

Full Length Research Paper

Interpretation of surface water quality using principal components analysis and cluster analysis

A. O. Ayeni* and A. S. O. Soneye

Department of Geography, University of Lagos, Lagos, Nigeria.

Accepted 27 February, 2013

Variety approaches are being used to interpret the concealed variables that determine the variance of observed water quality of various source points. A considerable proportion of these approaches are statistical methods, multivariate statistical techniques in particular. The use of multivariate statistical technique(s) is/are required when the number of variables is large and greater than two for easy and robust evaluation. By means of multivariate statistics of principal components analysis (PCA) and cluster analysis (CA), this study attempted to determine major factors responsible for the variations in the quality of 30 surface ponds used for domestic purposes in six (6) selected communities of Akoko Northeast LGA, Ondo State, Nigeria. The samples' locations were classified into mutually exclusive unknown groups that share similar characteristics/properties. The laboratory results of 20 parameters comprising 6 physicals, 8 chemicals, 4 heavy metals and 2 microbial from the sampled springs were subjected to PCA and CA for further interpretation. The result shows that 5 components account for 97.52% of total variance of the surface spring quality while 2 cluster groups were identified for the locations. Based on the parameters concentrations and the land uses impacts, it was concluded that domestic and agricultural waste strongly influenced the variation and the quality of ponds in the area.

Key words: Multivariate statistics, surface water, quality, interpretation.

INTRODUCTION

The complexity of water quality as a subject is reflected in various types of measurements. These measurements include simple (in situ), basic and more complex parameters (Laboratory). For instance, pH, temperature and DO could be measured with a portable in-situ pH meter, a mercury thermometer and M90 Mettler Toledo AG DO meter, respectively (USGS, 2006). BOD, TSS, Cu, Fe, Total bacterial counts, Total Coliforms etc. could be analyzed in the laboratory using standard methods for water samples examination (Ayoade, 1988; APHA, 2005; WHO, 2006; USGS, 2006).

The surface water quality assessment is a matter of serious concern today due to its role in servicing domestic water needs of water stress areas (Yerel, 2010; Ayeni

et al., 2011). The surface water quality is principally influenced by the natural and the anthropogenic processes particularly in the urban areas and agricultural activities around the rural areas (Ayeni, 2010; Ayeni et al., 2011). The level of water quality is relatively determined by the content of physical, chemical and biological parameters present in it. Relationship between two parameters may also lead to increases or decrease in the concentration of others. This relationship or association is usually achieved using multivariate statistical techniques (Ifabiye, 1997; Mazlum et al., 1999; Jaji et al., 2007). This is because some analysis is primarily concerned with relationships between samples, while others trepidation are largely with relationships between variables.

*Corresponding author. E-mail: aayeni@unilag.edu.ng.

According to Mazlum et al. (1999) and Yerel (2010), many multivariate statistical techniques have the capacity to summarize large data by means of relatively few parameters. Nonetheless, the choice of using any of the multivariate statistical techniques lies in the nature of the data, problem, and objectives of the study. In view of the fact that the daily drinking and domestic water needs of the majority of residents in the area are met by unsafe surface water, in particular surface springs (Ayeni, 2010), there is the need to understand the variables that control the variations in their quality. Principal Component Analysis (PCA) and Cluster Analysis (CA) of multivariate techniques are therefore adopted for the study. According Praus (2005) PCA is used to search new abstract orthogonal eigen-values which explain most of the data varies in a new harmonize structure. Each principal component (PC) is a linear combination of the original variables and describes different source of information by eigenvalue based on the decomposition of the covariance/correlation matrix (Geladi and Kowalski, 1986). PCA is designed to modify the observe variables into uncorrelated variables of linear combinations of the original variables called "principal components" (Praus, 2005; Yerel, 2010) as well as to investigate the factors which caused variations in the observed datasets (Mazlum et al., 1999). The principal component therefore provides information for inter-pretation and better understanding of the most meaningful parameters which describes the whole data set through data reduction with a minimum loss of the original information. Cluster analysis (CA) is an exploratory analysis technique for classifying a set of observations into two or more mutually exclusive unknown groups based on combinations of interval variables (Stockburger, 1997; Trochim, 2006; Murali-Krishna et al., 2008; Yerel, 2010). According to Yerel (2010), CA 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. In this study, observation and sampling location classification were done by the use of Hierarchical Cluster Analysis (HCA) procedure. HCA identify relatively homo-geneous groups of variables (cases) through dendrogram based on selected characteristics. Dendrogram clearly distinguished locations behaviours and interprets the description of the hierarchical clustering in a graphical format (Hastie et al., 2001; Ryberg, 2006).

Study area and sampling locations

The study area lies between longitude 5° 38' and 6° 04'E, and latitude 7° 26' and 7° 42'N in the northern senatorial part of Ondo State, SW - Nigeria (Figure 1). It is bounded

by the Akoko North West Local Government Area (LGA) to the north, Edo State to the east, Akoko South East and West LGAs to the south, and Ekiti State to the west. It is primarily characterized by undulating relief ranges between 149 and 671 m above sea level and located on basement complex rock formation (Ayeni, 2010). The complex composed mainly of granite, mischist, gneisses and metasediment (Barbour et al., 1982; Adekunle et al., 2007). The area falls within sub-tropical climate with average rainfall over 1500 mm per annum. The temperature ranging from about 30 to 38°C while the vegetation cover is dominated by derived secondary rain forest. The soil is classified as Ferric Acrisols with relatively higher cation profiles (Fasona et al., 2007; Nwachokor and Uzu, 2008).

METHODS

Twenty water quality parameters from 30 surface springs were monitored for 12 months. For each month, water sample from selected springs are collected and analysed in the laboratory using APHA (2005) standard methods for the examination of water and wastewater. The coordinates of sampled springs are interpolated on geo-rectified map of the study area (Figure 1). The selected surface water quality parameters for the study are pH, temperature, dissolved oxygen (DO), biochemical oxygen demand (BOD), total suspended solid (TSS), total dissolved solid (TDS), Turbidity, total hardness (TH), calcium hardness (Ca⁺), Magnesium hardness (Mg²⁺), Chloride (Cl⁻), Nitrate (NO₃⁻), Phosphate (PO₄³⁻), Oil and grease, Copper (Cu), Iron (Fe), Manganese (Mn), Zink (Zn), total bacterial counts (TBC) and total coliforms (TC).

The laboratory results were evaluated using multivariate statistical techniques of PCA for selected parameters and CA for sample locations. The principal component is thus given by the formula:

$$Z_{ij} = a_{i1}X_{1j} + a_{i2}X_{2j} + a_{i3}X_{3j} + \dots + a_{in}X_{nj} \quad (i)$$

Where, z = component score, a = component loading, x = measured value of variable, i = component number, j = sample number, and n = the total number of variables.

In the case of cluster analysis, the formula is given thus:

$$d_{ij}^2 = \sum_{k=1}^m (z_{ik} - z_{jk})^2 \quad (ii)$$

Where d_{ij}^2 = the Euclidean distance, z_{ik} = the values of variable k for object i ; z_{jk} = the values of variable k for object j , and m = the number of variables.

RESULTS AND DISCUSSION

Quality assessment

The pH of springs water ranges between 5.3 and 8.3 (Figure 2a). Findings reveal that 13 ponds (Omiidu, Agboomi, Isemo, Alapoti, Arae, Agbo-Ilepa, Isun, Imurun, Omi-Alagoke, Omi-Olokungboye, Odewo and Adangbara in Ikare, and Ajagun in Ise) are below the regulatory limits

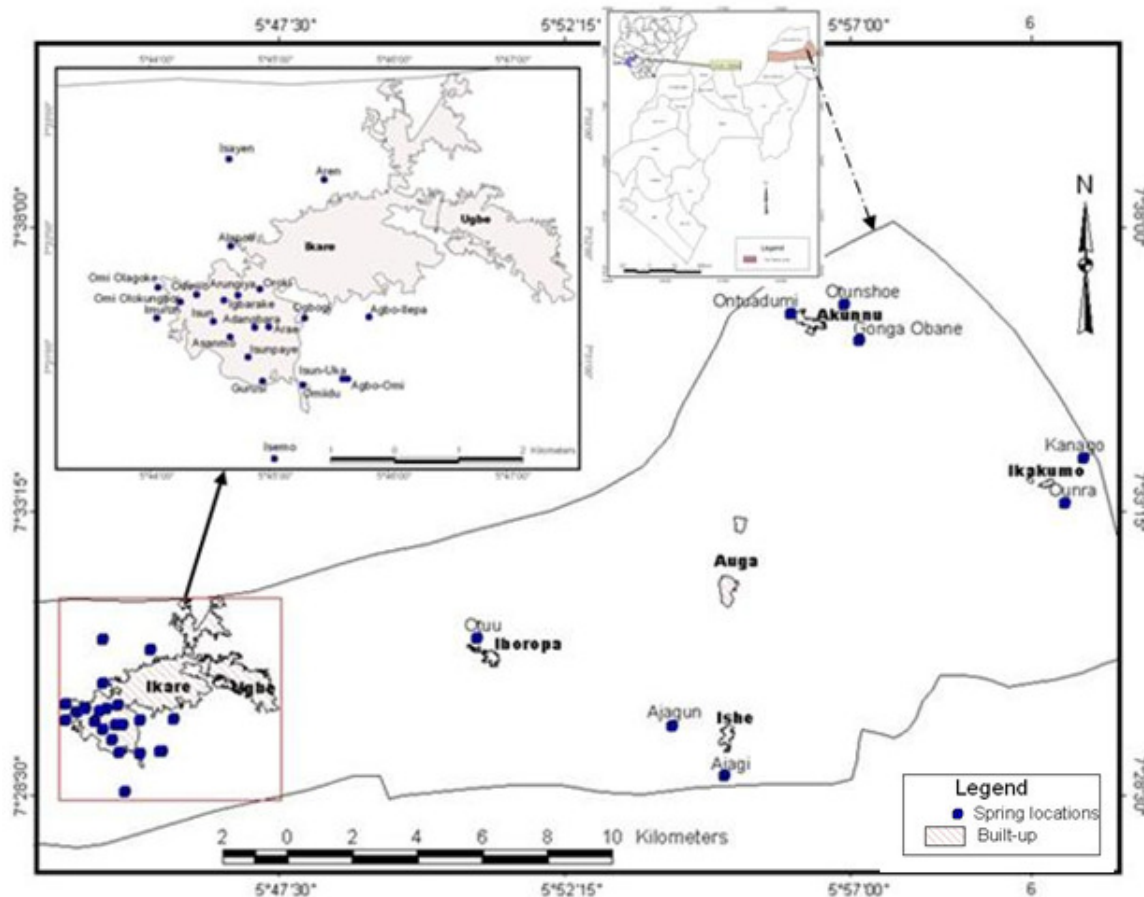


Figure 1. Selected spring locations in Akoko Northeast LGA of Ondo State, SW – Nigeria.

of 6.5 and 8.5. The spring water temperature ranges between 19.6 and 29.2°C with the lowest and highest recorded at Gurusi in Ikare and Ajagun in Ise respectively (Figure 2b). DO in the springs range from 1.13 to 10.0 mg/l. DO is generally lower when compared with their mean except Arae, Igbarake, Omi-Alagoke, Omi-Olokungboye, Asanmo, Ogbogi and Arae at Ikare, Ajagun at Ise, Otuu at Iboropa and, Otunadumi and Gongga Obane at Akunnu, which are above the regulatory limits. The lowest and highest values were recorded at Otunshoe and Otunadumi at Akunnu respectively (Figure 2c). TSS in the springs ranges between 2.0 and 75.0 mg/l with a mean value of 15.0 mg/l (Figure 2d).

TDS from the sampled springs ranges between 22 and 546 mg/l (Figure 2e). spring water turbidity ranges from 0.0 to 0.21 mg/l with a mean value of 0.03 mg/l (Figure 2f). BOD in the springs ranges between 2.20 and 15.20 mg/l with the lowest and highest recorded at Ajagun at Ise and Ounra at Ikakumo respectively (Figure 2g). TH in the sampled springs ranges between 12.0 and 376.0 mg/l (Figure 2h). Results show that in 4 springs (Arae, Igbarake, Oroki and Omi-Olokungboye), Ca^{2+} in the spring water ranges between 4.0 and 180.0 mg/l (Figure 2i). Results reveal that 20 springs (Agboomii, Isemo,

Alapoti, Agbo-Ilepa, Arungiya, Ajagi, Ajagun, Isun, Imurun, Omi-Alagoke, Omi-olokungboye, Odewo, Adangbara, Ajagi, Gongga Obane, Ounra, Kanago, Isun-uka, Gurusi, and Isunpaye) are outside WHO (2006) regulatory limits of 75.0 and 250 mg/l. Mg^{2+} in the spring water ranges between 7.00 and 196.0 mg/l (Figure 2j). Cl⁻ in all the sampled springs ranges between 8.0 and 144 mg/l (Figure 2k). Findings reveal that Cl⁻ in all sampled sources are generally low when compared with WHO (2006) minimum limit of 200 mg/l. NO_3^- in the springs water ranges between 0.10 and 0.91 mg/l (Figure 2l). When this is put into comparison with the WHO (2006) regulatory limits of 5 and 10 mg/l, NO_3^- in all springs are below limits.

PO_4^{3-} in the sampled springs ranges between 0.80 and 8.20 mg/l (Figure 2m). Findings reveal that PO_4^{3-} in all sampled sources exceed the regulatory limits of 0.30 and 0.05 mg/l for WHO (2006) and SON (2007) respectively. Oil and grease in all sampled springs range between 0.01 and 3.12 mg/l (Figure 2n). In comparison with WHO limit of 0.1 mg/l, fifteen springs (Isemo, Ajagun, Ounra, Isun, Imurun, Omi-Alagoke, Omi-Olokungboye, Odewo, Oroki, Asanmo, Arae and Adangbara in Ikare and, Otunshoe, Otunadumi, Gongga-Obane in Akunnu) are higher than the

limit while the remaining sampled sources fall below the limit. TBC in the sampled springs range between 1.10 and 2.40 cfu/ml with a mean value of 1.47 cfu/ml (Figure 2o). Spring water TC ranges between 0.00 and 3.50 cfu/ml with a mean value of 0.50 cfu/ml (Figure 2p).

Zn detected in sampled springs ranges between 0.01 and 0.66 mg/l (Figure 2q). The findings reveal that Zn contents in all sampled sources are generally low when compared to the WHO (2006) regulatory limit of 5.0 mg/l. Fe in all sampled springs ranges between 0.4 and 4.7 mg/l (Figure 2r). Fe detected in all sampled springs are higher than the WHO (2006) and the SON (2007) regulatory limit of 0.3 mg/l. Mn detected in the 17 springs with their values ranging between 0.01 and 1.15 mg/l (Figure 2s). Findings reveal that Mn in 3 springs (Agboomi, Isemo, Omi-Olokungboye and Oroki at Ikare) is higher than the WHO (2006) limits of 0.05 and 0.1 mg/l, and SON limit of 0.2 mg/l. Three springs (Otuu at Iboropa and, Gongga-Obane at Akunnu) are higher than the WHO (2006) limits of 0.05 and 0.1 mg/l. Cu in the sampled springs ranges between 0.0 and 2.10 mg/l (Figure 2t). From the findings, Cu in 7 springs (Gurusi, Arae, Agbo-llepa, Omi-Alagoke, Odewo and Ogbogi at Ikare, and Otunshoe at Akunnu), are above the WHO (2006) and SON (2007) regulatory limit of 1.0 and 0.1 mg/l respectively.

In summary, this finding corroborates Mallo et al. (2001) who reported that surface waters in the Tandil region of Argentina are highly polluted when compared to other sources such as pipe borne water system, wells, springs by analyzing pH, temperature, hardness, Cl, Ca, nitrates and bacteriology. Drinking water having microbial pollution posed a major threat to human health.

Principal component and cluster analyses

The result of principal components analysis in Table 1 shows that of the 20 components, only 5 had extracted eigenvalues over 1. This is based on Chatfield and Collin (1980) assumption which stated that components with an eigenvalue of less than 1 should be eliminated. The extracted 5 components were subsequently rotated according to varimax rotation in order to make interpretation easier and fundamental significance of extracted components to the water quality status of the selected springs. The result of rotation revealed further, the percentages of the total variances of the 5 extracted components when added account for 97.52% (that is their cumulative variance) of the total variance of the observed variables. This indicates that the variance of the observed variables had been accounted for by these 5 extracted components. The calculated components loadings, eigenvalues, total variance and cumulative variance are shown in Table 2 while the scree plot of the eigenvalues of observed components is depicted in Figure 3.

Based on the component loadings, the variables are grouped accordingly with their designated components as follows:

- Component 1: TH, Mg^{2+} , TDS, Cl^- and Ca^+
- Component 2: Zn, Turbidity, DO, pH, TSS and PO_4^{3-}
- Component 3: Cu, temperature, TC, TBC and NO_3^-
- Component 4: Oil and grease and Fe
- Component 5: Mn and BOD

Component 1, 2, 3, 4 and 5 explained 36.996, 22.058, 16.674, 12.908 and 8.882% of the variance respectively. Classifying the component loading according to Liu et al. (2003) the loading values greater 0.75 signifies "strong", the loading with absolute values between 0.75 and 0.50 indicate "moderate" while loading values between 0.50 and 0.30 denote as "weak". Using this classification, all variable in component 1 and component 2 had strong positive loading except PO_4^{3-} with moderate positive. Of the 5 variables in component 3, two had strong positive loading (Temperature and TC), NO_3^- had moderate positive loading while Cu and TBC were signified with strong and moderate loading respectively. All variables in component 4 and 5 explained strong positive loading.

An interpretation of the rotated 5 principal components is made by examining the component loadings noting the relationship to the original variables. Component 1 gives information about the variations in TH, Mg^{2+} , TDS, Cl^- and Ca^+ . In this component, loading indicates that organic matter and organic acids which could be attributed to various anthropogenic activities and geological formation and/or composition of the area greatly influenced the quality of selected springs. The same also interpreted for component 2 but considers its eigenvalue and total variance, it is quite lower compared with component 1.

Components 3 explained information about Cu, temperature, TC, TBC and NO_3^- . This component represents pollution from domestic and agricultural waste as well as the geological composition of the area. However, the significance of NO_3^- in the component 3 indicates that nitrification takes place in the vicinity of the springs. In the component 4, it can be understood that dissolved or emulsified oil and grease extracted from water especially unsaturated fats and fatty acids and Fe extracted from the parent rock of the area are of the significance in that component. In the component 5, Mn presence is an indication of the parent rock influence while BOD claimed that there is a high level of organic pollution, caused usually by poorly treated waste water.

The dendrogram of the observed locations dataset was generated using Euclidean distance of HCA for CA result (Figure 4 and Table 3). Based on Euclidean distance, two major clustering groups (cluster 1 and cluster 2) were observed. Cluster 1 characterized with low Euclidean distance corresponds to 23 locations (6, 9, 7, 3, 13, 19, 17, 22, 10, 30, 4, 5, 2, 12, 23, 27, 18, 26, 29, 21, 25, 28 and 24). Cluster 2 which has high Euclidean distance is

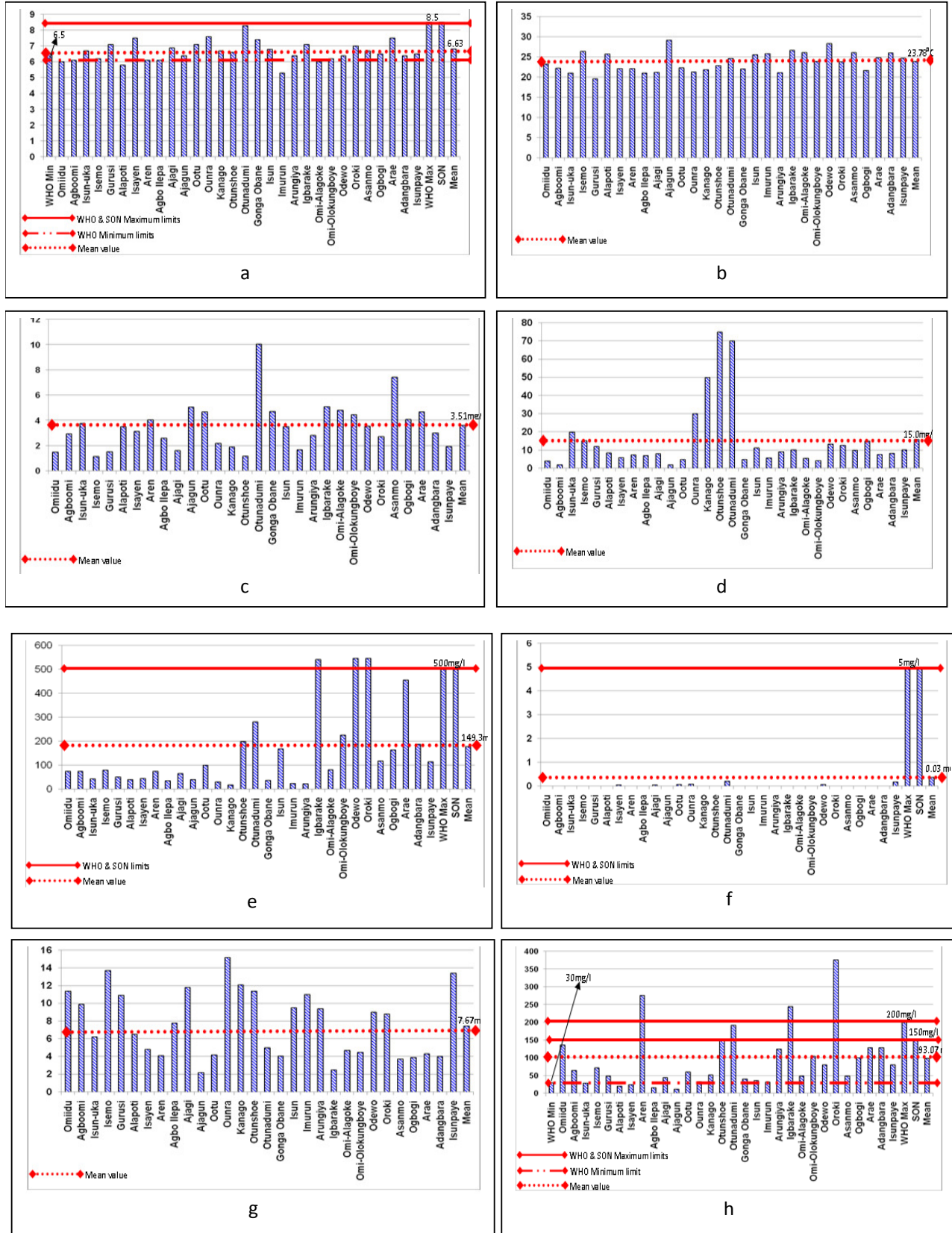


Figure 2. Concentration of: (a) pH; (b) Temperature; (c) DO; (d) TSS; (e) TDS; (f) Turbidity; (g) BOD; (h) TH; (i) Ca^{2+} ; (j) Mg^{2+} ; (k) Cl^- ; (l) NO_3^- ; (m) PO_4^{3-} ; (n) Oil and grease; (o) TBC; (p) TC; (q) Zinc; (r) Iron; (s) Manganese; (t) Copper.

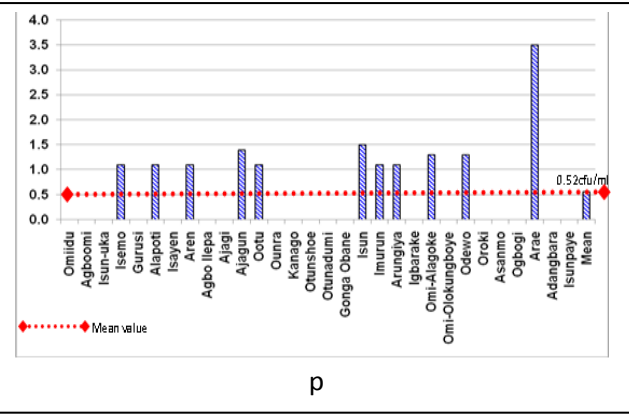
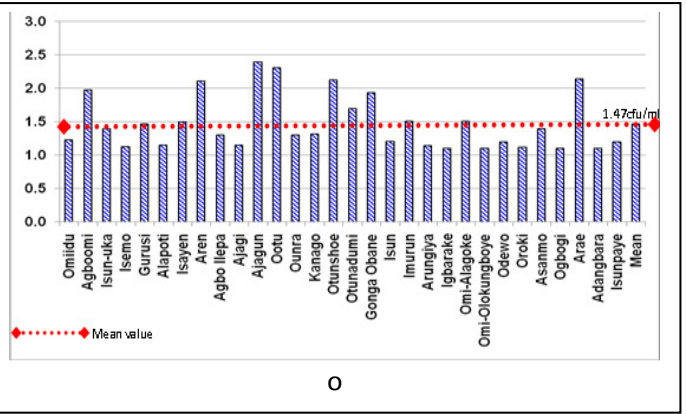
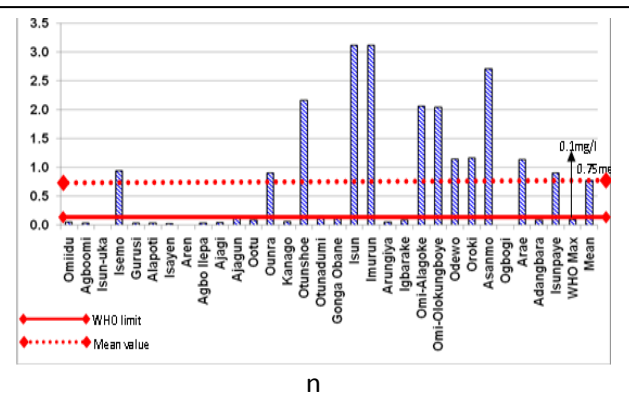
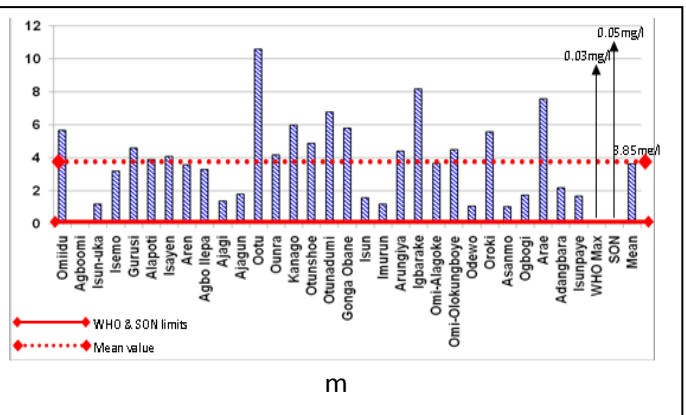
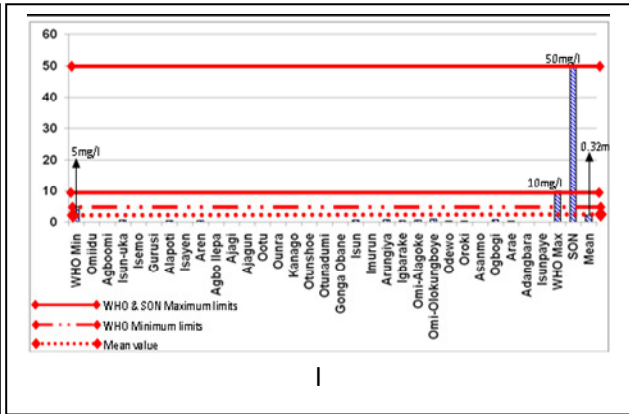
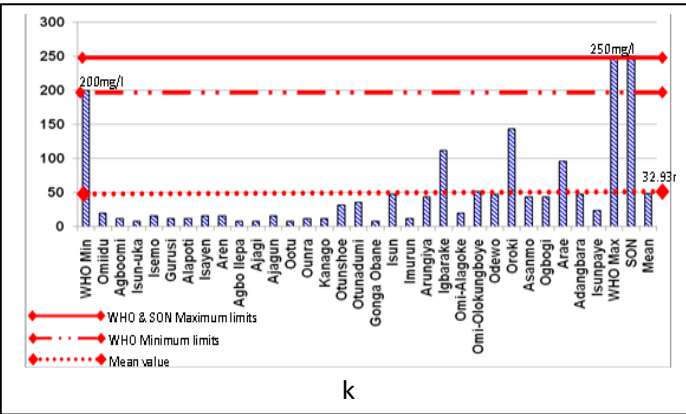
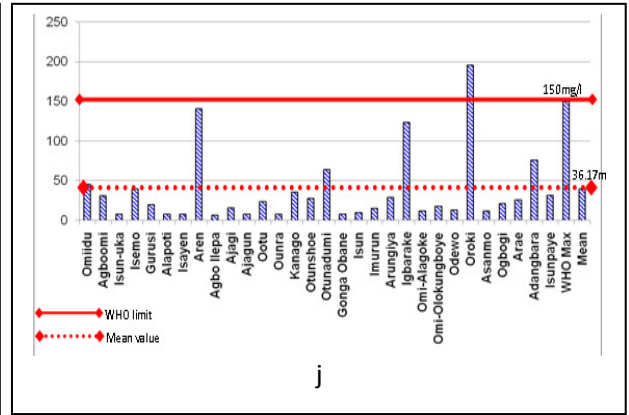
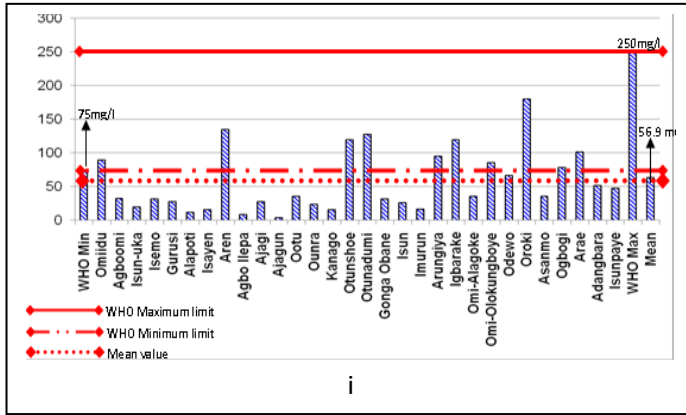


Figure 2. Contd.

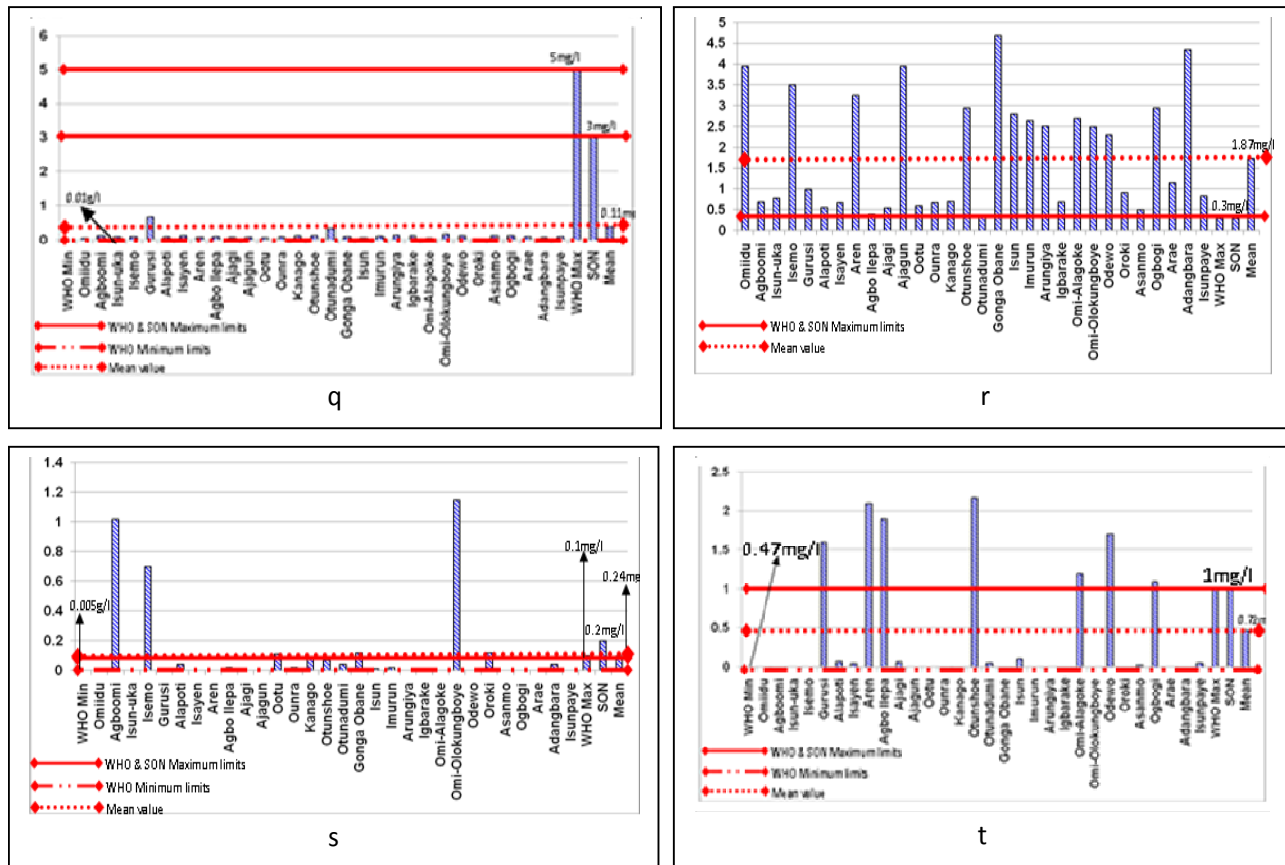


Figure 2. Contd.

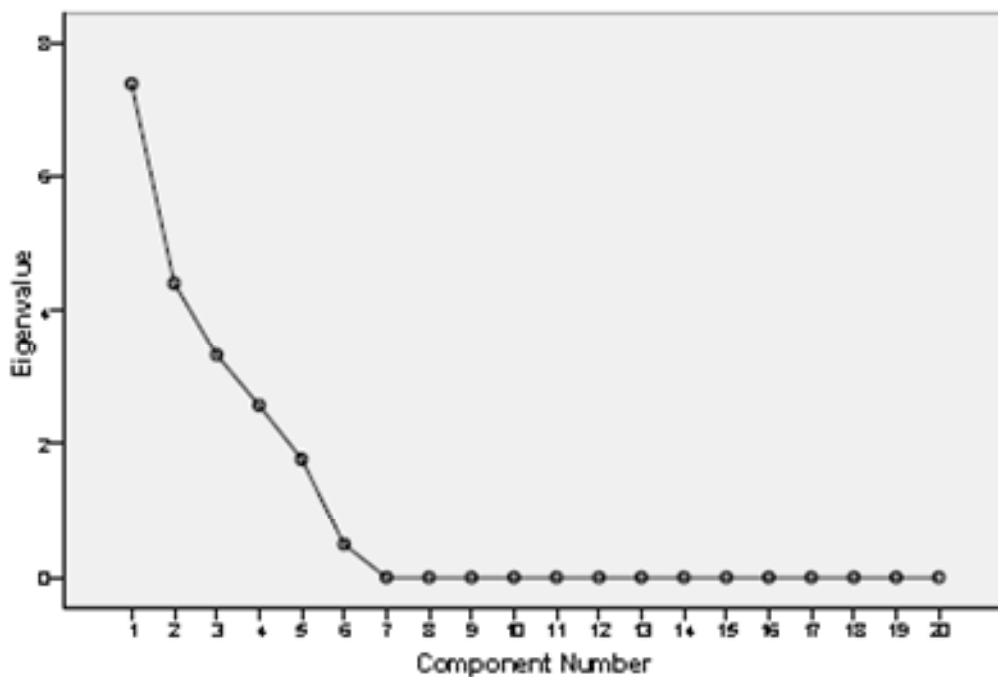


Figure 3. The scree plot of the eignvalues.

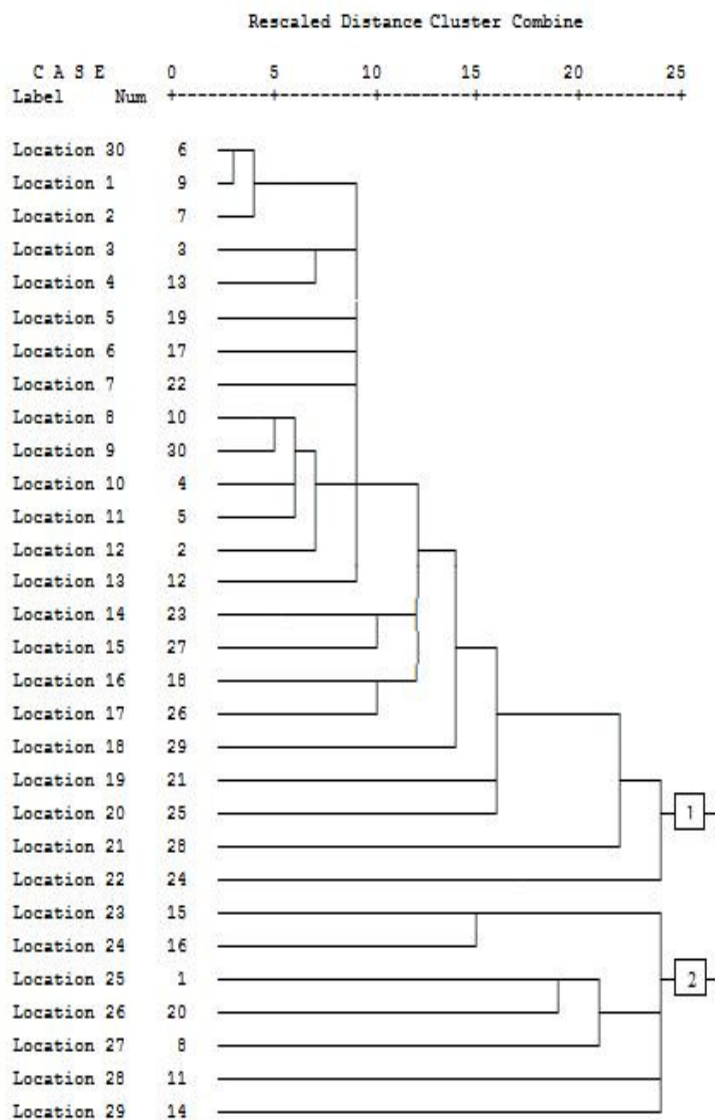


Figure 4. Dendrogram of hierarchical cluster analysis.

coherent to 7 locations (15, 16, 1, 20, 8, 11 and 14). Subgroup clusters were also clarified within the major cluster 1 and vary with significance Euclidean distance.

The dendrogram clarifies cluster 1 as the abnormality observation which had high variation in the concentration of the surface water quality parameters compared to cluster 2 surface water samples concentration. The variation in cluster 1 might due to low polluted effluents from non-point sources (agricultural and urban activities). Cluster 2 shows a high pollution from agricultural area which encompasses the springs.

Conclusion

This study presents the usefulness of multivariate statistical techniques of large and complex dataset in order to

obtain better information and interpretation concerning surface water quality. Principal component analyses helped in identifying the factors responsible for surface water quality variations in 6 selected communities. The result revealed that the percentages of the total variances of the 5 extracted components when added account for 97.52% (that is their cumulative variance) of the total variance of the observed variables. The variation in components 1 and 2 loadings indicates that organic matter and organic acids could greatly influence the quality of selected springs. Component 3 ascribed mainly to domestic and agricultural waste of the springs environment while component 4 and 5 respectively attributed to dissolved/emulsified poorly treated waste water. On the other hand, the result of cluster analysis revealed 2 major clustering groups resulting from the influences of agricultural and urban activities around the samples'

Table 1. Principal component matrix of eigenvalues less than 1 (5 components extracted).

Variable	Component				
	1	2	3	4	5
pH	.838	.001	-.195	.379	.171
Temperature	-.029	.273	-.589	0.649	-.067
Do	.687	-.349	-.592	.015	.115
BOD	-.535	.334	.581	.450	-.243
TSS	.641	-.361	.378	.484	.200
TDS	.688	.715	.050	-.008	.092
Turbidity	.755	-.399	-.380	.324	-.042
TH	.734	.654	.145	-.004	-.097
Ca	.825	.476	.258	.116	.035
Mg ²⁺	.576	.776	.017	-.126	-.222
Cl ⁻	.449	.871	.018	-.152	.104
NO ³⁻	-.361	.451	-.520	-.083	.578
PO ₄ ²⁻	.928	-.022	.045	.020	-.221
Oil & Grease	-.273	.408	.312	.427	.663
Zn	.621	-.641	-.224	.379	-.084
Fe	-.513	.230	.428	.701	.069
Mn	-.353	.209	.096	.517	-.744
Cu	-.026	-.449	.822	-.309	.161
TBC	.462	-.439	.596	.304	.324
TC	-.770	.126	-.490	.346	.173

Table 2. Rotated component loading matrix, eigenvalues, total variance and cumulative variance.

Variable	Component				
	1	2	3	4	5
TH	983				
Mg ²⁺	979				
TDS	979				
Cl ⁻	954				
Ca ⁺	900				
Zn		.959			
Turbidity		.947			
Do		.840			
pH		.818			
TSS		.779			
PO ₄ ²⁻		.621			
Cu			-.943		
Temperature			.872		
TC			.766		
TBC			-.699		
NO ³⁻			.661		
Oil & Grease				.953	
Fe				.745	
Mn					.952
BOD					.773
Eigenvalue	7.399	4.412	3.335	2.581	1.776
Total variance %	36.996	22.058	16.674	12.908	8.882
Cumulative variance %	36.996	59.054	75.728	88.634	97.517

Table 3. Pond names and their cluster membership.

Identification	Ponds' names	Cluster
Location 1	Omiidu	1
Location 2	Agboomi	2
Location 3	Isun-uka	2
Location 4	Isemo	2
Location 5	Gurusi	2
Location 6	Alapoti	2
Location 7	Isayen	2
Location 8	Aren	1
Location 9	Agbo-Ilepa	2
Location 10	Ajagi	2
Location 11	Ajagun	1
Location 12	Otuu	2
Location 13	Ounra	2
Location 14	Kanago	1
Location 15	Otunshoe	1
Location 16	Otunadumi	1
Location 17	Gonga Obane	2
Location 18	Isun	2
Location 19	Imurun	2
Location 20	Arungiya	1
Location 21	Igbarake	2
Location 22	Omi-Alagoke	2
Location 23	Omi-Olokungboye	2
Location 24	Odewo	2
Location 25	Oroki	2
Location 26	Asanmo	2
Location 27	Ogbogi	2
Location 28	Arae	2
Location 29	Adangbara	2
Location 30	Isunpaye	2

locations. Cluster 1 characterized with low Euclidean distance corresponds to 23 locations and clarifies with sub groups that varies with significance Euclidean distance while cluster 2 coherent to 7 locations and observed high Euclidean distance with sub group of insignificance Euclidean distance. Therefore, it is worthwhile to conclude that PCA and CA are better tools for better understanding of the concealed information about parameters variance and datasets discrete information in water quality assessment studies.

REFERENCES

- Adekunle IM, Adetunji MT, Gbadebo AM, Banjoko OB (2007). Assessment of Groundwater Quality in a Typical Rural Settlement in Southwest Nigeria. *Int. J. Environ. Res. Public Health* 4(4):307-318
- APHA (2005). Standard methods for examination of water and wastewater". American Public Health Association, 21st Edition, Washington DC pp.4-144
- Ayeni AO (2010). Spatial Access to Domestic Water In Akoko Northeast LGA, Ondo State, Nigeria. Unpublished Ph.D. Thesis, University of Lagos, Lagos, Nigeria
- Ayeni AO, Balogun II, Soneye ASO (2011). Seasonal Assessment of Physico-chemical Concentration of Polluted Urban River: A Case of Ala River in South western Nigeria. *Res. J. Environ. Sci.* 5(1):21-35.
- Ayoade JO (1988). *Tropical Hydrology and Water Resources*. London, Macmillan Publishers Ltd.
- Barbour KM, Oguntoyinbo JS, Onyemelukwe JOC, Nwafor JC, (1982). Nigeria in Map. London, Hodder and Stoughton p.209.
- Chatfield C, Collin AJ (1980). *Introduction to Multivariate Analysis*. Chapman and Hall in Association with Methuen, Inc. 733 Third Avenue, New York NY.
- Fasona AS, Omolayo FO, Falodun AA, Ajayi OS (2007). Granite Derived Soils in Humid Forest of Southwestern Nigeria - Genesis, Classification and Sustainable Management. *America – Eurasian J. Agric. Environ. Sci.* 2(2):189-195.
- Geladi P, Kowalski BR (1986). Partial least square regression: A tutorial. *Anal. Chim. Acta* 185:1-17.
- Hastie T, Tibshirani R, Friedman J (2001). *The elements of statistical learning – data mining, inference, and prediction*. New York, Springer Science+Business Media, Inc., p.533.
- Ifabiye IP (1997). Variation in Water Gravity with Rainfall Incidences: A Case Study of Ogbe Stream Ile-Ife, Ife. *Res. Publications Geogr.* 6(1&2):139-144.
- Jaji MO, Bamgbose O, Odukoya OO, Arowolo TA (2007). Water Quality Assessment of Ogun River, South West Nigeria. *Environ. Monitor. Assess.* 133:473-482
- Liu CW, Lin K H, Kuo YM (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *Sci. Total Environ.* 313:77-89.
- Mallo R, Lagato A, Erpelding A, Lopez R, Andres M, Tabera A, Murno G (2001). Research on water pollution in the Tandil Mountain basin, Argentina. *Inform. Technol.* 12(1):45-50
- Mazlum N, Ozer A, Mazlum S (1999). Interpretation of Water Quality Data by Principal Components Analysis. *Tran. J. Eng. Environ. Sci.* 23:19-26.
- Murali Krishna PSM, Moses SG, Murali Krishna VSGK (2008). Application cluster analysis, discriminate analysis and principal component analysis for water quality evaluation for the river Godavari at Rajahmundry region. *Int. J. Appl. Environ. Sci.* 2(3):195-209
- Nwachokor MA, Uzu FO (2008). Updated Classification of Some Soil Series in Southwestern Nigeria. *J. Agron.omy* 7(1):76-81
- Praus P (2005). Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA* 31(4):417-422
- Ryberg KR (2006). *Cluster Analysis of Water-Quality Data for Lake Sakakawea, Audubon Lake, and McClusky Canal, Central North Dakota, 1990-2003*, Prepared in cooperation with the Bureau of Reclamation, U.S. Department of the Interior Scientific Investigations Report 2006–5202, U.S. Geological Survey, Reston, Virginia p.47.
- SON (2007). Nigerian Standard for Drinking Water Quality. ICS 13.060.20, Approved by Standards Organization of Nigeria (SON) Governing Council, Lagos and Abuja, Nigeria.
- Stockburger DW (1997). *Multivariate Statistics: Concepts, Models, and Applications*. <http://www.psychstat.missouristate.edu/multibook/mlt04.htm> accessed on the 29/06/2011.
- Trochim WM (2006). *The research methods knowledge base (2nd ed.)*. Retrieved April 13, 2010, from: <http://www.Socialresearchmethods.net/kb/>.
- USGS (2006). Common water measurements Page Last Modified: Monday, 28-Aug-2006 14:56:21 EDT <http://ga.water.usgs.gov/edu/characteristics.html> Accessed on the 14/02/2009
- WHO (2006). *Guidelines for Drinking-Water Quality, Volume 1, Recommendations 1st Addendum to 3rd edition*. World Health Organization, Geneva (Electronic version). http://www.who.int/water_sanitation_health/dwq/gdwq3rev/en/index.ht_ml Accessed on the 14/02/2009.
- Yerel S (2010). Water Quality Assessment of Porsuk River. Turkey E-J. Chem. 7(2):593-599.