Quest Journals
Journal of Research in Environmental and Earth Science
Volume 2~ Issue 7 (2016) pp: 01-09
ISSN(Online) : 2348-2532
www.questjournals.org

Research Paper

Application of Multivariate Statistical Techniques in Hydro geochemical Process Mapping of Nairobi City and its Environs

Sosi Benjamin
Department of Natural Resources, Egerton University, P.O Box 536, Egerton, Kenya

Received 20 January, 2016; **A**ccepted 16 February, 2016

ABSTRACT:- *Published physico-chemical and microbiological datasets were analyzed to map hydro geochemical processes using Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). In order to eliminate effects of scale dimensionality, PCA decomposed the 19 individual metric parameters determined in water samples to 4 latent constructs explaining data variation at about 100 % of the cumulative variance; the first factor 52.01%, the second factor 22.97%, the third factor 15.69%, and the fourth 9.33% of the total variances explained respectively. These 4 components represented alkalinity, ionic/hardness, conductivity and iron respectively. By way of Ward's linkage CA of the samples characterized, the four clusters were confirmed. These agglomeration schedule interpreted in integrative approach with site geology, classified water samples according to process and pathways; fault permeability, water-rock interaction, weathering processes and fracture permeabilility. Multivariate statistical approaches are hence useful for analysis and interpretation of hydro chemical data, identification and understanding of pollution sources for management of the quality of water.*

Keywords:- *Multivariate Statistical Techniques, Hydro chemical Process, Mapping, Nairobi*

I. INTRODUCTION

Groundwater is one of the most important natural resources in the city of Nairobi. Its role in the overall water supply for Nairobi City is discussed by [1]. Urban planning and development challenges provoke sustainable protection of water and environmental quality within cities [2] and more specifically Nairobi [3]. Water resources quality deterioration due to salinity and contamination processes has been reported [4], [5], [6]. Potentially harmful microbes (e.g. *Escherichia coli*) and concentrations of certain element species (e.g. F, NO_3^-,) form the health-based criteria for water quality evaluations in all these studies. As observed by [7] and [8], hydrochemical mapping of the origin and mechanisms of the contamination process is fundamental to solve and manage the problem.

Multi-Variate Data Analysis (MVDA) techniques have been utilized broadly in this framework (e.g. [9], [10], [8], and [11]). Specific applications include: evaluation and interpretation of ground water quality [12], [10]), possible sources of pollution/polluting processes [8], coastal aquifers' assessment of hydrochemical processes, and interaction of surface water/groundwater [13]. PCA and HCA statistics have been utilized to identify and analyze quality parameters and pollutants sources in nearly all the fore-mentioned researches.

1.1 The study area

Nairobi city and its environs herein referred to as study area, lies on the eastern flank of the Kenyan rift. The study area is bounded by latitudes and 1° 10'S and 1° 25'S and longitude 36° 40'E and 36° 40'E with easterly gently slope at mean altitude 1,761 meters above sea level. The main physiographic features include the Ngong hills (towards the west), the rivers; Ngong, Nairobi and Mathare. Proximity of the study area to the Rift system causes occasional minor earthquakes and tremors.

The Water Resources Management Authority (WRMA) considers the Nairobi Aquifer System (NAS) to be a strategic aquifer in an alarm state [14]. The NAS covering surface area of approximately 6,500 km^2, and underlies the Nairobi city. The dominant aquifer type is the inter-montane valley-fill characterized by multilayered volcanic / volcanoclastic materials of Plio-Pleistocene origin (more specifically lavas and lake sediments). Regional flow regime ranges from intergranular to fissure. The NAS is recharge is from the eastern franks of Rift Valley at the 109 MCM/year with groundwater flowing eastwards [14]. The system becomes entirely confined in the eastern parts of the area where the aquiferous Upper Athi Series (Tuffs and Lake Beds) typically occurring at depths of 160±40 m bgl is wholly confined.

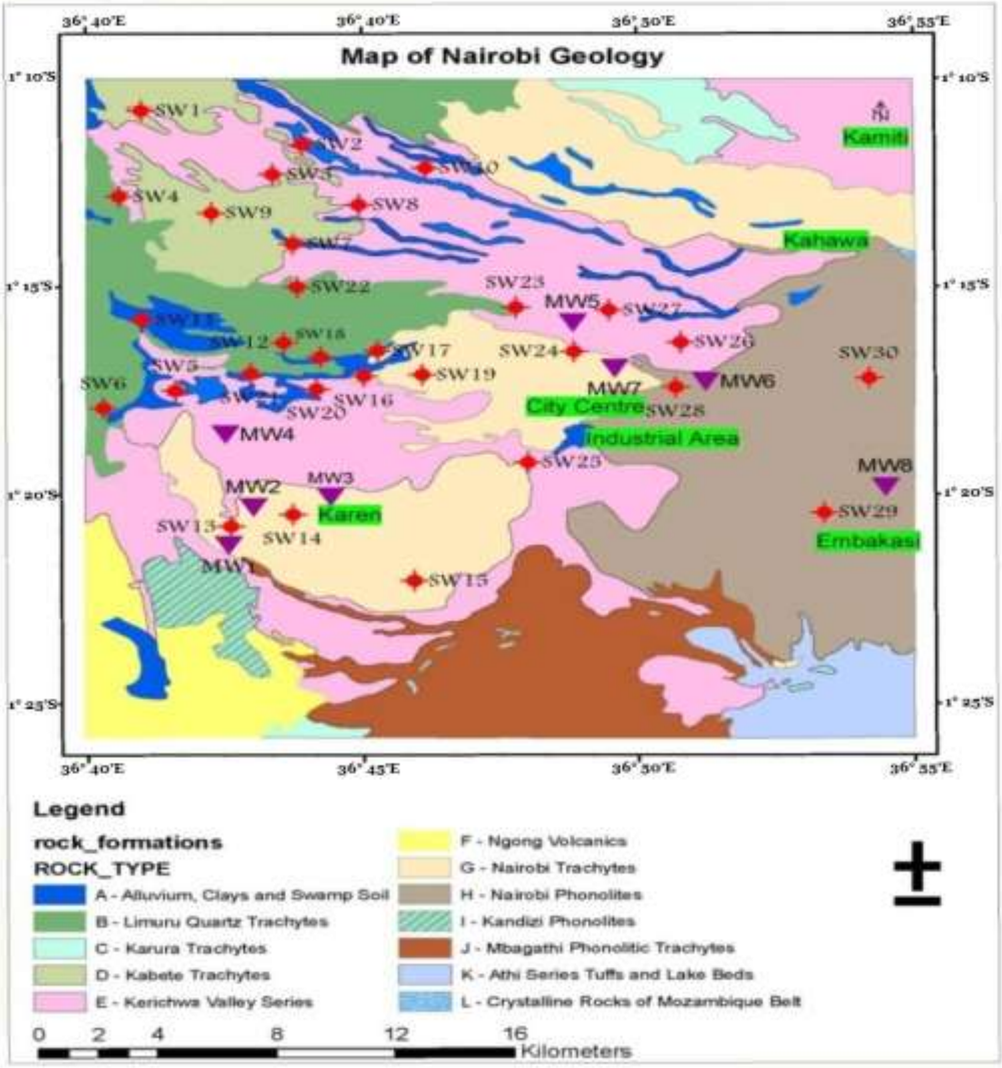

Fig. 1: Geological map of Nairobi area [15] showing Study Wells (SW) and Monitoring Wells (MW) {Source: [6]}

The NAS is being abstracted at 58 MCM/year and its seriously vulnerable to extensive pollution [14]. Water is mainly abstracted for domestic and commercial usage. Analysis of groundwater use in the area is fair (but not quantitative) and analyses of pollution drivers are incipient. The major catalyst for such analyses being the observation that groundwater development decisions are made based on the market forces of demand and supply. Developers are the main driving force behind groundwater development [14].

II. MATERIALS AND METHODS

2.1 Samples

Data sets used in this study are particulars of 19 water samples were collected at the wellhead for chemical and bacteriological analysis as reported by [6] and [5]. The data comprised 19 water quality parameters. The sampling was conducted at groundwater boreholes and streams occurring in varied geological formations. The sampling and analysis procedures are reported accordingly. The sampling locations are shown in Figure 1.

2.2 Statistical Analysis

The objective was to produce scale unidimensionality as well as visualize underlying experimental data structure and relationships between the data and sampling sites. The resultant factor solutions create parsimony in the interpretation.

2.1.1 Principal Component Analysis

PCA is an unsupervised linear and sequential technique in MVDA. It provides information on the most meaningful parameters by transforming a set of inter-correlated variables with high dimensionality into few new orthogonal (i.e. uncorrelated) constructs called latent abstract variables or eigen vectors or Principal Components (PCs) in a new coordinate system [16], [12], [10], [17]. A reduced number of latent constructs provides more comprehensive and more or less interpretable data focused on study objectives [18]. Interpretation of latent constructs can be reveal underlying data structure by reflecting background processes. As [16] observes each Principle Component (PC) is a linear combination of the measured variables and describes a different source of variation viz:

$$PC_1 = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \qquad (1)$$

where x_i and w_i are the measured variable and its corresponding weight, respectively. The principal component weights are used as computation of the correlation between the observed variables and the latent constructs.

2.1.2 Hierarchical Cluster Analysis

HCA is unsupervised agglomeration scheduling techniques that identifies a pattern in the datasets. Its algorithms yield nested partitions including natural grouping of samples (or sites) with similarities constructed sequentially. According to [16], unlike within the clusters, objects in different clusters are dissimilar. Hierarchical agglomerative clustering offers insightful similarity relationships between any one sample/site and the entire data set [19], [10]. Cluster membership is determined stage-wise; starting with the most similar pair of variables and sequentially forming higher clusters. Cluster process formation is repeated until a single cluster containing all the variables is obtained. The utilized methods, algorithms, and similarity/dissimilarity measures are described elsewhere (e.g. [20]). In this study, the commonly applied average group and the Ward's clustering methods were used. The Euclidean distance was used as a similarity measure.

III. RESULTS AND DISCUSSION

3.1 Data Structure Interpretation and Source Identification

The descriptive statistics of data sets taken for purposes of water quality screening of the study area characterized by 17 physico-chemical variables and 2 microbiological parameters are shown (Table. 1). Cronbach's Alpha based on standardized items of 0.698 indicates high internal consistency in the experimental data. From these data, 4 principal components, contributing to 100 % of the total variance, were extracted based on multiple decision rules; orthogonal varimax rotation with Kaiser Normalization (after [21]) at 6 iterations (Table 2), from a scree plot (after [22]) and criterion of the eigenvalues greater than 1.

A scree plot shows the eigenvalues arranged from large to small as a function of the latent constructs number (Fig. 2). Only eigenvalues greater than 1 are considered significant [23]. The inspection of the scree plot (Fig. 2) and eigenvalues produced a departure from linearity coinciding with a 4-construct result. Additional constructs provided marginally less explanatory capability and were therefore omitted. The correlation coefficients between the observed items and the principal components are the factor loadings. They explain the weights (w_i) of the observed variables in the principal components. The principal component weights envisaged in equation 1 by [16] are shown in Table 2. They provide precise guides towards weight significance considerations. The scale nomenclature of absolute construct weights of > 0.75, 0.75-0.50 and 0.50-0.30 as ‘strong’, ‘moderate’ and ‘week’, respectively corresponds to the classification of [24] cited in [13]. A benchmark value of 0.55 was used after rotation to produce a more parsimonious solution; reduction of multi-collinear effects and ensures that data fits in a logical factor structure. The eigenvectors, and their cumulative percent variances are summarized (Table 3). The PCA decomposes the overall variance of the observed variables to principal components which explain the inter-construct variation in a monotonically decreasing manner.

Table 1: Descriptive statistics for water quality parameters

	N	**Range (study station with SW prefix)**		**Mean**	**Std. Deviation**
		Minimum	Maximum		
TDS	18	59.00(25)	544.00 (17)	246.991	134.362
pH	10	7.18(27)	8.50 (26)	7.8280	0.500
NTU	18	1.00(24,26,27)	85.00(29)	29.111	33.034
EC	18	9.22(21)	879.00(17)	403.923	211.725
Mn^{-4}	6	3.16(15)	6.32(21)	4.067	1.263
Hard	10	6.00(24)	214.00(17)	41.200	62.341
TA	10	29.00(18)	223.00(21)	122.200	67.592

E.coli	10	0.00	649.00(17)	67.500	204.482
G.coli	10	0.00(24,27)	2440.00(17, 30)	486.600	1011.131
Fe	18	0.01	1.99(29)	0.519	0.662
Mn	12	0.01	1.0(30)	0.157	0.314
Ca	18	1.60(24)	39.20(17)	13.572	10.507
Mg	18	0.44(24)	28.20(17)	6.157	6.2531
Na	18	6.50(27)	203.40(21)	65.172	44.132
K	10	0.40(27)	7.00(17)	2.860	1.838
NO_3^-	10	0.01(18)	27.00(17)	3.981	8.123
SO_4^-	10	1.40(21)	8.86(15)	3.764	2.242
Cl	17	3.00(18)	167.00(21)	41.853	42.829
F	10	0.13(21)	20.00(25)	5.291	6.402

In the analysis only four PCs were retained using the criteria of multiple decision rules. PC_1 with the largest Eigen value accounted for maximum of the total variance (52.006%) in the data set (Table 3). It is oriented in the direction of largest variation of the observed variables and goes through center of the data (Fig. 3). PC_2 (ionic component) accounted for the total variation of 22.97% and corresponds in concept to the first PC. The third PC_3 and the PC_4 scores explained 15.69% and 9.33% respectively of the total variance. The third PC passing through the centre of the dataset is oriented towards the next largest variation, and is orthogonal to the preceding PCs, and so forth. The observed eigen value decomposition corresponds to earlier observations by [16] and [17] that after the first PC, the second PC explains the maximum of the remaining variance and so on.
As it is obvious, PC_1 can be called as alkalinity/Temperature component as the microbes (*General coli. and Escherichia coli)* survival and growth is a construct of the interaction of endogenic (water related) versus exogenic (environmental) constructs such as alkaline petrochemical conditions, nutrient ions (e.g. Na, K and Mn) and water activity/temperature (expressed by TDS). The positive correlation of Na, K, Cl^- and Mn (Table 4) reflect enrichment from deep-seated environment. The positive association of these nutrients with microbes strongly indicates surface derivation followed by fracture permeability and leaching transportation by volcanic rocks and pyroclastics. A pH value range of 8.7-9.5 reported at Kenya Polytechnic hostels (MW9) and Unilever Industries (MW7) could favour survival and growth of pathogenic microorganisms [25]. Intracellular acidification often resulting into disruption and damage of key biochemical processes is caused by severe acidic pH [26], [25].

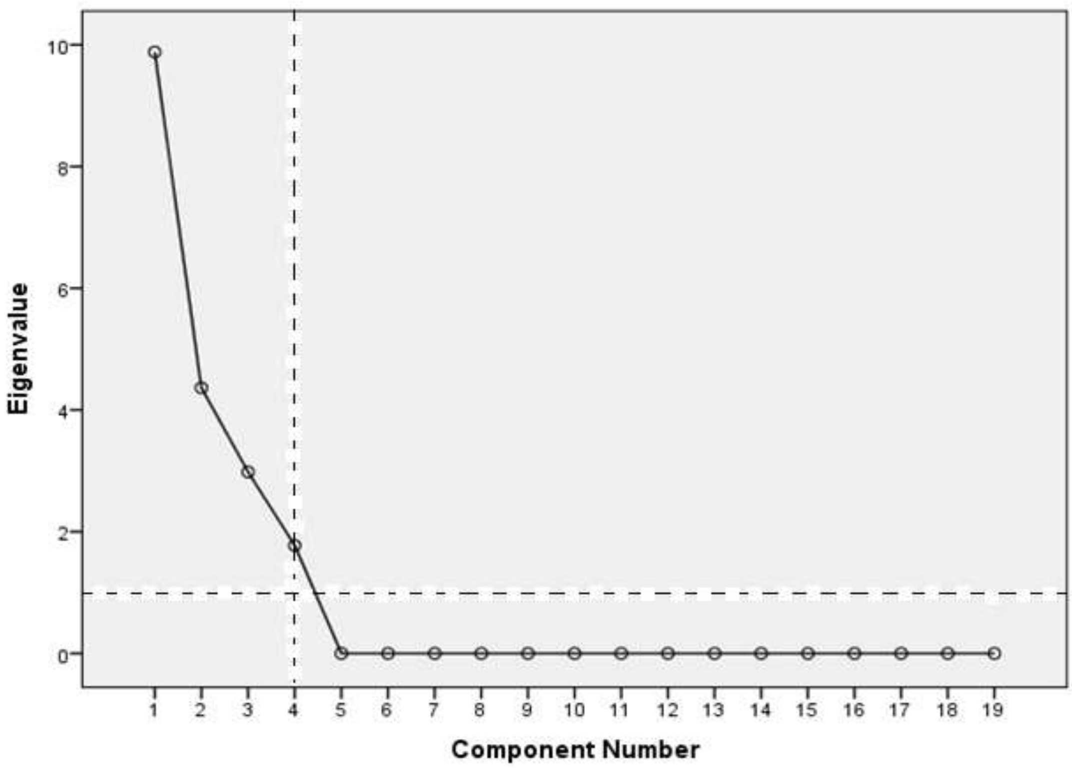

Fig. 2: Scree test with eigenvalues criteria{after [22]}

Table 2: The Principal Component loadings of the measured parameters which suited the provisions of Orthogonal Varimax rotation

Parameters	Component			
	PC_1	PC_2	PC_3	PC_4
TDS	0.988*	0.147	0.005	0.039
pH	0.362	0.879*	0.231	-0.206
NTU	0.042	-0.052	0.255	0.965*
EC	-0.600	0.356	0.669	0.258
Mn^{-4}	0.835*	0.290	-0.384	0.269
Hard	-0.153	-0.912*	0.360	-0.126
TA	0.955*	0.167	0.178	0.168
E.coli	0.967*	-0.023	-0.246	-0.067
G.coli	0.966*	-0.021	-0.250	-0.056
Fe	0.109	0.092	-0.104	0.984*
Mn	0.856*	0.073	-0.379	0.345
Ca	-0.161	-0.802*	0.575	0.010
Mg	-0.112	-0.954*	0.057	-0.271
Na	0.977*	0.203	-0.067	0.030
K	0.784*	-0.063	0.531	0.316
SO_4^-	-0.260	-0.354	0.898*	0.007
Cl	0.969*	-0.023	-0.243	-0.032
F	-0.378	0.910*	0.101	-0.136

*Represents strong item loadings {after [24]}

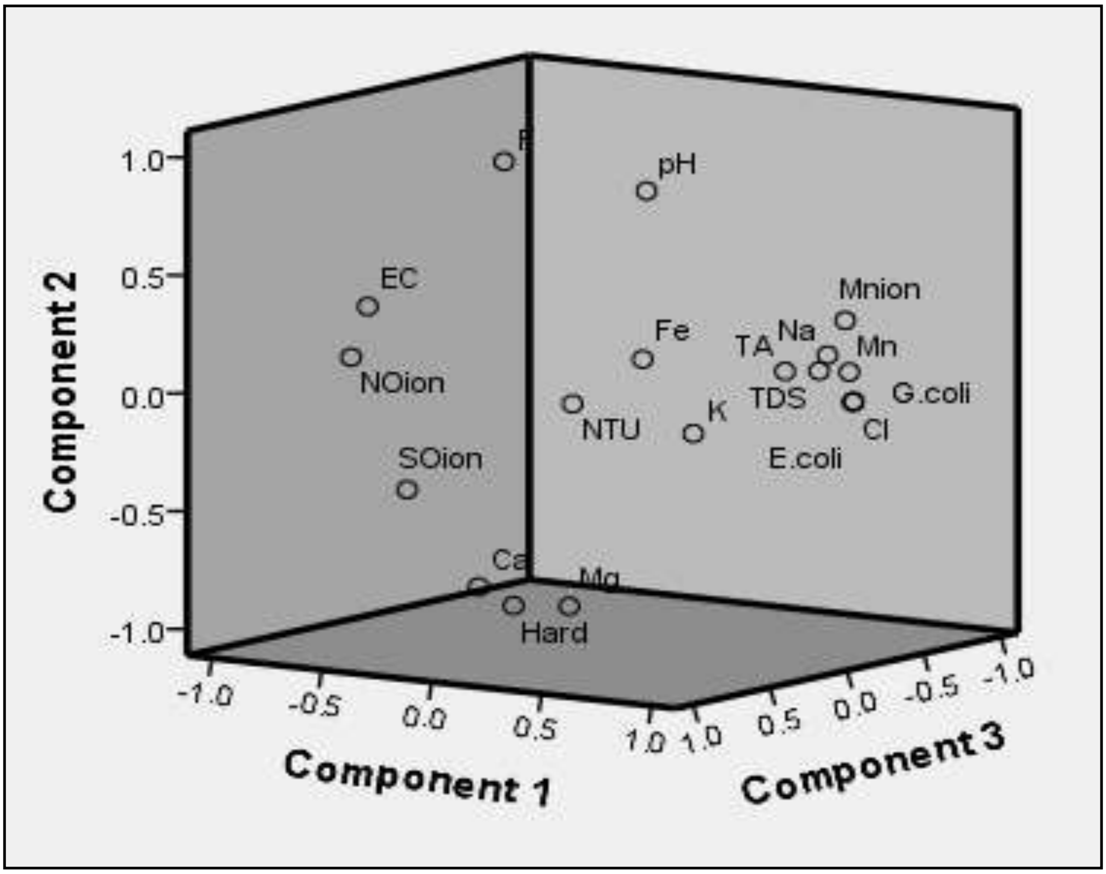

Fig. 3: Loading Component plot in rotated space

Table 3: Total variance explained by the latent constructs

Component	***Total variance explained before rotation***			***Total variance explained by varimax rotation***		
	Extraction Sums of Squared Loadings			*Rotation Sums of Squared Loadings*		
	Initial eigen values			*Final eigen values*		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.881	52.006	52.006	9.059	47.677	47.677
2	4.364	22.970	74.976	4.445	23.397	71.074
3	2.981	15.691	90.667	2.938	15.462	86.536
4	1.773	9.333	100.000	2.558	13.464	100.000

The second PC_2 is the ionic component because it is depicted chiefly by acidity (H^+) and hardness (including F, Mg^{2+} and Ca^{2+}) saturation. The hardness of water is in the form of $MgSO_4$ or $CaSO_4$ solutions (Table 4). The areas underlain by Tertiary phonolites towards the eastern franks of the study area have mean fluoride concentration of 7.6 ppm; an anomalous enrichment possibly due to derivation from feldspathoids and CaF_2 associated with volcanic petrochemistry.

The third principal component, PC_3 is Conductivity component which is associated with high concentrations of electrolytic ions (NO_3^- and SO_4^{2-}). Nitrate significantly contributes to conductivity because of its high concentrations at a site such as Kanungaga (SW 17) which is a shallow borehole in the vicinity of a leaking sewer line. Other sites with elevated EC and nitrate values are Unilever (MW7) and Trufoods (MW6) boreholes in Industrial Area and Kabansora (MW8) borehole in Embakasi [6]. Nitrate is therefore an extrinsic observed construct in regard to groundwater conductivity as it is introduced through diverse and widespread anthropogenic sources of contamination. This relation rather signalizes interaction and or mixing of groundwater and surface water. Within the hydrogeological framework, surface-groundwater interactions can be accounted for by fracture-fissure character of the Nairobi phonolites blanketing the aquiferous Athi-Series of the eastern segment. Surface-groundwater interaction is reported as significant [14] with base-flow 35 to 45 percent of total flow.

PC_4 unseen on the loading plot (Fig. 3) should be identified as the iron component. It's mainly composite of iron, and turbidity which are all associated with corrosion of water infrastructural utilities. Turbidity in water is caused by suspended particles or colloidal matter that may be caused by inorganic iron. As observed by [6], water boreboles with high turbidity are located either along or adjacent to faults (e.g. MW1 - Jorgen Mbagathi Ridge and MW5 - Boulevard Hotel) or within the Industrial Area (MW6 - Trufoods). Chlorine loading on PC_4 exhibits a negative component weight implying that NTU and Fe composition will decrease chlorination efficiency of blue water.

Table 4: Inter-Item Correlation Matrix

	TDS	**pH**	**NTU**	**EC**	**Mn_4^-**	**Hard**	**T. A**	**E.coli**	**G.coli**	**Fe**	**Mn**	**Ca**	**Mg**	**Na**	**K**	**NO_3^-**	**SO_4^{2-}**	**Cl^-**	**F^-**
TDS	1.000																		
pH	.480	1.000																	
NTU	.073	-.171	1.000																
EC	-.527	.197	.376	1.000															
Mn_4^-	**.876**	.413	.182	-.585	1.000														
Hard	-.288	-.748	.011	-.025	-.564	1.000													
TA	**.976**	.499	.239	-.351	**.822**	-.256	1.000												
E.coli	**.948**	.287	-.086	-.770	**.876**	-.207	**.864**	1.000											
G.coli	**.949**	.285	-.076	-.768	**.882**	-.212	**.866**	**1.000**	1.000										
Fe	.159	-.107	**.923**	.152	.423	-.262	.267	.063	.075	1.000									
Mn	**.868**	.215	.268	-.652	**.973**	-.377	**.820**	**.896**	**.901**	.479	1.000								
Ca	-.274	-.633	.192	.199	-.585	**.961**	-.184	-.280	-.283	-.141	-.411	1.000							
Mg	-.261	-.810	-.202	-.304	-.464	**.942**	-.302	-.082	-.087	-.373	-.280	**.813**	1.000						
Na	**.996**	.510	.043	-.550	**.907**	-.362	**.960**	**.954**	**.955**	.162	**.886**	-.358	-.314	1.000					
K	**.781**	.286	.476	-.056	**.517**	.088	**.886**	**.607**	**.609**	.336	**.574**	.233	-.083	**.727**	1.000				
NO_3^-	-.727	-.130	.421	**.943**	-.727	.173	-.563	-.899	-.896	.198	-.738	.353	-.074	-.753	-.217	1.000			
SO_4^{2-}	-.304	-.199	.244	**.633**	-.662	**.684**	-.146	-.465	-.469	-.147	-.586	**.842**	.415	-.385	.298	**.649**	1.000		
Cl^-	**.952**	.282	-.051	-.760	**.887**	-.211	**.873**	**.999**	**1.000**	.098	**.909**	-.278	-.092	**.957**	**.622**	-.887	-.463	1.000	
F^-	-.245	**.715**	-.169	**.583**	-.127	-.719	-.214	-.402	-.402	-.102	-.342	-.613	-.784	-.196	-.343	.394	-.134	-.407	1.000

3.2 Site Similarity

The hierarchical agglomerative clustering methods utilized here was Ward's method. The approach depends on computing the similarity between any two patterns using a distance method called squared Euclidean distance. The Ward's method is preferred as it is less sensitive to outliers and efficient. The technique utilizes an Analysis Of Variance (ANOVA). The approach compares average sample concentrations from experimental data to group the data based on similarity and inter-group dissimilarity. ANOVA calculates the total sum of squared deviations from the mean of each cluster. Each member of a cluster should generate the smallest plausible increase in the error sum of squares. The typical outcome of the agglomeration process is illustrated in a tree-like structure display, called a dendrogram (Fig. 4). The proximity in the dendrogram based on/ measured with a rescaled distance demonstrates high intra-cluster homogeneity and also high inter-cluster heterogeneity. The dendrogram is useful in discerning the data structure not only among observations, but also among variables.

In CA the dendrogram was partitioned at different levels to capitulate 4 clusters (or groups). However, the final cluster was fixed rather arbitrary. Based on the four similarity groups, it was concluded that:

Cluster 1: Generally comprising of stations SW 24, SW 20, SW 19, SW 17, SW 21, SW 15 and SW27. Hydrochemical investigations suggest relatively polluted water by anomalously high chemical ion values of Fe, Mn, F, Ca, Mg, K, SO_4^{2-} and NO_3^-. Majority of the sites lie in the highly faulted segment northwest of the city. Fault compartmentalization of the NAS in the near vicinity of these measurement locations accounts for the high turbidity, high levels of NO_3^- and hence vulnerability of water to pollution via surface water / groundwater interaction. High fluoride content in the groundwater is accounted for by high fault permeability and hence transport through the westerly striking faults and fault weathering of feldspathoids hosted within the Nairobi volcanic rocks [27] cited in [6]. Polluted groundwater is confirmed by enriched presence of inorganic (high hardness and pH) or organic matter (e.g. E.coli) or a mixture of the both [14]. High turbidity can protect microbes by decreasing of disinfection efficiency and can stimulate bacterial growth (e.g. E.coli). Microbial contamination sources can therefore be assessed by use of NTU as an indicator.

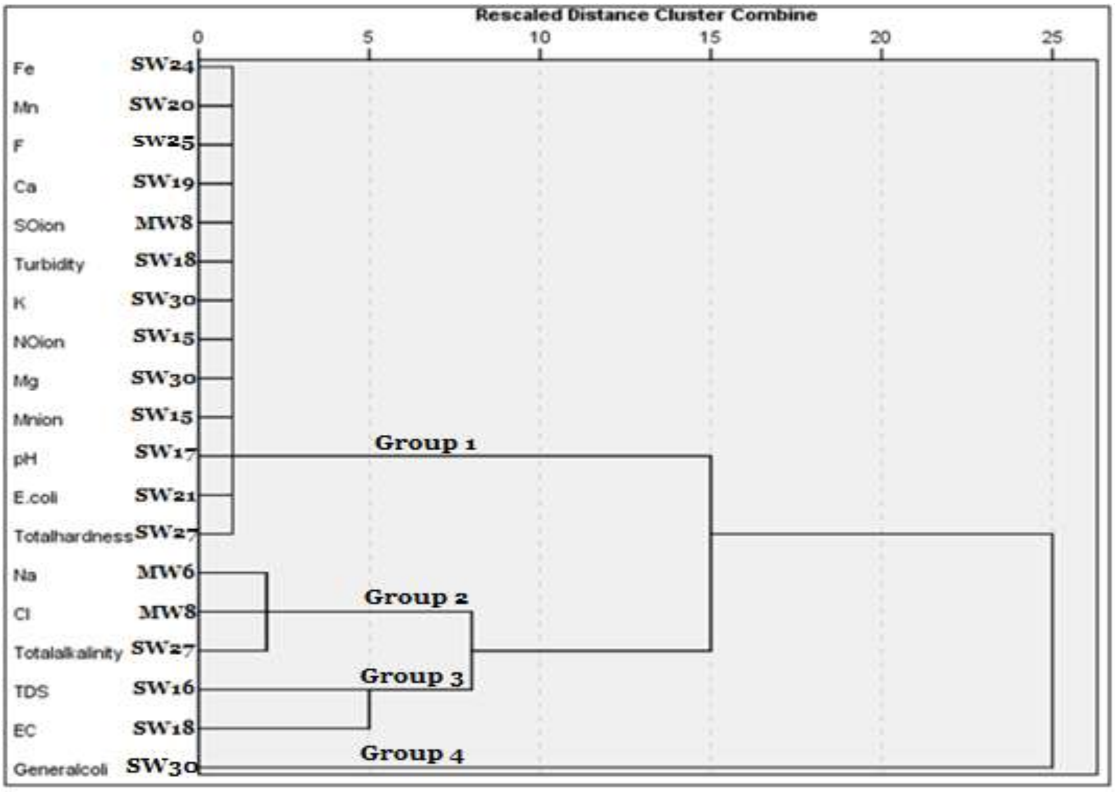

Fig. 4: Dendogram obtained by the hierarchical cluster analysis using water quality data

Cluster 2: These group of sampling sites are similar in terms of alkalinity component as indicated by enriched composition in Na and Cl. [28] had observed that all water struck in Tertiary formations (e.g at MW 6, MW 8 and SW 27) contains sodium bicarbonate and are usually alkaline. The three sites are similar to the first cluster locations in the provisos of fault permeability and water-rock interaction; judging from the comparatively smaller rescaled distance between the two clusters.

Cluster 3: The cluster of sampling stations includes SW 16 and SW 18 sited at the boundary between Kerichwa Valley Series/Nairobi Trachytes and alluvium/swampy clays. The inter phase in geology to clays is associated with flaking and scaling (weathering) alteration through infiltration of water. Such boundaries are dominated by clay minerals (kaolinite, illite and vermiculite). Clay mineral substrates are associated with resistivity values < 65Ω in Nairobi area [29]. High flow rates at the geological breaks increase dissolution of substances and therefore related directly to TDS.
Cluster 4: The final cluster lying at some Euclidean distance from others comprises of samples obtained from site SW 30 (Umoja area) with significant content of the general coli forms. The site is very distinct in terms of its mean and standard deviation (Table 1). The site is correlated to sites MW 6, MW 8 and SW 17 (having significant total alkalinity), SW 17 (with significant pH) and site SW 16 (with significant water activity expressed by TDS as indicated in PC loading plot Fig. 3). Studies by [26] on stress response of coli forms indicate that survival and growth level is a factor dependence on pH/TA and water activity/TDS. The Umoja site probably receives improperly treated effluents from the Dandora Sewage Treatment Plant and domestic garbage from informal settlements characterized by presence of earth drains, communal water points, pit latrines, and no systematic solid-waste disposal. Fracture permeability of the Nairobi phonolite is the most likely cause that accentuates the flow of gray waters into the aquifer.

IV. CONCLUSIONS

The primary aim of this study was to determine and evaluate the most meaningful hypothetical parameters responsible for groundwater processes. The second tier was to identify similarities/dissimilarities among the sites sampled. In the present study, a large data matrix was subjected to PCA and CA techniques. The study demonstrates the application of the unsupervised data reduction techniques (PCA and CA) in the task without any significant information-loss in the observed data. The integrated approach permitted mapping the hydro chemical signatures and hence characterization of various sources of ground water pollution. Simple representation of samples and sampling locations experimental data was based on inter-correlated variations and sample concentration averages to classify the data sets. The integration PCA and CA enabled parsimonious synthesis and inferences of processes and sources of groundwater pollution.

Major parameters of water quality from the study include inorganic or organic matter or a mixture of the both made available from domestic, industrial operations and geological processes. The difference in geochemistry of the groundwater at different sites relates different geological formations, localized geochemical processes, compartmentalization due to faulting and the presence of pollution and discharge of untreated sewage and household wastes. Analyses of the present study reveal that the NAS is exposed to sewage and waste water. The results call for more mitigation procedures to avoid more deterioration for the groundwater resources in the study area.

ACKNOWLEDGEMENTS

Water chemistry analyses obtained from the cited sources permitted application of the MVDA techniques possible. The utilization of trade names in the paper does not imply approval by these organizations or disapproval of those not mentioned.

Table 5: Abbreviations and Symbols

MVDA	**Multivariate Data Analysis**	**Cl**	**Chloride, mg L-1**
FA	Factor Analysis	NTU	Natural turbidity unit
PCA	Principal Component Analysis	Mg	Magnesium, mg L-1
CA	Cluster Analysis	NO_3^-	Nitrate nitrogen, mg L-1
PC	Principal Component	SO_4^{2-}	Sulphate, mg L-1
MW	Monitoring Wells	Ca	Calcium, mg L-1
SW	Study Wells	Fe	Iron, mg L-1
hard	Hardness, Mg L-1	K	Potassium, mg L-1
TDS	Total Dissolved Solids, Mg L-1	Ph	Potential Hydrogen ions
TA	Total Alkalinity, Mg L-1	EC	Electrical Conductivity
E. coli	*Escherichia Coli*	Mn	Manganese, Mn L-1
G. coli	General Coli forms	Mnion	Manganese ion, Mn L-1
Na	Sodium, Mg L-1	F	Fluoride, F L-1

REFERENCES

[1]. S.S.D. Foster, and A. Tuinhof, The Role of Groundwater in the Water Supply of Greater Nairobi, Kenya, 2005.

[2]. R. Reza, and G. Singh, Heavy metal contamination and its indexing approach for river water, *Int. J. Environ. Sci. Tech.*, 7(4), 2010, 785-792.

[3]. A. Tibaijuki, Nairobi and its Environment, In Barr, J. & Shisanya, C. (Ed.), *Cities can achieve more sustainable land use if municipalities combine urban planning and development with environmental management*, United Nation-HABITAT, (Nairobi, 2007).

[4]. S. M. Kithiia, and F.M., Mutua, Impacts of land use changes on sediment yields and water quality within the Nairobi River sub-basin, Kenya, (Ed. by J. Rowan, R. W. Duck & A. Werrity, *Proc. of a symposium On: Sediment dynamics and the Hydromorphology of Fluvial systems.* Wallingford, UK, 2006, 582-588.

[5]. S. M. Kithiia, Water Quality Degradation Trends in Kenya over the Last Decade, Water Quality Monitoring and Assessment, Dr. Voudouris (Ed.), *InTech*, 2012, Available from: http://www.intechopen.com/books (accessed on 8th August, 2015).

[6]. J. Onyancha and Z. Gitenga, Geochemistry of groundwater in the volcanic rocks of Nairobi city, *Global Journal of Science Frontier Research Environment & Earth Science,* 3(3), 2013, 1-8.

[7]. S. Khayat, A. Marei, B. Natsheh, and N. Abu-Khalaf, Mechanisms of groundwater pollutants transport in Tulkarm area / Palestine, *Resources and Environment*, 2(6), 2012, 281-290.

[8]. A. Nawaf, K. Saed, and N. Basel, Multivariate Data Analysis to Identify the Groundwater Pollution Sources in Tulkarm Area / Palestine, *Science and Technology,* 3(4), 2013, 99-104, DOI: 10.5923/j.scit.20130304.01.

[9]. R. Santos, L. Ribeiro, and C. Dill, The use of multivariate statistical analysis to evaluate spatial and temporal water contamination in Germunde coal mine (Portugal), *Proc. of the fourth international colloquium on Hydrology and Management of Water Resources,* Guimaraes, Portugal, 2005, Available at www.imwa.info/.../IMWA2005_063_...

[10]. M. Akbulut, H. Kaya, E.S. Çelýk, D.A. Odabasi, S.S. Odabasi, and K. Selvý, Assessment of Surface Water Quality in the Atikhisar Reservoir and Sarýçay Creek (Çanakkale, Turkey), *Ekoloji,* (19(74), 2010, 139-149.

[11]. S. Kiymaz, and U. Karadavut, Application of Multivariate Statistical Analysis in the Assessment of Surface Water Quality in Seyfe Lake, Turkey, *Journal of Agricultural Sciences,* 20, 2014, 152-163.

[12]. R. Koklu, B. Sengorur, and B. Topal, Water Quality Assessment Using Multivariate Statistical Methods—A Case Study: Melen River System (Turkey), *Water Resour Manage*, 24, 2009, 959–978, DOI 10.1007/s11269-009-9481-7.

[13]. Shivasharanappa, and P. Srinivas, Studies on Seasonal Variation of Ground Water Quality Using Multivariate Analysis for Bidar Urban & Its Industrial Area (Karnataka-State, India), *IJRETIC-RICE Conference Issue,* 2013, 252-261, Available @ http://www.ijret.org252 (accessed on 8th August, 2015).

[14]. WRMA, *Preliminary Water Allocation Plan of the Nairobi Aquifer Suite*, and *Long Term Water Resources Management Strategy*, Report Prepared by Norken Ltd, Nairobi, Kenya, 2010.

[15]. E.P. Saggerson, Geology of Nairobi Area. Rep. No. 98, Geol. Surv. Kenya. 1991.

[16]. P. Praus, Urban water quality evaluation using multivariate analysis, *Acta Montanistica Slovaca Ročník,* 12(2), 2007, 150-158.

[17]. M. Hossain, A. Patras, C. Barry-Ryan, A. Martin-Diana, and N. Brunton, Analysis of Principal Component Analysis, Hierarchical Cluster Analysis to classify different spices based on in vitro antioxidant activity and individual polyphenolic antioxidant compound, *Journal of Functional Food,* 3, 2011, 179-189.

[18]. Z. Shuguang, Z. Kefa, C. Yao, J. Wang, and J. Ding, Exploratory data analysis and singularity mapping in geochemical anomaly identification in Karamay, *Journal of Geochemical Exploration* 154, 2015, 171–179

[19]. J.E. McKenna, An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis of the Fuji River basin; Japan, *Environmental Modelling and Software,* 18(3), 2003, 205-220.

[20]. B. Everitt, *Cluster analysis (4th edn.) (*Hodder Arnold, London, 2001).

[21]. H., F. Kaiser, The application of electronic computers to factor analysis, *Educ. Psychol. Meas*. 20, 1960, 141-151

[22]. R. B. Cattel, The scree test for the number of factors, *Multivariate Behav. Res*., 1, 1966, 245-276.

[23]. A. Zare Garizi, V. Sheikh, A. Sadoddin, Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods, *Int. J. Environ. Sci. Tech.*, 8(3), 2011, 581-592.

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

[25]. WHO, *Guidelines for drinking water quality,* World Health Organization, 2011.

[26]. H.J. Chung, W. Bang, and M.A. Drake, Stress Response of *Escherichia Coli*, *Comprehensive Reviews in Food Science and Food Safety,* 5, 2006, 52-64.

[27]. J. M. Mailu, *Hydrogeology of the Athi Basin,* Doctoral thesis, University of Nairobi, Kenya, 1987.

[28]. E.A. L. Gevaerts, Hydrogeology of Nairobi Area. Technical Report, Survey of Kenya, 1964.

[29]. C.K. Onyancha, E. M. Mathu, S.K. Mwea, and W.M. Ngecu, Geophysical resistivity survey in subsurface characterization for underground excavation and heavy construction in Nairobi City, Kenya, (unpublished),Available at: http://www.uonbi.ac.ke (accessed.ke on 8th August, 2015).