Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.124–134.

HEAVY METAL SOURCE IDENTIFICATION AND ANALYSIS USING MULTIVARIATE STATISTICAL METHODS IN SOILS FROM AKUSE AREA, SOUTH-EASTERN GHANA

DANIEL ODURO BOATEY NUAMAH

Department of Geophysics, University of Miskolc
danielobnuamah54@gmail.com

Abstract: Multivariate statistical methods (principal component analysis, cluster analysis, and correlation analysis) have been applied to coastal soils of Akuse, Southeastern Ghana to determine heavy metal sources (Cu, Zn, Ni, Pb, Cr, Co and Ba). Thirty-four Composite soil samples were taken on a grid of 1km × 1km from the study area and analyzed using X-ray fluorescence analytical protocol. The study showed that: (i) Ni, Pb, Co and Ba had anthropogenic sources (ii) Zn and Cu were associated with parent materials and therefore had natural sources whilst (iii) Cr showed varied sources of both natural and anthropogenic. From the statistical mean estimation, Chromium (Cr) showed the highest mean concentration (mean = 442.8 ± 68.658 ppm). It was followed by Barium (Ba) (mean = 261 ± 137.88 ppm) and Nickel (mean = 83.94 ± 59.915 ppm). Elements such as copper, zinc and Cobalt showed average mean values ranging from (70.45 ppm – 47.47 ppm). Significant positive correlations were found between all metals except Cr and Cu (with r = –0.62). Element pairs such as Cu-Ni, Zn-Cr, Ni-Pb, Ni-Ba, Ni-Co and Cu-Zn showed strongly positive correlation coefficient "r", indicating their association in the study area. The principal component analyses (PCA) yielded 3 components which accounted for 84.73% of the total variance in the data set, indicating that the remaining 16.38% were not explained by these components. The first cluster (Ni, Pb, Co, Ba) accounts for 45.75% of the total variance, the second (Cu, Zn) for 22.01%, the third contains only Cr and suggests both natural and anthropogenic origins. The distribution of heavy metals in the soils had been greatly affected by soil formation, atmospheric deposition and human activities.

Keywords: *multivariate statistics, cluster analysis, principal component analysis, correlation, communalities*

1. INTRODUCTION

The characteristic of soil contaminated by heavy metals is commonly influenced by total heavy metal contents. The study area falls within the Dangme West District of the Greater Accra region in Southeastern Ghana. Although some studies have already extended to the investigation of heavy metal fractions within the Akuse area, it is still far from enough. These studies provided inadequate information about the bioavailability and toxicity of heavy metals. This study focuses on heavy metal sources, which is critical for the monitoring and assessment of soil contamination. There are two main sources of heavy metals in the soil: (i) natural background, which represents heavy metal concentration derived from parent rocks; (ii) anthropogenic contamination resulting from application of agrochemicals, addition of organic amendments, animal manure, mineral fertilizer, and sewage sludge. Generally, there are more heavy metals in soils originated from anthropogenic sources than natural sources. Due to human settlement, the area of farmland in Akuse has been rapidly increasing. Meanwhile, agro-environmental pollution caused by chemical fertilizers and pesticides, waste discharges such as animal wastes, sewage irrigation, and sludge application are so serious. Therefore, the identification of heavy metal sources in the present work would offer essential information on the monitoring and assessment process of agricultural soils in the Akuse area. The aims of the study were: (1) to determine average local concentrations of some heavy metals (Cu, Zn, Ni, Pb, Cr, Co and Ba); (2) to define their natural and/or anthropogenic sources; (3) to identify their local or exotic sources causing contamination in top soils. A multivariate statistic approach using principal component analysis, cluster analysis and correlation analysis were adopted to assist the interpretation of geochemical data and to distinguish different sources of heavy metals.

2. REGIONAL GEOLOGIC SETTINGS

The study area *(Figure 1)* form part of the southeastern segment of the Trans-Saharan belt exposed in southeastern Ghana and adjoining parts of Togo and Benin known as Dahomeyide orogenic settings [1, 2, 3] precisely the Shai Hills area. The principal tectonic elements of the Dahomeyide orogen are: (1) the deformed edge of the West African Craton with its cover rocks consisting of craton verging nappes and thrust sheets bounded by ductile shear zones; (2) the suture zone representing the eastern boundary of the autochthonous West African Craton; (3) exotic rocks that form the granitoid gneiss complexes east of the suture zone *(Figure 1)*. Metamorphism is generally in the amphibolite facies. Gneisses with garnet, pyroxene and scapolite occur among more ordinary quartzo-feldspathic, biotite and hornblende-bearing varieties. Eclogites (high-pressure garnet-pyroxene rocks chemically equivalent to basalt) have been recorded from among large masses of mafic gneisses that include amphibolites and pyroxenites and contain much garnet. In southeastern Ghana, the High Pressure mafic granulites have been referred to as Shai Hills gneiss [1], and are tectonically juxtaposed with the alkaline gneiss complex in the suture zone. Rocks of the Shai Hills Gneiss unit are folded into west and southwest verging nappes and crop out in inselbergs of the Accra Plains, which include the northsouth trending Shai Hills. Most of the isolated hills are asymmetrical with steep, west-facing scarp slopes such as Krobo Hills and the prominent Osu Yongwa. Several quarries in Krobo, Shai Hills and in the low hills south of Shai Hills provide access to fresh outcrops for detailed sampling. In these quarries, the distinctive features of the rocks exposed are the prominent modal layering and extensive veining. The layering is discontinuous and consists of alternating garnet-rich and hornblende-rich zones, which give the rock a streaky appearance and are interpreted to be shear induced. The veins occur in all sizes and orientated to the tectonic layering in these rocks.

2.1. Geology of study Area

Strongly metamorphosed ancient sediments occur within the study area with Dahomeyan gneiss and schist's occupying most of the plains cover. Basic gneiss forms a number of large inselbergs (isolated rocky hills) in the north and center of the area. Small rock outcrops are also common in the north close to the inselbergs but are rare in south and southeast. The eastern belt of acidic gneiss consists mainly of the grained metamorphosed rocks rather richer in minerals than the rocks in the western part and with many fewer quartz veins. The soil type is largely clayey and fine grain in texture. There are no known mineral deposition of commercial and economic value in the area, except for clays of various types occurring in different places used for pottery and bricks.

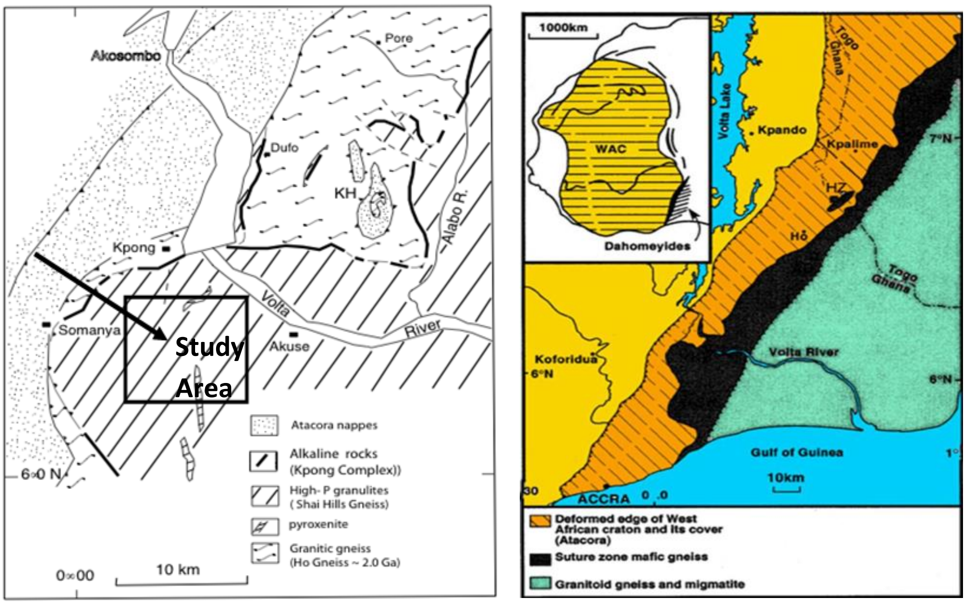

Figure 1. *Location maps showing the study area and the Dahomeyide Tectonic Settings, South-eastern Ghana (Attoh et al., 1997)*

3. MATERIALS AND METHODS

3.1. Sample Collection and analysis

The sampling protocol employed in this work is soil sampling. This was selected based on the time and resources available, the size of the program, the indicators to be measured and data sources to collect information from. A total of 34 samples were taken by six groups. All samples were taken from precise and accurate sample points except E4 which deviated as a result of inaccessible point of location. Using the earth chisel, the samples were taken at a depth range of 50–90 cm.This was intended to remove the top soil. The samples were dried, crushed and sieved using a 108 and 106 micrometer sieves. They were shaken continuously for three minutes producing finer samples. Five grams (5 g) of each sample were placed in a paper bag, labeled and taken to the laboratory for analyses. Pellets were prepared from the samples and analyzed with X-Ray Fluorescence spectrometer (XRF). This system provided a complete elemental analyses and qualitative characterization. In all, 34 soil samples were analyzed.

3.2. Statistical analysis

Multivariate statistical analysis was performed with the aid of Microsoft excel and IBM statistical software SPSS version 16.0. Pearson's correlation coefficient between the variables were calculated in the forms of matrix and used as a measure of similarity and relationship between the six elements. Cluster analysis has most frequently been employed as a classification tool and has been used by some researchers as a means of representing the structure of data via the construction of dendrogram [4, 5, 6] or overlapping clusters [7]. Cluster analysis was also performed in order to evaluate the metal pollution and distinguish the sources. Between groups average linkage method was tested during the classifying procedure. Factor analysis, using principal component solution was also carried out on the data. During factor extraction, the shared variance of a variable is partitioned from its unique variance and error variance to reveal the underlying factor structure, thus, only shared variance appears in the solution [8]. Principal component analysis does not discriminate between shared and unique variance. When the factors are uncorrelated and communalities are moderate, it can produce inflated values of variance accounted for by the components [9, 10]. Since factor analysis only analyzes shared variance, factor analysis should yield the same solution while also avoiding the inflation of estimates of variance accounted for [8]. Principal component analyses (PCA) is designed to transform original variable into new uncorrelated variables called components. It was used to explain the variance observed in the data, and to understand the relationship between the different elements. Thus, Factor analysis is a multivariate analytical technique, which derives a subset of uncorrelated variables called factors that explain the variance observed in the original dataset [11, 12, 13]. It was used to unearth the latent structure of a set of variables and attempted to identify few factors that were responsible for the correlation among a large number of observed variables. Its application was to describe, if possible, the covariance relationship among many variables in terms of a few underlying, but unobservable, random quantities called factors.

4. RESULTS AND DISCUSSION

4.1. Correlation between total heavy metal contents

Table 1 shows the total metal content in the soil calculated in mean values with their corresponding standard deviations. Chromium (Cr) showed the highest mean concentration (mean = 442.8 ± 68.658 ppm) followed by Barium (Ba) (mean = 261 ±

137.88 ppm). *Table 2* shows the Pearson's correlation coefficient of heavy metals in the samples. The Pearson's correlation coefficient measures the strength of a linear relationship between any two variables on a scale of –1 (perfect inverse relation) through 0 (no relation) to +1 (perfect sympathetic relation). Moderate positive correlations were found between all metals except Cr and Cu (with r = –0.62). There exists poorer correlation between metals such as Cu-Pb, Cu-Ba, Zn-Ba and Cr-Pb showing their low association in the field. This implies that the pollution sources of these metals has no significant effect on their chemical properties and hence are not closely related in the field of survey. Relatively strong positively correlated metals included Cu-Zn, Cr-Ni, Co-Ni, Ba-Ni, Ba-Co and Pb-Ni *(Figure 3)*. Others such as Zn-Ni and Co-Cu showed weaker positive correlations. The above relations are further showed with scattered diagrams below.

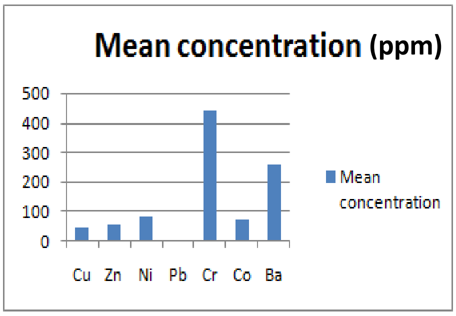

Figure 2. Plot showing mean concentrations of heavy metals

Table 1
Descriptive Statistics (ppm)

	Mean (ppm)	Std. Deviation (ppm)	N
Cu	47.47	18.221	34
Zn	57.26	13.274	34
Ni	83.94	59.915	34
Pb	2.14	1.028	34
Cr	442.85	405.658	34
Co	70.45	38.296	33
Ba	261.13	137.880	34

Table 2
Correlation Matrix

Elements	Cu	Zn	Ni	Pb	Cr	Co	Ba
Cu	1.000						
Zn	0.697	1.000					
Ni	0.188	0.404	1.000				
Pb	0.093	0.106	0.515	1.000			
Cr	-0.062	0.216	0.672	0.011	1.000		
Co	0.435	0.327	0.646	0.504	0.196	1.000	
Ba	0.049	0.057	0.664	0.510	0.182	0.712	1.000

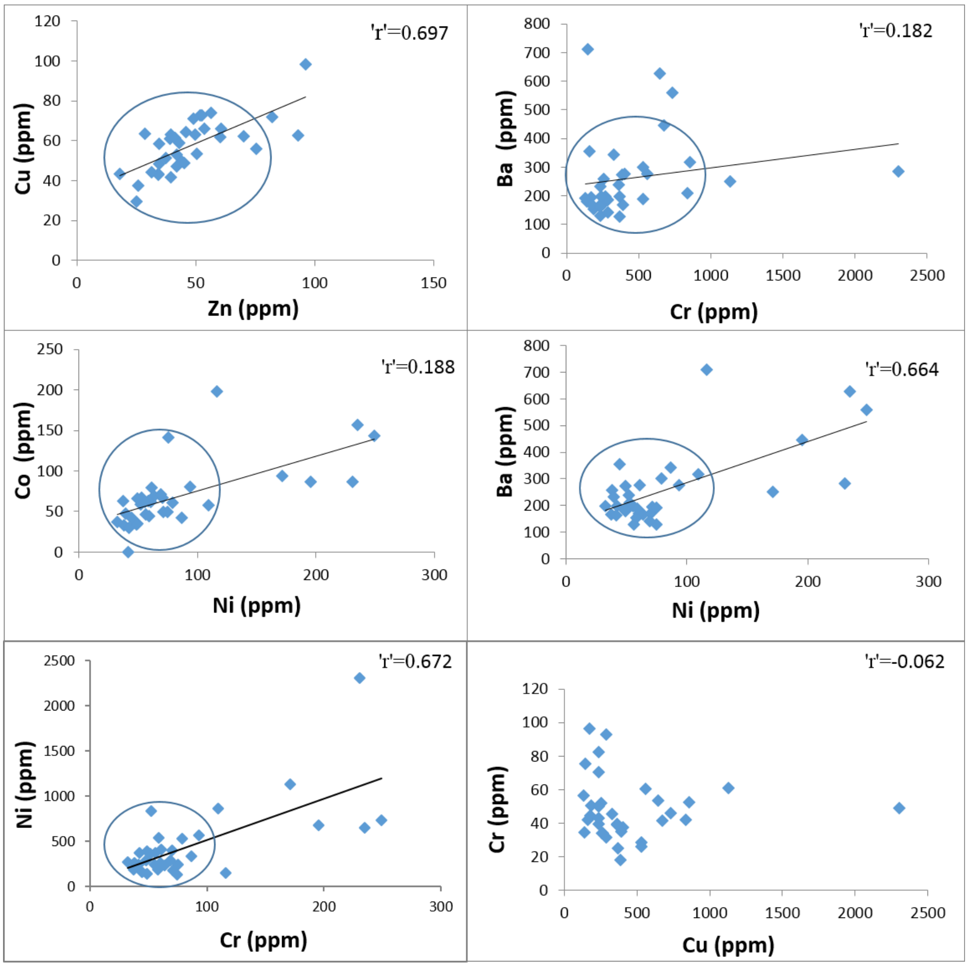

Figure 3. *Correlation plots of heavy metal in the study area*

4.2. Principal Component Analysis

Table 3 shows the Total variance explained by principal components in cumulative percentages whilst *Table 4* shows the principal component coefficient for heavy metals in the Akuse area extracted in three components. The PCA yielded 3 components which explained 84.73% of the total variance in the data set, indicating the remaining 16.38% were not explained by these axes. The first component accounts for 45.75% of the total variance and contains Ni, Pb, Co and Ba. Component 1 showed significant positive load for all the heavy metals. Ni showed the highest load in this factor (load value 0.901). This was followed by Co, Ba, Pb, Zn, Cr and Cu respectively. Although Zn was not classified in this component, it showed significant loading in the component. The metals Ni, Pb, Co and Ba were also significantly correlated *(Table 2)*, indicating that they are from the same source.

Table 3
Total variance explained by principal components

Com-ponent	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.203	45.754	45.754	3.203	45.754	45.754
2	1.541	22.010	67.764	1.541	22.010	67.764
3	1.187	16.963	84.726	1.187	16.963	84.726
4	0.544	7.775	92.501			
5	0.283	4.041	96.542			
6	0.165	2.361	98.903			
7	0.077	1.097	100.000			

Table 4
Principal component coefficients

	Component		
	1	2	3
Cu	0.444	0.817	-0.188
Zn	0.528	0.740	0.176
Ni	0.901	-0.184	0.308
Pb	0.629	-0.294	-0.434
Cr	0.456	-0.176	0.846
Co	0.859	-0.006	-0.256
Ba	0.757	-0.419	-0.238

This factor suggested an anthropogenic source since Ba, Ni and Co showed a higher mean *(Table 1)* value exceeding permissible background values. The second component accounts for 22.01% of the variance and contains Cu and Zn. The close association of Cu and Zn shows that they are from possibly naturally related source. The two heavy metals were also highly correlated with r = 0.697 *(Table 2)* showing their close association in the area. The remaining metal exhibited significant negative loads in this component showing their disassociation with Cu and Zn in the study area. This accession is further explained by lower correlation values showed by Cu and Zn with the other heavy metals *(Table 2)*. The third component accounts for 16.963% of the variance and contains only Cr (load value 0.846). This factor suggests a source of mixed origins of natural and anthropogenic as chromium (Cr) showed the highest mean value *(Table 1)*.

4.3 Cluster Analysis

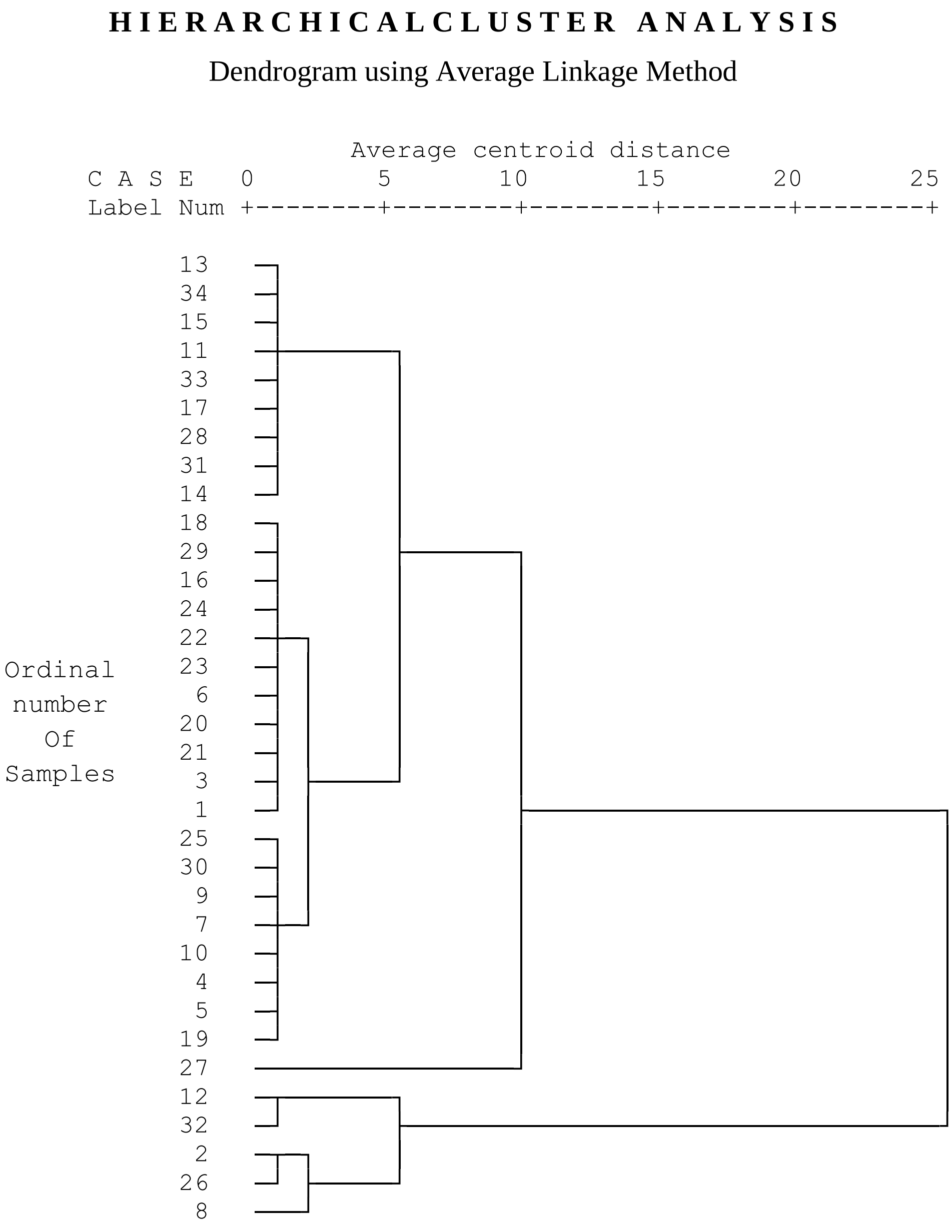

Figure 4.

Connections and relations for interdependence among the different metals in the Akuse area were also established by means of cluster analysis. Hierarchical cluster analysis created a hierarchy of clusters which were represented in a tree structure called dendrogram. The dendrogram of the hierarchical cluster analysis of metal contents is illustrated below *(Figure 4)*. Dissimilarities between clusters of objects are defined in several ways including the maximum dissimilarity (complete linkage), minimum dissimilarity (single linkage) and average dissimilarity (average linkage) [14]. Group-average linkage (agglomerative) clustering as applied in this study evaluates cluster quality based on all similarities between data including pairs from the same cluster but self-similarities are not included in the average. Group average linkage satisfies the reducibility principle which ensures that the procedure cannot result in inversion in the dendrogram representing the hierarchy [15]. It is also known to minimize the distortion imposed on the inter-object similarity matrix when a hierarchic classification is generated [16].

The results revealed four clusters of elements, the first cluster included elements that had previously been interpreted as anthropogenic (Ni, Pb, Co and Ba) while the second clusters contained the natural source elements (Cu and Zn). Cr concentrations were represented by cluster 3. Cluster 1 ranges from 13-14 on the dendrogram, Cluster 2 ranges from 1-18, Cluster 3 ranges from 19–25 with the rest scattered in cluster 4. These group distributions also corresponded to the locations of samples. Cluster 4 contained local abnormal points and could be neglected in analysis. Based on earlier discussions, the results suggested again that Ni, Pb, Co and Ba were mainly controlled by anthropogenic sources, while, Cu and Zn had natural sources. Cluster analysis gave similar results to correlation and factor analyses, enabling the identification of the two sources of heavy metals. Therefore, this analysis re-affirmed that the elements studied come from two different sources in the soils except Cr which has a mixed source.

5. CONCLUSION

Multivariate statistical methods applied in this study proved useful in the characterization of heavy metal sources in soils from Akuse area. Relatively homogenous groups of metals were identified, and varied sources or origins of metals could be distinguished based on their elemental association, level of accumulation and chemical properties. Similar geochemical properties such as ionic size and radius accounted for the positive correlation observed among heavy metals Ni, Cr, Co, Zn and Cu which belong to the transition trace elements group. Among the factors affecting the distribution of heavy metals in the soil are the abundance of the element in the parent rock, the nature of the weathering in the study area and human activities. Chromium and barium showed excessively high concentrations in the study area. The sequences of geo-accumulation of heavy metal in a decreasing order are Cr, Ba, Ni, Co, Zn, Cu and Pb.

The study showed that: (i) Ni, Pb, Co and Ba had anthropogenic sources (probably from over use of chemical fertilizers and pesticides, industrial and discharges, animal wastes, sewage irrigation, etc.), (ii) Zn and Cu were associated with parent materials and therefore had natural sources, thus, the weathering of parent materials and subsequent pedogenesis due to the alluvial deposits. These elements are largely associated with high pressure acidic gneisses complexes scattered across the north and western parts of the study area. (iii) Cr showed varied sources of both natural and anthropogenic. The effect of heavy metals in the soils was greatly affected by soil formation, atmospheric deposition and human activities. These findings provided essential information on the possible sources of heavy metals without assessing their extent of pollution in the study area but would still go a long way to contribute to the monitoring and assessment process of agricultural soils in the area.

ACKNOWLEDGEMENT

I want to thank the entire lecturers and staff of the Earth Science Department (University of Ghana), who helped in the provision of geological equipment and assistance during the field work which was exceedingly courteous and helpful. Also, my sincere gratitude to the entire staff of the Geological Survey Department (Accra branch) for their help during the preparation and analysis of samples.

I am greatly indebted to Dr. Frank K. Nyame (University of Ghana) and Dr. Boamah (Geological Survey Department) for their technical advice. I would like to express my sincere gratitude to the academic staff, Department of Geophysics (University of Miskolc) for their assistance and encouragement to make this paper a success.

REFERENCE

[1] ATTOH, K.–DALLMEYER, R.D.–AFFATON P. (1997)
Chronological nappe assemblage in the Pan African Dahomeyide orogeny, West Africa. Evidence from 40Ar/39Ar mineral ages. *Precambrian Research*, Vol. 82, pp 82.

[2] AFFATON, P.–RAHMAN, M. A.–TROMPETTE, R.–SOUGY, J., (1991)
The Dahomeyide orogeny; Tectonothermal Evolution and Relationships with the Volta Basin. Springer Berlin Heidelberg.

[3] CASTAING, C.–FEYBESSE, J.L–THIEBLEMONT, D.–TRIBOULET C.–CHEVREMONT, P. (1994)
Palaeogeographical Reconstructions of the Pan, African/brasilianoorogen: Closure of an Oceanic Domain or Intracontinental Convergence between Major blocks? *Precambrian Research*, 67, 327–344.

[4] BERTIN, J. (1967)
Sémiologie Graphique. Paris: Gauthier-Villars.

[5] Hartigan, J. A. (1967),
Representation of Similarity Matrices by Trees, *Journal of the American Statistical Association*, 62, 1140–58.

[6] SZABÓ, N. P.–DOBRÓKA, M.–KAVANDA, R. (2013)
Cluster Analysis Assisted Float-Encoded Genetic Algorithm for a More Automated Characterization of Hydrocarbon Reservoirs. *Intelligent Control and Automation*, 2013, 4, 362–370.

[7] SHEPARD R. N.–P. ARABIE (1979)
Additive Clustering: Representation of Similarities as Combinations of Discrete Over-Lapping Properties, *Psychological Review*, 86, 87–123.

[8] COSTELLO, A. B.–OSBORNE, J. W. (2009)
Best practices in exploratory factor analysis; four recommendations for getting the most from your analysis. *The Pan-African Management Review*, 2009, Vol. 12, No. 2, 131–146.

[9] GORSUCH, R. L. (1997)
Exploratory factor analysis: its role in item analysis. Journal of Personality Assessment, 68:532-560.

[10] McArdle, J.J (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, 25, 81–87.

[11] ANAZAWA–OHMORI, (2005)
The hydrochemistry of surface waters in andesitic volcanic area, Norikura volcano, central Japan. *Chemosphere*, Volume 59, Issue 5, April 2005, pp. 605–615.

[12] BROWN, (1998)
Applied Multivariate Statistics in Geohydrology and Related Sciences Springer. New York, (1998)

[13] YIDANA, S. M–OPHORI, D.–BANOENG-YAKUBO, B. (2008)
A multivariate statistical analysis of surface water chemistry – the Ankobra Basin, Ghana. *Environ. Manag*. 86, 80–87.

[14] CHRISTOPHER D. MANNING–PRABHAKAR RAGHAVAN–HINRICH SCHÜTZE
Introduction to Information Retrieval. Cambridge University Press, 2008.

[15] F. MURTAGH (1983)
A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26, 354–35

[16] A. EL HAMDOUCHI–P. WILLET (1989)
Comparison of hierarchic agglomerate clustering methods for document retrieval. *The Computer Journal*, Vol. 32, No. 3.