A Multivariate Statistical Approach for Groundwater Classification along the Virunga Volcanic Range and the Northern Kivu Rift Area, Rwanda

Eugene Karangwa

Energy Development Corporation Limited/ Rwanda Energy Group

P.O. Box: 3855 Kigali, Rwanda

eukarangwa@edcl.reg.rw

eukarangwa@gmail.com

Keywords:

Groundwater, Multivariate Analysis, Kivu Rift, Virunga Volcanic Range

ABSTRACT

Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) are applied to some physico-chemical characteristic properties of 26 groundwater samples collected along the Virunga Volcanic Range and the northern Kivu Rift area in Rwanda during two different campaigns by BGR in 2008 and JICA in 2013, respectively. The aim of the study was to extract principal factors corresponding to the different sources of variation in hydrochemistry, with the objective of defining the main controls on the hydrochemistry of different groundwater sources in the study area. The analysis indicated three main hydro-chemical facies; namely thermal groundwater (group 1), young groundwater (group 2) and old groundwater mixed with surface water (group 3). Besides each group contains sub-clusters based on differences in hydro-chemical properties. The results of this study clearly demonstrate that multivariate methods offer significant improvement for integrated characterization of groundwater types and mixing trends deciphering for geothermal resource exploration and development.

1. Introduction

The East African Rift System is one of the most important zones of the world where the heat energy of the interior of the earth escapes to the surface in the form of volcanic eruptions, earthquakes and the upward transport of heat by hot springs and natural vapour emanations. As a consequence, it appears to possess a remarkable geothermal energy potential (Teklemariam, 2006). However, the development of indigenous geothermal resource in the western branch of the EARS region is hindered by various challenges including the lack of appropriate exploration approaches taking into consideration the importance of the diversity of geologic and hydrologic conditions of the western branch vis à vis that one of the eastern branch and associated potential geothermal resource ("Technical Workshop on the Geologic Development and Geophysics of the Western Branch of the Greater East African Rift System with Emphasis on Factors that influence the Development of their Geothermal Systems ,'' 2016).

The recharge of any geothermal system in the area of the present study is likely to occur through infiltration of rainfall through highly permeable volcanic rocks on the slopes of the volcanoes down to the basement. Recent faults should provide pathways for infiltration into the basement. Given the regional hydraulic gradient, the expected topography of the basement, and the orientations of faults and dykes, greater NE-SW fluid flow is expected than suggested in existing models and dykes could potentially compartmentalise convective cells (Peskett,2014). Permeability in the basement rocks appears to be controlled primarily by fracture networks, which are observed in surface outcrops but are likely change with depth. Recent N-S and SW-NE fault zones related to rifting are likely to have the largest permeability (Peskett,2014).
The drainage system is strongly controlled by the geology, leading to define three main types of drainage system:

- Sub-dendritic system on the basement rocks
- Radial system on the volcanoes
- N-S system along the rift axis (Egbert et *al*., 2009).

Moreover, a number of springs exist in the area. These include cold, hot and mineralized springs. The hot springs all appear to outflow from the basement rocks, while the cold springs are associated with both basement rocks and the volcanic sequences (Shalev et *al*., 2012).

Literature shows the always-increasing potentiality of multivariate statistical techniques in obtaining useful information from environmental data, which hardly could be otherwise correlated and interpreted. In particular many examples can be found of the application of multivariate analysis to sets of variables collected for surface and ground waters (Belkhiri, Boudoukha & Mouni, 2011). Hierarchical cluster analysis (HCA) and principal component analysis (PCA) are effective means of manipulating, interpreting and representing data concerning groundwater physico-chemical characteristics (Venkatramanan, S., Ramkumar, T. and Anithamary, 2012). They are frequently employed to characterize the quality and type of groundwater as well as to discuss geochemical evolution, mineralization and groundwater contamination and to interpret the hydrogeological data for trends identification and patterns recognition (Brown, 1998 and Hesel & Hirch, 2002).

The present research expects to contribute to water samples classification and mixing trends deciphering through the application of multivariate statistical techniques to existing hydrochemistry data of springs. HCA and PCA are used to evaluate the structures and relationships in the measured quantities for an improved understanding of the regional fluid flow by dividing water samples into water groups with similar characteristics. The studied area is located in the north-western part of Rwanda.

2. Materials and Methods

A total of 26 spring samples with 11 common hydro-chemical variables (EC, pH, K, Na, Mg, K, Cl, B, SiO_2, HCO_3 and SO_4) from two different campaigns were considered for the present study. The samples were collected and analyzed by BGR (Egbert *et al.*, 2009) and JICA ("Electricity development plan for sustainable geothermal energy development in Rwanda," 2016) respectively, at the sampling points shown in Figure 1. The groundwater measured parameters are summarized in Table 1.

Data correlations and statistical analyses were performed using the IBM SPSS statistical package to determine the variables that represent the controlling factors behind the geochemistry of all water samples considered. Two multivariate statistical approaches; hierarchical cluster analysis (HCA) and principal component analysis (PCA) were used to group the water samples and variables analyzed.

PCA is a factor analysis method that is used to extract components representing the information contained in the data that explain the pattern of correlations and differences within a group of variables (Reghunath., Murthy & Raghavan, 2002). The number of factors or components was extracted by PCA with varimax rotation based on Kaiser criterion. Only factor with eigenvalues greater than or equal to one were accepted as possible sources of variance in the data (Kaiser, 1960).

HCA is a classification method that reveals natural groupings or clusters within a data set by reorganizing the data into homogeneous groups and linking the two most similar clusters until all of the variables are joined in a complete classification tree (Ward, 1963). The results of HCA are presented in a dendrogram, which was constructed using Ward's method with the Euclidean distance as a measure of similarity between the samples. Ward's method is one of the most widespread hierarchical clustering methods for the classification of hydro-geochemical data by using the minimum variance to evaluate the distances between the clusters and it produces the most distinctive groups where each member within the group is more similar to its fellow members than to any member outside the group (Belkhiri et al., 2011). All the 11 hydro-chemical variables selected (consisting of EC, pH, Ca, Mg, Na, K, Cl, SO_4, HCO_3, SiO_2 and B) were utilized for statistical analysis. The data were standardized by subtracting the sample mean from each variable and dividing the resulting value by the standard deviation (z-score standardization) prior to cluster analysis to ensure that each variable was weighted equally (Güler *et al*., 2002).

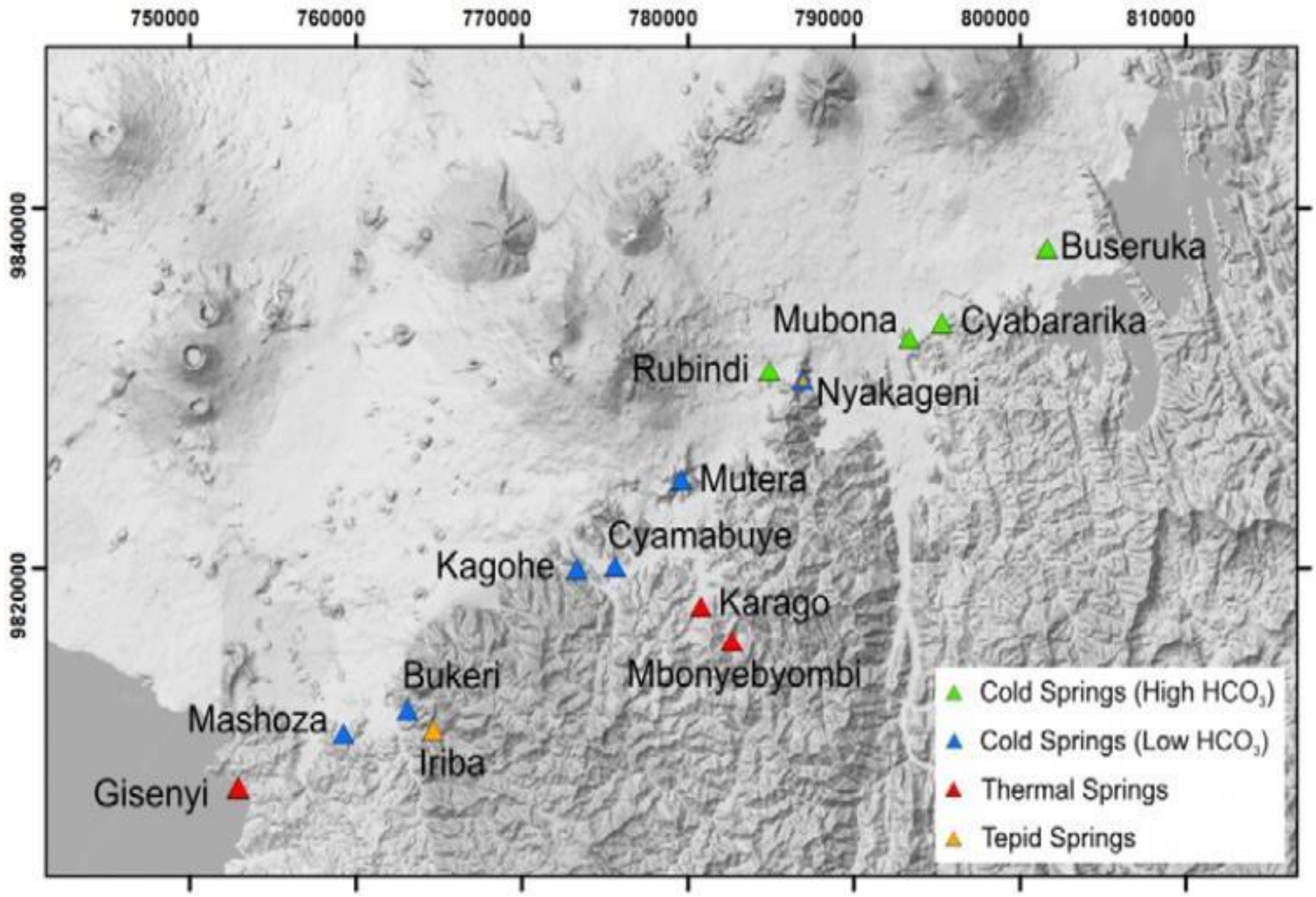

Figure1: Locations of sampling sites for springs waters

Table1: Hydro-chemical composition of groundwater samples from thermal and cold springs

Sample Name	EC (mS/m)	pH	Na (mg/L)	K (mg/L)	Ca (mg/L)	Mg (mg/L)	Cl (mg/L)	SO_4 (mg/L)	HCO_3 (mg/L)	B (mg/L)	SiO_2 (mg/L)
Gisenyi BGR	245	7.4	495	38.2	36.3	11.2	195	55.8	1140	2.14	56.2
Karago BGR	125	7.2	263	14.7	21.2	2.4	76.6	77.9	537	0.34	84
Mbonyebyombi BGR	92.1	7	187	11.5	20.2	2.4	51.8	44.2	414	0.22	60.3
Iriba BGR	229	7	394	17.2	76.6	23.2	287	67	846	0.42	58.3
Buseruka BGR	408	6.5	239	226	149	356	12.9	69.4	3200	0,43	99.8
Mubona BGR	305	6.7	160	180	112	225	22.3	41.4	2200	0.28	72.1
Cyabararika BGR	288	6,4	157	159	121	204	7.6	44.6	2100	0.25	69.2
Rubindi BGR	196	7.1	105	114	32.9	145	14.7	12.2	1320	0.17	94.3
Mutera BGR	215	7.4	4.5	5.1	21.6	9.9	2.6	6.51	91.2	0.02	55.7
Bukeri BGR	247	7.5	9.9	9.3	21.3	9.4	5.6	16.1	98.3	0.02	57.4
Cyamabuye BGR	442	7	13.9	21.3	34.9	17.4	12.5	24.6	158	0.04	47.7
Kagohe BGR	456	7	19.2	6.2	46.6	15.3	11.7	34	166	0.03	41.3
Gisenyi-1 JICA	243	7.2	555	46	36.8	11	230	60.7	1130	0.63	58
Gisenyi-2 JICA	248	7.19	553	46.9	36	10.7	220	59.9	1180	0.57	59

Karago JICA	126	7.64	255	15.3	21.2	2.35	79.3	78.1	555	0.08	91
Mbonyebyombi-1 JICA	91.2	7.21	178	11.8	19.1	2.34	54	44.7	415	0.05	62
Mbonyebyombi-2 JICA	15.4	6.6	5.72	1.06	12.7	5.41	5.1	6.7	24	0.005	31
Nteranya JICA	39.8	6.63	9.59	8.48	40.5	15.6	8.9	10.1	220	0.005	48
Iriba JICA	248	6.79	439	19.1	86.7	25.1	320	75	970	0.11	58
Rubindi-1 JICA	238	6.61	149	150	44.2	206	23.6	16.8	1660	0.05	110
Rubindi-2 JICA	28.1	7.06	12.5	17	16.2	12	6.6	7.1	144	0.005	46
Mubona JICA	294	6.51	172	185	117	243	22.5	42	2140	0.06	76
Cyabararika JICA	286	6.5	175	171	130	230	16.9	44.2	2090	0.06	75
Mata JICA	20.7	6.8	5.85	5.13	16.9	10.4	5	3.1	109	0.005	35
Buseruka JICA	305	6.56	184	175	131	253	15.6	57.5	2260	0.07	80
Kagohe JICA	46.7	6.84	18.9	5.94	49.3	14.5	12.1	35.7	159	0.005	38

Figure 2: Dendrogram of the spring water samples

3. Results and Discussions

The groundwater samples of the study area have pH values ranging from 6.40 to 7.64, which indicate that the groundwater is slightly alkaline. The data were classified in a simple and direct manner with results presented as a three clusters dendrogram in Figure 2. Based on an imaginary horizontal line (linkage line) on the cluster scale, three distinct groups or clusters were displayed.

The water type 1 (Group1 on Figure 2) encompasses thermal spring samples from Gisenyi, Iriba, Karago and Mbonyebyombi characterized by higher content of Na, Cl, B and SO_4 as compared to the remaining water samples (Table 1). That seems to be associated with some geothermal activity. However, the above-mentioned parameters show a decrease from Gisenyi to Karago and Mbonyebyombi through Iriba in a SW-NE trending direction. This might indicate diminishing contribution of the deep geothermal fluids by dilution. Besides, Egbert and co-workers (2009) reported low tritium values (0.03 - 0.43 TU) indicating old groundwater as well as old groundwater mixed with surface water. Moreover, volcanic activity in the northern part of the Lake Kivu basin developed on the western (DRC) side as well as along the eastern side (Rwanda); in 2002 an eruption revealed the existence of an open fissure linking Nyiragongo volcano and the Kivu Lake rift axis, including the development of normal faults and emplacement of lava flows at surface, magma injection along dikes, soil deformation and the formation of fumaroles and gas (CO2 and CH4) emissions (Ross *et al*., 2014). Furthermore, the water samples, in this cluster are basically characterized by temperature greater than average in the study area.

The water type 2 (Group 2 on Figure 2) comprises cold springs samples characterized by the lowest concentration for most of the analytical variables. As far as trace tritium content is concerned, this water category has the highest values (1.63 – 1.92 TU) characteristic of recent meteoritic water.

The water type 3 (Group 3 on Figure 2) consist of cold spring samples from Rubindi, Mubona, Cyabararika and Buseruka. It is bicarbonate dominated with relatively high silica content (Table 1) compared to the majority of thermal spring waters in the region. This water category is also characterized by low tritium values (0.05 - 0.3 TU) indicating old groundwater as well as old groundwater mixed with surface water.

Correlation coefficient was used to measure and establish the relationship between each two variables and show the degree of dependency of one variable to the other. The correlation matrix of eleven variables has been presented in Table 2. The values of HCO_3 exhibited high positive correlation with K, EC, Mg, Ca, and SiO_2. This positive strong correlation may suggest a common source. Similarly, Na displayed strong positive correlation with, Cl, B and SO_4, possibly associated with a common natural process.

Under the Kaiser criterion, only factors with eigenvalues greater than or equal to 1 were accepted as possible sources of variance in the data, with the highest priority ascribed to the factor that has the highest eigenvector sum. The rationale for choosing 1 is that a factor must have a variance at least as large as that of a single standardized original variable to be acceptable (Kaiser, 1960). Two principal components (PC) were extracted and rotated using the varimax normalization. The results show that the two PC account for more than 80.50% of the total variance (Table 3), which is quite good and can be relied upon to identify the main sources of variation in the hydro-chemistry. PC1 represents about 49.75% of the variance and has high loadings for K, Mg, Ca, HCO_3, EC and SiO_2 and probably shows the result of mineralwater reactions in the area. But, PC 2 which accounts for about 30.76% of the total variance contains high loadings for Na, Cl, SO_4 and B possibly connected with some natural processes associated with geothermal activity. Thus, PC1 and PC2 are assumed to be indicative of the natural processes and water-rock interaction in the study area. Figure 3 clearly displays the distinction between two sets of variables and suggests that all the water samples in their respective groups have similar chemistries hence similar flow paths or sources.

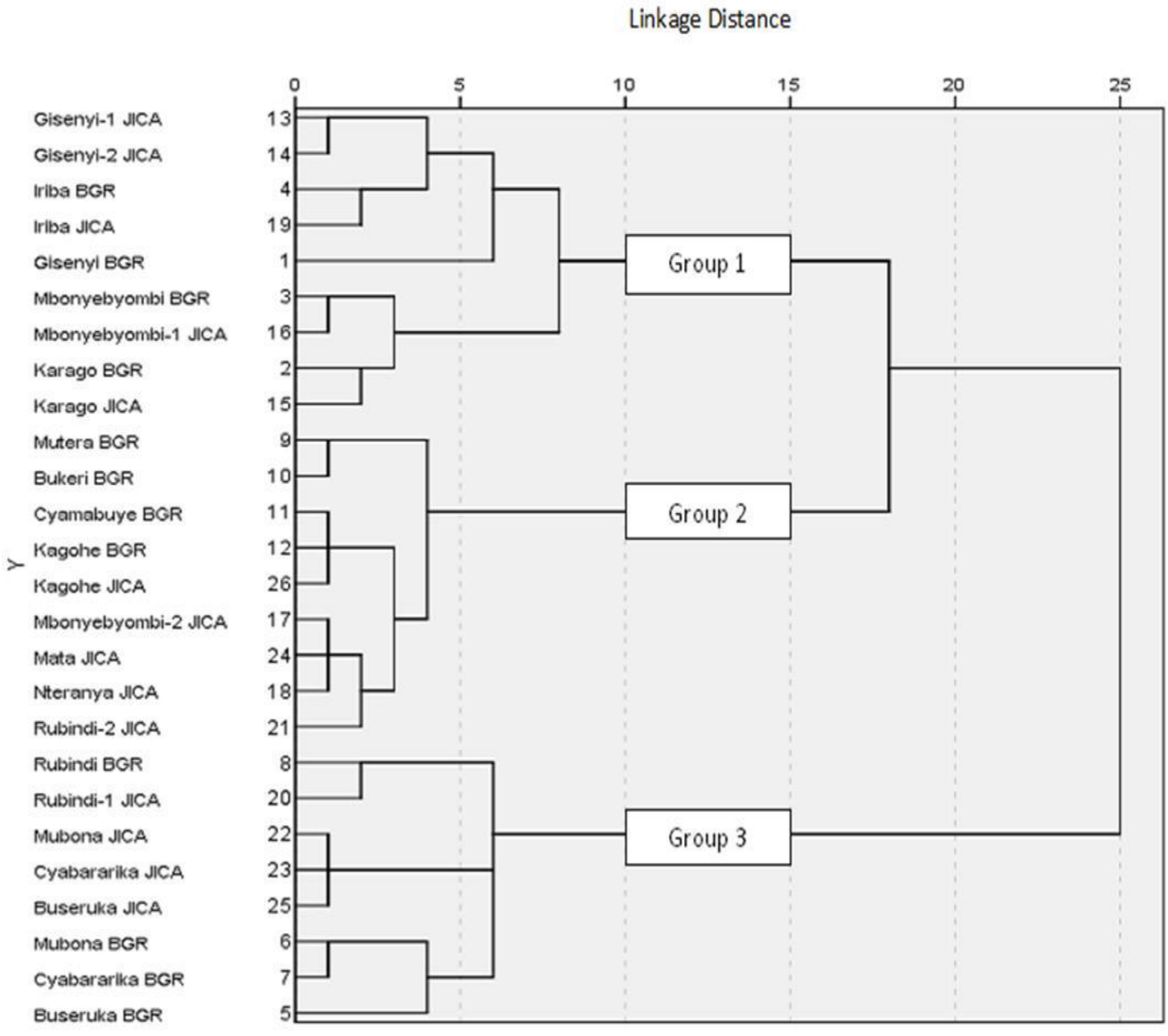

Figure 2: Dendrogram of the spring water samples

Table 2: Correlation coefficient among groundwater hydro-chemical variables

	EC	pH	Na	K	Ca	Mg	Cl	SO_4	HCO_3	B	SiO_2
EC	1										
pH	-0.421	1									
Na	0.610	0.207	1								
K	0.830	-0.632	0.093	1							
Ca	0.811	-0.693	0.167	0.844	1						
Mg	0.772	-0.683	-0.018	0.984	0.860	1					
Cl	0.323	0.277	0.870	-0.239	-0.013	-0.311	1				
SO_4	0.602	0.106	0.769	0.185	0.391	0.136	0.619	1			
HCO_3	0.951	-0.554	0.354	0.955	0.873	0.924	0.020	0.420	1		
B	0.364	0.284	0.654	0.023	0.001	-0.068	0.494	0.377	0.205	1	
SiO_2	0.652	-0.106	0.247	0.680	0.408	0.659	-0.054	0.383	0.689	0.039	1

Table3: Varimax rotation PCA loading matrix

	EC	pH	Na	K	Ca	Mg	Cl	SO_4	HCO_3	B	SiO_2	Eigen-value	Variance (%)	Cumulative (%)
PC1	0.857	-0.694	0.130	0.981	0.906	0.982	-0.167	0.284	0.967	0.010	0.673	5.473	49.754	49.754
PC2	0.503	0.366	0.966	-0.042	0.076	-0.141	0.884	0.799	0.232	0.724	0.196	3.383	30.756	80.510

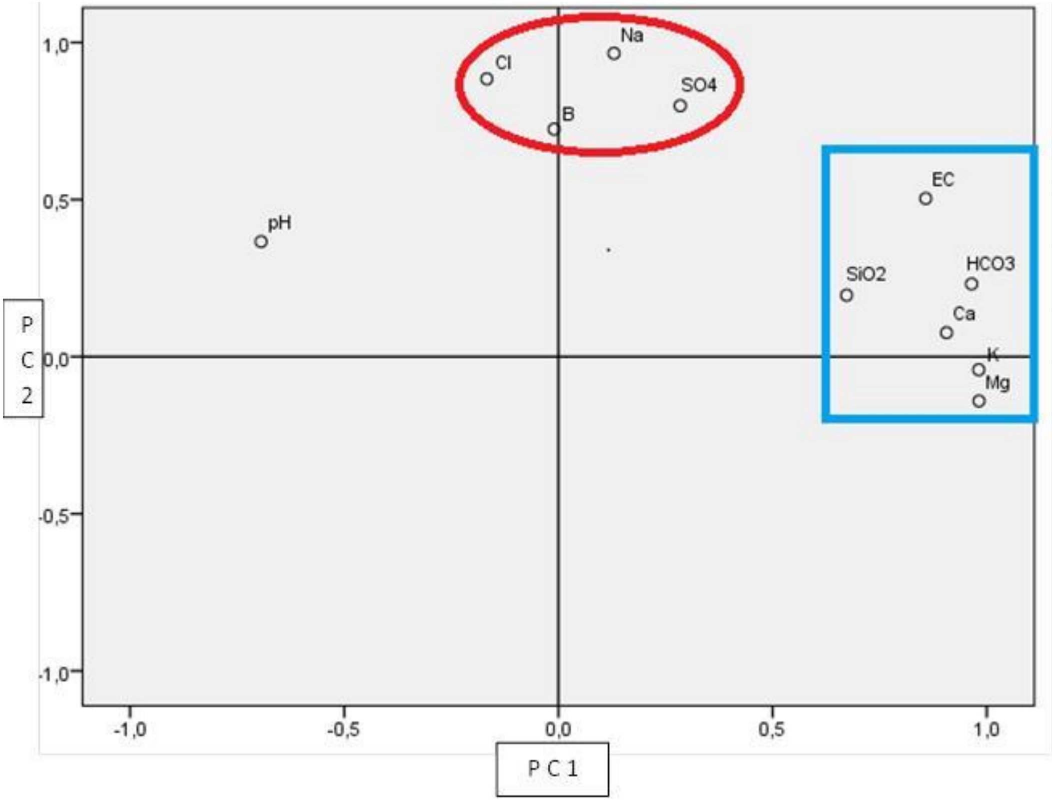

Figure 3: PCA loading/loading plot in the first and second principal component (PC1 vs PC2) respectively

4. Conclusions

In this research, multivariate statistical techniques were used to evaluate variations of groundwater parameters in the northern Kivu Rift System and the Virunga Volcanic Range, Rwanda. Interpretation of analytical data showed that the HCO_3 values exhibited high positive correlation with K, EC, Mg, Ca, and SiO_2 probably associated with rock-fluid interaction in the region. On the other hand, Na displayed strong positive correlation with, Cl, B and SO_4, possibly related to a common natural process such as geothermal activity. Three major groundwater categories are suggested by the cluster analysis.

The samples from the study area were classified as thermal groundwater mixed with surface water (Group 1), young groundwater (Group 2) and old cold groundwater water mixed with surface water (Group 3). In principal component analysis, the first 2 factors explain 80.51% of the total variance, their loadings allowing the interpretation of hydro-chemical processes that take place in the area.

The results of this study clearly demonstrate that multivariate methods offer significant improvement for integrated characterization of groundwater types and mixing trends deciphering for geothermal resource exploration and development.

REFERENCES

Belkhiri, L., Boudoukha, A. and Mouni, L.: A multivariate Statistical Analysis of Groundwater Chemistry Data, *Int. J. Environ. Res.*, 5(2) (2011), 537-544.

Brown, C. E: Applied Multivariate Statistics in Geohydrology and Related Sciences, *Springer*, New York (1998).

Egbert, J, Gloaguen, R., Wameyo, P., Ármannsson, H. and Hernández-Pérez, P.A.: Geothermal Potential Assessment in the Virunga Geothermal Prospect, Northern Rwanda. *Report* Geotherm I, Federal Institute for Geosciences and Natural Resources (2009).

Güler, C., Thyne, G. D., McCray, J. E. and Turner, A. K.: Evaluation of graphical and multivariate statistical methods for classification of water chemistry data, *Hydrogeology Journal*, 10, (2002), 455-474.

Helsel, D. R., and Hirch, R. M.: Statistical methods in water resources: U. S. Geological Survey Techniques of Water-Resources Investigations, *Book 4*, Hydrological Analysis and Interpretation (2002).

Kaiser, H. F.: The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20, (1960) 141-151.

Peskett, L.: Geothermal energy potential in Rwanda; analyzing structural controls on permeability in the northern geothermal prospect, *MSc Thesis*, Imperial College London (2014).

Reghunath, R., Murthy, T. R. S. and Raghavan, B. R.: The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India, *Water Research* 36, (2002) 2437–2442.

Resolutions: Technical Workshop on the Geologic Development and Geophysics of the Western Branch of the Greater East African Rift System with Emphasis on Factors that influence the Development of their Geothermal Systems, Kigali, Rwanda 9^{th} – 11^{th} March 2016 (2016).

Ross, K. A., B. Smets, M. De Batist, M. Hilbe, M. Schmid, and F. S. Anselmetti, Lake-level rise in the late Pleistocene and active subaquatic volcanism since the Holocene in Lake Kivu, East African Rift, *Geomorphology*, 221, (2014) 274-285

Shalev, E., Browne, P., Wameyo, P., Hochstein, M., Palmer, J., and Fenton R.: Geoscientific Surveys of the Rwandan Karisimbi, Gisenyi and Kinigi Geothermal Prospects, *Final Report,* Institute of Earth Science and Engineering, University of Auckland (2012).

Teklemariam, M.: Overview of Geothermal Resource Utilization and Potential in East African Rift System, Geothermal Resource Council, *Transaction*, 30, San Diego, California, USA (2006).

Venkatramanan, S., Ramkumar, T. and Anithamary, I.: A Statistical approach on hydrogeochemistry of groundwater in muthupet coastal region, Tamilnadu, India. *Carpathian Journal of Earth and Environmental Sciences*, 7, (2012) 47 – 54.

Ward, J. H.: Hierarchical grouping to optimize an objective function, *J Am Stat Assoc*, 69, (1963) 236-244.

West Japan Engineering Consultants, Inc., Kyushu Electric Power Co., Inc. and Mitsubishi Material Techno Corporation: Electricity development plan for sustainable geothermal energy development in Rwanda, *Final report*. JICA, Rwanda (2016).