Environmental Research Journal 6 (3): 141-150, 2012
ISSN: 1994-5396

Application of Multivariate Statistical Methods to Assessment of Water Quality in Selected Locations of the Lagos Lagoon, Nigeria

[1]M.K. Ladipo, [2]V.O. Ajibola and [3]S.J. Oniye
[1]Department of Polymer and Textile Technology, Yaba College of Technology, Lagos, Nigeria
[2]Department of Chemistry, [3]Department of Biological Sciences,
Ahmadu Bello University, Zaria, Nigeria

Abstract: Multivariate statistical methods, i.e., Cluster Analysis (CA), Principal Component Analysis (PCA) and Discriminant Analysis (DA) were used to assess temporal and spatial variations in the water quality of the Lagos Lagoon during the wet period (July 2007 and 2008) anddry seasons (February 2008 and 2009). The study was focused on nine locations of the lagoon, specifically to describe the distribution of water physicochemical parameters and identify the parameter (s) that most influence the distributions observed. Physicochemical parameters (pH, EC, salinity, turbidity, DO, BOD_5, COD, TSS, TDS, alkalinity, NO_3, PO_4 and SO_4) were used to study spatial and temporal variations in water quality of these locations. The descriptive statistics of the average values obtained for each location during the period of study were discussed. The results obtained from the detailed chemical analysis of water from the different sections of the lagoon confirmed the dynamic nature and diverse chemistry of the water. Multivariate analysis of obtained data during the periods of study further reflects this diversity during each of the periods samples were collected. The loading pattern of principal components showed some variations during each of the period of sample collection. The processes or sources associated with the principal components obtained during the different sampling periods are highly localized and contributed mainly by anthropogenic sources. Hierarchical CA grouped the nine locations into three based on the water characteristics during each period of sample collection. Hierarchical CA and PCA did not give a clear trend in temporal distribution of the parameters. As a result it was difficult to determine a constant similarity between locations during these periods however, DA showed EC and TDS were the only good predictors or discriminant variables in all the locations during the period of investigation.

Key words: Water quality, spatial and temporal differences, multivariate analysis, Lagos Lagoon, Nigeria

INTRODUCTION

Lagos is the commercial nerve centre and one of the most densely populated cities in Nigeria and therefore has long been under pressure due to industrial, commercial and population growth. The Lagos Lagoon which is a large body of shallow water extends from the Republic of Benin in West towards the Niger Delta in the East. The lagoon has been dump for treated and untreated industrial wastes, refuse, domestic wastes, sewage and oil spills, etc. and it is important to monitor the water quality and interpret the temporal and spatial variations in the water quality since, the lagoon still serves a source of food to the people.

One very difficult task facing environmental managers is to convert complex data to information for better defining the sources and typology of the pollution. Environmental data is characterized by high variability because of a variety of natural and anthropogenic influences. The best approach to avoid misinterpretation of environmental data is the application of multivariate statistical methods for environmental data classification and modelling (Reisenhofer *et al*., 1996; Boyacioglu and Boyacioglu, 2008). Several researchers have used these statistical methods in the interpretation of water quality data (Mazlum *et al*., 1999; Grande *et al*., 2003; Simeonova *et al*., 2003; Liu *et al*., 2003; Koonce *et al*., 2006; Zhou *et al*., 2007; Sojka *et al*., 2008; Kumar and Riyazuddin, 2008; Michalik, 2008; Gupta *et al*., 2009; Belkhiri *et al*., 2011). In this study, various multivariate statistical methods such as Principal Component Analysis (PCA), Cluster Analysis (CA) and Discriminant Analysis (DA) were used to help in the interpretation of the complex data sets, obtained during rainy and dry seasons water quality monitoring programs allowing the extraction of latent information about the similarities or dissimilarities among the monitoring sites or periods and also identify water quality variables responsible for temporal and

Corresponding Author: M.K. Ladipo, Department of Polymer and Textile Technology, Yaba College of Technology, Lagos, Nigeria

spatial variations in water quality. The selected parameters for the estimation of water quality characteristics were: pH, EC, salinity, turbidity, DO, BOD_5, COD, TSS, TDS, alkalinity, NO_3, PO_4 and SO_4.

MATERIALS AND METHODS

Sample collection and analytical procedures: Water samples were collected from nine locations of the lagoon in June 2007 and 2008 and February 2008 and 2009, representing rainy and dry seasons. Sampling locations were chosen in order to cover various anthropogenic activities including waste disposal. Grab samples were collected at 20 cm below the water level using a water sampler. Samples for major ions and other inorganics were collected in cleaned plastic bottles. The samples were immediately transported to the laboratory under low-temperature conditions in ice-boxes and stored in the laboratory at 4°C until analysis. A wide range of water quality parameters, namely pH, EC, salinity, turbidity, DO, BOD_5, COD, TSS, TDS, alkalinity, NO_3 PO_4 and SO_4 at these sites which reasonably represent the water quality in the study area were measured. All the samples were analysed for these parameters according to the standard methods of APHA-AWWA-WEF (APHA, 1995). Time lapse between sample collection and analysis was short.

Statistical techniques: Multivariate statistics was to evaluate the large amount of data in order to decipher patterns within the dataset that otherwise was not observed. The multivariate techniques used in this study which include hierarchical Cluster Analysis (CA), Principal Component Analysis (PCA) and Discriminant Analysis (DA) were performed using the Statistical Package for the Social Sciences (SPSS) Version 18.0 Software.

Factor analysis is used to uncover the latent structure of a set of variables. In technical terms, common factor analysis represents the common variance of variables excluding unique variance and is thus a correlation-focused approach seeking to reproduce the inter-correlation among the variables. On the other hand, components (from PCA) reflect both common and unique variance of the variables and may be seen as a variance-focused approach that reproduces both total variable variance with all components as well as the correlations. PCA is more commonly used than factor analysis however, it is common to use factors interchangeably with components in multivariate analysis. Principal Component Analysis (PCA) is a variable reduction technique which reduces analytical data of each sample and then inter-correlates them into a smaller set of factors that are then interpretable. The method consists of three steps, namely data standardization, factor extraction and rotation of factor axes. PCA starts by building a correlation matrix for the data and rearranges them in a manner that better explains the structure of the underlying system that produced the data. This is followed by the generation of a new group of variables from the initial data set (factors or principal components) that are a linear combination of the original variables (Chatfield and Collins, 1980). Then, the component loadings matrix is rotated to according to some rotation techniques namely, varimax, equamax or quartimax. The idea is that each variable should be heavily loaded on as few components as possible. One of the most commonly used techniques for accomplishing this transformation is the varimax rotation. This technique tends to eliminate medium-range correlations between components and original variables thus simplifying the decision as to which of the original variables to include in the components extracted. PCA reduces the large data matrix into two smaller matrices called Principal Component (PC) loadings and PC scores which are obtained through, the process of eigen analysis. In order to determine the number of components to be retained the Kaiser criterion is followed. The components which best describe the variance of the analysed data (eigenvalue >1) and can be reasonably interpreted are accepted for further analysis. Because PCA is simply the generation of pairs of eigenvalues and eigenvectors, the data do not need to be normally distributed (Johnson and Wichern, 2002). The first factor or component has the highest eigenvector sum and represents the most important source of variation in the data, i.e., explains the biggest part of the variance. The last factor is the least important process contributing to the chemical variation. Factor loadings on the factor loadings tables are interpreted as correlation coefficients between the variables and the factors. Component loadings show how the factors characterize the variables. High factor loadings (close to 1 or -1) indicate strong relationship (positive or negative) between the variable and the factor describing the variable. The measure of how well the variance of a particular variable is described by a particular set of factors is called communality. Finally, factor scores are calculated for each sample and plotted as a scatter diagram. Extreme positive factor scores (>+1) reflect sampling stations most affected by the process and extreme negative <-1) scores reflect those unaffected by the process explained by the factor. Near zero scores reflect sampling stations affected to an average degree by the process.

Cluster analysis is a pattern recognition technique that reveals intrinsic structure of a data set without making a priori assumption about the data in order to

classify the objects of the system into relatively similar or homogeneous groups. The agglomerative hierarchical cluster analysis which is the most common approach where clusters are formed sequentially, starting with the most similar pair of objects and forming higher clusters step by step was used. It was performed according to Ward's Method with squared Euclidean distances to detect the multivariate similarities in the lagoon water qualityand a distance can be represented by the difference between analytical values from both the samples. The method uses an analysis of variance approach to evaluate the distances between clusters, attempting to minimize the sum of squares of any two clusters that can be formed at each step. The different measures for similarity with respect to distance between parameters and different algorithms for finding a cluster are applied and displayed in form of a dendrogram. The dendrogram provides a visual summary of the clustering process, presenting a picture of the groups and its proximity with a dramatic reduction in dimensionality of the original data. In this study hierarchical clustering analysis was used to group similar locations within the lagoon into separate clusters based on the physicochemical parameters that were measured in separating sampling periods.

Discriminant Function Analysis (DA) undertakes the same task as multiple linear regression and one of its objectives is to determine the significance of different variables which can allow the separation of two or more naturally occurring groups thus by predicting an outcome. DA was applied to the raw dataset using the standard mode to construct Discriminant Functions (DFs) to evaluate spatial and temporal variations in water quality. The location (spatial) and season (temporal) were the grouping (dependent) variables while measured parameters constitute the independent variables. In this study, four groups for temporal (J07, F08, J08 and F09) and nine groups for spatial (nine locations) evaluation were selected.

RESULTS AND DISCUSSION

Data obtained from nine observation stations in the study area during the periods of investigation were processed. Monitoring stations are shown in Fig. 1. The selected parameters for the estimation of surface water quality characteristics were: pH, electrical conductivity, turbidity, alkalinity, salinity, Biochemical Oxygen Demand (BOD_5), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), Total Dissolved Solids (TDS), Total Suspended solids (TS), Phosphate (PO_4^{3-}), Nitrate-Nitrogen (NO_3-N) and Sulphate (SO_4^{2-}). The descriptive statistics of average values for all the parameters characterising the water quality at the different sections of the Lagos Lagoon obtained over the period of this investigation is shown in Table 1. The pH was found to range from slightly acidic to slightly alkaline (6.37-7.54) with little variation at each location having a mean of 6.99. Electrical conductivity and total dissolved solids which were found to correlate well were high and with very large variations. Salinity was found to vary from 2.18-12.02 ppt, the water being mainly brackish with salinity increasing in sections very close to the sea. Turbidity was highly varied ranging from 3.38-23.50 NTU. DO was found to be low in certain sections of the lagoon ranging between 2.26 and 6.00 mg L^{-1}. Elevated concentrations were observed of Biochemical Oxygen Demand (BOD_5), from 1.74-28.25 mg L^{-1}, Chemical Oxygen Demand (COD), from 16-191.25 mg L^{-1}, alkalinity from 69.32-120.38 mg L^{-1}. Concentrations of NO_3 and PO_4 were found in concentrations that would support the growth of planktons, i.e. from 3.70-7.88 mg L^{-1} and 5.27-15.74 mg L^{-1}, respectively. Sulphates were found to vary from 19.50-1292 mg L^{-1}. The analysis of variance showed that there were temporal and spatial variations in the levels of the quality parameters measured over the study period with statistically significant differences at 95% confidence level.

Principal Component Analysis (PCA): The correlation matrix for each sampling period showed not too consistent levels of correlation between the parameters measured. This shows the dynamic nature of the lagoon and characteristics of the different sections of lagoon during the period of this investigation. Principal component analysis was employed to investigate the factors which caused variations in the observed quality data at the nine locations of the Lagos Lagoon during the different sampling periods. In this study the principal component analysis was separately applied to the physicochemical dataset pertaining to J07, F08, J08 and F09 sampling periods. The principal components were derived from correlation matrix R and the variables were standardized prior to analysis so as to have unit variance. The use of the R matrix involved the decision that variables have been considered to be equally important (Chatfield and Collins, 1980) and also because the parameters (variables) are in widely different units, i.e., mg L^{-1}, pH, mS cm^{-1}, etc. (Karpuzcu and Sene, 1987). In order to determine the number of factors to be retained, the Kaiser criterion is followed. The factors which best describe the variance of the analysed data (eigenvalue >1) and can be reasonably interpreted are accepted for further analysis.

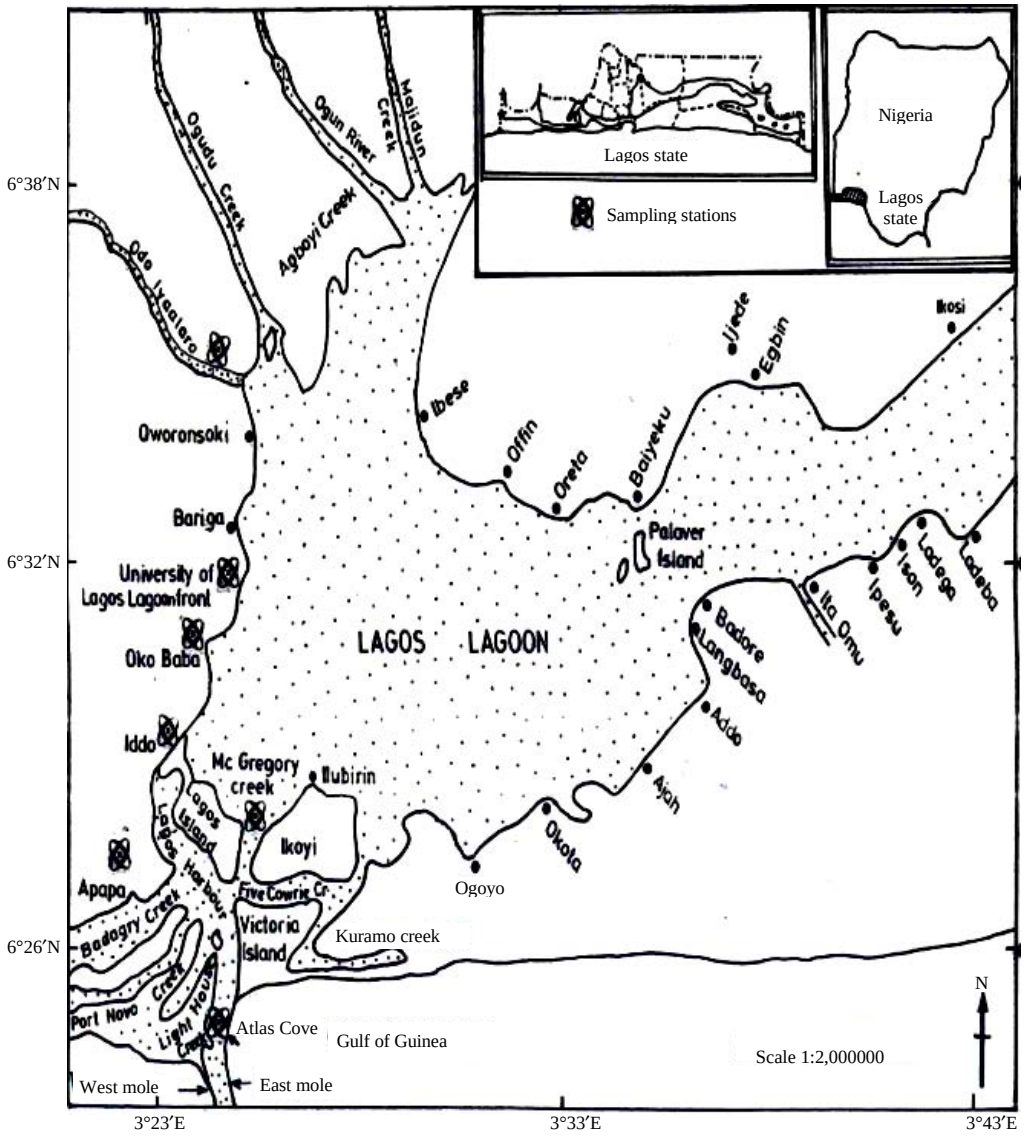

Fig. 1: Map of the Lagos Lagoon showing sampling areas

Table 1: Descriptive statistics of mean water quality parameters different sections of the lagoon throughout periods of investigation

Parameters	Min.	Max.	Range	Mean	SD	Variance	Skewness	Kurtosis
pH	6.37	7.54	1.17	6.99	0.36	0.13	-0.13	-0.21
EC (mS cm^{-1})	14179.53	31412.00	17232.47	18904.59	5353.64	2.87E07	1.83	3.82
TDS (mg L^{-1})	2310.50	13532.50	11222.00	9723.81	3178.79	1.01E07	-1.67	4.01
Salinity (ppt)	2.18	12.02	9.84	8.25	2.77	7.66	-1.19	2.71
TSS (mg L^{-1})	12.50	143.50	131.00	43.28	51.57	2659.87	1.63	0.95
Turbidity (NTU)	3.38	23.50	20.12	6.57	6.46	41.73	2.82	8.14
DO (mg L^{-1})	2.26	6.00	3.74	4.98	1.08	1.17	-2.36	6.48
BOD_5 (mg L^{-1})	1.74	28.25	26.51	13.04	7.46	55.59	0.74	1.71
COD (mg L^{-1})	16.75	191.25	174.50	78.12	49.14	2414.95	1.59	3.67
Alk (mg L^{-1})	69.32	120.38	51.06	84.97	16.53	273.24	1.37	1.64
PO_4 (mg L^{-1})	5.27	15.74	10.47	9.75	3.24	10.49	0.94	0.53
SO_4 (mg L^{-1})	19.50	1292.00	1272.50	849.47	378.10	142959.68	-1.33	2.35
NO_3 (mg L^{-1})	3.70	7.88	4.18	4.86	1.59	2.52	1.45	0.54

The results of principal components analysis of the data obtained by rotation according to varimax rotation together with the communalities for the variables during sampling periods are shown in Table 2-5.

Table 2: Rotated component matrixfor water quality parameters of J07 samples

	Component extraction			
Parameters	1	2	3	Communalities
EC	0.938	0.123	-0.239	0.951
TSS	0.846	0.296	-0.273	0.878
PO4	0.764	0.451	-0.439	0.980
TUR	0.757	0.555	-0.344	0.999
TDS	0.050	-0.974	0.163	0.977
COD	0.532	0.832	0.126	0.990
ALK	0.603	0.778	-0.006	0.969
DO	-0.472	-0.758	0.395	0.954
NO_3	0.589	0.723	-0.201	0.910
BOD	0.678	0.689	-0.255	0.999
Sal	-0.073	-0.182	0.912	0.869
pH	-0.417	0.094	0.859	0.920
SO_4	-0.504	-0.468	0.629	0.869
Eigenvalues	9.309	1.899	1.057	-
Variance (%)	37.410	36.090	20.850	-
Cumulative (%)	37.410	73.500	94.350	-

Extraction Method: Principal component analysis; Rotation Method: Varimax with Kaiser Normalization; rotation converged in 5 iterations

Table 3: Rotated component matrix for water quality parameters of F08samples

	Component extraction		
Parameters	1	2	Communalities
PO_4	0.961	0.032	0.924
pH	-0.884	0.108	0.793
EC	0.864	0.363	0.878
TSS	0.816	0.541	0.958
SO_4	-0.806	-0.393	0.804
TUR	0.790	0.601	0.984
NO_3	0.703	0.614	0.871
TDS	-0.061	-0.932	0.872
Sal	-0.010	-0.931	0.867
COD	0.294	0.908	0.911
ALK	0.417	0.866	0.923
BOD	0.675	0.725	0.982
DO	-0.656	-0.722	0.952
Eigenvalues	9.485	2.286	-
Variance (%)	46.420	46.420	-
Cumulative (%)	43.740	90.160	-

Extraction Method: Principal component analysis; Rotation Method: Varimax with Kaiser Normalization; rotation converged in 3 iterations

Table 4: Rotated component matrixfor water quality parameters of J08 samples

	Component extraction				
Parameters	1	2	3	4	Communalities
EC	0.978	0.152	-0.072	-0.004	0.986
TDS	0.978	0.083	-0.054	0.122	0.981
SO_4	0.975	0.079	-0.016	0.172	0.987
Sal	0.973	0.070	-0.061	0.126	0.971
BOD	0.960	0.061	0.004	0.237	0.981
TUR	0.906	0.181	0.081	-0.004	0.860
COD	0.792	0.184	0.151	0.214	0.729
NO_3	0.037	-0.908	0.207	-0.187	0.904
DO	0.463	0.732	0.350	0.175	0.904
pH	0.533	0.712	0.020	-0.076	0.797
PO_4	-0.206	0.545	-0.537	-0.488	0.866
TSS	-0.103	-0.020	0.949	-0.122	0.926
ALK	0.223	0.185	-0.112	0.916	0.936
Eigenvalues	7.392	2.052	1.299	1.087	-
Variance (%)	52.180	17.780	10.830	10.200	-
Cumulative (%)	52.180	69.960	80.790	90.990	-

Extraction Method: Principal component analysis; Rotation Method: Varimax with Kaiser Normalization; rotation converged in 5 iterations

Table 5: Rotated component matrixfor water quality parameters of F09 samples

	Component extraction			
Parameters	1	2	3	Communalities
EC	0.956	0.145	0.097	0.945
TDS	0.956	0.145	0.097	0.945
Sal	0.951	0.105	0.155	0.939
TSS	-0.874	-0.118	-0.239	0.834
NO_3	-0.805	0.073	-0.401	0.815
PO_4	0.721	0.073	-0.251	0.588
DO	0.155	0.880	0.231	0.852
SO_4	0.240	0.790	-0.327	0.790
COD	0.082	0.749	0.476	0.794
BOD	0.517	0.735	0.329	0.916
ALK	-0.383	0.692	0.192	0.663
TUR	0.173	0.165	0.889	0.847
pH	0.187	0.572	0.724	0.886
Eigenvalues	6.321	3.139	1.352	-
Variance (%)	40.240	26.240	16.690	-
Cumulative (%)	40.240	66.480	83.170	-

Extraction Method: Principal component analysis; Rotation Method: Varimax with Kaiser Normalization; [a]Rotation converged in 5 iterations

Principal Component Analysis (PCA) confirmed the significant differences in the composition of the analysed physicochemical parameters of water in the different locations of the Lagos Lagoon during the periods of study which were clearly affected by point and non-point sources. When the percentages of the total variances of the 3 extracted components (with eigenvalues >1) for J07 variables are accumulated, it can be seen that these 1st 3 principal components explained 94.35% of the total variance and the communalities show that variances of all the variables have been described well (either positive or negative) by the three components (Table 2). Component loading PC1 explained 37.41% of the total variance while PC2 explained 36.09% and PC3 explained 20.85%. In general, components larger than 0.60 were taken into consideration in the interpretation.

In other words, the most significant variables in the components are represented by high loadings. In addition to the high significance of high loading values, there exists a difference between the components; the components with larger variances give more information about the data. When the variances (eigenvalues) of the table are examined, it can be seen that principal components are in decreasing order of importance with respect to their variances. An interpretation of the rotated principal components in Table 5 was made by examining the component loadings noting the relationship to the original variables. The first component gives information about the variation in Phosphates (PO_4), pH, Electrical Conductivity (EC), Total Suspended Solids (TSS) and turbidity. In this component, these quality parameters indicates that discharges contain substantial amounts of

detergents from car washing facilities and colloidal materials that could affect the turbidity of the water during the period of sampling.

This could be as a result of point and non-point discharges during the rainy season. Significant loading of BOD_5 also indicates that domestic discharge also existed during this period however, the contribution is not as important as the other parameters. In the second component PC2, Total Dissolved Solids (TDS), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), Nitrates (NO_3), Biochemical Oxygen Demand (BOD_5) and alkalinity are important. The fluctuations and relatively low DO during this period is demonstrated by the relationship between BOD and COD resulting from dumping of organic and inorganic wastes into the lagoon. In the third component PC3, salinity and pH have strong loadings but are not having major controlling effect on the quality of the water during this period.

PCA rendered two significant components (eigenvalue >1) for the F08 sampling period, explaining 90.16 % of the total variance of the parameters analysed (Table 3). This was slightly different from what was observed in J07 with SO_4 now as an important contributor. The values of communalities show that the variances of all the variables were described well. The first component loading (PC1) accounts for as much as 46.42% of the total variance and was strongly correlated with the PO_4, pH, EC, TSS, SO_4, turbidity and NO_3 in this decreasing order of importance. In this component these quality parameters indicate the presenceof substantial levels of organic and inorganic wastes that containing colloidal materials being discharged into the lagoon. It was noted that there is a considerable overlapping variables; NO_3 and turbidity also form part of the significant components that make up the second component (PC2). The second component which accounts for 43.74% of the total variance strongly correlated with TDS, salinity, COD, alkalinity, BOD and DO and also demonstrated the relationship between DO and BOD/COD as observed during the J07 sampling period.

PCA of the measured water quality parameters for J08 sampling period rendered four components (with eigenvalues >1) explaining 90.99% of the total variance (Table 4). PC1 explained as much as 52.18% of the variance and has strong loadings of EC, TDS, SO_4, salinity, BOD, turbidity and COD. In this component there is an indication of discharge of both organic and inorganic wastes into the lagoon. PC2 explained 17.78% of the total variance and has strong loadings of NO_3, DO and pH. It will be noted that PO_4 has a moderate loading in PC2. PC3 and PC4 explained 10.83 and 10.20% of the total variance and have strong loadings for TSS and alkalinity, respectively. The contribution of TSS was not as important as in J07 and F08 periods, suggesting that turbidity may be mainly as a result of colloidal materials in the water.

PCA rendered three significant components for the F09 sampling period explaining 83.17% of the total variance (Table 5). PC1 gives information about the variation of EC, TDS, salinity, TSS, NO_3 and PO_4 explaining 40.24% of the total variance. PC2 explains about 26.24% of the total variance and has high loadings for DO, SO_4, COD, BOD and alkalinity. PC3 explains about 16.69% of the total variance and is characterized by high loadings for turbidity and pH.

Cluster analysis: Cluster analysis was performed on the standardized (z-scale) dataset (Guler *et al*., 2002) for the four sampling periods separately by Ward's Method (Ward, 1963) using square Euclidean distance as similarity measurebetween two samples. The sample locations are grouped on the vertical axis and the linkage distances, representing the relative differences between clusters are shown on the horizontal axis. The dendrogram resulting from agglomerative hierarchical CA showing the spatial clustering for each sampling period is in Fig. 2a-d. CA was also performed on variables for each period separately to determine the temporal similarity ofparameters the number of groupings obtained differs for each of the periods. The results obtained for all the sampling periods agreed with those obtained from PCA. The dendrogram for each sampling period is shown in Fig. 3a-d.

The dendrogram obtained for J07 sampling period (Fig. 2a) detected the similarity groups consisting of three statistically significant clusters. Based on this result Cluster I consisted of locations L1, L2, L3, L4, L5 and L6 with L6 being the only singleton within the cluster; Cluster II include L7 and L8; Cluster III consisted of only L9. The first level of aggregation is established in Cluster I with locations L3, L5, L4 and L2 at a distance of about1.0; locations L7 and L8 (Cluster II) were then added and associated with Cluster I at a distance of about 5.0. Finally, location L9 (Cluster III) is associated with the cluster at a larger distance thus presenting very small similarity with any other group. CA performed on variables indicates the groups of variables that behave similarly and/or have similar origin (Fig. 3a). The dendrogram shows two major clusters; Cluster I (corresponding to the variables having high positive loadings in PC1 and PC2) comprising of turbidity, BOD, PO_4, NO_3, EC, TSS, COD and alkalinity; Cluster II (corresponding to the variables having high negative loadings in PC2 and the variables having high positive loadings in PC3) comprised of pH, salinity, DO, SO_4 and TDS.

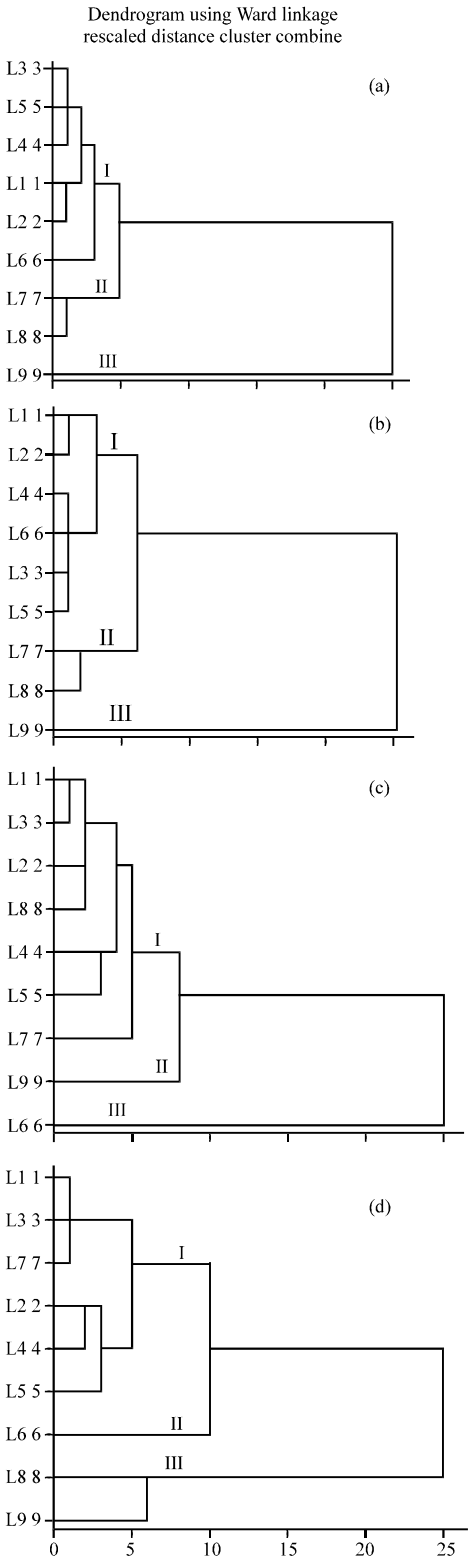

Fig. 2: Dendrogram showing spatial clustering during a) J07; b) F08; c) J08 and d) (F09) sampling periods

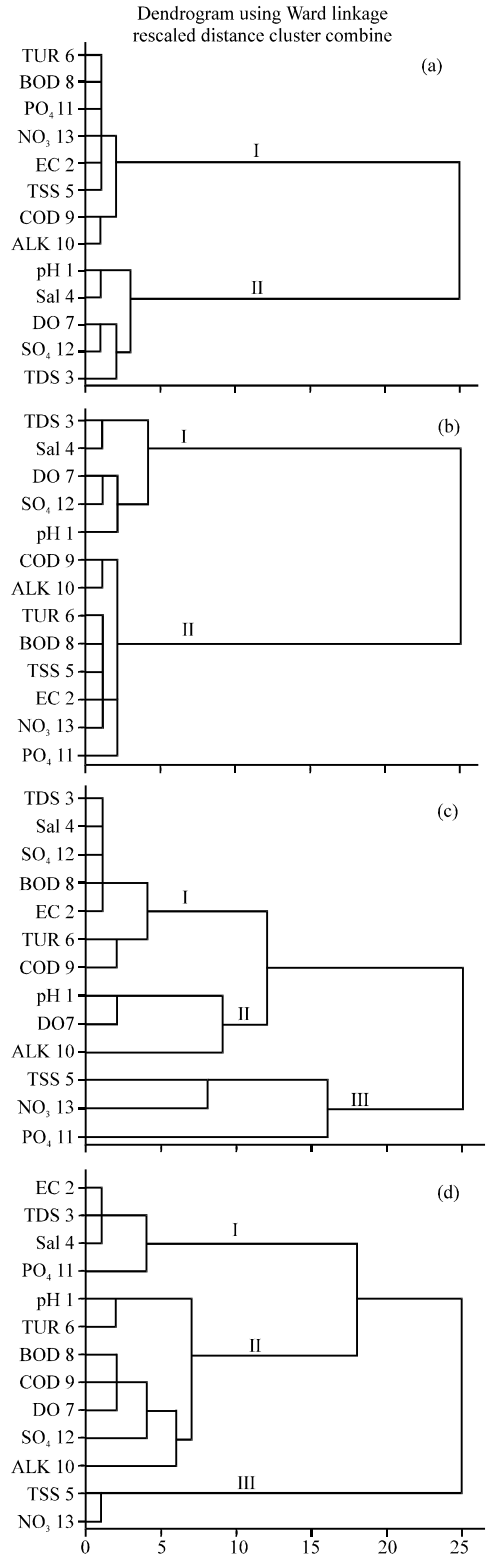

Fig. 3: Dendrogram showing clustering of variables during; a) J07; b) F08; c) J08 and d) (F09) sampling periods

The dendrogram obtained for the period F08 (Fig. 2b) also detected the similarity groups yielding consisting of three statistically significant clusters. The grouping of the locations was similar to the J07 period however with slightly different levels. Here the first level of aggregation is established in Cluster I with locations L1-L6 at a distance of about 1.0; Cluster II was formed by L7 and L8 and associated with Cluster I at about 5.0. Again location L9 (Cluster III) presents very small similarity with any other group. The grouping of variables was similar to the J07 period however, the similarity levels of the variables were changed in the F08 samples (Fig. 3b). Here Cluster I is made up of components having high negative loadings in PC1 and PC2 while Cluster II is made up of components with high positive loading in PC1 and PC2.

The dendrogram resulting from hierarchical agglomerative CA for the sampling period J08 for clustering of cases (locations) displayed the presence of three major clusters; Cluster I includes L1, L2, L3, L4, L5, L7 and L8; cluster includes only L9 and Cluster III only L6. The groupings as well as the similarity levels of each of the locations during this period was different from that observed during the J07 sampling period. Grouping of variables during this period also showed a large difference from that obtained during the J07 sampling period. The three major clusters showed very large similarity distances. Cluster I (corresponding to the variables in PC1) comprised of TDS, salinity, SO_4, BOD, EC (which formed the first level of aggregation at a distance of about 1.0), turbidity and COD (which associated later at a distance of about 4.0), Cluster II (corresponding to components having high positive loadings in PC2 and PC4) comprised of DO, pH and alkalinity which joined the Cluster I at a distance of about 9.0. Cluster III (corresponds to components having negative loadings in PC2 and PC3) was made up of TSS and NO_3 and then PO_4 associating with them at a distance of about 16.0 thus showing little affinity with any of the group.

The result obtained by CA for F09 sampling period also detected the similarity groups yielding a dendrogram into three major clusters for the locations. Again the grouping of the locations showed some differences from the F08 sampling period. Cluster I comprised locations L1, L2, L3, L4, L5 and L7. The first level of aggregation was established in Cluster I with locations L1, L3 and L7 at similarity distance of about 1.0. Cluster II is made up of a singleton L6 and Cluster III comprised of L8 and L9 which seem to show little semblance with other locations. The groupings of variables were again very different from those obtained for F08. Here the similarity levels were changed with much higher similarity distances observed during this period compared to the period F08. Three major clusters were identified similar to that obtained for the J08 sampling period. Cluster I (components with high positive loading in PC1) comprised of EC, TDS and salinity at a distance of about 1.0 with PO_4 associating later at a distance of a about 4.0; Cluster II (components having high positive loadings in PC2 and PC3) comprised of pH, turbidity, BOD, COD, DO and the singletons SO_4 and alkalinity; Cluster III (components with high negative loadings in PC1) comprised of TSS and NO_3.

Discriminant Analysis (DA): Temporal DA was conducted on the raw dataset which comprised of the thirteen parameters for the combined periods of sample collection after grouping the cases according to period (J07 = Group 1; F08 = Group 2; J08 = Group 3; F09 = Group 4) to predict similarity in water quality during the periods of study. The standard mode for building discriminant function coefficient based on entering all the independent variables together was used.

Significant mean differences were observed for most of the predictors on the Discriminant Variables (DV). The log determinants were quite similar; Box's M test indicated that the assumption of equality of covariance matrices was violated. However, given the large sample, this problem is not regarded as serious. The Discriminant Functions (DFs) were tested using Wilks' lambda and χ^2-tests. The values of Wilk's lambda and Chi-square (χ^2) for each discriminant function varied from 0.00-0.401 and 24.20-209.08 respectively and $p<0.05$ indicating that the spatial-DA in this study had discriminatory ability of the function and was reliable. The Standard Mode DA formed three Discriminant Functions (DFs). The DFs revealed a significant association between groups and the predictors.

The first function accounts for 69.7% of the variation in the discriminant variables (water parameters measured) during the four sampling periods while function 2 and 3 account 28.2 and 2.1%, respectively. Analysis of the structure matrix revealed EC and TDS as the only significant predictors while other parameters are poor predictors. Thus, suggesting that TDS and EC were needed to account for most of the expected spatial/temporal variations in water quality of these locations. Salinity though not as important as TDS and EC could also be used as a predictor. Cross validation result revealed that 97.2% of the original grouped cases were classified correctly and 88.9% of the cross-validated grouped cases were correctly classified for the four periods samples were taken. Figure 4 shows the

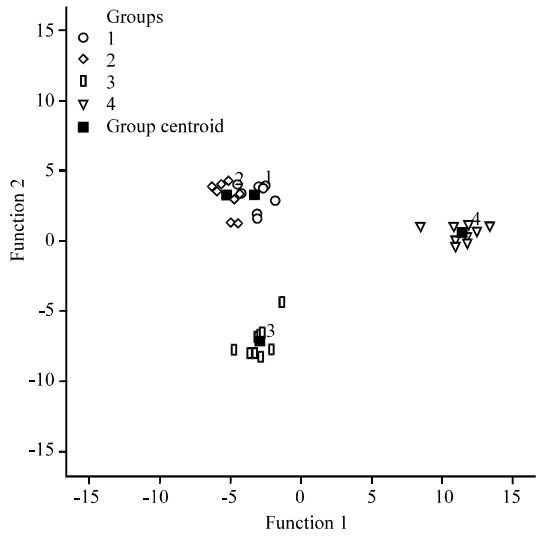

Fig. 4: Bivariate plot of discriminant functions 1 and 2 during the J07 (Group 1); F08 (Group 2); J08 (Group 3) and F09 (Group 4) sampling periods

projection of the DFs 1 and 2 for the dataset for the four periods in a 2D scatter plot. Among the four periods J08 and F09 are clearly separated from each other and from J07 and F08 whereas some overlap exist between J07 and F08.

CONCLUSION

In this study, different multivariate statistical methods were used to assess temporal and spatial variations in water quality of different sections of the lagoon. The principal component and cluster analysis of obtained dataset reflects great diversity in the water quality of the different sections of the Lagos Lagoon. This diversity in water quality was not necessarily dependent on season but more on anthropogenic pollution (point and non-point sources) and dredging activities. Hierarchical CA grouped the nine locations into three groups based on the similarity of water quality characteristics. During the entire period of this investigation, CA revealed that locations L1 to L5 showed consistently similarities in water quality and locations L9 was found to consistently show little semblance with the other locations. PCA distributions however, showed that water samples within groups differ slightly suggesting differences in the chemistries hence processes such as dilution, source/type of contamination, mixing ability of the lagoon, etc., EC and TDS among the variables also showed consistently high loadings also indicating similarity in the source.

Moreover, DA rendered an important reduction in the required amount of data for the three groups of monitoring sites because two parameters (TDS and EC) were found to be the predictor for the spatial/temporal analysis. Therefore, DA allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters responsible for large variations in water quality that could reduce the number of sampling parameters. Hence, this study further strengthens the fact that multivariate statistical methods are excellent exploratory tool for interpreting complex water quality data sets and for understanding spatial and temporal variations which are useful and effective for water quality management.

REFERENCES

APHA, 1995. Standard Methods for the Examination of Water and Wastewater. 19th Edn., American Public Health Association, Washington, DC., USA.

Belkhiri, L., A. Boudoukha and L. Mouni, 2011. A multivariate statistical analysis of groundwater chemistry data. Int. J. Environ. Res., 5: 537-544.

Boyacioglu, H. and H. Boyacioglu, 2008. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environ. Geol., 54: 275-282.

Chatfield, C. and A.J. Collins, 1980. An Introduction to Multivariate Analysis. Chapman and Hall, London.

Grande, J.A., J. Borrego, J.A. Morales and M.L. De-La-Torre, 2003. A description of how metal pollution occurs in the Tinto-Odielrias (Huelva-Spain) through application of cluster analysis. Mar. Pollut. Bull., 46: 475-480.

Guler, C., G.D. Thyne, J.E. McCray and A.K. Turner, 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol. J., 10: 455-474.

Gupta, I., S. Dhage and R. Kumar, 2009. Study of variations in water quality of Mumbai Coast through multivariate analysis techniques. Indian J. Mar. Sci., 38: 170-177.

Johnson, R.A. and D.W. Wichern, 2002. Applied Multivariate Statistical Analysis. 5th Edn., Prentice Hall Inc., Upper Saddle River, New Jersey.

Karpuzcu, M. and S. Sene, 1987. Design of monitoring systems for water quality by principal component analysis and a case study. Proc. Int. Symp. Environ. Manage., 1: 673-690.

Koonce, J.E., Z. Yu, I.M. Farnham and K.J. Stetzenbach, 2006. Geochemical interpretation of groundwater flow in the southern Great Basin. Geosphere, 2: 88-101.

Kumar, A.R. and P. Riyazuddin, 2008. Application of chemometric techniques in theassessment of groundwater pollution in asuburban area of Chennai city, India. Curr. Sci., 94: 1012-1022.

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

Mazlum, N., A. Ozer and S. Mazlum, 1999. Interpretation of water quality data by principal components analysis. Turkey J. Eng. Environ. Sci., 23: 19-26.

Michalik, A., 2008. The use of chemical and cluster analysis for studying spring water quality in Swietokrzyski National Park. Polish J. Environ. Stud., 17: 357-362.

Reisenhofer, E., G. Adami and A. Favretto, 1996. Heavy metals andnutrients in coastal, surface seawaters (Gulf of Trieste, NorthernAdriatic Sea): An environmental study by factor analysis. Fresenius J. Anal. Chem., 354: 729-734.

Simeonova, P., V. Simeonov and G. Andreev, 2003. Waterquality study of the Struma River Basin, Bulgaria (1989-1998). Central Eur. J. Chem., 1: 136-212.

Sojka, M., M. Siepak, A. Ziola, M. Frankowski, S. Murat-Blazejewska and J. Siepak, 2008. Application of multivariate statistical techniques to evaluation of water quality in Mala Welna River (Western Poland). Environ. Monit. Assess., 147: 159-170.

Ward, Jr. J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc., 58: 236-244.

Zhou, F., Y. Liu and H. Guo, 2007. Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern new territories, Hong Kong. Environ. Monit. Assess., 132: 1-13.