6	0.5	2.3	0.2	28.2
			Stdev	0.3	1.3	2.7	301.6	606.8	0.6	1.9	0.1	2.6
28	Ulhas creek Kolshet Reti bunder	133	Mean	7.4	4.8	9.4	508.4	738.2	0.5	36.8	0.3	27.4
			Stdev	0.4	1.2	2.7	524.4	592.7	0.5	136.9	0.2	2.0
29	Uttan sea	139	Mean	7.9	4.6	10.6	140.9	387.2	0.1	1.7	0.1	27.6
			Stdev	0.4	1.0	3.5	76.9	262.3	0.2	1.0	0.1	1.3

contd

Table 4— Mean and Standard Deviation of significant parameters at Monitoring Locations (contd…)

Serial No	Monitoring Site	Station code		pH	DO mg/l	BOD mg/l	FC/100 ml	TC/100 ml.	Ammonia-N mg/l	Nitrate-N mg/l	Phophate	Temp. ºC
			Stdev	0.5	1.0	3.7	403.3	498.5	0.3	0.9	--	2.9
31	Vashi creek at vashi bridge	2185	Mean	7.5	4.5	10.9	301.0	601.1	0.3	1.8	--	28.2
			Stdev	0.3	0.8	3.5	261.1	445.1	0.3	0.7	--	3.5
32	Versova	2169	Mean	7.8	4.7	10.9	205.5	333.5	0.3	1.1	--	31.6
			Stdev	0.3	1.0	3.7	354.1	375.5	0.4	0.6	--	1.6
33	Worli seaface	2167	Mean	7.7	4.8	20.4	351.9	653.7	0.3	1.2	3.0	32.0
			Stdev	0.3	1.1	33.8	310.3	501.5	0.7	0.5	8.6	2.7

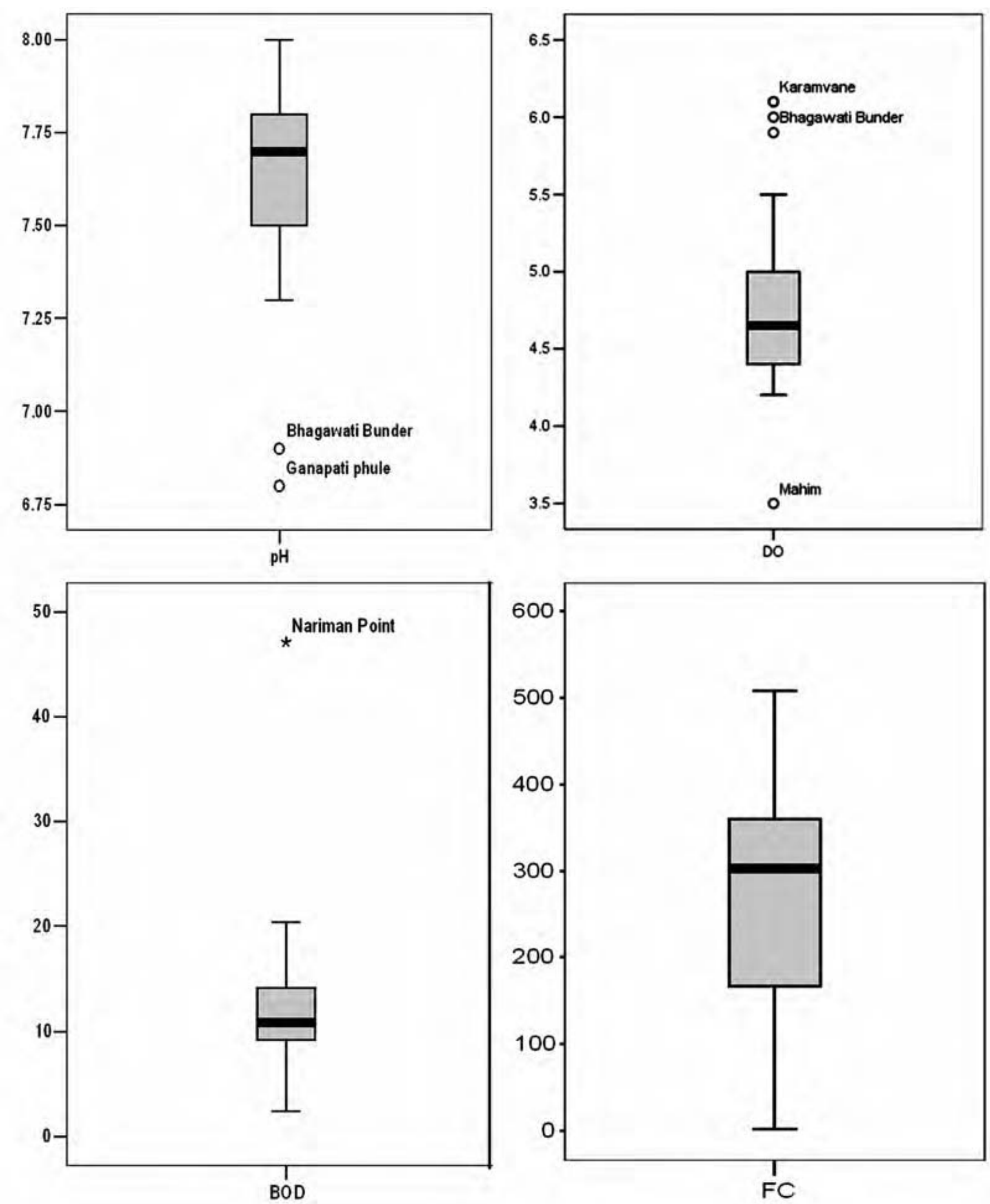

Fig. 2—Box plots of pH, DO, BOD, FC, TC, Ammonia, Nitrate and Temperature

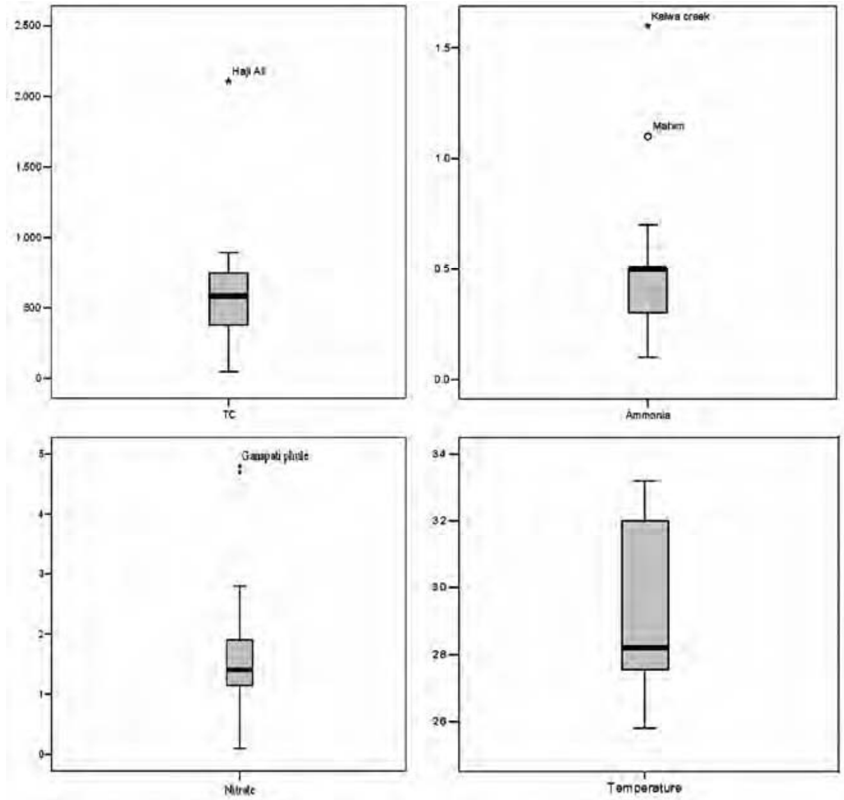

Fig. 2—Box plots of *p*H, DO, BOD, FC, TC, Ammonia, Nitrate and Temperature (Contd…)

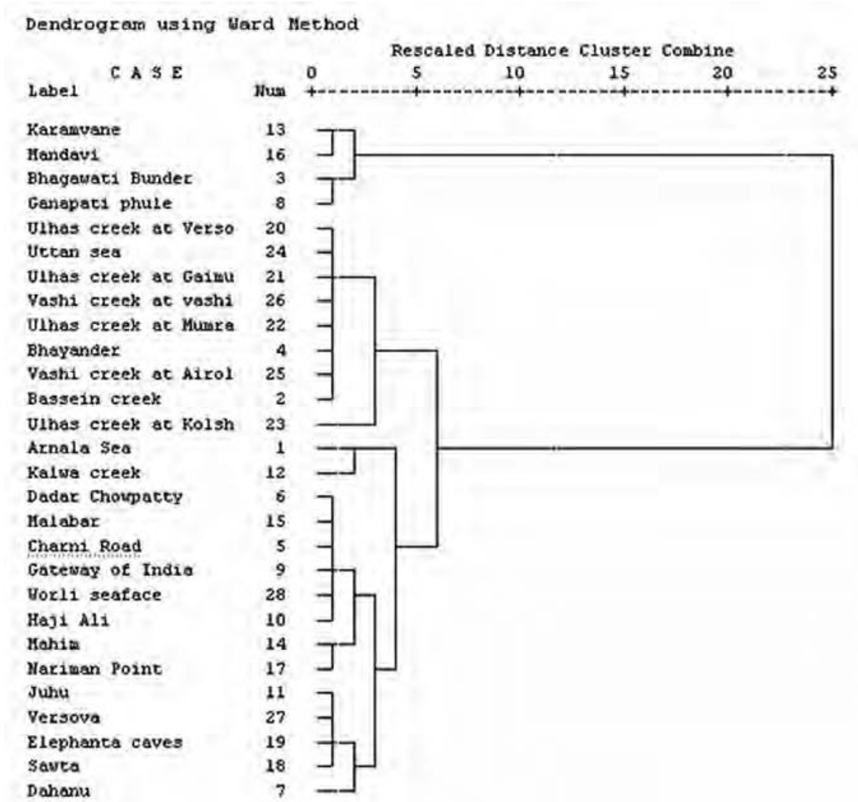

Fig. 3—Dendrogram showing clustering of Monitoring sites

Table 5 — Yearly and Average Water Quality Index of Sea and Creeks of Maharashtra

Serial No.	Station Name	2007	2008	2009	Avg WQI	Category
1	Bhagawati Bunder		78	77	78	GE
2	Ganapatipule		79	80	80	GE
3	Gateway of India		50	46	48	Bad
4	Arnala Sea		55	45	50	MG
5	Bassein creek	49	57	58	55	MG
6	Bhyander Creek		53	49	51	MG
7	Charni road choupathy		52	50	51	MG
8	Dadar Choupati		49	48	48	Bad
9	Dahanu Creek		53	48	50	MG
11	Haji Ali		52	50	51	MG
12	Juhu Beach		54	45	49	Bad
13	Kalwa Creek		52	45	48	Bad
14	Karambhane Creek		82	84	83	GE
15	Mahim creek	44	56	55	52	MG
16	Malbar Hill		49	48	49	Bad
17	Mandavi Bunder		80	80	80	GE
18	Nariman Point		48	48	48	Bad
19	Sawta Creek		59	51	55	MG
20	Thane creek at Elephanta Island	48	61	59	56	MG
21	Ulhas creek at Gaimukh		54	56	55	MG
22	Ulhas creek at Kolshet Reti bunder		55	50	53	MG
23	Ulhas creek at Mumbra Reti bunder		58	50	54	MG
24	Ulhas creek at Versova bridge.		60	45	53	MG
25	Uttan Sea		54	45	50	Bad
26	Vashi creek at Airoli bridge		56	52	54	MG
27	Vashi creek at Vashi bridge		50	53	52	MG
28	Versova		57	52	54	MG
29	Worli seaface		55	50	53	MG

Table 6 —Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
*p*H	.379	20.517	2	25	.000
DO	.298	29.463	2	25	.000
BOD	.660	6.430	2	25	.006
FC	.470	14.090	2	25	.000
TC	.667	6.240	2	25	.006
Ammonia	.950	.653	2	25	.529
Nitrate	.899	1.411	2	25	.263
Temperature	.267	34.398	2	25	.000

Table 7 —Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Temperature	1.00	34.40	
2	Temperature	0.97	17.16	0.30
	DO	0.97	14.09	0.27
3	Temperature	0.63	29.10	0.29
	DO	0.92	13.09	0.17
	Ammonia	0.64	5.88	0.12
4	Temperature	0.62	26.43	0.19
	DO	0.92	7.44	0.09
	Ammonia	0.62	5.83	0.08
	pH	0.97	5.12	0.08
5	Temperature	0.54	28.79	0.14
	DO	0.91	3.59	0.05
	Ammonia	0.62	5.21	0.06
	*p*H	0.87	6.55	0.06
	FC	0.78	4.89	0.06

Table 8— Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.415(a)	71.4	71.4	.930
2	2.570(a)	28.6	100.0	.848

a First 2 canonical discriminant functions were used in the analysis

Table 9— Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.038	75.349	10	.000
2	.280	29.268	4	.000

Table 10— Classification Results

	Group	Predicted Group Membership			Total
		1	2	3	
Count	1	4	0	0	4
	2	0	9	0	9
	3	0	1	14	15
%	1	100	0	0	100
	2	0	100	0	100
	3	0	6.7	93.3	100

96.4% of original grouped cases correctly classified.

Table 11— Rotated Component Matrix (a)

	Component 1	2	3
ln pH	.701	.553	-.107
ln DO	-.896	-.042	-.162
ln BOD	.938	.156	.011
ln FC	.956	.039	-.080
ln TC	.941	.074	-.057
ln NH_3	-.001	.021	.994
ln NO_3^-	-.051	-.978	-.051
ln Temp	.673	.399	-.021

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 4 iterations.

essential. The first factor which explained 55.3% of the total variance had strong positive loadings on *p*H, BOD, FC, TC and temperature, and negative loadings on DO. Thus, the first factor represented organic pollution from domestic wastewater and sewage treatment works. The inverse relationship between temperature and DO is a natural process because warm water easily becomes saturated with oxygen, and thus, can hold less dissolved oxygen. The second factor (18.2% of the total variance) included *p*H, which may be interpreted as industrial pollution. The third factor which explains 12.98% of the variance had strong positive loading on ammonia and thus represented nitrogenous nutrient/fecal pollution.

Conclusions

The results of the spatial clustering shows Cluster I comprises of Karamvane, Mandavi, Bhagawati Bunder and Ganapati phule which were very clean beaches. Group 2 comprised of Ulhas creek at Versova, Uttan sea, Ulhas creek at Gaimukh, Vashi creek at vashi, Ulhas creek at Mumra, Bhayander, Vashi creek at Airol, Bassein creek and Ulhas creek at Kolshet which were moderately polluted. Group3 consists of Arnala Sea, Kalwa creek, Dadar Chowpatty, Malabar Hill, Charni Road, Gateway of India, Worli seaface, Haji Ali, Mahim, Nariman Point, Juhu, Versova, Elephanta caves, Sawta and Dahanu which were heavily polluted. DA provided an important data reduction and used only five parameters (DO, FC, Ammonical nitrogen, *p*H and temperature) affording 96.4% correct assignment in spatial analysis. 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 three factors explaining 86.5% of the total variance. First factor obtained represents organic pollution. Second factor may be interpreted as industrial pollution. Third factor represented nitrogenous nutrient/fecal pollution.

References

1 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

2 Abbasi S A, 2002, Water Quality Indices State–of–the– Art, Pondicherry: Pondicherry University, Centre for Pollution Control & Energy Technology.

3 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* 54 (2007), 745–756

4 CPCB, 2001, Environmental Atlas of India, New Delhi: Central Pollution Control Board.

5 McKenna Jr., J E, 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling & Software 18 (3), 205e220.

6 Otto M, 1998. Multivariate methods. In: Kellner R, Mermet J M, Otto M, Widmer H M (Eds.), Analytical Chemistry. WileyeVCH, Weinheim.

7 Johnson R A, Wichern D W, 1992. Applied Multivariate Statistical Analysis.Prentice-Hall, Englewood Cliffs, NJ.

8 Tomasz Kowalkowski, Radosław Zbytniewski, Jacek Szpejna, Bogusław Buszewski Application of Chemometrics in river water classification, Water *Research*, *Volume 40, Issue4*, *February2006*, *Pages* 744- 752

9 Astel M, Biziuk A, Przyjazny J, Namiesnik, Chemometrics in monitoring spatial and temporal variations in drinking water quality Water Research 40 (2006) 1706 – 1716

10 Zhi-Ping Lou, Cui-Ci Sun, Mei-Lin Wu, Shu-Hua Han, Multivariate statistical analysis of water quality and phytoplankton characteristics in Daya Bay, China, from 1999 to 2002, Oceanologia, 48 (2), 2006. pp. 193–211.

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