International Journal of Geology, Agriculture and Environmental Sciences Volume – 3 Issue – 6 December 2015
Website: www.woarjournals.org/IJGAES *ISSN: 2348-0254*

The Use of Agglomerative Hierarchical Cluster Analysis for the Assessment of Mangrove Water Quality of Okoro River Estuary, Southeastern Nigeria

John Ukpatu[1], Enenwan Udoinyang[2] and James P. Udoh[3]

[1]Department of Zoology, Faculty of Science, University of Uyo, P.M.B 1017, Uyo, Akwa Ibom State , Nigeria.

[2]Department of Fisheries and Aquatic Environmental Management,
Faculty of Agriculture, University of Uyo, P.M.B 1017 Uyo, Nigeria.
Corresponding Author: +2348027716902

[3]Department of Fisheries and Aquatic Environmental Management,
Faculty of Agriculture, University of Uyo, P.M.B 1017 Uyo, Nigeria.

Abstract: *The agglomerative hierarchical cluster analysis was employed to evaluate the tropic status of the mangrove water quality of Okoro River Estuary, Southeastern Nigeria. Cluster analysis using the ward method with squared Euclidean distance measure was performed, which indicated the distribution of parameters and depicted different hydrochemical facies. The results yielded a dendrogram with four statistically significant clusters of similar behavour with the biggest cluster of seven (7) parameters. The water parameters were classified into mutually exclusive unknown groups that share similar characteristics and properties. Dendrogram clearly distinguished parameter bahaviours and interprets the description of the hierarchical clustering in a graphical format. The dendrogram showed a moderate cophentic correlation coefficient (C = 0.72) among physicochemical parameters of mangrove water quality. The behavioural pattern of parameters and the agglomeration of nearest neighbour cluster indicated that the process of salinization, organic decomposition, mineralization, metal accumulation and surface runoff characterized the mangrove water quality of the area. The Euclidean distance revealed the extent of similarity and dissimilarity between water quality characteristics. The result of month-wise hierarchical cluster analysis showed discrimination of the months with similar environmental and physicochemical behaviour. This study illustrates the benefit of agglomerative hierarchical cluster analysis for the characterization, evaluation and interpretation of complex dynamism in the mangrove dominated area of southeastern Nigeria and recommends regular monitoring and assessment of mangrove water quality.*

Keywords: Agglomerative hierarchical cluster analysis, Dendrogram, Cophentic correlation coefficient Icicle plot, Okoro River estuary, Southeastern Nigeria.

1. Introduction

The agglomerative hierarchical cluster analysis is a method of cluster analysis which seeks to build a hierarchy of clusters using agglomerative function. Hierarchical agglomerative clustering is the most common approach [23], [24], [15] which provides intuitive relationships between any one sample, variable and the entire data set, and is typically illustrated by a tree-like diagram called dendrogram. The method organizes parameters into groups based on the similarities inside of the group and dissimilarities outside of different groups. It is also a major technique for classifying a 'mountain' of information into manageable meaningful piles. It is a data reduction tool that creates subgroups that are more manageable than individual datum. [26] explained that the groups are divided by their unique characteristics, and often, it helps in interpreting the complex data set. The agglomerative hierarchical cluster analyses according to Ward's method with squared Euclidean distances are often applied to detect multivariate homogeneity and multivariate heterogeneity in respect to seasons and water quality parameters [25]. Again, agglomerative hierarchical cluster analysis can be used as an important tool for analyzing water quality data and to understand the relationship among parameters and among seasons [8].

The technique is gaining popularity in the analysis of water quality data due to their simplicity when handling a large number of variables simultaneously and capable of producing more easily interpretable results for the evaluation and assessment of water quality. For instance, [10] applied the agglomerative hierarchical cluster analysis to analyze the correlation between the physicochemical, environmental, biological parameters and grouped them into tree-like cluster diagram called the 'dendrogram'. Agglomerative hierarchical cluster analysis with dengrogram was reported by [17], [22], [21], [1]. [16] stated that the analysis of waters only on the basis of their quality and chemistry is not fully substantiated and required further statistical technique such as agglomerative hierarchical cluster analysis. According to [19] agglomerative hierarchical cluster analysis shows the distances (similarities or dissimilarities) between the cases being combined to form clusters. [23] used the dendrogram from the agglomerative hierarchical cluster analysis to identify the behavioural pattern of the mangrove water quality in Pondicherry coast of India. [15] also carried out hierarchical cluster analysis of hydrochemical variables of Karwan-sengar sub-basin, central Ganga Basin, India and identified six agglomerative hierarchical clusters or groups named as A, B C, D, E and F. [6] further explained that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise.

Evidences abound on the application of agglomerative hierarchical cluster analysis. However, in Nigeria, the application of this technique for the identification, classification and characterization of mangrove water quality are scarce. Thus, this method was adopted for this study to uncover natural grouping of physicochemical parameters and also to diagnose the behavioural patterns of the mangrove quality of the Southeastern region of Nigeria.

2. Materials and Methods:

2.1 Description of the study area:

The study was carried out in Okoro River estuary, Southeastern Nigeria located (4^o 33' N - 4^o 55' N; 7^o45 E -7^0 55' E about 650 m above sea level in the tropical mangrove forest belt east of the Niger Delta between the lower Imo and Qua Iboe River estuaries (Figure 1). The tidal range in the area is about 0.8 m at neap tides and 2.20m during spring tides with little fresh water input joined by numerous tributaries as they empty into the Atlantic Ocean [18] The climate of the area is tropical with distinct rainy (April to October) and dry seasons (October to May) with a high annual rainfall averaging 2500 mm [11], [2]. The mean water temperature of the study area is 28.2^{o}C [27].

2.2 Vegetation and Mineral Resources

The area is characterized by an expensive mangrove swamp dominated with mangrove species: *Rhizohoraracemosa, Avicenniagerminas, Conocarpuseractus*, interspersed with *Nypafruticans* with inter-tidal mud flats influenced by the semi-durnal tidal regime of the estuary. Fishing and farming are the main economic activities in this study area. Oil palm (*Elaeisguineensis*) and coconut palm (*Cocusmueifera*) are also widely distributed in the surrounding villages. The area is also an oil-producing area with several oil exploration wells and oil pipelines.

2.3 Physicochemical Analysis of Water

Surface water temperature, pH (hydrogen ion concentration), salinity and total dissolved solids (TDS) were measured insitu with the multi-parameters monitoring instrument (Model- EXTECH, EC, 500).Dissolved Oxygen was measured with DO meter – Model – EXTECH 11, DO 600, while biochemical oxygen demand (BOD_5) was determined with DO –meter Model – EXTECH 11, DO 600 after 5 days incubation @ 20^{o}C. Similarly, total suspended solids (TSS) was determined using Gravimetric method..Water hardness was determined by complexometric titration- with EDTA. Total hydrocarbon content (THC) was extracted with carbon tetrachloride (CCL4) in a separating funnel at pH 5 and absorbance read from the Fisher Electrophotometer at 450 nm wave length after appropriate treatment and digestion. Other tests using standard methods and procedures included: sodium and potassium while copper, zinc, nickel was determined using atomic absorption spectro- photometer (AAS)- Perkin- Elmer Model 2380 (APHA, 1998). Rainfall data was collected from the Meteorological Unit, Department of Geography, University of Uyo.

2.3 Physicochemical Analysis of Water

Surface water temperature, pH (hydrogen ion concentration), salinity and total dissolved solids (TDS) were measured insitu with the multi-parameters monitoring instrument (Model- EXTECH, EC, 500).Dissolved Oxygen was measured with DO meter – Model – EXTECH 11, DO 600, while biochemical oxygen demand (BOD_5) was determined with DO –meter Model – EXTECH 11, DO 600 after 5 days incubation @ 20^{o}C . Similarly, total suspended solids (TSS) was determined using Gravimetric method..Water hardness was determined by complexometric titration- with EDTA.

Total hydrocarbon content (THC) was extracted with carbon tetrachloride (CCL4) in a separating funnel at pH 5 and absorbance read from the Fisher Electrophotometer at 450 nm wave length after appropriate treatment and digestion. Other tests using standard methods and procedures included: sodium and potassium while copper, zinc, nickel was determined using atomic absorption spectro- photometer (AAS) - Perkin- Elmer Model 2380 [5]. Rainfall data was collected from the Meteorological Unit, Department of Geography, University of Uyo.

2.4 Statistical Analysis

Agglomerative hierarchical cluster analysis was done by means of ward's method using squared Euclidean distances as a measure of similarity. Ward's method enables an analysis of variance approach to evaluate the distances between the clusters in order to minimize the sum of squares of any two clusters (that can be formed at each step) [16]. The levels of similarity of which observations are merged were used to construct a dendrogram.

The method was used to parameterize physicochemical variables into their similar or dissimilar domains. Pairs of variables in close proximity were amalgamated using the linkage function in cluster analysis. The linkage function used the Euclidean Distance generated to measure the distance between pairs of variables and also to determine the proximity of variables to each other when they are amalgamated in an ecological system.

Month-wise CA was carried on the data matrix of physicochemical parameters using ward's method to group months based on similarity distance

3. Results

The use of agglomerative hierarchical cluster analysis for studying of mangrove water quality in Okoro River estuary, Southeastern Nigeria yielded a dendrogram (agglomerative bottom-up tree-like structure), with four statistically significant clusters (Fig.1). The result revealed homogeneity within groups and heterogeneity between the physicochemical parameters. The dendrogram also showed four different cluster groups (Group A = 7 parameters, Group B = 2 parameters, Group C = 5 parameters and Group D = 3 parameters) with biggest cluster of seven parameters.

Fig. 1 Okoro River estuary, Southeastern Nigeria (Source: Google earth)

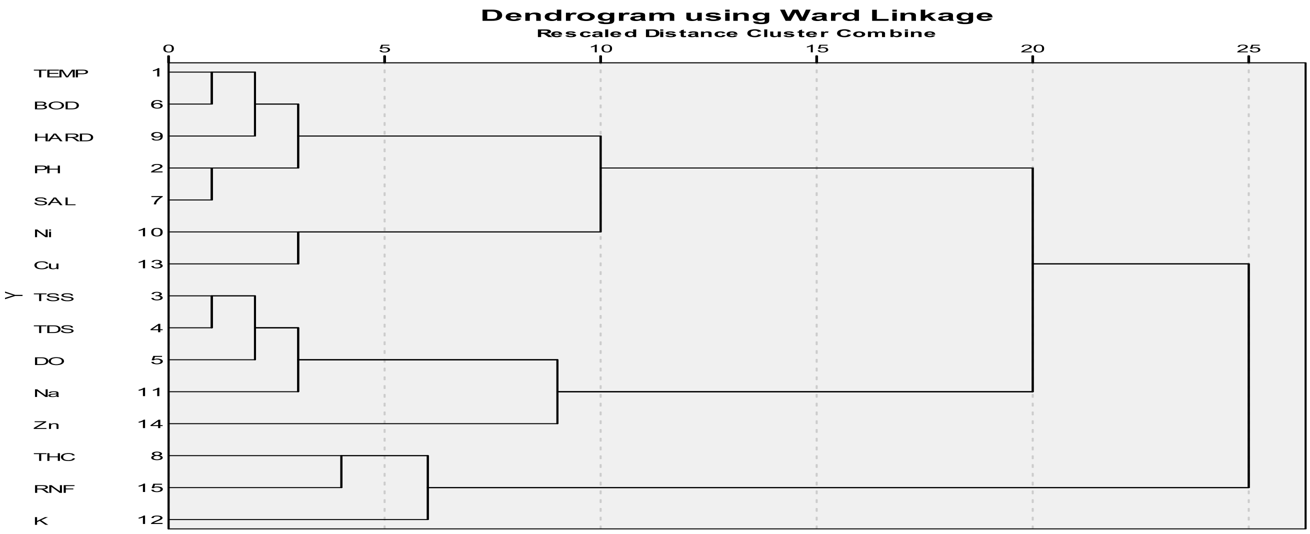

Figure 1: Dendrogram of Similarity and Dissimilarity Clusters showing Similar Physicochemical Behaviour and the Amalgamation of Various Parameters into Domains

The Cophenetic correlation coefficient or degree of accuracy by which the dendrogram preserves the pair wise distances between the physicochemical parameters was recorded as (C= 0.72). The cophenetic correlation

coefficient assessed the cluster base and measured how faithfully the dendrogram preserves the Euclidean distances between the original data points. It showed how well the dendrogram measures the resemblance in the metrics match.

The distance between two clusters was computed as the distance between the two closest cases in the two clusters using the agglomerative linkage method. The result as shown in the agglomerative schedule (Table 1) revealed the agglomeration coefficients and the Euclidean distance between clusters which indicated that variables are formed into natural system of similar and dissimilar groups.

Table 1: Agglomeration Schedule for Physicochemical and Environmental Parameters of Okoro River Estuary, South Eastern Nigeria

Cluster No	1st item	2nd item	Euclidean Distance	Agglomeration Coefficients
1	Temp	BOD_5	0.016	0.008
2	TSS	TDS	0.019	0.017
3	Temp (Cluster 1)	Hardness	0.037	0.043
4	TSS(Cluster 2)	DO	0.050	0.076
5	pH	Salinity	0.051	0.111
6	Temp (Cluster 1)	pH (Cluster 5)	0.082	0.173
7	TSS (Cluster 2)	Na	0.095	0.237
8	Ni	Cu	0.128	0.306
9	THC	Rainfall	0.184	0.398
10	Temp(Cluster 1)	TSS (Cluster 2)	0.192	0.526
11	THC(Cluster 9)	K	0.241	0.937
12	Temp(Cluster 1)	Ni (Cluster 8)	0.291	0.937
13	pH(Cluster 5)	Zn	0.485	1.388
14	Temp(Cluster 1)	THC(Cluster 9)	0.536	1.970

The cluster groups, their members and their nearest neighbours are presented in (Table 2). The neighbour joining clusters showed that the community structure is connected with ecosystem function.

Table 2: Cluster Groups and their Members

Cluster Groups	Parameter Nos	Cluster Members	Nearest Neighbours
A	1,6,9,2,7,10,13	Temp,BOD_5, Hardness, pH, Salinity, Nickel & Copper	Temp&BOD_5(1&6); Temp &Hardness (1&9; Temp, Hardness & pH (1,9&2); pH & Salinity(2&7); Nickel & Copper (10&13)
B	10.13	Nickel & Copper	Nickel &Copper (10&13)
C	3,4,5,11,14	TSS, TDS, DO, Na &Zn	TSS &TDS (3&4); DO &TSS(5&3); Na, TSS,DO (11,3&5); Zn &Na (14&11)
D	8,15,12	THC, Rainfall, K	THC & Rainfall (8&15); K&THC(12&8)

The Icicle plot indicated existing gap between parameters (Fig 2). It also revealed an alignment and mutual alliance within and between parameters

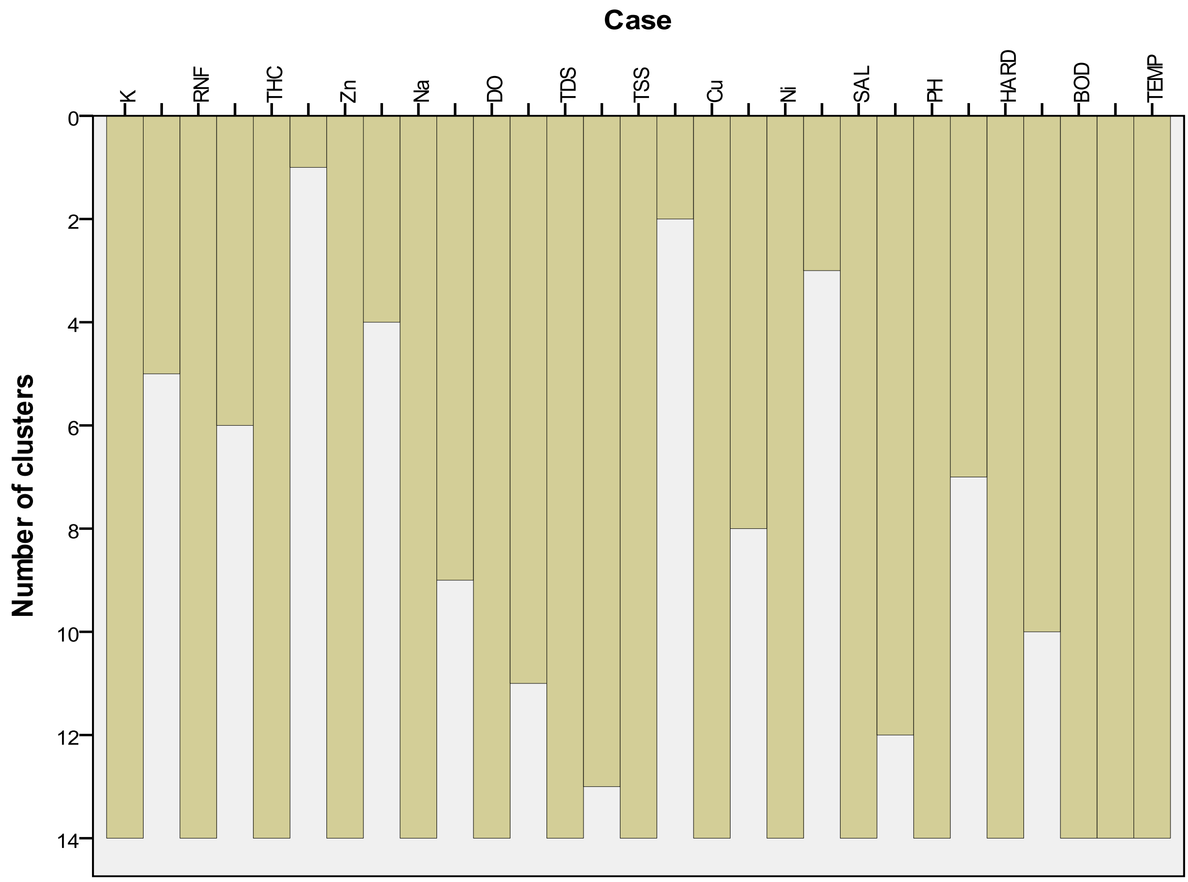

Figure 2: Icicle Plot of Parameters

The result of agglomerative hierarchical cluster analysis showed the discrimination of physicochemical parameter of mangrove ecosystem of Okoro River into the two seasons operating in the study area; rainy season (April-October) and dry season (November – March) (Fig 3).

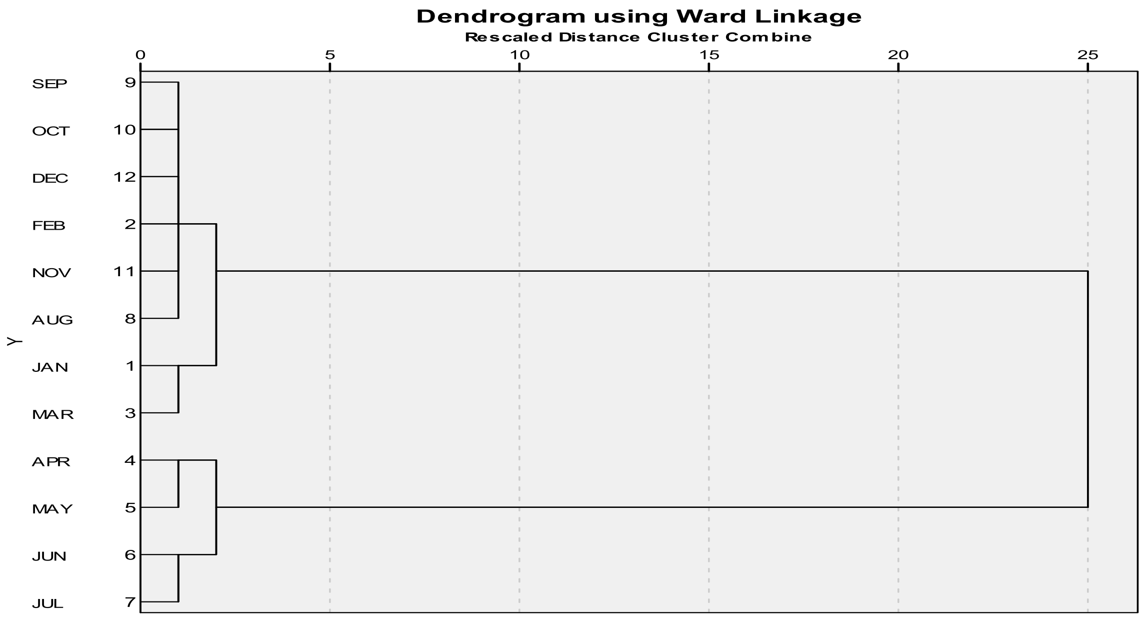

Figure 3: Dendrogram of Month-wise Cluster Analysis with Similar Behaviour

4. Discussion

The dendrogram showed four different cluster groups with biggest cluster of seven parameters in mangrove ecosystem of Okoro River estuary. The four cluster groups is different from other reports. [10] reported three different cluster groups while studying the physical and chemical parameters of Ootacamund Lake of India but with the biggest cluster of seven parameters. The biggest cluster of seven parameters is similar to the result of this study but different from the three cluster groups recorded by [16] in the Kelce, Poland. It is also different from the six cluster groups with the biggest cluster of six parameters recorded by [15] in the Central Ganga Basin, India. [20] also reported four different cluster groups while studying the physical, chemical and environmental parameters of Munj Sugar Talab, India, but with the biggest cluster of nine parameters. However, the use of agglomerative hierarchical cluster analysis had been used by several researchers to study water quality in different locations: Ootacamund Lake by [10] Lingambudhi Lake, Mysore, Karnataka by [17] Mengkabong Sabah, Malaysia by [21] and Kelce, Poland by [16].

Cophenetic correlation coefficient lies between -1.0 to 1.0 and any value above 0.6 measures moderate degree of accuracy in hierarchical cluster [12], [14]. The measure of degree of accuracy by which the dendrogram preserves the pair wise distances between the physicochemical parameters as recorded in this study is above 0.6. Previous researchers have recorded C > 0.6. [20] recorded 0.76 in Madhya Pradesh water of India and [15] recorded 0.62 in the Central Ganga Basin, India. According to [21] higher values of cophenetic correlation (C > 0.6) indicates good similarity between data matrix of parameters and dendrogram and higher cophenetic correlation coefficient justified the authenticity of hierarchical clustering of water parameters.

The Euclidean distance was selected as the measure of similarity and dissimilarity between water parameters. [28] also found that using the Euclidean distance as a distance measure and Ward's method as a linkage rule produced the most distinctive groups. Short Euclidean distance indicated homogeneous clusters whereas large Euclidean distance indicated heterogeneous clusters. The smaller the Agglomeration coefficient, the stronger is the Amalgamation bonding between the mangrove water parameters. This is similar to the findings of [18] in the Ootacamund Lake of India. The Euclidean distance as reported by [6] gave the similarity between two samples or variables. [15] had explained that the closer the parameter, the smaller the Euclidean distance. The result of this study is similar to the findings of [21] and [6] that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. According to [28], hierarchical clustering organizes sampling entities into discrete clusters, such that within-group similarity is maximized and among-group similarity is minimized according to some objectives criteria. Its purpose is to discover a system of organizing observations and sort them into groups so that it is statistically easier to predict behavior of such observations based on group membership that share similar identities/properties. Similarly, [24] pointed out that small distance showed that similar cases are being clustered whereas large distance indicate dissimilarity between the cases being clustered based on the average distances between all pairs of cases. This is similar to the result of this study.

The Agglomeration coefficients and the Euclidean distance recorded in this study support the assertion by [13] who had reported that the large coefficients and distance indicated that quite dissimilar members were combined while small coefficients and distance indicated relatively similar members. Cluster groups identified in this study have a high degree of natural association and also depict the agglomeration of cluster groups, their members and their nearest neighbours as showed in (Table 2). [24] reported similar natural selection in the coastal waters of India. The Icicle plot confirmed the extent of the relationship or existing gaps between parameters and showed the similarities or dissimilarities between the cases being combined to form clusters. The smaller the existing gap, the closer is the relationship and *vice versa*. This result is in strong agreement with the reports of [19], [20] that Icicle plots showed the distances (similarities or dissimilarities) between the cases being combined to form clusters and demonstrated the existing gaps between parameters being combined to form cluster in an hierarchical order. The months of April to October in Okoro River estuary usually coincide with the rainy season when high volumes of freshwater are discharged into estuarine waters that lower or dilute the salinity of mangrove water systems. This is similar to the pattern in other West African Zones. For instance, the temporal pattern of Okoro River estuarine water quality was consistent with the two seasons as reported in Nigeria by [8]. This supports the classification of seasons as dry and (November - March) and rainy (April –October) seasons as reported by [9] in the Cross River system, Nigeria. However, the result is different from the findings of [13] who recorded three clusters (winter season, summer season and rainy season) for Mumbai Coast of India. The differences may be due to weather pattern and geographical peculiarity.

5. Conclusion

Agglomerative hierarchical cluster analysis is a powerful tool for analyzing mangrove water quality. The results also showed a high degree of natural selection and association among physicochemical and environmental parameters and seasons as indicated in the Dendrogram (down-top agglomerative tree diagram). The temporal pattern of Okoro River estuary was consistent with the two major seasons segmented (dry and rainy seasons) as widely obtained in the Coast of West Africa. The present baseline information on the hierarchical structure of the plysicochemical properties of Okoro River estuary would form a useful tool for further ecological assessment and monitoring of the mangrove water system of the South eastern coast of Nigeria.

6. References

[1] F. M. Abbas, M. D. Al-karkhil, I. Azmal Hossain. and I. Norli. Application of Cluster Analysis for Water quality Parameters-Juru Estuary (MALAYSIA), International Conference on Environmental Research and Technology (ICERT 2008) pp.486-490,2008.

[2] AKUTEC). Final Report for the Implementation of Akwa Ibom State University, 202pp. 2006.

[3] S. E. Allen '*Chemical Analysis of Ecological Materials'* Blackwell Scientific Publications Oxford, pp.32. 1974.

[4] American Public Health Association (APHA) *Standard Method for the Examination of Water and Waste Water.*14th Edition. American Public Health Association, Washington DC. 1393 P. 1980.

[5] American Public Health Association (APHA)). *Standard Methods for the Examination of Water and Waste Water* 20th ed. APHA, Washington DC, pp.1193,1998.

[6] M. Asif; M. Wagas;, W. M Mhammadand A. Favooq. Application of Multivariate Statistical Techniques for the Characterization of Ground Water Quality of Labhore, Gujranwala and Sialkot (Pakistan). *Pakistan Journal for Analytical Environment and Chemistry*, 12 (1 & 2) pp. 102 -112, 2011.

[7] K. Chan, J. J. Jiao, J. and R. Huang. "Multivariate Statistical Evaluation of Trace Element in Groundwater in a Coastal Area in Schenzheu,"*China Environmental Pollution*, 147 (3) :771-780, 2007.

[8] C. A. Edokpayi., J. K. Saliu and; O. J. Eruteya. "Assessment of Temporal Fluctuation in Water Quality of the Coastal Water of Training Mole, Tarkwa Bay, Nigeria" *Journal of American Science*, 6 (10):1179-1185, 2010.

[9] U. K. Eyenihi, O. A. Adeyemi-Wilson and, A. I. Obiekezie.). Ecological Parameters of the Mangrove Swamp Forest of the Cross River State-Baseline Study. In the Petroleum Industry and the Nigeria Environment. Fed. Min. Works and Housing and the Petroleum Inspectorate NNPC, Lagos, Nigeria, pp. 228 -239.1988.

[10] P. Geetha, K. Shenthi, R. Jaganathan. and, S. Balasubramanian (1999). Modeling seasonal fluctuation of aquatic fauna with reference to the environment factors in a high altitude lake of peninsular index (The Ootacamund Lake).In Limnological research in India (Ed.S.R.Mishra). Daya Publishing House, Delhi-110035, and pp 225-240

[11] A. E. Gibo, Relationship between rainfall trends and flooding in the Niger-Benue River basin. J. Meterol., 13: 132-133, 1988.

[12] S. Gregory, *Statistical Methods and the Geographer*. Longman, London, 1974.

[13] I., Gupta, S. Dhage, and R. Kumar, Study of Variation in Water Quality of Mumbai Coast through Multivariate Analysis Techniques.*India Journal of Marine Sciences,* 38(2) :170 -177, 2009.

[14] B., Helena, R., Pardo, M., Vega, E. Barrado, J. Fernandez, and L. Fernandex, Temporal Evolution of Groundwater Composition in an Alluvial Aquifer (Pisuerga River, Spain) by Principal Component Analysis. *Water Resources*, 34 (3): 807-816, 2000.

[15] T. A. Khan, Multivariate Analysis of Hydrochemical Data of the Groundwater in Parts of Karwan-Sengar Sub-basin, Central Ganga Basin, India. *Global NEST Journal*, 13 (3):229-236, 2011.

[16] A. Michalik, The Use of Chemical and Cluster Analysis for Studying Spring Water Quality in Swietokrzyski National Park, Poland . *Polish J. of Environ. Stud*, 17 (3): 357-362. 2008.

[17] T. B. Mruthrunjaya, and S. P. Hosmani, Application of cluster analysis to evaluate pollution in Lingambudhi Lake in Mysore, Karnataka. *Nature Environment and Pollution Technology*, 3(4): 463-466, 2004.

[18] NEDECO, The waters of the Niger Delta. Reports of investigation by NEDECO (Netherlands Engineering Consultants), The Hague. 1961.

[19] U. K., Pradhan, P. V. Shirodkar, and B. K. Sahu, Physico – Chemical Characteristics of the Coastal Water Off Devi Estuary, Orissa and Evaluation of its Seasonal Changes using Chemometric Techniques. *Current Science*, 96 (9): 1203-1209, 2009.

[20] M. M. Prakash, and A. Dagaonkar, Application of Cluster Analysis to Physico-chemical Parameters of MunjSagarTalab, Dhar (Madhya Pradesh, India). *Recent Research in Science and Technology*, 3(1): 41:50 2011.

[21] S. M., Praveena, A., Ahmed, M., Radojevic, M. H., Abdullah, A. Z. Aris, Factor-cluster analysis and enrichment study of mangrove sediments and examples from Mengkabong Sabah. *The Malaysian Journal of Analytical Sciences*, 11(2): 421-430, 2007.

[22] M. K., Rana, K. V. Bhat, RAPD markers for genetic diversity study among Indian cotton cultivators. *Current Science*, 88(12): 1956-1961, 2005.

[23] P. Satheeshkumar and A. B. Khan, Identification of Mangrove Water Quality by Multivariate Statistical Analysis Methods in Pond cherry Coast, *India Environmental Monitoring and Assessment* DOL 10.1007/S 10661 -011 -2222-4. 2011.

[24] T. Sukarma, and P. Trivedi, Assessment of Water Quality of Bennithora River in Karnataka

through Multivariate Analysis *Nature and Science*, 8 (6): 51 -56. 2010.

[25] S. Thareja, and P. Trivedi,. Assessment of Water Quality of Bennithora River in Karnataka through Multivariate Analysis. *Nature and Science*, 8(6):51-56, 2010.

[26] E. P. Udofia, Applied Statistics with Multivariate Methods. Immaculate Publications Limited, Enugu, Nigeria, pp.406 – 408, 2011.

[27] O. M. Udoidiong, Studies on the Impact of Nipa Palm (*Nypafruticanswurmb*) on the Epibenthic Communities of Mangrove Swamps of Eastern Obolo Local Government Area, Akwa Ibom State, Nigeria. Ph.D. Thesis: Department of Animal and Environmental Biology, University of Port Harcourt, Nigeria, 2005.

[28] S. Yerel, Water Quality Assessment of Porsuk River. Turkey E-J. Chem. 7(2):593-599 2010.