Hierachical Cluster Analysis (HCA) of Water Sources: Case Study of Some Water Sources in Bui Division, Northwest Region of Cameroon

Fondzenyuy Vitalis Fonfo[1*] and Fondzenyuy Lionel Nyanchi[2]

[1]University of Dschang, Faculty of Science, Department of Earth Sciences, Cameroon
[2]University of Bamenda, College of Technology, Cameroon

*Corresponding author

Fondzenyuy Vitalis Fonfo, University of Dschang, Faculty of Science, Department of Earth Sciences, Cameroon.

ABSTRACT

Complex data with multiple variables can be simplified to deduce the trends and relationship between the variables using the Principal Component Analysis (PCA), which further submits the usage of the variables to establish a Hierachical Cluster Analysis (HCA). Sample bottles were cleaned and rinsed with water from the sampling sites and filled to the brim to avoid any atmospheric oxidation. Five water catchments, 29 boreholes. 16 Open wells, 16 rivers and 16 streams were sampled in this work, and were taken immediately to the Laboratory for physico-chemical analysis. The physico-chemical data in application of the Software Package for Social Sciences (SPSS), converted the multiple variables into interpretable data in a correlation matrix. Pearson's correlation matrix used in this work, reduced the data in correlation circles that were a range of geometrical projections that ranged from +1 to -1 called the correlation ratios. Variables, with correlation ratios greater than 0.5 or (-0.5) and closer to +1 or (-1) are considered significant, and are used in the Hierachical Cluster Analysis. The HCA is in form of a tree diagram called a dendrogram that arranged the variables in an increasing order of significance, with similar correlated variables put in clusters. HCA is innovative and compliments the PCA that already exists and focuses on analysis of ions expected to greatly impact an effect on water quality from their significance in the water sources. The Hierachical Component Analysis is derived from and viewed as an extension of the Principal Components Analysis..

Introduction

An evaluation of water sources based on modern analysis in communities have become increasingly important and has same implications in Bui division of the North West Region of Cameroon as elsewhere. Population growth, economic activity and urbanization of most settlements, has an impact on water sources. Many water related studies with human activity and influence on physico-chemical variables that can compromise water quality are done [1,2]. Water quality and its suitability for diverse purposes has remained preoccupying worldwide [2-10]. Water analysis with an innovative aspect such as a Hierachical Cluster Analysis (HCA) merits a particular attention and was done in this work for some water sources in Bui division, Northwest Region of Cameroon.

The ground water Chemistry of most water sources, is a result of chemical interactions such as; soil- water interactions, dissolution of primary minerals, rock- water interactions and anthropogenic sources [11-13]. The importance ground water has initiated some detailed works on its geochemical evolution [14-17]. The evaluation and establishing a correlation between the variables in water sources is important in understanding water quality.

The objective of this study was to establish a Hierachical Cluster Analysis (HCA) for Bui water sources. Specifically, the physico-chemical results obtained were used in Principal Component Analysis (PCA) to have a Pearson's correlation matrix with correlations coefficients between the variables. The variables with significant correlation coefficients were used in the HCA to construct dendrograms for the studied water sources. The variables; Temperature (oC), pH, EC, Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3, NO^-, Cl^-, SO^{2-}, SiO were needed and obtained from the sampled sites [18].

Location of the Study Zone

Bui division has boundaries, with Donga Mantung to the North, Ngohketunjia to the South, Boyo to the West, and Noun division to the East. The location of some sampling points are presented in Figure 1.
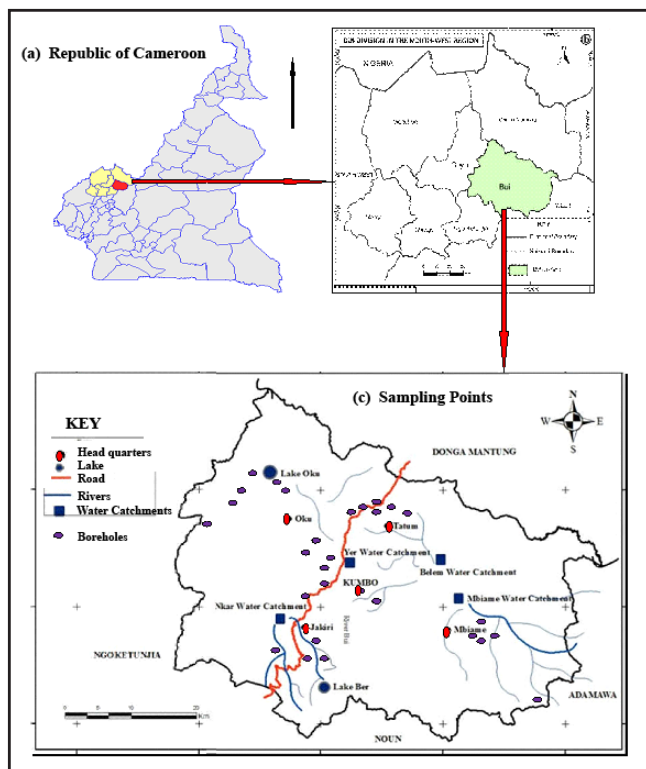
Figure 1: Location (a) Republic of Cameroon (b) Bui in the NWR (c) Sampling points

Hydrography

The Kilum Mountain range commonly referred to as Mount Oku is an important watershed in the division. It has a dendritic hydrographic network to the basins surrounding the massif. These basins include Ndop in Ngohketunjia division and Mbam Oku in Bui. Another important watershed is the Belem highlands from where water drains dentritically to Kumbo central subdivision, Nkum (Tatum) and Mbven (Mbiame) subdivisions. The Bui Plateau of the Bamenda Highlands is divided into six drainage basins that provide a major watershed for the Niger and Sanaga river systems. The streams in Bui division are mainly of first and second order as shown in Figure 2.
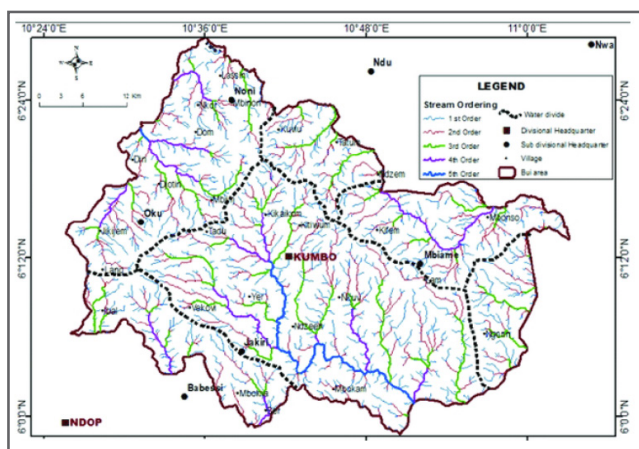

Figure 2: Hydrologic map of Bui Division [19]

Geology
Lithology

The Bamenda highland in which Bui division is part has a granitic basement covered by basalts and trachytes from tertiary volcanicity in the Cameroon Volcanic Line [20]. The division is part of the Bamenda highlands, a northward extension of the Bambouto Mountain part of the continental Cameroon Volcanic Line (C.V.L). The dominant geologic formations of Bui are basalts and trachytes similar to those of the Bamenda Mountain as reported by kamgang. A synthetic review of the geological setting of the Bambouto Mountain, presented by Kagou et al., indicated that, the first stage 21 million years ago corresponded to the building of an initial basaltic shield volcano [21]. The second stage from 18.5 to 15.3 million years ago was marked by the collapse of the caldera linked to the pouring out of ignimbrites, rhyolites and trachytes. The third stage from 15 to 4.5 million years ago renewed with basaltic effusive activity together with post-caldera extrusions of trachytes and phonolites. In these works, the Bambouto Mountain is reported to be of volcanic origin similar to its northern extension the Bamenda highlands in which Bui the study area is found. These results are in agreement with the field observations indicating that trachytes and basalts are the dominant rock types in Bui. To these rocks are often associated phonolites, rhyolites and ignimbrites. A summary of the minerals in the volcanic rocks of Bui division are presented In Table 1.

Table 1: Summary of the minerals in the volcanic rocks of Bui Division

Rock	Minerals
Basalt	Plagioclase Olivine, pyroxene
Trachyte	Plagioclase Alkaline feldspar, pyroxene, amphibole, biotite
Rhyolite	Plagioclase Alkalinefeldspar, quartz, amphibole, biotite
Phonolite	Pyroxene
Ignimbrite	Plagioclase Quartz, biotite, alkaline, feldspar, pyroxene, rocky inclusions

Soils

Mostly the parent rock, climate, vegetation and relief influence the formation of soils and its nature. The soils resulted from the weathering of basalts and trachytes. Basalts weather to produce soils with reddish to brown colour, as seen within areas of water sources in basaltic environments.

Materials and Methods
Materials

Sample bottles were cleaned and rinsed with the water sample from the collection site and finally filled to the brim to avoid any atmospheric oxidation. The samples were collected by means of a water pumping mechanism into an insulating bucket from which the sample bottles were filled for each borehole. Groundwater from active wells was collected using a rope tied to a bucket. Surface water (streams and rivers) was sampled as far as possible from the edges of the water bodies and as deep as possible along the flow path. The samples were stored in a special Thermos vaccine carrier that prevented possible modifications. The samples were taken immediately to the Laboratory for physico-chemical analysis.

Methods

The physico-chemical data was submitted to treatment in a Software Package (SPSS) from which the Principal Components Analysis (PCA) was done to obtain Pearson correlation matrix. From the input data into the software, an output was produced that established, a correlation matrix that gave the correlation coefficient between the variables that ranged from - 1 to +1.

The Principal Components Analysis (PCA)

A correlation between the analysed parameters from the water sources was done to find out the relation between the variables in a bid to establish the closeness or distance between these parameters in water. The correlation coefficient is one of the indices that evaluates the strength of a relation between many variables, Kim et al., and Bulut et al., [22,23]. A high positive correlation between elements generally indicated the possibility of a common source of origin.

Results

Tables, II and III presented the physico-chemical results of significant variables from Pearson's correlation matrix for water catchments (springs) and boreholes respectively. Table IV and V had the significant variables from Pearson's correlation matrix for open wells, and rivers/streams respectively [24]. These variables were used in the Hierachical Cluster Analysis of the water samples.

The Hierachical Cluster Analysis (HCA)

A Hierachical Cluster Analysis of the variables was obtained in form of a tree diagram; dendrogram that arranged the variables in an increasing order of significance of each variable, with similar correlated variables put in clusters [25,26]. In order to construct a dendrogram we made some assertions and took some steps.

- It's assertive that the variables were from the analysed water samples from the alternative water sources studied in Bui Division, North West Region of Cameroon.
- The variables used all had significant correlation strengths
- The variables were linked with a bottom to top approach to construct the dendrograms that related the variables in clusters. A vertical line had the height according to the Pearson's correlation coefficient value, while the horizontal line presented the distancing between the clusters or their dissimilarities.
- The pairing of the variables was based on equal or close strengths (2 variables with identical correlation coefficients were paired up, or an average for very close values). Next, the cations of same charge were paired (Na^+/ K^+. Ca^{2+} /Mg^{2+}) as well as the anions (HCO_3^-/ Cl^-/ NO_3^-) followed by physical parameters (pH/temperature, EC/ TDS). Independent parameters such as NO_3^-, K^+, and SO_4^{2-} due to their anthropogenic source unlike other parameters were identified and handled independently in the construction of dendrograms.

All the sampled water sources had at least 11 variables with 4 factors in springs, rivers and streams, 5 factors for boreholes and open wells. The significant variables were put in clusters or as independent variables in some cases. The variables were expressed in dendrograms to correlate them and reveal their interrelation in Figures; 4,5,6, and 7 for springs, boreholes, open wells, rivers and streams respectively. Tables; 2,3,4, and 5 had data for dendrogram construction for springs (catchments), boreholes, open wells, rivers and streams respectively.

Table 2: Data for construction of dendrograms for springs (catchments)

Variable	SiO_2	Ca^{2+}	K^+	HCO_3^-	Na^+	NO_3^-	EC	TDS	pH	Mg^{2+}	T°C
FI & F2	0.5	0.5	0.5	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.9
Variable	SiO_2	Ca^{2+}	HCO_3^-	Na^+	TDS	EC	pH	Mg^{2+}	NO_3^-		
FI & F3	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.7		
Variable	SiO_2	Ca^{2+}	Na^+	TDS	EC	Mg^{2+}	pH	K^+	NO_3^-		
FI & F4	0.5	0.5	0.6	0.7	0.7	0.8	0.8	0.5	0.6		
Variable	K^+	Mg^{2+}	NO_3^-	TDS	EC	HCO_3^-	Na^+	Cl^-	T°C		
F2 & F3	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.8	0.9		
Variable	Mg^{2+}	K^+	EC	TDS	HCO_3^-	Ca^{2+}	T°C				
F2 & F4	0.5	0.5	0.6	0.6	0.7	0.7	0.9				
Variable	Na^+	HCO_3^-	K^+	Ca^{2+}	Cl^-						
F3 & F4	0.5	0.5	0.5	0.6	0.7						

Table 3: Data for the construction of Dendrograms for Boreholes

Variable	Na^+	EC	NO_3^-	K^+	Mg^{2+}	SiO_2	Ca^{2+}	HCO_3^-
FI & F2	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.8
Variable	Na^+	EC	Mg^{2+}	K^+	pH	T°C	HCO_3^-	
FI & F3	0.5	0.5	0.6	0.6	0.6	0.7	0.8	
Variable	Na^+	EC	NO_3^-	K^+	HCO_3^-			
FI & F4	0.5	0.5	0.6	0.6	0.8			
Variable	Mg^{2+}	EC	Na^+	NO_3^-	K^+	Cl^-	HCO_3^-	
F1 & F5	0.5	0.5	0.6	0.6	0.6	0.6	0.8	

Variable	Mg^{2+}	SiO_2	pH	T^oC	Ca^{2+}
F2 & F3	0.6	0.7	0.6	0.6	0.8
Variable	EC	Mg^{2+}	SiO_2	Ca^{2+}	
F2 & F4	0.5	0.5	0.5	0.6	

A total number of three or four factors were used in each water source with three or four clusters each, with the variables correlated within the cluster and the clusters are further linked up from bottom to top to reveal the branching present in the relationship. The number of variables present in each cluster could be read directly from the dendrogram and the numbers varied depending on the strength (percentage contribution) of the factors concerned as expressed in the dendrograms. Figure 4 and 5 presents the dendrogram for springs (catchment) and boreholes respectively. Factors; F1/F2, F1/F3, and F1/F4 were used to construct dendrograms for springs (catchment). Temperature and NO3- exist as independent variables in Figure 4, clusters 1 and 3. Temperature is an environmental factor not dependent on other variables while NO_3^- has as source fertilizer application with no dependence nor correlation with other variables.
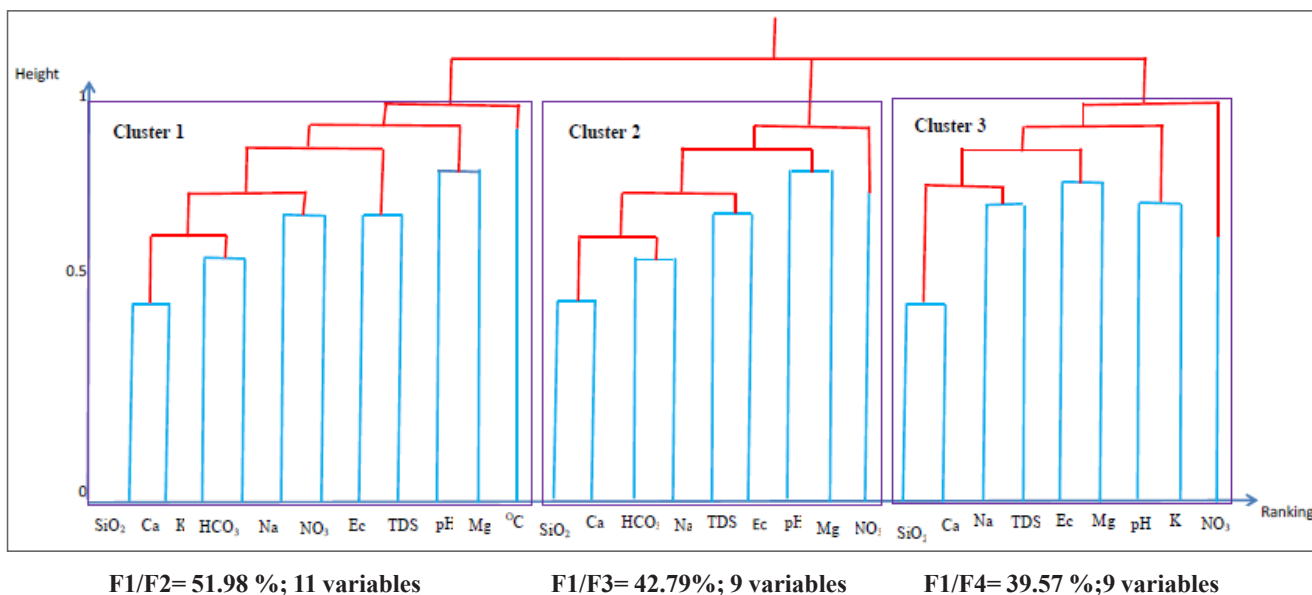

F1/F2= 51.98 %; 11 variables   F1/F3= 42.79%; 9 variables   F1/F4= 39.57 %;9 variables

Figure 3: Dendrograms for Springs (Catchments)

F1/F2= 35.56 %; 8 variables   F1/F3= 32.57 %; 7 variables   F1/F4= 30.30 %; 5 variables   F1/F5= 29.05 %; 7 variables
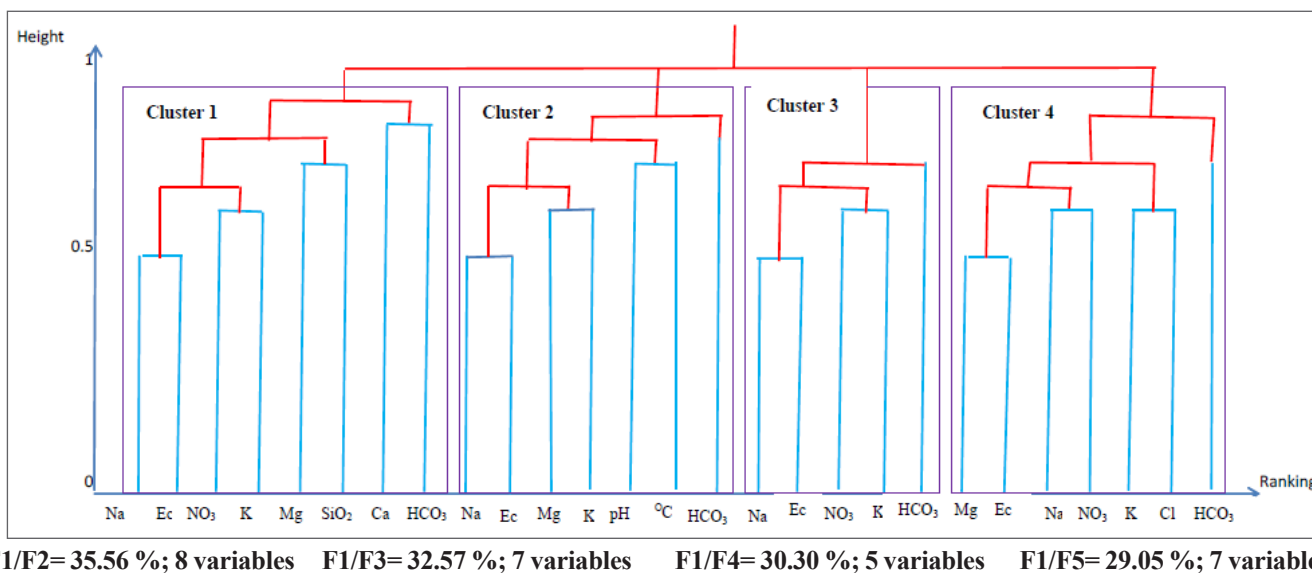
Figure 4: Dendrograms for Boreholes

The coefficients for construction of dendrograms for open wells, Streams/ Rivers in Figures 6 and 7 are in Tables 4 and 5 respectively

**Table 4: Data for the construction of Dendrograms for open wells**

Variable	SO_4^{2-}	Mg^{2+}	SiO_2	NO_3^-	K^+	$T°C$	HCO_3^-	pH
FI & F2	0.5	0.6	0.7	0.7	0.7	0.7	0.7	0.7
Variable	EC	$T°C$	Ca^{2+}	NO_3^-	Na^+	HCO_3^-	K^+	pH
FI & F3	0.6	0.7	0.7	0.7	0.7	0.7	0.8	0.9
Variable	Ca^{2+}	Mg^{2+}	$T°C$	HCO_3^-	NO_3^-	pH	K^+	
FI & F4	0.5	0.5	0.6	0.7	0.7	0.8	0.8	
Variable	EC	$T°C$	SiO_2	Mg^{2+}	Ca^{2+}	Na^+	Cl^-	
F2 & F3	0.6	0.6	0.7	0.7	0.7	0.7	0.7	
Variable	SO_4^{2-}	Ca^{2+}	$T°C$	Mg^{2+}	SiO_2	Cl^-		
F2 & F4	0.5	0.5	0.6	0.6	0.7	0.7		
Variable	Mg^{2+}	EC	SO_4^{2-}	Ca^{2+}	Na^+			
F3 & F4	0.5	0.6	0.6	0.7	0.7			

Open wells had variables in 4 clusters defined by corresponding factors as expressed in Figure 6, with independent variables; K and Cl^- in clusters 3 and 4. Unlike other variables, K could originate from application of potassium fertilizers within the vicinity of the open wells, and Chlorine from chlorination of some wells. This is different from the origin of most ions which are from weathering of rock minerals.
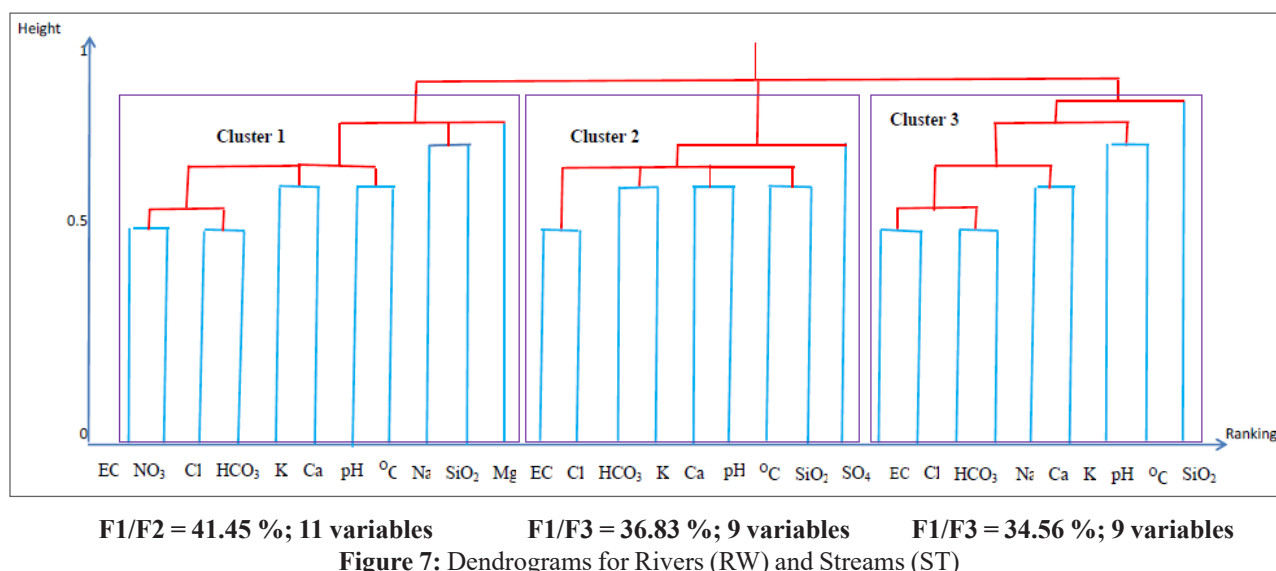
Table 5: Data for the construction of Dendrograms for Rivers and Streams

Variable	EC	NO_3^-	Cl^-	HCO_3^-	K^+	Ca^{2+}	pH	$T°C$	Na^+	SiO_2	Mg^{2+}
FI & F2	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.7	0.8
Variable	EC	Cl^-	HCO_3^-	K^+	Ca^{2+}	pH	$T°C$	SiO_2	SO_4^{2-}		
FI & F3	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.8	0.8		
Variable	EC	Cl^-	HCO_3^-	Na^+	Ca^{2+}	K^+	pH	$T°C$	SiO_2		
FI & F4	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.8		
Variable	EC	Cl^-	HCO_3^-	K^+	Ca^{2+}	pH	$T°C$	SiO_2			
F1 & F5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.7			
Variable	Ca^{2+}	Cl^-	NO_3^-	Na^+	Mg^{2+}						
F2 & F4	0.5	0.5	0.5	0.7	0.7						
Variable	NO_3^-	Na^+	Mg^{2+}	SO_4^{2-}							
F3 & F4	0.5	0.7	0.8	0.8							

Rivers and Streams water sources are also presented in dendrograms with 3 clusters with their corresponding factors as expressed in Figure 7.
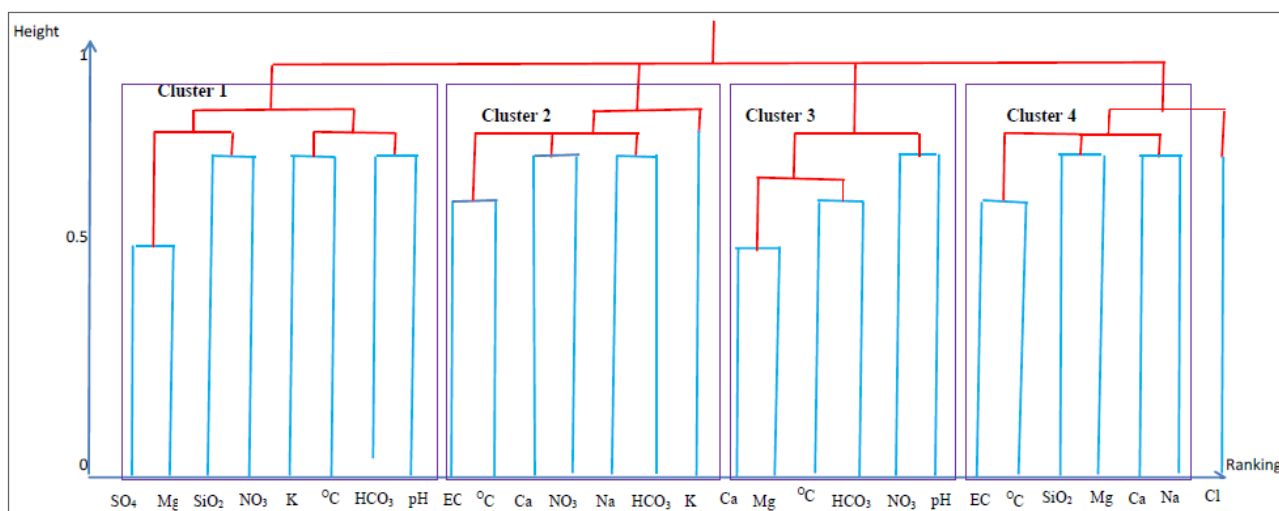
F1/F2= 46.93%; 8 variables   F1/F3= 39.87 %; 7 variables   F1/F4= 37.14 %; 6 variables   F2/F3= 35.66%; 7 variables
Figure 5: Dendrograms for Open wells

F1/F2 = 41.45 %; 11 variables   **F1/F3 = 36.83 %; 9 variables**   **F1/F3 = 34.56 %; 9 variables**

Figure 7: Dendrograms for Rivers (RW) and Streams (ST)

In all, each cluster had a corresponding number of variables correlated. These variables and the responsible factors are expressed in tables for each water source from where a dendrogram is constructed as a branching diagram that represents the relationships or similarity among the variables.

A Hierachical Clustering of the sampled water bodies was done on the basis of the significant correlation coefficients obtained at each site and values used to get a background for establishing a relationship between the variables expressed in the different dendrograms. Unfortunately, due to the heterogenous aspects within the different sites, the correlation coefficients obtained from analysis of all the water samples jointly are too low to give a tangible reflection of clustering for the entire area of study. Consequently, a Hierachical Cluster Analysis (HCA) is limited to the studied water sites and not the entire Bui division. This is because an analysis using variables with weak and insignificant correlation coefficients will reflect weak and insignificant outcomes. On the strength of the insignificant correlation coefficients for the entire division in this work, the strength of the PCA and HCA were limited to the studied water sources and not to the entire division.

General Conclusion

It was crystal clear that the greatest percentage contribution of the variables in the definition of the factors, were given by factors 1 and 2 as opposed to other factors with lower percentage contribution. These factors; F1 and F2 consequently, contributed dominantly in the data analysis. In all, the positively correlated variables are grouped together and evolve concomitantly, while the negatively correlated variables are ungrouped and evolve in an antagonistic manner. The Hierachical Cluster Analysis uses data from PCA. Hence, the HCA is viewed as an extension of the Principal Component Analysis.

Perspectives

Vulnerability on parameters inherent in water quality envisages regular sampling and analysis to monitor evolution and suggest strategies to mitigate any disastrous outcome linked to water quality to the local authorities.

Conflict of Interest

In course of this work no conflict of interest was encountered. The authors had total collaboration with support of the Earth science department of the University of Dschang- Cameroon.

Data Availability

The data from this work is available in the article and includes; physico-chemical data from analysis of the water samples from water catchments, boreholes, open wells, rivers and streams.

Reference

1. Dechao C, Acef E, Hualian X, Xinliang X, Zhi O. A Study in the Relationship between Land Use Change and Water Quality of the Mitidja watershed in Algeria Based on GIS and RS. Sustainable Water Quality Management in the Changing Environment. 2020. 12: 3510.

2. Mohammed H, Yangun S, Mohamed M, Weijin J, Majid K, et al. Assessment of Groundwater Resources in Coastal Areas of Pakistan for Sustainable Water Quality Management Using Joint Geophysical and Geotechnical Approach : A Case Study. Sustainable Water Quality Management in the Changing Environment. 2020. 12: 9730.

Sajil KPJ, Elango L, James EJ. Assessment of Hydrochemistry and water quality in south Chennai coastal area. Arabian Journal of Geosciences. 2014. 7: 2641-2653.

Tyagi S, Singh P, Sharma B, Singh B. Asseeement of Water Quality for Drinking Purposes in District Pauri of Uttarakhand, India. Applied Ecology and Environmental Sciences. 2014. 2: 94-99.

Njoyim EBT, Mofor NA, Niba MLF, Sunjo J. Physicochemical and bacteriological quality assessment of the Bambui Community drinking water in the North West Region of Cameroon. African Journal of Enviromental Science and Technology. 2016. 10: 181- 191.

Inam E, Inoh GG, Offiong NO, Etim BB. Physico-che0mical Characteristics and Health Risk Assessment of Drinking Water Sources in Okoroette Community, Eastern Coast of Nigeria. American Journal of Water Resources. 2017. 5: 13-23.

Germain KNG, Jules MOM, Narcisse KA, Aristide GD, Lanciné DG. Caractérisation hydrogéochimique des eaux souterraines du bassin versant de la Baya, Est Côte d'Ivoire. International Journal of Biological and Chemical Sciences. 2019. 13.

3. WHO. Guidelines for Drinking- Water Quality. 4th Edition. Incorporating the First Addendum, Geneva. 2017.

WHO. Developing drinking-water quality regulations and standards. Water sanitation hygiene. Publication on water sanitation and health. 2018.

4. WHO. UN-Water Global Analyses and Assessment of Sanitation and Drinking-Water (GLASS) 2019 Report. 2019.

Ramesh K, Elango L. Groundwater quality and its suitability for domestic and agricultural use in Tondiar river basin, TamilNadu. India. Environ. Monit. Assess. 2012. 184: 3887-3899.

Akoachere RA, Eyong TA, Egbe SE, Wotany RE, Nwude MO, et al. Geogenic Imprint on Groundwater and Its Quality in Parts of the Mamfe Basin, Manyu Division Cameroon. Journal of Geoscience and Environmental Protection. 2019. 7: 184-211.

Frommen T, Groeschke M, Noelscher M, Schneider M. Anthropogenic and geogenic impacts on peri-urban aquifers in India - Insight from a Case Study in the Northeast of Jaipur (in prep). 2019.

5. Edmunds WM, Smedley PL. Ground water geochemistry and health with reference to developing countries. Geological Society Special Publication No. 1996. 113: 91-105.

6. Tanyileke GZ, Kusakabe, Evans WC. Chemical and isotopic characteristics of fluids along the Cameroon Volcanic Line. 1996. 433-441.

7. Deutsch WJ. Groundwater geochemistry. Fundamentals and applications to contamination. CRC Press. Florida. 1997. 221.

8. Appelo CAJ, Postma D. Geochemistry, groundwater and pollution (2nd edition). Balkama, Amsterdam 635. 2005.

9. Fondzenyuy VF, Kengni L. Hydrogeochemistry of alternative water sources in Bui division, North west region of Cameroon: Implications on drinking, domestic and agricultural uses. PhD Thesis, University of Dschang, Faculty of sciences, department of Earth sciences. 2021.

Suiven JPT. Standardised Precipitation Valuation of Water Resources Vulnerability to Climate variability on the Bui Plateau, Northwest Cameroon. Environment and Ecology Research of. 2020. 2: 83-92.

Ngako V, Affaton P, Njonfang E. Pan-African tectonic in North–Western Cameroon implication for history of Western Gondwana. Gondwana Res. 2008. 14: 509- 522.

Kagou D, Nkoathio D, Pouclet A, Bardintzeff JM, Wandji K, et al. The discovery of late Quatenary basalts on Mount Bambouto: Implications for recent widespread volcanic activity in the Southern Cameroon line. Journal of the African Earth Science. 2010. 57: 96-108.

Kim JG, Ko KS, Kim TH, Lee GH, Song Y, et al. Effect of mining and geology on the chemistry of stream water and sediment in a small watershed. Geoscience Journal. 2007. 11: 175- 183.

10. Bulut VN, Bayram A, Gundogdu A, Soylak M, Turfekei M. Assessment of water quality parameters in the stream Galyan, Trabzon, Turkey. Environmental Monitoring Assessment. 2009. 165: 1-13.

11. Kaiser HF. The varimax criterion for factor analytic rotation in factor analysis. Educational and Psychological Measurement. 1958. 23 770-773.

Johnbosco E. Ground water quality assessment using pollution index (PIG), Ecological Risk Index (ERI) and Hierachical ClusterAnalysis (HCA): A case study. Ground waterfor sustainable development. 2020. 10: 1000292.

Zolfaghari F, Khosravi H, Shariyari A, Jabbari M, Abolhasani A. Hierachical Cluster Analysis to identify the homogenous desertification management units. PLoS ONE. 2020. 10: 20226355.