Journal of Hazardous Materials Advances 18 (2025) 100688

Contents lists available at ScienceDirect

Journal of Hazardous Materials Advances

journal homepage: www.elsevier.com/locate/hazadv

Patterns and source apportionment of potentially toxic elements distribution in the soils of the Nangodi area, Northeast Ghana: A multivariate and machine learning approach

Raymond Webrah Kazapoe [a], Daniel Kwayisi [b,*], Seidu Alidu [c], Obed Fiifi Fynn [d], Samuel Dzidefo Sagoe [e], Ebenezer Ebo Yahans Amuah [f], Emmanuel Nyavor [b]

[a] *Department of Geological Engineering, University for Development Studies, Nyankpala, Ghana*
[b] *Department of Earth Science, University of Ghana, Legon-Accra, Ghana*
[c] *Ghana Geological Survey Authority, P.O. Box M80, Accra, Ghana*
[d] *Department of Geological Sciences, University of Energy and Natural Resources, Sunyani, Ghana*
[e] *Department of Environment and Sustainability Sciences, University for Development Studies, Ghana*
[f] *Department of Civil Engineering, Takoradi Technical University, P. O. Box 256, Takoradi, Ghana*

ARTICLE INFO

Keywords:
Mining
Pollution
Self organising maps
Pollution indices
Ecological risk

ABSTRACT

This study assessed the sources, distribution and pollution status of heavy metals in the Nangodi area of Northwestern Ghana. Cr (120.86 mg/kg) and Co (30.92 mg/kg) had respective average values of 2.4 and 1.2 times higher than their Continental Crustal Averages of 100 mg/kg and 25 mg/kg. The Potential Toxic Elements (PTE) displayed a decreasing trend in the order Ba > Cr > V > Sr > Cu > Zn > Co > Mo. The Metal Index assessment highlighted the significant effect of galamsey on the soil health of the area. The samples were ranked as slightly (26.45 %), moderately (25.18 %), Strongly (21.20 %) and seriously (23.91 %) affected. The positive Matrix Factorization identified three Factors as controlling PTEs in the area. Factor 1/anthropogenic (V = 84 %, Cu = 84 %, Co = 75.5 % and Zn = 58.9 %). Factor 2/geogenic (Ba = 87.5 %, Sr = 83.1 %, Pb = 57.8 %). Factor 3/mixed source (Cr = 91.8 % and Mo = 43.4 %). The Pearson correlation matrix outlined two groups of PTEs; (1) PTEs with moderate to strong correlation (V, Co, Cu and Zn) and (2) PTEs with weak to moderate correlation (Sr, Mo, Ba and Pb). The first group occurs at the southwestern boundary of the study area, reflecting the influence of local geology and mining practices on the levels of potentially toxic elements (PTEs) in the soil. The Self Organising Map (SOM) identified three higher concentration clusters, V, Zn, Cu, and Co, inferred to be the mining activities. Geogenic-sourced Sr and Ba are located centrally. Pb, Mo, and Cr show distinct distributions, suggesting mixed factors affecting their spread. The study identified systematic heavy metal pollution, which could pose a deleterious risk to the environment and inhabitants of the area.

1. Introduction

Soil is a very important element in the delivery of ecosystem services to human beings and their environment. Because soil is a vital resource that facilitates industrialisation, urbanisation, and intensive agricultural production, its sustainable use has become an issue of economic and environmental concern (Zahoor et al., 2022; Amuah et al., 2024). Environmental degradation, particularly the release of potentially harmful heavy metals, has become a growing concern (Ukaogo et al., 2020; Verma et al., 2021). Ghana's fast urbanisation in recent years has increased the risk of soil pollution, which is frequently related to poor environmental management techniques. The nation is heavily exposed to pollution containing heavy metals as a result of energy production, transportation, urbanisation, industrialisation, and mechanised agriculture (Cobbinah et al., 2017; Songsore,2020; Kazapoe et al., 2024; Obiri-Nyako et al., 2024). Mining (especially illegal mining known as Galamsey in Ghana) is one of the most significant anthropogenic activities that alter soil chemistry, contributing to elevated concentrations of heavy metals in affected areas (R.W. Kazapoe et al., 2021).

Studies conducted in the southern mining district (Bempah and Ewusi, 2016; Gyamfi et al., 2021; Arhin et al., 2019) and some parts of Northern Ghana (Arhin et al., 2017; Darko et al., 2019; Akoto et al.,

---

* Corresponding author.
E-mail address: dkwayisi@ug.edu.gh (D. Kwayisi).

https://doi.org/10.1016/j.hazadv.2025.100688
Received 31 December 2024; Received in revised form 13 March 2025; Accepted 14 March 2025

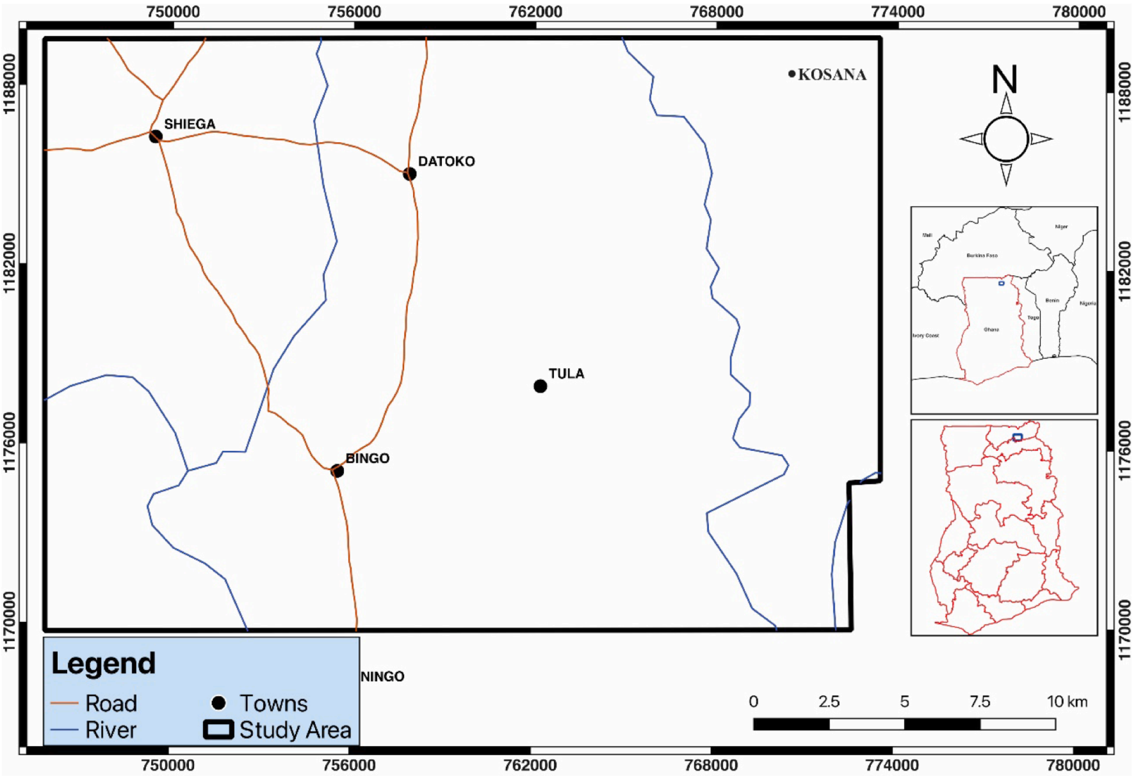

Fig. 1. Map of the study area.

2023) have revealed elevated elemental concentrations in the soils in these communities. In the last decade, there has been significant growth in mining companies and illegal small-scale mining (galamsey) in the northern region of Ghana, including the study area. Additionally, population growth has dramatically increased new settlement formation zones. This has led to the transformation of some wastelands, especially former mining zones into arable land and living villages (Owusu-Nimo et al., 2018). This is a public health concern within the affected population as they come into direct or indirect contact with contaminated soils and mining tailings. It is, therefore, vital for metal and metalloid pollution levels, sources and risks in agricultural soils, as well as their effects on human health, to be closely monitored and assessed. Besides, knowing the sources of the metals and metalloids in the soils is important in managing soil pollution.

To address this situation, the use of soil pollution indices has been adopted globally to evaluate soil contamination level (Tang et al., 2019; Kowalska et al., 2018). Thakkar et al. (2024) leveraged pollution indices in the assessment of heavy metals in soils and the ecological risk associated with it. Different techniques include Pearson's correlation matrix (PCM), principal component analysis (PCA), self-organizing maps (SOMs) and hierarchical cluster analysis (HCA) for source identification while receptor models like chemical mass balance (CMB) and positive matrix factorization (PMF) popularised by the United States Environmental Protection Agency (USEPA) are used for quantifying the contribution of sources of metals in soils (Kim et al., 2020; Chakraborty et al., 2023). Of these, PMF is preferred because of its reliability, and it is established that an ensemble of models provides better outcomes (Uranishi et al., 2017; Lv et al., 2021). In addition, the geostatistical analysis assisted by GIS helps in determining the distribution of metals and mapping the areas affected by contamination (Arhin et al., 2019; Khan et al., 2022).

This study focused on soils in Nangodi and its environs in Upper East Region, Ghana, to evaluate the environmental quality of the area with respect to heavy metal pollution. The main objectives are to determine (i) concentrations of heavy metal pollutants, (ii) sources and interactions of these pollutants by applying machine learning and multivariate statistical techniques, (iii) spatial distribution of the heavy metals, and (iv) ecological risk related to the heavy metals. The study addresses a critical knowledge gap by integrating machine learning techniques, such as Self-Organizing Maps (SOM) and Positive Matrix Factorization (PMF), with traditional multivariate statistical methods to comprehensively assess the sources, distribution, and ecological risks of potentially toxic elements (PTEs). While previous studies have highlighted heavy metal contamination in Ghana's mining regions, few have systematically combined these advanced analytical approaches to differentiate between anthropogenic and geogenic pollution sources. This study's innovation lies in its use of a multi-method approach to enhance the accuracy of pollution source apportionment and ecological risk assessment, providing a more robust framework for environmental monitoring and policy formulation.

2. Materials and methods

2.1. Area of study

The Nangodi area (Fig. 1) is located in the Talensi-Nabdam district of the Upper East Region of Ghana, with a population of 94,650 (Ghana Statistical Service, 2010). It borders Bolgatanga District, West Mamprusi District, Kassena-Nankana District, and Bawku West District. The district lies between latitudes 10°35′00″N and 10°50′00″N and longitudes 0°30′00″W and 0°45′00″W. The economy is primarily agricultural, with 85.9 % of households engaged in crop cultivation, animal rearing, and tree planting. About 49.3 % of agricultural workers are male, and 50.7 % are female (Ministry of Food and Agriculture, n.d.). The discovery of gold has led to an influx of illegal mining activities, affecting the area (Tom-Dery et al., 2012; R.W. Kazapoe et al., 2021).

The topography is generally low-lying, with elevations ranging from 100 m to 300 m above sea level. The area is drained by the White Volta, Red Volta, Ayedama, and Kulumasa rivers, with the Vea dam and nine smaller dams providing water. The climate is semi-arid with a single

Table 1
Summary statistics of PTEs in the study area.

Elements	V	Cr	Co	Cu	Zn	Sr	Mo	Ba	Pb
Min	9.45	11.6	3.3	3.25	9.8	6.9	0.4	40.5	1.7
Max	528	1294	311.8	249.9	228.2	567	14.2	2735	35.8
Mean	120.86	242.42	30.92	43.8	35.42	98.17	0.95	337.62	10.07
Standard Deviation (SD)	106.11	229.65	26.94	43.77	21.58	90.5	1.38	265.02	4.4
Skewness	1.6	1.65	3.8	1.86	2.51	2.1	5.94	2.64	1.26
Kurtosis	1.85	2.79	26.04	3.2	12.76	4.83	42.46	13.81	3.49
Coefficient of Variation (CV%)	87.79	94.74	87.13	102.45	60.93	92.19	144.66	78.50	43.71
The number of Samples exceeding CCA	153	351	215	137	46	12	67	146	145
Percentage of Samples exceeding CCA (%)	27.72	63.59	38.95	24.82	8.33	2.17	12.14	26.45	26.27
CCA (Taylor, 1965)	135	100	25	55	70	375	1.5	425	12.5
*World**		100		30	50				42.5
*EU**		75			1				
*USEPA**		11		270	1100				200
Arhin and Kazapoe (2017)		80.74	6.44						6.99
McLaughlin et al. (2000)	200	100		100					150
Crommentuijn et al. (2000)	1.1	3.8	24	3.5				9	55

rainy season from May to October, receiving between 600 mm and 1400 mm of rainfall annually. Temperatures range from 33 °C to 42 °C in March-April, and from 12 °C to 26.5 °C during the dry season (Dickson & Benneh, 1988).

The area falls within the Sudan savannah zone, with vegetation consisting of short deciduous trees and grasses. Economic trees include Shea, dawadawa, baobab, and acacia. The district has three gazetted forest reserves, including the Red Volta Forest Reserve, covering 455.21 sq. km (Ministry of Food and Agriculture, n.d.).

The Nangodi area is part of the Paleoproterozoic Birimian gold-bearing belt, which includes greenstones, granitoids, and metasedimentary rocks (Sakyi et al., 2019; 2024; Kazapoe, 2014).

2.2. Soil sampling

A total of 552 soil samples were systematically collected from the Nangodi area using a geographic coordinate system to ensure uniform spatial distribution. The study area was divided into 50 equal grids using ArcGIS, with 11 samples randomly selected from each grid to maintain representativeness. Additionally, from the 25th and 50th grids, an extra sample was collected for quality control. Sampling was conducted by digging up to a depth of 20 cm from the surface using soil augers and shovels. To minimize contamination, non-soil materials such as stones, roots, and plant debris were carefully sieved out in the field. The cleaned soil samples were then securely contained in Ziploc bags, properly labelled, and transported for laboratory analysis. The data used in this study was obtained in September 2024.

2.3. Analytical procedures

Composite samples were prepared by mixing three homogenized samples from each sampling point. The samples were washed, dried, and sieved to <2 mm in the field, with a further 75-micron sieve at the Ghana Geological Survey Authority (GGSA). A 50 g portion of sieved soil was ground to a fine powder using a mechanical agate mortar and pestle, then sealed in containers for further analysis. Elemental determination was performed using an Energy Dispersive X-ray Fluorescence (ED-XRF) spectrometer, calibrated with standard samples. Approximately 10 grams of each soil sample were pressed into pellets under 10,000 kN for one minute, then analyzed in triplicate to minimize errors.

The spectrometer operated at 50 kV excitation voltage, 1 mA current, and 300 seconds counting time. Blank measurements and control samples were included to detect contamination and ensure accuracy. Calibration was performed regularly with Certified Reference Materials (CRMs), and XRF performance was monitored by comparing results with certified values. The relative error for most elements was within 5 %, which was deemed acceptable for the study. The methodology followed Kazapoe et al. (2022) and Amuah et al. (2022).

2.4. Statistical analysis

A multivariate statistical approach was used to identify the source of elemental enrichments in the soil. Principal Component Analysis (PCA), Factor Analysis (FA), and Hierarchical Cluster Analysis (HCA) were applied to determine the source of element concentrations. For FA, factors with extraction commonality above 0.5 (50 %) or common variances <0.5 were selected, reflecting 13 out of 17 variables (Reimann et al., 2002). This method has been successfully used in similar studies (Temple, 1978; Subbarao et al., 1996; Arhin and Kazapoe, 2017). Prior to analysis, the concentration data were log-transformed after testing for normality using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Descriptive data for elements like Cu, Pb, Zn, Ag, Ni, Co, Mn, As, Cr, Ba, V, and Sr are provided in Table 1, which compares the variance of components from regulated baselines in Ghana, Africa, and globally. ArcGIS Pro was used to analyze the spatial distribution of the heavy metals.

2.5. Assessment of heavy metal pollution in the soil

Two (2) indices were analysed to assess the extent of heavy metal pollution in the study area (Table 1). These included the Metal Index and the Integrated Potential Ecological Risk Index.

2.5.1. Potential ecological risk index (RI)

Introduced by (Hakanson, 1980), RI is a widely used method for assessing the environmental impact of heavy metal contamination in various ecosystems. It integrates the concentrations and toxic response factors of the heavy metals in evaluating the potential risk posed by them. The index is denoted by the Eq. (1):

$$RI = \sum_{i=1}^{n} E_r = \sum_{i=1}^{n} T_r \times C_f \tag{1}$$

Where RI represents the integrated potential ecological risk index, E_r represents the potential ecological risk index for each heavy metal, T_r represents the toxic response factor, and C_f represents the concentration factor.

2.5.2. Metal index (MI)

The Metal Index is an important quantitative tool employed in the assessment of overall heavy metal contamination levels in an environmental sample. This is achieved through the aggregation of concentrations of various heavy metals and comparing them to established maximum allowable concentrations (MAC) (Goher et al., 2014). It is denoted by Eq. (2):

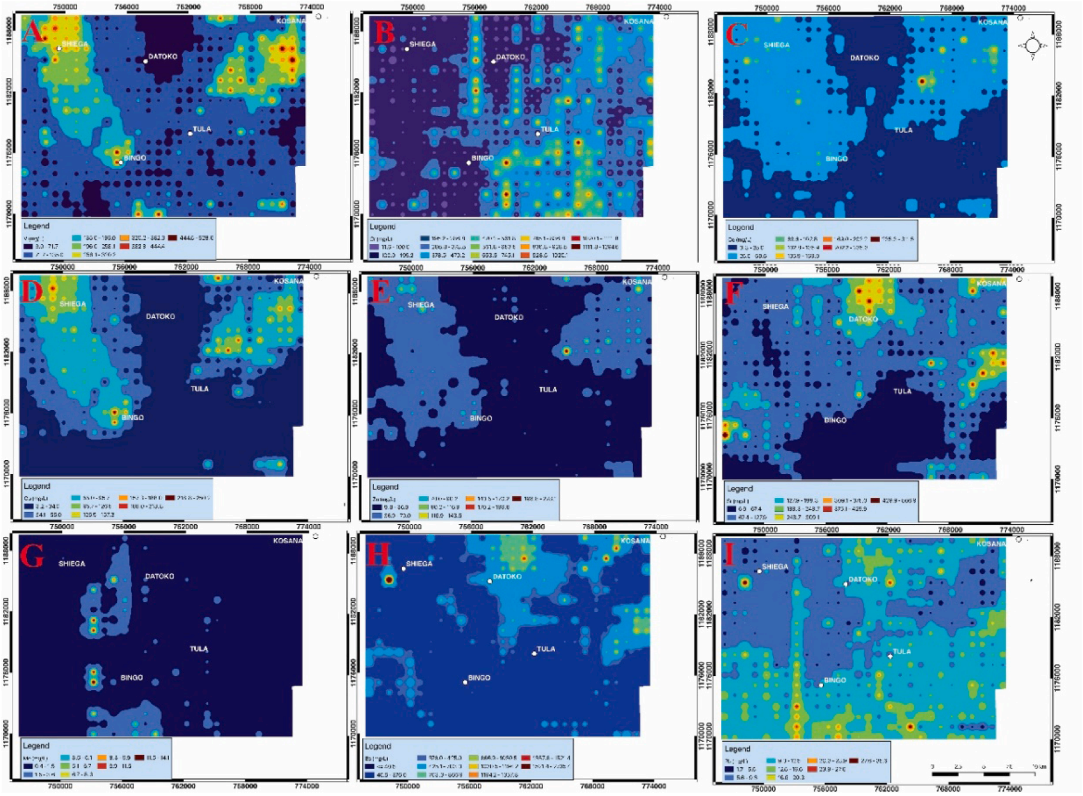

Fig. 2. Interpolation maps of (A) V (B) Cr (C) Co (D) Cu (E) Zn (F) Sr (G) Mo (H) Ba (I) Pb.

Table 2
Summary of Metal Index results.

Class	Characteristics	MI Values	Number of samples	% of Samples
1	Very pure	<0.3	0	0
2	Pure	0.3 - 1	18	3.26
3	Slightly affected	1 - 2	146	26.45
4	Moderately affected	2 - 4	139	25.18
5	Strongly affected	5 - 6	117	21.20
6	Seriously affected	>6	132	23.91

$$MI = \sum_{i=1}^{n} \frac{Ci}{(MAC)} \tag{2}$$

Where MI represent the metal index and Ci the concentration for each heavy metal, and MAC, the maximum allowable concentration.

2.6. Positive matrix factorization (PMF)

In this study, the Positive Matrix Factorization (PMF) model, a multivariate receptor model recognized by the U.S. Environmental Protection Agency, was applied to identify and assess the concentration and distribution of potentially toxic elements (PTEs) in environmental samples. PMF is well-regarded for its ability to decompose complex environmental datasets into a smaller set of factors, which represent potential pollution sources, making it a reliable tool for source apportionment (Bhuiyan et al., 2021; Chakraborty et al., 2023). Following Zhao et al. (2024a, b), this was mathematically represented in Eq. (3) as:

$$X_{iy} = \sum_{j=i}^{a} g_{ia}\, f_{ay} + e_{iy} \tag{3}$$

Where, x_{iy} = concentration of y^{th} HM in the i^{th} sample. g_{ia} = contribution of the a^{th} PTE in the i^{th} sample. f_{ay} = factorization of the a^{th} PTE that is nearby to PTE (k). e_{iy} = analytical error. These data were achieved until the lower Q value was achieved, where Q is defined as Eqn. 4. $Q(A) =$ (3). Here the α_{iy} indicates the "uncertainty" of y^{th} PTE for the i^{th} samples.

The raw data for the PMF analysis consisted of concentration data for various potentially toxic elements (PTEs) across different environmental samples. Prior to analysis, the data was pre-processed to address missing values and outliers to meet the model's requirements. The standard errors of each measurement were also calculated and provided to the PMF model, as these error margins are important for determining the weighting of each observation during factorization.

The number of factors was assessed using diagnostic tools, including the Q-value, residual analysis, and comparisons with data from existing literature on source profiles. The appropriate number of factors was determined by fitting the results to the model while maintaining interpretability.

The source profiles derived from the PMF analysis were considered accurate, as they were consistent with known sources and 'a priori' expected sources in the study area. The sensitivity of the PMF model was also tested by adjusting the number of factors and allowing the measured uncertainty estimates to vary. The final model outputs identified the sources and their contributions, providing valuable insights for developing management strategies and mitigation measures for the PTEs in the environment.

2.7. Machine learning model

2.7.1. Self-organizing map (SOM)

Self-Organizing Maps (SOMs) fall under a type of unsupervised Artificial Neural Network developed by Teuvo Kohonen in 1982 in an effort to visualize and analyse high-dimensional data in lower dimensional space mainly two-dimensional (Kohonen, 1982). SOMs work by arranging input data into a grid of neurons with each neuron representing a representative data vector. The map undergoes adjustments during the training process where the neurons are arranged based on

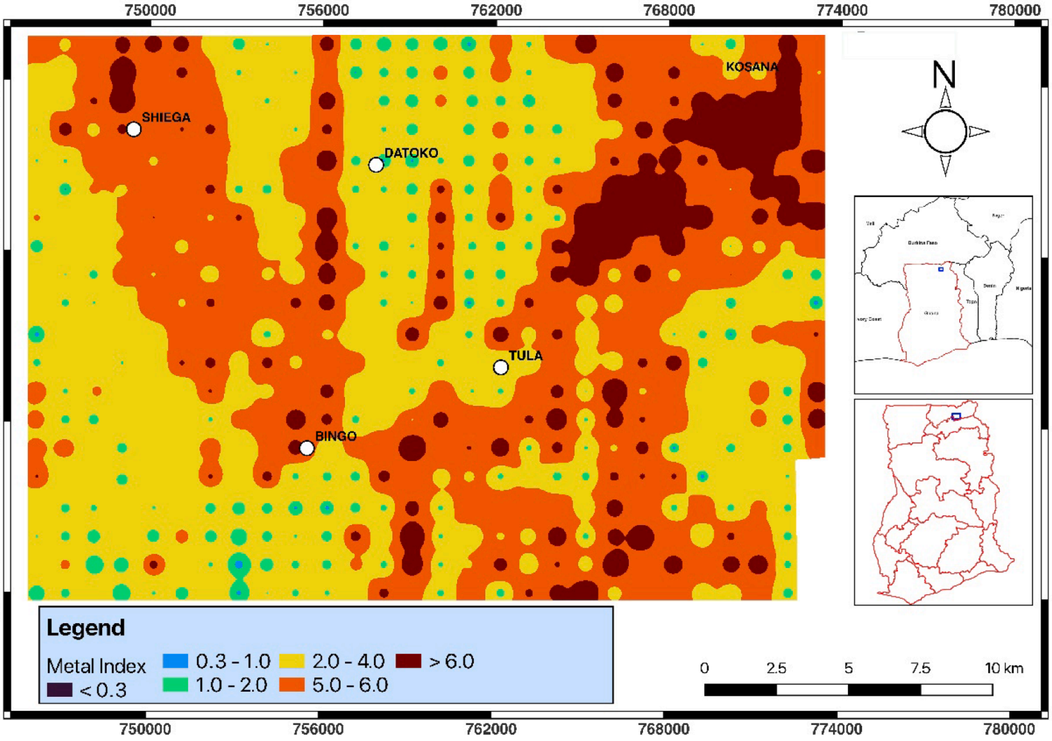

Fig. 3. Interpolation map showing the spatial representation of the Metal Index across the Study area.

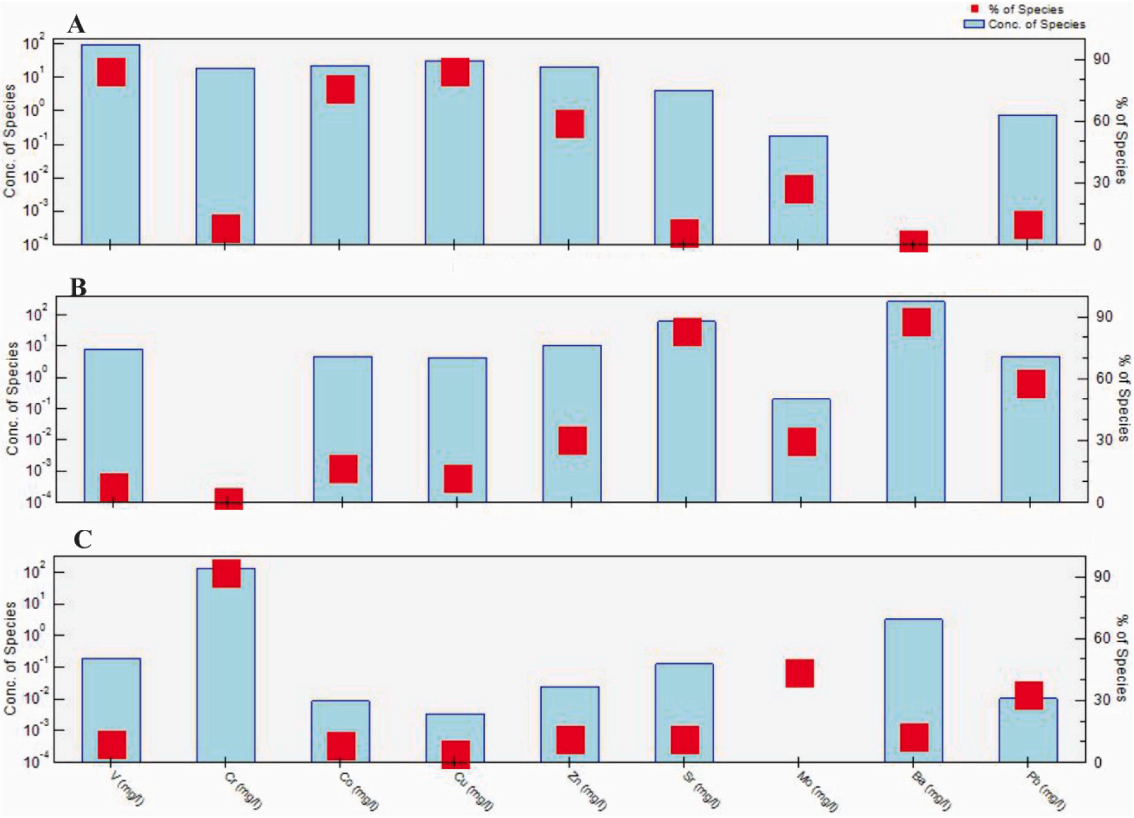

Fig. 4. Element profile source and contribution from PMF (A) Factor 1 (B) Factor 2 (C) Factor 3.

their similarity with the input data essentially clustering similar data points together (Kohonen, 1997,). Within the field of environmental science, SOMs are increasingly employed in identification of patterns and correlations among pollutants, their sources, and environmental factors associated with them. Chakraborty et al.(2023) employed SOMs effectively in the identification of hazardous elements in croplands in

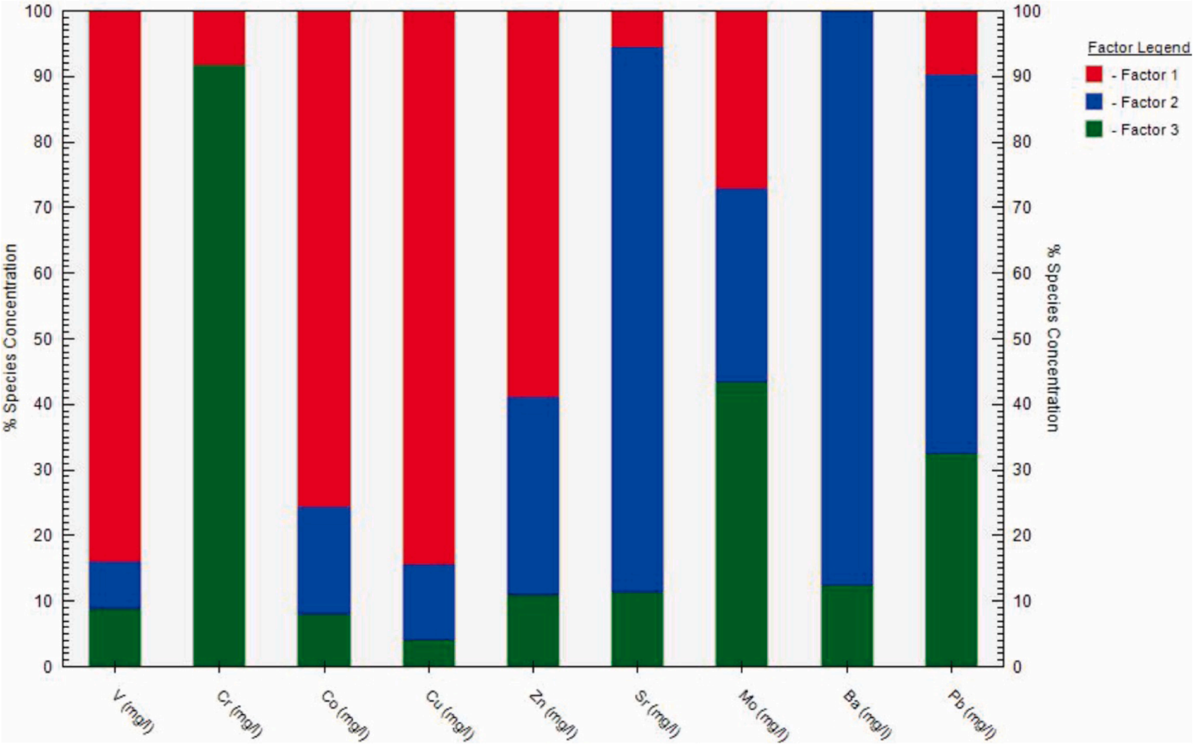

Fig. 5. Element source and factor fingerprint from PMF.

Bangladesh. Licen et al. (2023) also used SOMs in the assessment of spatial and temporal patterns of pollutants in environmental compartments. This study applies SOMs in the identification of relationships that exist amongst PTEs in soil samples.

3. Results and discussion

3.1. Pollutants, elemental concentrations and spatial patterns

Table 1 presents the general distribution of elemental values in the study area. The high skewness and kurtosis values show that the data does not follow a normal distribution. It also follows that there are points of localised high spots which suggest elemental anomalies either interpreted as pollution or enrichment hotspots. The relatively high value of the Coefficient of Variation (>60 %) except Pb (43.71 %) signifies a high level of alteration of the background concentration of the area, suggesting anthropogenic influence. V ranges from 9.45 to 528 mg/kg with a mean of 120 mg/kg. The results show that 27.72 % of samples have V values which exceed the continental crustal average (CCA) of 135 mg/kg. The most significant hotspots are located to the northwest of the area around Shiega trending downwards towards the central part of the area to Bingo (Fig. 2A). This aligns with the parts of the study area where some of the most intensive mining activities take place. These values are higher than the average distribution of V set as the contaminant level in comparison to values from the EU, USEPA and other parts of Europe. Amuah et al. (2021) reported anomalous values of V in northeastern and southeastern Ghana which coincides with the Birimian zones where mining occurs. The authors noted that V in these areas show positive correlations between V and PTEs with systematic toxicity such as Sr, Zn and Zr.

The average concentration of Cr recorded in the area was 242.42 mg/kg. This value was 2.4 times over the CCA value of 100 mg/kg, with the majority of the samples (63.59 %) surpassing this CCA limit. This was also higher than the contamination level reported by Arhin and Kazapoe (2017) for the Birimian of Ghana (80.74 mg/kg). The values recorded for Cr were much higher than those obtained by R.W. Kazapoe et al (2021) in the Southwestern region of Ghana (17.00–331.00 mg/kg). Anomalous Cr values cover about two-thirds of the study area and can be found east of Datoko, around Tula and Kosana (Fig. 2B). According to the World Health Organization (2000), crops with high Cr concentrations due to uptake from the soil when ingested may cause allergic eczematous and acute irritative dermatoses, chrome ulcers and other adverse health effects.

The concentration of Co in the soils of the study area fell between 3.3–311.8 mg/kg with an overall mean of 30.92 mg/kg. The average value of Co in the area was 1.2 times more than the CCA value of 25 mg/kg. According to Crommentuijn et al. (2000), the maximum contaminant level for the nutrient is 24 mg/kg and 40 % of the samples lie above this level. Additionally, 38.95 % of the samples fall above the CCA with most of the anomalous values recorded around the northwestern boundary of the study area and the central part in Bingo (Fig. 2C). Comparatively, 24.82 % of the Cu values were above the CCA values of 55 mg/kg. Twelve per cent of the samples were above the average of 100 mg/kg set by McLaughlin et al. (2000) in their study. The recorded mean of 43.8 mg/kg is nearly double what Arhin and Kazapoe (2017) reported as the background values of Cu (25 mg/kg) in the Birimian of southwestern Ghana.

The range of Zn was from 9.8 and 228.2 mg/kg. Only 8.33 % of the samples exceeded the CCA values. Hotspots of the Zn align with hotspots of V, Co and Cu, located in a northwestern trend from Bingo towards Shiega and also at the Northeastern boundary of the area around Kosana (Fig. 2E). These areas host the majority of mining communities. Sr and Mo presented average concentrations of 0.95 mg/kg and 337.62 mg/kg, respectively. The majority of samples had Sr and Mo values below the CCA of 375 mg/kg and 1.5 mg/kg respectively. Isolated hotspots of Sr were located around Dakoto and Kosana (Fig. 2F). Ba was found to have values ranging from 40.5 and 2735 mg/kg. This shows that all the samples fell above the recommended level of 9 mg/kg as put by Crommentuijn et al. (2000). However, only 26.45 % of the samples reported values above the CCA (425 mg/kg). These values were reported around Shiega, Dakoto and Kosana. The concentration of Pb (Fig. 2I) was also quantified and found to be varying from 1.7 to 35.8 mg/kg. The average

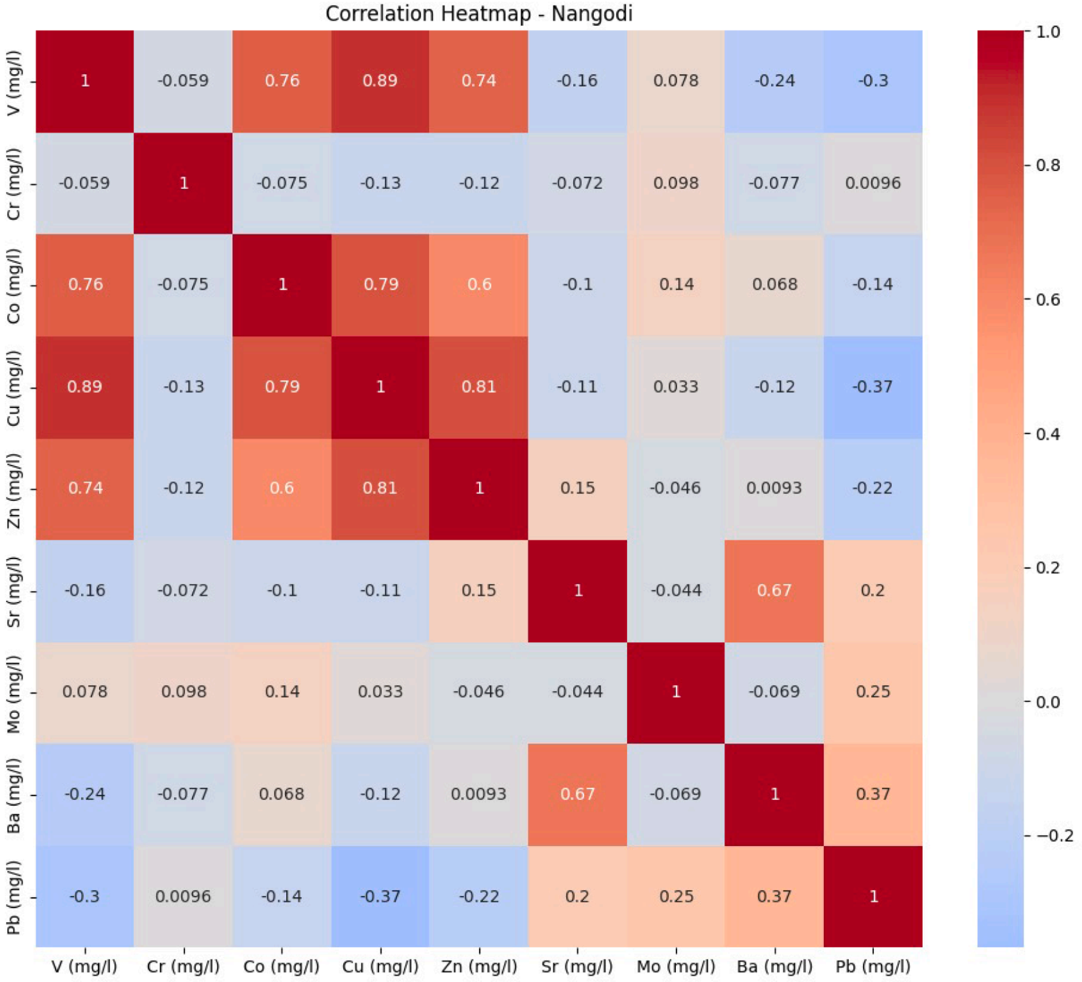

Fig. 6. Correlation heat map of the PTEs in the study area.

value was 10.07 mg/kg which is below CCA, the global values and Crommentuijn et al. (2000) values. The average value of Pb is only slightly higher than the background value of Pb (6.99 mg/kg) determined by Arhin and Kazapoe (2017) for Birimian of Ghana. The anomalous values of Pb are reported around Shiega, and Dakoto and clustered to the southern half of the area (Fig. 2I). This is per the findings of Darko et al. (2019) in Gbani, Ghana and R.W. Kazapoe et al. (2021) in southwestern Ghana. The PTEs displayed a decreasing trend in the order Ba > Cr > V > Sr > Cu > Zn > Co > Mo which closely mirrors what was reported by Rudnick and Gao (2003) for the upper continental crust (Ba > V > Cr > Zn > Cu > Co > Pb) except for the V vs Cr and Zn vs Cu. This is suggestive of a degree of human interference.

3.2. Pollution assessment indices

The Integrated Potential Ecological Risk Index (RI) and Metal Index (MI) were employed to determine the concentration of heavy metals in the soil of the study area and the ecological risks that are associated with them (Table 2). The Integrated Potential Ecological Risk Index estimated the potential ecological risks related to the heavy metal concentrations in the soil, by taking into account the toxic levels of the metals and the extent to which they are present in the environment. The Metal Index was calculated to estimate overall heavy metals pollution status of the given soil sample. It provided a single index that represented the overall concentration of several heavy metals in the soil as compared to their initial concentrations in the environment. Results from the RI showed that all the samples fell within the “low risk” category. However, the MI painted a more severe situation. Under the MI, only 18 samples representing 3.26 % of the samples ranked within the “Pure” category, indicating that these samples were related not impacted by pollution. The majority of the samples were ranked as slightly to moderately affected (26.45 % and 25.18 % respectively) located around Tula, Dakoto and southwest of the area (Fig. 3). Additionally, 21.20 % of the samples (117 samples) were ranked as “Strongly affected”, while 132 samples representing 23.91 % of all considered samples were categorised as “Seriously affected”. These samples are located around Shiega, Kosana and southeast of Tula, again highlighting the significant effect of galamsey on the soil health of the area.

3.3. Relationships among the pollutant and trace elements

3.3.1. Positive matrix factorization (PMF) model

The EPA PMF model (V. 5) was used with nine elements found in the soil samples to identify and quantify the probable origins of PTEs in the study area as well as the effect of every element. In the present research, the PMF model was used 20 times for which the chosen factors were 3 or 4. In this work, three factors were chosen according to the level of pollutant for each of the heavy metals. The value of R2 (fitness) of predicted and observed concentration for values above 0.94 mark out the fitness of the model. The concentration and contribution rate of every factor are shown in Figs. 4 and 5 for the elements.

Factor 1 is characterized by high contributions from V (84 %), Cu (84 %), and Co (75.5 %), along with a moderate contribution from Zn (58.9 %). This factor likely represents an external source or process that

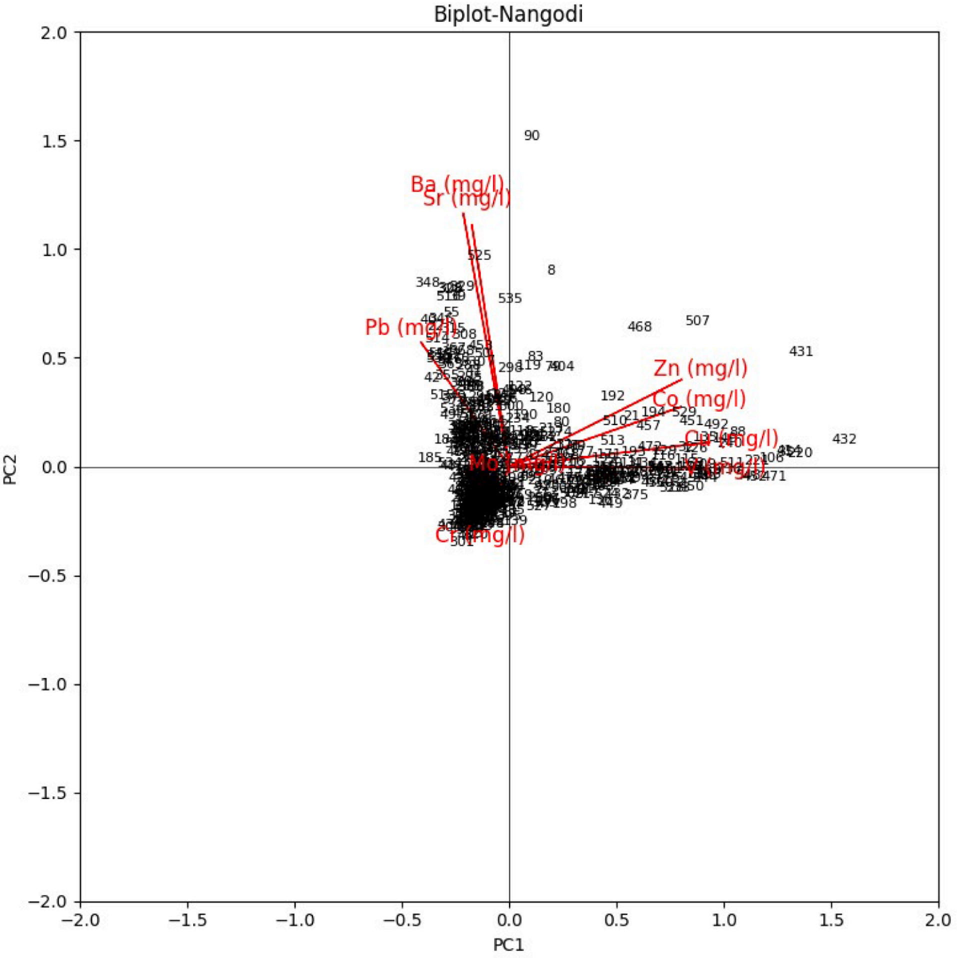

Fig. 7. Biplot of elements in the study area.

heavily influences the presence of these metals in the soil. The significantly high CV% recorded for the contributory elements to this factor (e.g. Cu-102.45 %, V-87.79 %; Table 1) underscores this assertion. The recorded mean value of V in the area (120.86 mg/kg) is similar to findings by R.W. Kazapoe et al. (2021) in an area with a similar geological setting and the presence of mining activities in Southwestern Ghana. the researchers in that study attributed the relatively high values to the underlying geology, which has been described as composed of granitic intrusions high in alkaline (Kesse, 1985; Kazapoe, 2014). Additionally, a likely source of Cu is through the application of agrochemicals containing agrochemicals; fertilizers, biocides, and biofertilizer as described by Chakraborty et al. (2023). These elements are also characteristically associated with gold (Kazapoe, 2023) and may suggest an association of Factor 1 with metallurgical processes or mining operations.

Factor 2 shows a strong association with Ba (87.5 %) and Sr (83.1 %), with moderate contributions from Pb (57.8 %). This factor might represent a geological or natural source, such as the weathering of specific mineral-rich rocks or soils. The moderate variation of Pb (CV% of 43.71 %) suggests a relatively low level of anthropogenic influence. This aligns with the assertions of Arhin et al. (2019) and Akoto et al. (2023) who indicate the high level of Pb to be associated with the gold-bearing rocks of the area. The results point to a mixed source for factor 2 which allows for contribution from agricultural inputs, as Ba and Sr are sometimes associated with fertilizers or pesticides. This suggests possible contributions from other mixed sources, such as low-level agricultural pollution.

Factor 3 is dominated by a very high contribution from Cr (91.8 %), with a lower but still significant contribution from Mo (43.4 %). According to the WHO (1996) Cr concentrations in soils are often controlled by rock weathering and erosion and only rarely by anthropogenic activities. However, the relatively high values recorded for Cr and its associated CV% (94.74 %) show anthropogenic influence. Similarly, Mo is related to the suite of elements known to be associated with gold in the Birimian terrain of Ghana and also shows a high CV% (144.66 %). Following this, factor 3 could be indicative of the effect of the waste from the mining activities, influencing soil chemistry in the area.

3.3.2. *Principal component analysis, pearson correlation analysis and heat map*

Pearson correlation analysis has been used to define the interdependence of the indicated PTEs and to define their sources in the Nangodi area (Fig. 6). Two groups of PTEs were formed in the correlation matrix: (1) PTEs with moderate to strong correlation (V, Co, Cu and Zn) and (2) PTEs with weak to moderate correlation (Sr, Mo, Ba and Pb). This similarly aligns with the deductions made in the PMF analysis which suggested an interplay between weathering and erosion-controlled elemental distribution and the effects of mining activities in the area. The biplot (Fig. 7) additionally corroborates this where we identify three (3) clusters; cluster 1-V, Co, Cu and Zn, Cluster 2- Ba, Sr and Pb and cluster 3- Cr and Mo. These cluster closely mirrors the Factors deducted in the PMF as well as the significant correlations shown in the heat map. The first and third clusters are interpreted as originating mainly from the mining activities prevalent in the area with a moderate contribution from the weathering of the local lithologies.

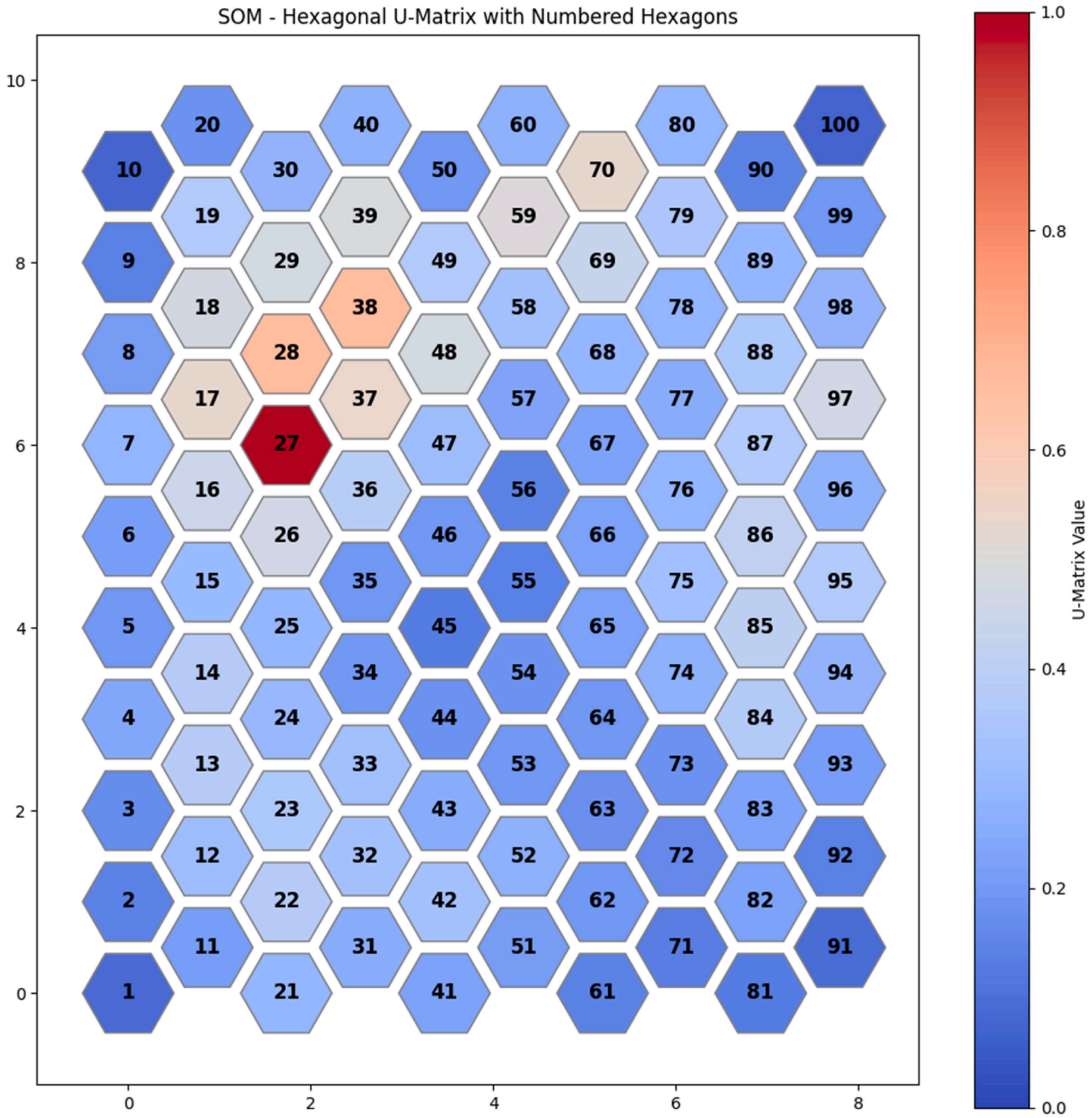

Fig. 8. SOM U Matrix representation.

The second cluster is interpreted as sourced primarily from the weathering and/or erosion of the underlying rocks.

4. Machine learning

4.1. Self-organizing map (SOM) analysis

In this study, Self-Organizing Map (SOMs) was employed in the exploration of cluster patterns that may exist within various heavy metals (V, Cr, Co, Cu, Zn, Sr, Mo, Ba, Pb) based on their concentrations across different samples. The findings from the SOM are presented in two Figs: the U-Matrix and the component plane of each heavy metal.

Fig. 8 shows the Unified Distance Matrix (U-Matrix). It illustrates the distance between neighbouring neurons. Hexagonal cells with high U-Matrix values are characterized by the red regions while cells with low values are characterized by the blue regions. The red regions indicate greater distance between neighbouring neurons suggesting distinct clusters. The upper left region of Fig. 8 shows a red cluster around hexagon 27 signifying a strong variability in the concentration levels of the analysed heavy metals in this region when compared to the concentration levels of more homogeneous regions this is the blue clustered areas. The majority of the map is characterized by the deep blue or light blue regions. The deep blue regions indicate lower U-Matrix values and suggest homogeneity amongst samples thus samples from these regions may share similar characteristics. The light blue areas indicate low U-Matrix values as well but not as low as the deep blue regions. Samples in this region may share some similar attributes with samples in the deep blue regions but not similar enough to be clustered with them. From Fig. 9 Red clusters are closely populated south of Tula and to a lesser extent, around the southwestern boundary of the study area. these zones could be inferred as areas where the local geology and mining practices control the level of PTEs in the soil. These areas align with spots with known mining activities as has been indicated on Fig. 9.This pattern is similarly visible in Cr, whose concentration has been interpreted in subSection 3.2 to be influenced by an interplay of the underlying rocks and the mining activities prevalent in those areas.

Fig. 10 presents the component planes for each of the heavy metals analysed in the study. Each subplot represents the distribution of

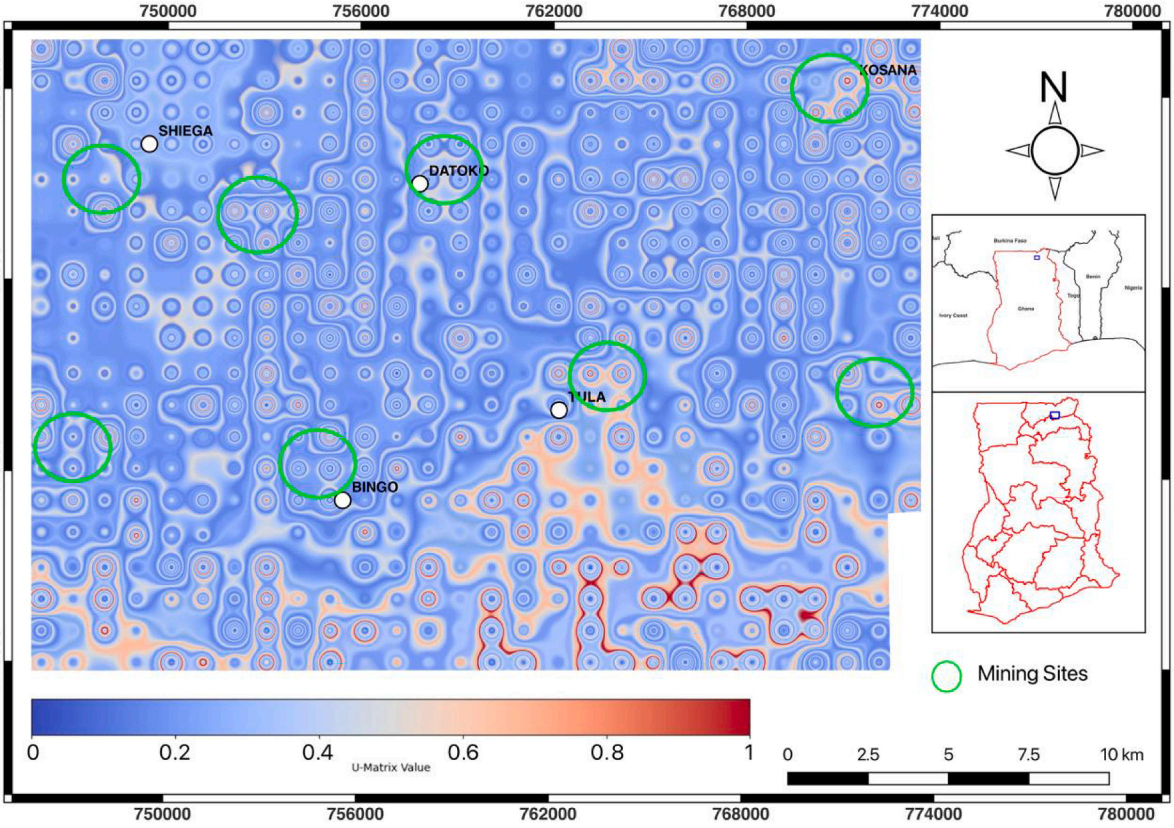

Fig. 9. Spatial representation of SOM U Matrix across the study area.

concentrations of a particular heavy metal across Fig. 10 helping us have a greater understanding of the role each heavy metal plays in the grid. Red regions in the subplot characterizes higher concentration levels while blue indicates lower concentration levels. Fig. 10 shows V, Zn, Cu, and Co with higher concentrations, this is red regions towards mostly the upper right and bottom left regions of the SOM (Fig. 10). This suggests that samples with higher concentrations of V, Zn, Cu, and Co are clustered together potentially indicating a common source or influence by similar environmental conditions. The component planes for Sr and Ba indicate a clustering pattern with high concentrations distributed centrally but more towards the uppermost and bottom right regions of the SOM also suggesting a common source or influence by similar environmental factors. The rest of the heavy metals (Pb, Mo, and Cr) show unique distributions across the SOM. Pb appears towards the upper left of the SOM, Cr bottom, and Mo mostly central left. The unique separation from these heavy metals indicates varying contributing factors to their distributions as compared to the other elements.

5. Conclusion

This study assessed the extent, sources, and spatial distribution of heavy metal pollution in the Nangodi area, North-Eastern Ghana, using machine learning and multivariate statistical techniques. The findings revealed significant anthropogenic influence, primarily from mining activities With hotspots of Zn, V, Co, and Cu identified, concentrated along a northwestern trend from Bingo to Shiega and at the northeastern boundary near Kosana. The PMF, Correlation and cluster analyses identified three major pollution sources: (1) mining-related emissions (V, Cu, Co, Zn), (2) geogenic influences (Ba, Sr, Pb), and (3) mining waste impact (Cr, Mo). The clustering of pollution sources from the SOM aligned with PMF results, further reinforcing the dominance of mining activities as the primary contributor to soil contamination. The study confirms systematic heavy metal pollution in the Nangodi area, posing potential environmental and human health risks. These findings underscore the need for urgent mitigation measures, including stricter regulation of mining activities, remediation of contaminated soils, and continuous environmental monitoring to prevent further degradation.

5.1. Recommendations

Based on the findings from the study on heavy metal pollution in the Nangodi area in North eastern Ghana, several recommendations can be made:

- to enhance and implement environmental laws that focuses on mining especially the galamsey around Shiega, Kosana and Tula. This entails the enhancement of measures in the fight against unauthorized mining activities.
- to establish a policy of frequent and consistent sampling of heavy metal levels in the soil especially in the hot spots regions to evaluate the temporal trends and the efficiency of the measures put in place.
- work with the local miners to encourage them on the need to embrace better and environmentally friendly mining methods. This means that, roping in training and support for cleaner and more efficient technologies in mining could help in minimizing the adverse effects while still realizing the economic gains.
- carry out an intensive appraisal of the agricultural lands to find out the degree of soil pollution by heavy metals and its threat to crops. The areas with high levels of contamination should be controlled to avoid the intake of heavy metals by the plants.
- further studies should be done to determine other causes of heavy metal pollution with emphasis on the anthropogenic sources of pollution. This can help to direct more precise interventions.
- conduct longitudinal surveys to assess the impacts of heavy metal pollution on natural ecosystems and the human body. This will be

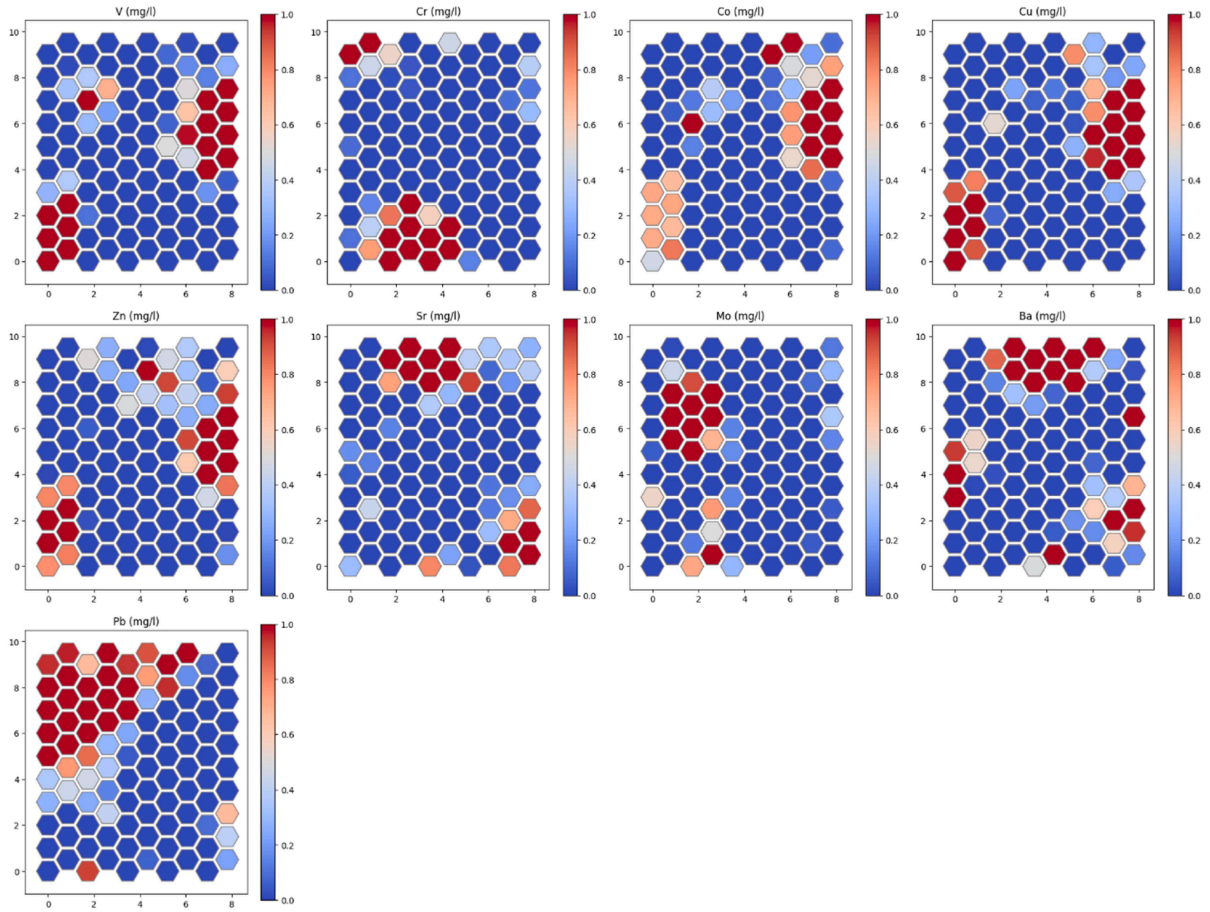

Fig. 10. Map showing the component planes for each of the heavy metals.

useful in identifying trends and the effectiveness of remediation operations with time.

5.2. *Limitation*

- The study examines only the top 0–20 cm of soil, potentially missing contamination in deeper layers or groundwater.
- Heavy metal concentrations may vary over time due to weathering, leaching, and human activities, which were not assessed.

Funding information

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

CRediT authorship contribution statement

Raymond Webrah Kazapoe: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Formal analysis, Data curation, Conceptualization. **Daniel Kwayisi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Conceptualization. **Seidu Alidu:** Writing – review & editing, Project administration, Methodology, Investigation. **Obed Fiifi Fynn:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Samuel Dzidefo Sagoe:** Writing – review & editing, Validation, Software, Methodology. **Ebenezer Ebo Yahans Amuah:** Writing – review & editing, Writing – original draft, Software, Methodology, Data curation. **Emmanuel Nyavor:** Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Akoto, O., Yakubu, S., Ofori, L.A., Bortey-Sam, N., Boadi, N.O., Horgah, J., Sackey, L.N., 2023. Multivariate studies and heavy metal pollution in soil from gold mining area. Heliyon 9 (1).

Amuah, E.E.Y., Fei-Baffoe, B., Kazapoe, R.W., 2021. Emerging potentially toxic elements (strontium and vanadium) in Ghana's pedological studies: understanding the levels, distributions and potential health implications. A preliminary review. Environ. Chall. 5, 100235.

Amuah, E.E.Y., Fei-Baffoe, B., Kazapoe, R.W., Dankwa, P., Okyere, I.K., Sackey, L.N.A., Kpiebaya, P., 2024. From the ground up: Unveiling Ghana's soil quality crisis and its ecological and health implications. Innov. Green Dev. 3 (1), 100097.

Amuah, E.E.Y., Fei-Baffoe, B., Sackey, L.N.A., Douti, N.B., Kazapoe, R.W., 2022. Understanding the distribution, source-pattern and geochemical controls of soils in an artisanal mine site during a ban on illegal mining activities: Is a ban an absolute solution? Soil Secur. 9, 100078.

Arhin, E., & Kazapoe, R. (2017). Selenium in locally produced food crops and implications on healthy eating: A case study at the Talensi district of Ghana.

Arhin, E., Mouri, H., Kazapoe, R., 2017. Inherent errors in using continental crustal averages and legislated accepted values in the determination of enrichment factors (EFs): A case study in northern Ghana in developing environmental policies. J. Geogr. Nat. Disast. 7 (204), 2167–0587.

Arhin, E., Zhang, C., Kazapoe, R., 2019. Medical geological study of disease-causing elements in Wassa area of Southwest Ghana. Environ. Geochem. Health 41 (6), 2859–2874.

Bempah, C.K., Ewusi, A., 2016. Heavy metals contamination and human health risk assessment around Obuasi gold mine in Ghana. Environ. Monit. Assess. 188, 1–13.

Bhuiyan, M.A.H., Karmaker, S.C., Bodrud-Doza, M., Rakib, M.A., Saha, B.B., 2021. Enrichment, sources and ecological risk mapping of heavy metals in agricultural soils of dhaka district employing SOM, PMF and GIS methods. Chemosphere 263, 128339. https://doi.org/10.1016/j.chemosphere.2020.128339.

Chakraborty, T.K., Mobaswara, M.Z., Nice, M.S., Islam, K.R., Netema, B.N., Rahman, M. S., Khan, A.S., 2023. Application of machine learning and multivariate approaches for source apportionment and risks of hazardous elements in the cropland soils near industrial areas in Bangladesh. Ecol. Indic. 154, 110856.

Cobbinah, P.B., Poku-Boansi, M., Peprah, C., 2017. Urban environmental problems in Ghana. Environ. Dev. 23, 33–46.

Crommentuijn, T., Sijm, D., De Bruijn, J., Van den Hoop, M.A.G.T., Van Leeuwen, K., Van de Plassche, E., 2000. Maximum permissible and negligible concentrations for metals and metalloids in the Netherlands, taking into account background concentrations. J. Environ. Manag. 60, 121–143.

Darko, G., Boakye, K.O., Nkansah, M.A., Gyamfi, O., Ansah, E., Yevugah, L.L., Acheampong, A., Dodd, M., 2019. Human health risk and bioaccessibility of toxic metals in topsoils from Gbani mining community in Ghana. J. Health Pollut. 9 (22), 190602.

Dickson, K.B., Benneh, G., 1988. A New Geography of Ghana. Longman Group UK Limited. Longman House, Burnt Mill, Harlow, Essex, England.

Ghana Statistical Services, 2010. Nabdam District. Provisional Results of the 2010 Population and Housing Census, p. 33, 826Retrieved from. https://unstats.un.org/unsd/demographicsocial/census/documents/Ghana/Provisional_results.pdf.

Goher, M.E., Hassan, A.M., Abdel-Moniem, I.A., Fahmy, A.H., El-Sayed, S.M., 2014. Evaluation of surface water quality and heavy metal indices of Ismailia Canal, Nile River, Egypt. Egypt. J. Aquat. Res. 40 (3), 225–233.

Gyamfi, O., Sørensen, P.B., Darko, G., Ansah, E., Vorkamp, K., Bak, J.L., 2021. Contamination, exposure and risk assessment of mercury in the soils of an artisanal gold mining community in Ghana. Chemosphere 267, 128910.

Ha°kanson, L., 1980. An ecological risk index for aquatic. Pollution control: sedimentological approach. Water Res. 14, 975–1001.

Kazapoe, R.W., 2014. MPhil thesis. University of Ghana.

Kazapoe, R.W., Addai, M.O., Amuah, E.E.Y., Sagoe, S.D., 2024. Fluid boundaries: a cross-country exploration of groundwater quality amid threats from climate change. Environ. Chall., 100953

Kazapoe, R.W., Amuah, E.E.Y., Dankwa, P., Ibrahim, K., Mville, B.N., Abubakari, S., Bawa, N., 2021a. Compositional and source patterns of potentially toxic elements (PTEs) in soils in southwestern Ghana using robust compositional contamination index (RCCI) and k-means cluster analysis. Environ. Chall. 5, 100248.

Kazapoe, R.W., Arhin, E., Amuah, E.E.Y., 2021b. Known and anticipated medical geology issues in Ghana. Ecofeminism Clim. Change 2 (4), 169–184.

Kazapoe, R.W., Okunlola, O., Arhin, E., Olisa, O., Kwayisi, D., Dzikunoo, E.A., Amuah, E. E.Y., 2023. Compositional characteristics of mineralised and unmineralised gneisses and schist around the Abansuoso area, southwestern Ghana. Appl. Earth Sci. 132 (1), 36–51.

Kazapoe, W.R., Okunlola, O., Arhin, E., Olisa, O., Harris, C., Kwayisi, D., Torkorno, S., Amuah, E.E.Y., 2022. Geology and isotope systematics of gold deposits in the abansuoso area of the sefwi belt, southwestern Ghana. Geol. Ecol. Landsc. 1–22.

Kesse, G.O. (1985). The mineral and rock resources of Ghana.

Khan, J., Singh, R., Upreti, P., Yadav, R.K., 2022. Geo-statistical assessment of soil quality and identification of heavy metal contamination using integrated GIS and Multivariate statistical analysis in industrial region of Western India. Environ. Technol. Innovation 28, 102646.

Kim, K.H., Yun, S.T., Yu, S., Choi, B.Y., Kim, M.J., Lee, K.J., 2020. Geochemical pattern recognitions of deep thermal groundwater in South Korea using self-organizing map: identified pathways of geochemical reaction and mixing. J. Hydrol. 589, 125202.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43 (1), 59–69.

Kohonen, T., 1997. Exploration of very large databases by self-organizing maps. In: In Proceedings of international conference on neural networks (icnn'97), 1. IEEE, pp. PL1–PL6.

Kowalska, J.B., Mazurek, R., Gąsiorek, M., Zaleski, T., 2018. Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–A review. Environ. Geochem. Health 40, 2395–2420.

Licen, S., Astel, A., Tsakovski, S., 2023. Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: a review. Sci. Total Environ. 878, 163084.

Lv, L., Chen, Y., Han, Y., Cui, M., Wei, P., Zheng, M., Hu, J., 2021. High-time-resolution PM2. 5 source apportionment based on multi-model with organic tracers in Beijing during haze episodes. Sci. Total Environ. 772, 144766.

Ministry of Food and Agriculture. (n.d.). Nabdam District. Retrieved from http://mofa.gov.gh/site/sports/district-directorates/upper-east-region/269-talensi-nabdam.

McLaughlin, M.J., Hamon, R.E., McLaren, R.G., Speir, T.W., Rogers, S.L., 2000. A bioavailability-based rationale for controlling metal and metalloid contamination of agricultural land in Australia and New Zealand. Soil Res 38 (6), 1037–1086.

Obiri-Nyarko, F., Quansah, J.O., Asare, S.V., Fynn, O.F., Okrah, C., Debrah, S.K., Karikari, A.Y., 2024. Determination of threshold values and heavy metal pollution assessment of soils in an industrial area in Ghana. Environ. Monit. Assess. 196 (6), 546.

Owusu-Nimo, F., Mantey, J., Nyarko, K.B., Appiah-Effah, E., Aubynn, A., 2018. Spatial distribution patterns of illegal artisanal small scale gold mining (Galamsey) operations in Ghana: A focus on the Western Region. Heliyon 4 (2).

Reimann, C., Filzmoser, P., Garrett, R.G., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem. 17 (3), 185–206.

Rudnick, R.L., Gao, S., 2003. Composition of the continental crust. Treatise Geochem. 3, 1–64.

Sakyi, P.A., Kwayisi, D., Nunoo, S., Ocran, E., Su, B.X., Malaviarachchi, S.P., 2024. Tectonic evolution of alternating paleoproterozoic belts and basins in the Birimian terrane in Southeastern West African Craton. J. Afr. Earth Sci. 105449.

Sakyi, P.A., Manu, J., Su, B.X., Kwayisi, D., Nude, P.M., Dampare, S.B., 2019. Geochemical and Sm–Nd isotopic evidence for the composition of the Palaeoproterozoic crust of the West African Craton in Ghana. Geol. J. 54 (6), 3940–3957.

Songsore, J., 2020. The urban transition in Ghana: Urbanization, national development and poverty reduction. Ghana Soc. Sci. J. 17 (2), 57. -57.

Subbarao, C., Subbarao, N.V., Chandu, S.N., 1996. Characterization of groundwater contamination using factor analysis. Environ. Geol. 28, 175–180.

Tang, J., Zhang, J., Ren, L., Zhou, Y., Gao, J., Luo, L., Yang, Y., Peng, Q., Huang, H., Chen, A., 2019. Diagnosis of soil contamination using microbiological indices: A review on heavy metal pollution. J. Environ. Manage. 242, 121–130.

Temple, J.T., 1978. The use of factor analysis in geology. J. Int. Assoc. Math. Geol. 10, 379–387.

Thakkar, D., Valand, M., Vachhrajani, K., 2024. Assessment of seasonal variations in soil heavy metal concentrations and potential health risks in Gujarat. Environ. Geochem. Health 46 (10), 391.

Tom-Dery, D., Dagben, Z.J., Cobbina, S.J., 2012. Effect of illegal small-scale mining operations on vegetation cover of arid northern Ghana. Res. J. Environ. Earth Sci. 4 (6), 674–679.

Ukaogo, P.O., Ewuzie, U., Onwuka, C.V., 2020. Environmental pollution: causes, effects, and the remedies. Microorganisms For Sustainable Environment and Health. Elsevier, pp. 419–429.

Uranishi, K., Ikemori, F., Nakatsubo, R., Shimadera, H., Kondo, A., Kikutani, Y., Asano, K., Sugata, S., 2017. Identification of biased sectors in emission data using a combination of chemical transport model and receptor model. Atmos. Environ. 166, 166–181.

Verma, S., Bhatt, P., Verma, A., Mudila, H., Prasher, P., Rene, E.R., 2021. Microbial technologies for heavy metal remediation: effect of process conditions and current practices. Clean Technol. Environ. Policy 1–23.

World Health Organization (WHO), 2000. Air Quality Guidelines for Europe. WHO. Available at https://www.euro.who.int/en/publications/abstracts/air-quality-guidelines-for-europe.

WHO, 1996. Permissible limits of heavy metals in soil and plants. Geneva World Health Organization, Switzerland.

Zahoor, Z., Latif, M.I., Khan, I., Hou, F., 2022. Abundance of natural resources and environmental sustainability: the roles of manufacturing value-added, urbanization, and permanent cropland. Environ. Sci. Pollut. Res. Int. 29 (54), 82365–82378.

Zhao, Z., Li, S., Li, Y., 2024a. Determining the priority control factor of toxic metals in cascade reservoir sediments via source-oriented ecological risk assessment. J. Hydrol. 631, 130755.

Zhao, Z., Li, S., Li, Y., 2024b. Controlling factors and sources-specific ecological risks associated with toxic metals in core sediments from cascade reservoirs in Southwest China. Sci. Total Environ. 924, 171570.