Journal of Geoscience and Environment Protection, 2022, 10, 303-317
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Application of Principal Component Analysis, Cluster Analysis, Pollution Index and Geoaccumulation Index in Pollution Assessment with Heavy Metals from Gold Mining Operations, Tanzania

Caren Anatory Kahangwa

National Environment Management Council, Dodoma, Tanzania
Email: caren.kahangwa@gmail.com

How to cite this paper: Kahangwa, C. A. (2022). Application of Principal Component Analysis, Cluster Analysis, Pollution Index and Geoaccumulation Index in Pollution Assessment with Heavy Metals from Gold Mining Operations, Tanzania. *Journal of Geoscience and Environment Protection, 10,* 303-317.
https://doi.org/10.4236/gep.2022.104019

Received: January 18, 2022
Accepted: April 26, 2022
Published: April 29, 2022

Abstract

Gold mining is now widely acknowledged as one of the significant sources of soil pollution in developed countries. In developing countries, the sources and levels of soil contamination have not been thoroughly addressed. Thus, this study was intended to determine the source of soil pollution and the level of contamination in the active and closed gold mining areas. The research paper presents the pollution load of heavy metals (lead-Pb, chromium-Cr, cadmium-Cd, copper-Cu, arsenic-As, manganese-Mn, and nickel-Ni) in 90 soil samples collected from the studied sites. Multivariate statistical analysis, including Principal Component Analysis (PCA) and Cluster Analysis (CA), coupled with correlation coefficient analysis, was performed to determine the possible sources of pollution in the study areas. The results indicated that Pb, Cr, Cu and Mn come from different sources than Cd, As and Ni. The results obtained from the metal pollution assessment using the Pollution Index (PI) and the Geoaccumulation Index (Igeo) confirmed that soils in the mining areas were contaminated in the range from moderately through strongly to highly contaminated soils. This study verified that soil contamination in the gold mining areas results from natural and anthropogenic processes. The current study findings would enhance our knowledge regarding the soil contamination level in the mining areas and the source of contamination. It is recommended to use PCA, CA, PI and Igeo to assess and monitor the heavy metal contaminated soil in gold mining areas.

Keywords

Heavy Metals Contamination, Principal Component Analysis, Cluster Analysis, Pollution Index, Geoaccumulation Index

1. Introduction

Heavy metals are among the essential pollutants spread in the environment, including air, water and soil. The common sources of heavy metals in the environment include wastewater from industries, deposited wastes, mineral exploration and extraction, urbanization and vehicles (automobiles), all referred to as anthropogenic pollutants (Mahugija & Sheikh, 2018). Mining is the extraction of valuable minerals or other geologic material from the earth, usually from an ore, resulting in contamination of the environment. Mining activities have resulted in waste rock dumps and tailings storage facilities, which generally present a constant danger of transforming toxic heavy metals into the environment (Mganga, 2014). Toxic heavy metals signify one of the probable environmental hazards from mines, which affect many countries with historic mining industries (Fazekasov & Fazekas, 2020). The soil naturally contains heavy metals, although in small amounts. However, heavy metals increasingly pollute the soil, whether from the air or spreading waste.

In some cases, high levels of heavy metals occur due to environmental pollution caused by anthropogenic activities (Fashola et al., 2016). Heavy metals in soils depend on natural processes and anthropogenic influences such as mining and its associated activities. Mining and milling operations, together with the disposal of tailings, provide apparent sources of soil contamination. Gold mining is one of the essential point sources of heavy metals in the environment. The extent of heavy metal contamination around gold mines depends upon geochemical characteristics and the mineralization of tailings. Specifically, open-pit mining has a severe environmental impact on soils and water streams (Krishna et al., 2013). Gold mining, whether open-pit or deep shaft, is associated with other heavy metals such as lead (Pb), chromium (Cr), cadmium (Cd), copper (Cu), silver (Ag) and arsenic (As) (Fashola et al., 2016).

In mining areas, intensive mineral extraction has resulted in a considerable volume of waste minerals and tailings, which contain heavy metals. These heavy metals have significant accumulation in the soil, leading to environmental pollution due to their toxicity, non-degradability and persistence in nature, causing a severe threat to human health. Without proper management, the abandoned waste dumps and tailings generate highly polluting acid mine drainage, which causes contamination of the water, sediments and agricultural soils in the vicinal areas (Ma et al., 2015). The risk herein is the potential accumulation of these contaminants by plants, resulting in phytotoxicity and human diseases like diarrhoea, cancer, stomach cramps, nausea, anaemia, kidney damage and even brain damage (Nkansah & Belford, 2017).

Substantial efforts have been made to remediate soils contaminated by heavy metals, including on-site management or encapsulation. However, neither of these methods solve decontamination but instead introduces secondary contamination at the dumpsite (Nkansah & Belford, 2017). In multivariate statistical analysis, Principal Component Analysis (PCA) and Cluster Analysis (CA) can be used to identify contamination sources, be they natural or anthropogenic (Dolezalova et al., 2015). In addition, the Geoaccumulation Index (Igeo) and Pollution Index (PI) can be used to classify the contamination status of the soils into low, moderate, strong and high contamination levels. The Lake Victoria goldfield has a long history of gold mining, and several mines are distributed along mineralized zones. Golden Pride Gold Mine (GPGM) and Geita Gold Mine (GGM) are among the gold mines in the Lake Victoria goldfield. Thus, it is essential to determine the source and level of pollution in these gold mines.

Numerous research has been published on heavy metal pollution and its sources in metalliferous soils (Bern et al., 2019; Chileshe et al., 2019; Chunhacherdchai et al., 2011; Demkova et al., 2017; Giri et al., 2017; Huang et al., 2017; Krishna et al., 2013; Masindi & Muedi, 2018; Wu et al., 2011; Yang et al., 2014). The studies reveal that mining activity is a chief contributor to soil contamination in mining areas. In Tanzania, different scholars assessed heavy metal pollution in gold mining areas (Gomezulu et al., 2018; Mganga et al., 2011; Mganga, 2014; Mkumbo, 2012; Mkumbo et al., 2012; Nkuli, 2008; Sangu, 2014). However, most of these studies have primarily focused on the phytoremediation of heavy metals, with only a limited assessment of the sources and levels of pollution in the mining areas. Therefore, this study aimed to assess sources and levels of soil heavy metal pollution with PCA, CA, PI and Igeo. The findings of this study provide valuable information to policymakers and mining operators that will help monitor and develop proper management strategies to reduce metal pollution in the mining areas. It is recommended that the mining companies put control measures to prevent pollution and ensure that heavy metal contaminated soil is monitored regularly.

2. Materials and Methods

2.1. Study Areas

This study was carried out at the GPGM and GGM in Tanzania (**Figure 1**). The GPGM is a closed mine located at Lusu Ward in Nzega District, Tabora Region. It is approximately 18 km north of the township of Nzega and 200 km south of the regional centre of Mwanza. The geographical location is between latitude 4˚5'0"S and 4˚47'0"S and longitude 33˚10'50"E and 33˚13'20"E. The altitude of the study area ranges from 1130 m to 1162 m. GPGM carried out operations between 1998 and 2013 and operated six pits and two waste rock dumps (Henckel et al., 2016). A special mining lease covered an area of 1064 ha. The Golden Pride gold deposit, characterized by sulfide and oxidized ore, is located in the central part of Tanzania in the Nzega greenstone belt on the southern margin of the Lake Victoria Goldfields (Vos et al., 2009).

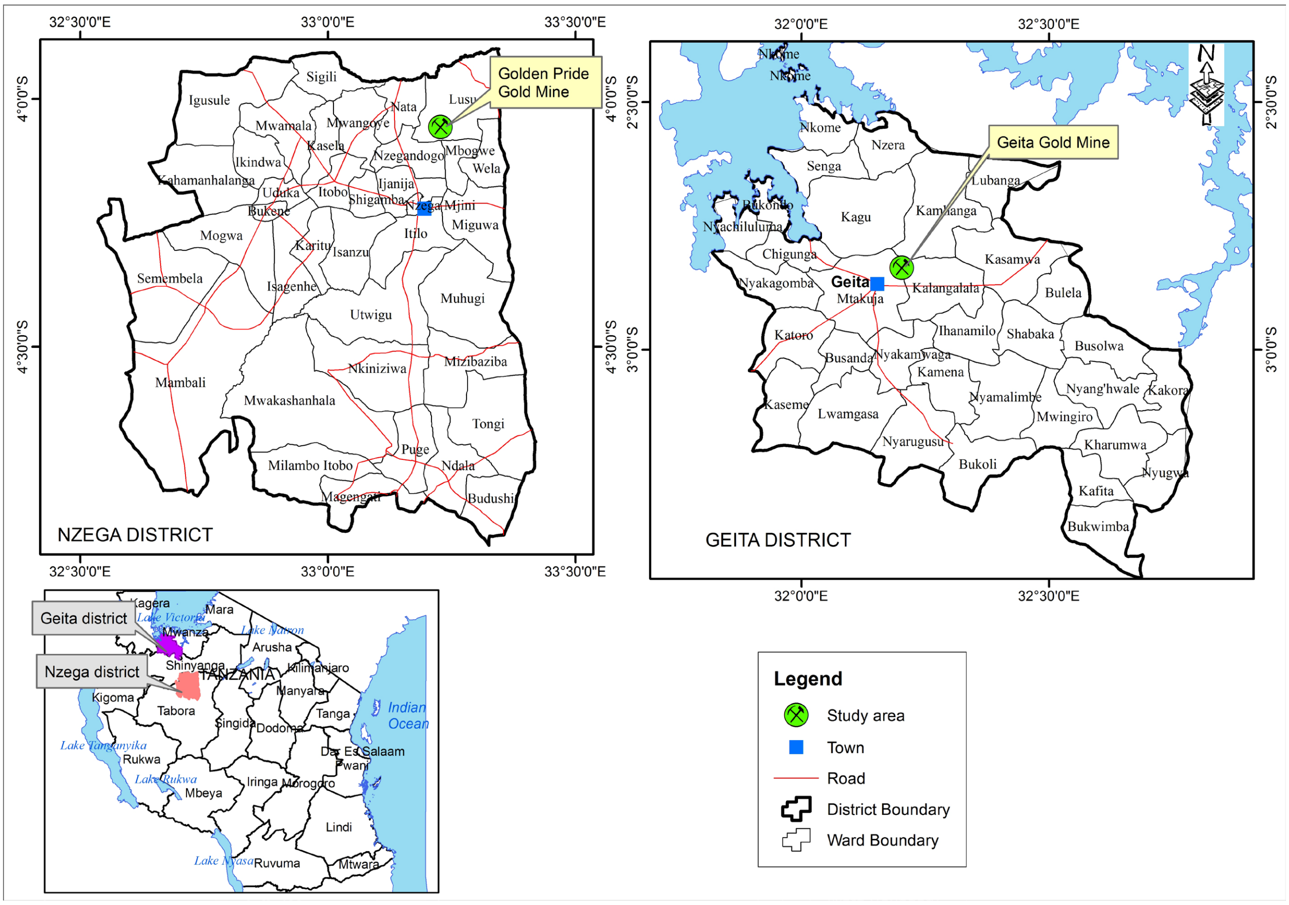

Figure 1. Map of Tanzania showing locations of Golden Pride and Geita Gold Mines (Kahangwa et al., 2020).

The GGM is an active mine located at Mtakuja and Kalangalanga Wards in Geita District, Geita Region, in the northern part of Tanzania (**Figure 1**). GGM is approximately 4 km west of Geita Town and 90 km southwest of Mwanza City in north-western Tanzania. The mine is situated at the headwaters of the Mtakuja River, which drains its water into Lake Victoria, approximately 20 km northwest of the mine (Sibilski & Stephen, 2010). The geographical location is between latitude 2˚50'38"S and 2˚52'36"S and longitude 32˚8'50"E and 32˚12'30"E. The altitude of the study area ranges from 1205 m to 1275 m. GGM mining operations began in 2000 and are expected to end in 2029 (Henckel et al., 2016; Stephen et al., 2017). The special mining lease covers an area of 19,627 ha. The GGM operates seven pits and seven waste rock dump stations. GGM is in the greenstone belt east of the Lake Victoria Goldfield, typically characterized by sulfide ore assemblages rich in gold and heavy metal deposits.

2.2. Data Collection

Soil sampling was conducted in December 2019 during a short rain. Three transect lines of 160 m each were established along the tailings storage facilities (TSF) of the GPGM and GGM (at the crest, middle and toe) with a distance between transects of 50 m. Three sampling plots of 20 × 20 square meter quadrats in each transect were established every 50 m. Thus, nine sampling points, three from each transect, were systematically established. Ninety (90) soil samples were collected from the studied sites. Soil samples were collected from four spots at a depth of 20 cm using a soil auger at each sampling point. The collected soil samples were mixed to constitute composite samples per sampling point. The mixed soil samples were kept in airtight plastic bags.

A mixed acid procedure was used to digest soil samples to determine total arsenic. Two grams (2 g) of each sample were heated gently with a mixture of 1 ml of conc. Sulphuric acid (H_2SO_4) and 10 ml of conc. Nitric acid (HNO_3). After dissolving most organic matter, the solution was only faintly yellow and 1 ml of perchloric acid ($HClO_4$) was added and the heat increased until dense fumes of sulphur (S) appeared. The resultant solutions were then used to quantify the As in soil samples using molybdenum blue methods described by Allen (1989). A colorometric procedure was based on molybdenum blue methods, and the quantification was done using a UV-VIS spectrophotometer. The solutions obtained were used to determine total Pb, Cr, Cd, Cu, Mn and Ni by Atomic Absorption Spectrometry (AAS, 240 Varian).

2.3. Calculation of Pollution and Geoaccumulation Indices

The pollution index was calculated using Equation (1).

$$\text{PI} = \frac{\text{Conc of metal in soils}}{\text{geochemical background conc}} \tag{1}$$

where; PI is the evaluation score corresponding to each sample, Ci is the measured concentration of the metal in the soils, and Si is the geochemical background concentration of the metals (Usman et al., 2012). Thus, the PI values of each metal contamination can be classified as either low contamination (PI ≤ 1.0), moderate contamination (1.0 > PI ≤ 3.0), high contamination (PI > 3.0). The background concentrations of heavy metals are presented in **Table 1**.

Index of geoaccumulation (Igeo) was calculated using Equation (2).

$$\text{Igeo} = \log\left(\frac{C_n}{1.5B_n}\right) \tag{2}$$

Table 1. Background concentrations of heavy metals (μg/g).

Heavy metals	The normal range in soil
Cd	0.03 - 0.3
Cr	10 - 200
Cu	5 - 80
Ni	5 - 500
Pb	2 - 20
Mn	200 - 2000
As	0.5 - 30

where; C_n is the measured concentration of an element in the samples; B_n is the background concentration, and the constant 1.5 allows analysis of the natural fluctuations in the content of a given substance in the environment and minimal anthropogenic influences. Igeo < 0 (0) = uncontaminated, Igeo 0 - 1 = Uncontaminated to moderately contaminated, Igeo 1 - 2 = Moderately contaminated, Igeo 2 - 3 = Moderately to strongly contaminated, Igeo 3 - 4 = Strongly contaminated, Igeo 4 - 5 = Strongly to extremely contaminated and Igeo > 5 = Extremely contaminated (Krishna et al., 2013; Wei et al., 2011).

2.4. Statistical Analysis

All data input and fundamental descriptive statistical analysis, including the mean, standard deviation, and coefficient of variation for various heavy metal variables, were computed using IBM SPSS Version 23 and PAST. The statistical significance of the Pearson correlation between the heavy metal content was determined by the bivariate correlation in IBM SPSS Version 23. PCA accomplished exploratory factor analysis of the elemental concentrations with varimax rotation to reduce data and increase the interpretability of the identified factors (Makupa, 2013). By applying Varimax rotation with Kaiser normalization, PCA was carried out to ascertain the possible contributing factors toward the elemental concentrations and thereby determine which elements have a common origin in soils. By extracting the eigenvalues and eigenvectors from the correlation matrix, the number of significant factors the percent of variance explained by each was calculated using IBM SPSS 23. PCA aids in optimizing the number and type of data that are best for carrying out the heavy metals contamination of the soil (Gergen & Harmanescu, 2012). A PCA using a correlation cross-product matrix was used to examine the grouping among soil heavy metals (Barona & Romero, 1996). CA using a hierarchical cluster was used to examine the classification of the heavy metals in the soil.

3. Results

3.1. Mean, Standard Deviation and Coefficient of Variation

The means, standard deviations and coefficient of variation of the variables are presented in **Table 2**. The coefficient of variation of As was lower than 20%, while for some others, like Cd, Cu, Mn and Ni were higher than 20% but lower than 30%, and Pb and Cr were higher than 30%.

3.2. Correlation Coefficients

The correlation coefficients for all the analysed elements are presented in **Table 3**. Results indicated that Pb levels were strong and positively correlated (p < 0.01) with Cr (r = 0.768), Cu (r = 0.673) and Mn (r = 0.557) but negatively correlated with As (r = −0.534). Furthermore, a strong and positive correlation was found between Cr and other elements such as Cu, Mn and Ni; Cu and Mn; and Mn and Ni. Positive correlations with significance (p < 0.05) were found between

Table 2. Mean, standard deviation (S.D.) and coefficient of variation (C.V.) of variables.

	Mean (mg/g)	S.D.	C.V
Pb	2.58	1.27	49.40
Cr	3.85	1.46	38.02
Cd	2.13	0.62	28.94
Cu	4.08	0.85	20.72
As	3.88	0.70	18.14
Mn	4.29	0.92	21.52
Ni	5.33	1.24	23.24

Table 3. Correlation coefficient between the soil variables.

		Pb	Cr	Cd	Cu	As	Mn	Ni
Pb	Pearson Correlation	1	0.768**	0.087	0.673**	−0.534**	0.557**	0.315*
	Sig. (2-tailed)		0.000	0.570	0.000	0.000	0.000	0.035
	N	45	45	45	45	45	45	45
Cr	Pearson Correlation	0.768**	1	0.061	0.803**	−0.601**	0.760**	0.552**
	Sig. (2-tailed)	0.000		0.689	0.000	0.000	0.000	0.000
	N	45	45	45	45	45	45	45
Cd	Pearson Correlation	0.087	0.061	1	−0.035	−0.285	−0.074	−0.124
	Sig. (2-tailed)	0.570	0.689		0.817	0.057	0.629	0.417
	N	45	45	45	45	45	45	45
Cu	Pearson Correlation	0.673**	0.803**	−0.035	1	−0.502**	0.598**	0.295*
	Sig. (2-tailed)	0.000	0.000	0.817		0.000	0.000	0.049
	N	45	45	45	45	45	45	45
As	Pearson Correlation	−0.534**	−0.601**	−0.285	−0.502**	1	−0.588**	−0.339*
	Sig. (2-tailed)	0.000	0.000	0.057	0.000		0.000	0.023
	N	45	45	45	45	45	45	45
Mn	Pearson Correlation	0.557**	0.760**	−0.074	0.598**	−0.588**	1	0.588**
	Sig. (2-tailed)	0.000	0.000	0.629	0.000	0.000		0.000
	N	45	45	45	45	45	45	45
Ni	Pearson Correlation	0.315*	0.552**	−0.124	0.295*	−0.339*	0.588**	1
	Sig. (2-tailed)	0.035	0.000	0.417	0.049	0.023	0.000	
	N	45	45	45	45	45	45	45

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Pb-Ni and Cu-Ni. In the present study, Cd has a weak positive correlation with Pb and Cr and is negatively correlated with Cu, As, Mn and Ni. Moreover, As has a negative correlation with all measured heavy metals.

3.3. Principal Component Analysis

The results in Table 4 show two factors whose eigenvalues are >1 suggesting a two-factor solution. Thus, these two factors were selected for further analysis. Other small eigenvalues of <1 were not used to obtain a probable number of contributing source factors. This indicates that the first and second components (PC1 and PC2) are the right choices. The first principal component (PC1) contains the greatest amount of variance (55.41%), including Pb, Cr, Cu, As, Mn and Ni, indicating the anthropogenic source of contaminants. The second principal component (PC2) accounts for 17.07% of the variance and has high Cd loadings, suggesting the dominance of both natural and artificial contaminants. It can be seen that the cumulative variance contribution of the first two principal components has been 72.48%, which can explain why the total variance is 72%. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) of 0.777 and Bartlett's test ($p < 0.01$) demonstrated that the PCA approach was valid.

Results show that the most important variables for the first PC1 were Pb, Cr, Cu, As, Mn and Ni (Table 5). The principal component loading shows seven

Table 4. Total variance explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.88	55.46	55.46	3.88	55.46	55.46	3.88	55.41	55.41
2	1.19	17.03	72.48	1.19	17.03	72.48	1.2	17.07	72.48
3	0.78	11.18	83.66						
4	0.45	6.37	90.03						
5	0.33	4.65	94.68						
6	0.27	3.81	98.49						
7	0.11	1.51	100						

Table 5. Factor extraction and component loading.

Rotated Component Matrixa			
	Component Loading		
	1	2	Communality
Pb	0.81	0.15	0.678
Cr	0.94	0.02	0.891
Cd	0.03	0.93	0.858
Cu	0.82	0.01	0.675
As	−0.73	−0.39	0.684
Mn	0.86	−0.16	0.760
Ni	0.63	−0.37	0.528

Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization; [a]Rotation converged in 3 iterations.

components, then extracted and rotated to two components (Table 5). So, the first PC is defined by soil contamination by six heavy metals: Pb, Cr, Cu, As, Mn, and Ni. The results also indicated that the particular variables have higher communalities, implying that the extracted components can explain much of the variance in the variables.

Based on the overall results of different analyses, PCA proved to be a beneficial method for identifying the most influential variables and quickly pointing out the relationship among the elements in the first two principal components, as illustrated in the score plot in Figure 2. The plot shows the location of the objects in the multivariate space of the first and second principal component score vectors. The plot of the heavy metals encountered showed that most of the elements were found clustered together, except for Cd, As and Ni, which clearly distinguished themselves from others. In particular, the heavy metals Pb, Cr, Cu and Mn placed to the right in the loading plot are close together and, therefore, positively correlated, indicating the common source of contamination. Ni deviated slightly from other heavy metals in PC1. As is placed to the left in the loading plot, it is negatively correlated with other heavy metals in the first PC. Finally, the PC2 is characterized by one variable, Cd, placed far right in the loading plot. Thus, As, Cd and Ni are from the same source of contamination.

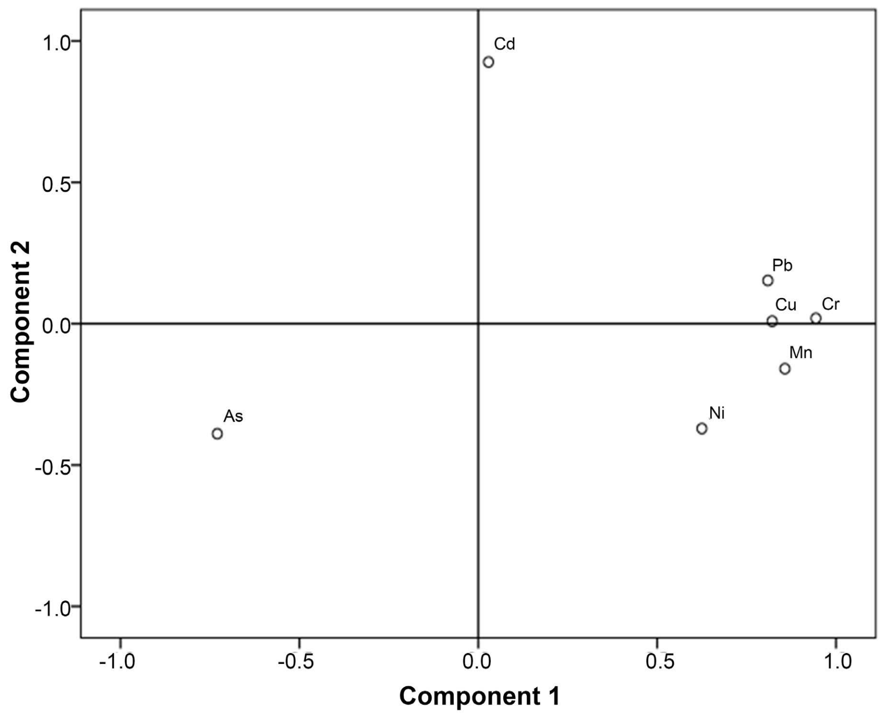

Figure 2. Plot of the first and second PC loading vectors Lead (Pb), Chromium (Cr), Cadmium (Cd), Copper (Cu), Arsenic (As), Manganese (Mn), Nickel (Ni).

3.4. Hierarchical Cluster Analysis

The results of the hierarchical cluster analysis are presented in the dendrogram. This directly reflects the correlation between soil metals, revealing the sources of heavy metal contaminants in soil. Variables are grouped into clusters using hierarchical cluster analysis with the average linkage between groups and the absolute values of the correlation matrix as a similarity measure. Strongly interrelated variables clustered together, regardless of the positive or negative sign of the relationship. For example, **Figure 3** displays two clusters: (C1) As, Cd, and Ni, and (C2) Pb, Cu, Cr and Mn.

3.5. Pollution Index (PI)

Based on the pollution index (PI), the soil contaminated with heavy metals is highly contaminated with Pb, Cr, Cd, Cu, As, and Ni. On the other hand, Mn had low contamination (**Table 6**). However, Cd, with a pollution index above Pb, Cr, Cu, As, Mn and Ni, is a chief contamination element of soil in the studied area.

3.6. Index of Geoaccumulation (Igeo)

Igeo indicated that the soils in the study area fall into moderately to highly contaminated soils (**Table 7**). Specifically, the results showed that the mean geoaccumulation index for Cd (3.035) showed moderately to strongly contaminated surface soils. As (4.485) and Cu (4.983) were strongly to extremely contaminated.

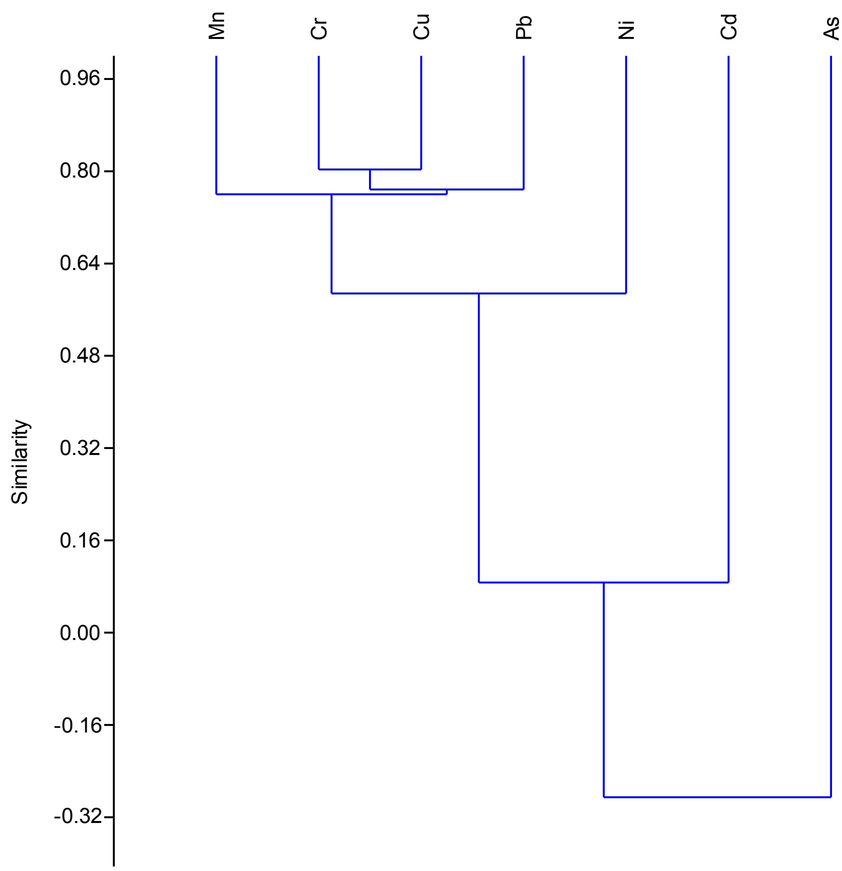

Figure 3. Dendrogram using average linkage (between groups) obtained by hierarchical cluster analysis.

Table 6. Pollution index for classification of soil contaminated with heavy metals.

Metal	PI	Classification
Pb	6.96	High contamination
Cr	3.33	High contamination
Cd	731.48	High contamination
Cu	17.42	High contamination
As	7.70	High contamination
Mn	1.00	Low contamination
Ni	5.41	High contamination

Table 7. Mean geoaccumulation index and standard deviation.

		Pb	Cr	Cd	Cu	As	Mn	Ni
N	Valid	45	45	45	45	45	45	45
	Missing	0	0	0	0	0	0	0
Mean		5.1130	5.3786	3.0347	4.9827	4.4838	7.0486	5.0803
Std. Deviation		0.21306	0.16430	0.13828	0.08708	0.08825	0.09344	0.12161

In the cases of Pb (5.133), Cr (5.379), Mn (7.049), and Ni (5.080), Igeo indicated extremely contaminated soils. Nevertheless, the Igeo value increased in the following order: Cd (3.035), As (4.485), Cu (4.983), Ni (5.080), Pb (5.133), Cr (5.379), and Mn (7.049) (**Table 7**).

4. Discussion

In this study, the strong positive correlation among Cu, Mn, Ni, Pb and Cr could be explained by the dependence of these heavy metals on one another. According to Krishna et al. (2013) the positive correlation among heavy metals indicates common anthropogenic pollution levels and sources. Anju & Banerjee (2012) also observed a significant correlation between Cu-Mn, Cu-Ni, Cu-Pb, Mn-Ni, Mn-Pb and Ni-Pb in a multivariate statistical analysis of the heavy metals in the soils of the Pb-Zn mining area, India.

The results show that Cd had a weak positive correlation with Pb and Cr and a negative correlation with Cu, As, Mn and Ni implying that the level of Cd slightly depends on the level of Pb and Cr, indicating the same processes on their levels in the soil. However, the level of Cd does not depend on the level of Cu, As, Mn and Ni indicating different processes are responsible for their levels in the study mines. The results contradict those of Anju & Banerjee (2012), who stated clearly that there is a positive correlation between Cd-Cu, Cd-Mn and Cd-Ni. Also, the study conducted by Benavides et al. (2005) on cadmium toxicity indicated that metals such as Pb, Cd and Cu eventually delivered from a growing number of anthropogenic activities, including mining operations.

The study also indicated negative correlations between As-Pb, As-Cr, As-Cd, As-Cu, As-Ni and As-Mn. Ma et al. (2015) observed similar correlations in heavy metal contamination of agricultural soils affected by mining activities around the Ganxi River in Chenzhou, Southern China. In addition, there is a significant negative relationship between Ni and As. The study negates the study conducted by Krishna et al. (2013) that found a strong correlation between As and Pb in the soil around the mining area.

According to the initial eigenvalue results, PCAs were considered and grouped into two component models, accounting for 72.48% of the total variance. The greatest variance (55.41%) indicated that Pb, Cr, Cu, As, Mn and Ni were strongly associated with PC1, which originated from mining and related activities. The results in PC1 are in line with the study conducted by Fernandez-Caliani et al. (2009) on the heavy metal pollution in soils around the abandoned mine sites of Iberian, showing that the Pyrite belt has strong loadings of Pb, Cu, As and Ni in the single factor, indicating mining is the source of contamination. The results are also in line with the study conducted by Gao et al. (2017) on the spatial distribution and accumulation characteristics of heavy metals in steppe soils around three mining areas in Xilinhot, Inner Mongolia, China, indicating that the PC1 mainly reflects the enrichment information of Cr, Cu, and Ni, and the contribution rate is 46.28%. On the other hand, Cd is distributed in PC2 originating from local natural and anthropogenic sources. The results of PC2 contradict the studies conducted by Fernandez-Caliani et al. (2009) and Gao et al. (2017) that showed Cd is not from the same source as other elements.

The cumulative variance contribution of the first two principal components has been 72.48, which can explain the total variance of 72.48%. These results align with Gao et al. (2017) that, in the gold mine area, all the heavy metals in the soil are extracted as two main components, with a cumulative contribution of 79.99%. The elements in PC1 mainly come from anthropogenic sources, such as mining and its associated activities. However, Cd, As and Ni deviated from other elements in the score plot, indicating a different source of pollution. The results also showed two clusters. Cluster 1 probably suggests a common source of contamination and cluster 2 may suggest different sources of contamination. Thus, the CA results in gold mining areas are almost consistent with what was observed in the PCA.

The present study indicated that the soils in the studied site were moderately to highly contaminated by Pb, Cr, Cd, Cu, As, Mn and Ni, as evidenced by PI and Igeo. The main contaminated components of soil in the examined areas are Pb, Cr, Mn, and Ni, which have Igeo that are essentially higher than Cd, As, and Cu. These findings contradict Wei et al. (2011), who found that Igeo of Cu and Cd are the most common contaminants, followed by Zn, Ni, Pb, and Cr.

5. Conclusion

In this study, the PCA and CA provided that Pb, Cr, Cu, and Mn come from different sources of pollution than Cd, As and Ni. Therefore, the study concludes that soil contamination results from both natural and anthropogenic processes, as evidenced by PCA and CA. The PI and Igeo confirmed that soils in the mining areas were contaminated in a range of moderately through strongly to highly contaminated soils. Thus, PCA, CA, Igeo and PI are essential in assessing and monitoring heavy metal contaminated soil.

Acknowledgements

The author acknowledges anonymous reviewers for their valuable comments and constructive criticism in shaping the article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The author declares to have no conflicts of interest regarding the publication of this paper.

References

Allen, S. E. (1989). *Chemical Analysis of Ecological Materials*. Blackwell Scientific Publications.

Anju, M., & Banerjee, D. K. (2012). Multivariate Statistical Analysis of Heavy Metals in Soils of a Pb-Zn Mining Area, India. *Environmental Monitoring and Assessment, 18,* 4191-4206. https://doi.org/10.1007/s10661-011-2255-8

Barona, A., & Romero, F. (1996). Distribution of Metals in Soils and Relationships among Fractions by Principal Component Analysis. *Soil Technology, 8,* 303-319. https://doi.org/10.1016/0933-3630(95)00029-1

Benavides, M., Gallego, S., & Tomaro, M. (2005). Cadmium Toxicity in Plants. *Brazilian Journal of Plant Physiology, 17,* 21-34. https://doi.org/10.1590/S1677-04202005000100003

Bern, C. R., Walton-Day, K., & Naftz, D. L. (2019). Improved Enrichment Factor Calculations through Principal Component Analysis: Examples from Soils near Breccia Pipe Uranium Mines, Arizona, USA. *Environmental Pollution, 248,* 90-100. https://doi.org/10.1016/j.envpol.2019.01.122

Chileshe, M. N., Syampungani, S., & Sandell, E. (2019). Physico-Chemical Characteristics and Heavy Metal Concentrations of Copper Mine Wastes in Zambia: Implications for Pollution Risk and Restoration. *Journal for Restoration,* 1-11.

Chunhacherdchai, L., Chotpantarat, S., & Tongcumpou, C. (2011). Investigating of Heavy Metals in Different Depths of Soil Tailings from Akara Gold Mine, Thailand Using Three-Steps Modified BCR Sequential Extraction. In *2nd International Conference on Environmental Science and Technology* (pp. 28-31).

Demkova, L., Jezny, T., & Bobulska, L. (2017). Assessment of Soil Heavy Metal Pollution in a Former Mining Area—Before and after the End of Mining Activities. *Soil and Water Research, 12,* 229-236. https://doi.org/10.17221/107/2016-SWR

Dolezalova, H., Pavlovský, J., & Chovanec, P. (2015). Heavy Metal Contaminations of Urban Soils in Ostrava, Czech Republic: Assessment of Metal Pollution and Using

Principal Component Analysis. *International Journal of Environmental Research, 9,* 683-696.

Fashola, M. O., Ngole-Jeme, V. M., & Babalola, O. O. (2016). Heavy Metal Pollution from Gold Mines: Environmental Effects and Bacterial Strategies for Resistance. *International Journal of Environmental Research and Public Health, 13,* 1047. https://doi.org/10.3390/ijerph13111047

Fazekasov, D., & Fazekaš, J. (2020). Soil Quality and Heavy Metal Pollution Assessment of Iron Ore Mines in Nizna Slana (Slovakia). *Sustainability, 12,* 2549. https://doi.org/10.3390/su12062549

Fernandez-Caliani, J. C., Barba-Brioso, C., González, I., & Galán, E. (2009). Heavy Metal Pollution in Soils around the Abandoned Mine Sites of the Iberian Pyrite Belt (Southwest Spain). *Water, Air, and Soil Pollution, 200,* 211-226. https://doi.org/10.1007/s11270-008-9905-7

Gao, Y., Liu, H., & Liu, G. (2017). The Spatial Distribution and Accumulation Characteristics of Heavy Metals in Steppe Soils around Three Mining Areas in Xilinhot in Inner Mongolia, China. *Environmental Science and Pollution Research, 24,* 25416-25430. https://doi.org/10.1007/s11356-017-0113-0

Gergen, I., & Harmanescu, M. (2012). Application of Principal Component Analysis in the Pollution Assessment with Heavy Metals of Vegetable Food Chain in the Old Mining Areas. *Chemistry Central Journal, 6,* 1-13. https://doi.org/10.1186/1752-153X-6-156

Giri, S., Singh, A. K., & Mahato, M. K. (2017). Metal Contamination of Agricultural Soils in the Copper Mining Areas of Singhbhum Shear Zone in India. *Journal of Earth System Science, 126,* 1-13. https://doi.org/10.1007/s12040-017-0833-z

Gomezulu, E. S., Mwakaje, A., & Katima, H. Y. (2018). Heavy Metals and Cyanide Distribution in the Villages Surrounding Buzwagi Gold Mine in Tanzania. *Tanzania Journal of Science, 44,* 107-122.

Henckel, J., Poulsen, K. H., Sharp, T., & Spora, P. (2016). Lake Victoria Goldfields. *Episodes, 39,* 135-154. https://doi.org/10.18814/epiiugs/2016/v39i2/95772

Huang, S., Yuan, C., Li, Q., Yang, Y., Tang, C., Ouyang, K., & Wang, B. (2017). Distribution and Risk Assessment of Heavy Metals in Soils from a Typical Pb-Zn Mining Area. *Pollution Journal of Environmental Studies, 26,* 1105-1112. https://doi.org/10.15244/pjoes/68424

Kahangwa, C., Nahonyo, C., & Sangu, G. (2020). Monitoring Land Cover Change Using Remote Sensing (RS) and Geographical Information System (GIS): A Case of Golden Pride and Geita Gold Mines, Tanzania. *Journal of Geographic Information System, 12,* 387-410. https://doi.org/10.4236/jgis.2020.125024

Krishna, A. K., Mohan, K. R., Murthy, N. N., Periasamy, V., Bipinkumar, G., Manohar, K., & Rao, S. S. (2013). Assessment of Heavy Metal Contamination in Soils around Chromite Mining Areas, Nuggihalli, Karnataka, India. *Environmental Earth Sciences, 70,* 699-708. https://doi.org/10.1007/s12665-012-2153-6

Ma, L., Sun, J., Yang, Z., & Wang, L. (2015). Heavy Metal Contamination of Agricultural Soils Affected by Mining Activities around the Ganxi River in Chenzhou, Southern China. *Environmental Monitoring and Assessment, 187,* 1-9. https://doi.org/10.1007/s10661-015-4966-8

Mahugija, J. A. M., & Sheikh, H. M. (2018). Status of Selected Heavy Metals Dispersion from Topsoil in and around Automobile Workshop Areas in Zanzibar Municipality, Tanzania. *Tanzania Journal of Science, 44,* 12-23.

Makupa, E. E. (2013). *Conservation Efforts and Local Livelihoods in Western Serengeti,*

Tanzania: Experiences from Ikona Community Wildlife Management Area. Department of Geography, University of Victoria.
http://ir.obihiro.ac.jp/dspace/handle/10322/3933

Masindi, V., & Muedi, K. L. (2018). Environmental Contamination by Heavy Metals. In *Heavy Metals* (pp. 115-133). IntechOpen.
https://doi.org/10.5772/intechopen.76082

Mganga, N. D. (2014). The Potential of Bioaccumulation and Translocation of Heavy Metals in Plant Species Growing around the Tailing Dam in Tanzania. *International Journal of Science and Technology (IJST), 3,* 690-697.

Mganga, N., Manoko, M. L. K., & Rulangaranga, Z. K. (2011). Classification of Plants According to Their Heavy Metal Content around North Mara Gold Mine, Tanzania: Implication for Phytoremediation. *Tanzania Journal of Science, 37,* 109-119.

Mkumbo, S. (2012). *Development of a Low Cost Remediation Method for Heavy Metal Polluted Soil*. TRITA LWR Licentiate, Thesis 2067.

Mkumbo, S., Mwegoha, W., & Renman, G. (2012). Assessment of the Phytoremediation Potential for Pb, Zn and Cu of Indigenous Plants Growing in a Gold Mining Area in Tanzania. *International Journal of Environmental Sciences, 2,* 2425-2434.

Nkansah, F. K., & Belford, E. J. D. (2017). Heavy Metals Accumulation by Indigenous Plants Growing in Contaminated Soil in a Gold Mining Area in Ghana. *Journal of Natural Sciences Research, 7,* 40-41.

Nkuli, G. (2008). *Effects of Mining Activities at Bulyanhulu Gold Mine (BGM) on the Water Quality of Bulyanhulu River: Shinyanga-Tanzania*. University of Zimbabwe.
https://www.waternetonline.org

Sangu, G. (2014). *Study on the Phytoremediation of Arsenic (As) Contaminated Soils in Geita*. PhD Thesis, University of Dar es Salaam.

Sibilski, U., & Stephen, R. (2010). ARD Management at Geita Gold Mine. In *International Conference on Acid Rock Drainage,* (pp. 2000-2010).

Stephen, R., Msalangi, C., Schanknecht, A., & Witcomb, A. (2017). *Geita Gold Mine: Statutory Mine Closure Plan (Issue 7)*.

Usman, A. R., Soo Lee, S., Awad, Y. M., Jae Lim, K., Yang, J. E., & Sik Ok, Y. (2012). Soil Pollution Assessment and Identification of Hyperaccumulating Plants in Chromated Copper Arsenate (CCA) Contaminated Sites, Korea. *Chemosphere, 87,* 872-878.
https://doi.org/10.1016/j.chemosphere.2012.01.028

Vos, I. M. A., Bierlein, F. P., Standing, J. S., & Davidson, G. (2009). The Geology and Mineralisation at the Golden pride Gold Deposit, Nzega Greenstone Belt, Tanzania. *Mineralium Deposita, 44,* 751-764. https://doi.org/10.1007/s00126-009-0245-3

Wei, Z., Wang, D., Zhou, H., & Qi, Z. (2011). Assessment of Soil Heavy Metal Pollution with Principal Component Analysis and Geoaccumulation Index. *Procedia Environmental Sciences, 10,* 1946-1952. https://doi.org/10.1016/j.proenv.2011.09.305

Wu, Y., Xu, Y., Zhang, J., Hu, S., & Liu, K. (2011). Heavy Metals Pollution and the Identification of Their Sources in Soil over Xiaoqinling Gold-Mining Region, Shaanxi, China. *Environmental Earth Sciences, 64,* 1585-1592.
https://doi.org/10.1007/s12665-010-0833-7

Yang, P., Yang, M., Mao, R., & Shao, H. (2014). Multivariate-Statistical Assessment of Heavy Metals for Agricultural Soils in Northern China. *The Scientific World Journal, 2014,* Article ID: 517020. https://doi.org/10.1155/2014/517020