



Multi-Technique Analytical and MATLAB-Based Evaluation of Elemental Composition in Alluvial, Black, Red, Lateritic, and Sandy Soils

Ajit Kumar Singh¹

(Research Scholar Department of Physics VKSU, Ara, Bihar)

Dr. Shikha Awasthi²

(Assistant Professor, Department of Physics M.M.M College, Ara, Bihar)

Dr. Ratnesh Kr. Pandey³

(Assistant Professor, Department of Physics U.P.E.S Dehradun, Uttarakhand)

Abstract: The current research works on different soil type to analysis its elements present in it by using diversified analytical and computational methods interfaced with MATLAB R2017b. The soil samples for Alluvial, Black, Red, Lateritic and Sandy were characterized for major and trace elements. Elemental quantification was performed with advanced techniques such as X-Ray Fluorescence (XRF), Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Atomic Absorption Spectroscopy (AAS) and MATLAB tools were used for data normalization, correlation analysis, Principal Component Analysis (PCA) and K-means clustering. The findings showed that the major elements (Si, Al, Fe, Ca, and Mg) dominated in the soils, while spatial variability in the trace metals (Zn, Cu, Pb, Cd, and Ni) depended on factors at the geological and human scales. PCA showed that only a small number of components were able to describe most of the variance of the dataset and clustering was able to put soils in four compositional groups. A Boxplot analysis show that Pb accumulation was particularly higher in lateritic and black soils (vs. sandy soils). Conclusions The analysis with MATLAB brings the ability to analyse with more precision the chemistry of soil which can be beneficial in the sustainable management of soil as well as in monitoring the contamination and agricultural planning.

Keywords--Soil analysis, Elemental composition, X-Ray Fluorescence (XRF), Principal Component Analysis (PCA), K-means clustering, Soil fertility, Environmental monitoring.

1. Introduction

Soil is one of the most important natural resources that contributes to plant growth, ecosystem sustainability, and biogeochemical cycling. Soil quality is essential for agriculture productivity and environmental quality, as it relates with composition, structure and nutrient content (Lal, 2020). The fertility state and pollution of soil systems are directly related to the chemical and elemental compositions, and mineralogical origins of soil. Macronutrients for plants include elements nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and iron (Fe), as well as many trace elements (e.g., zinc (Zn), copper (Cu), manganese (Mn), boron (B)), and both macro- and micronutrients are important for soil health and crop productivity (Alloway, 2013). On the other hand, the large and continuous dosages of some heavy metals like lead (Pb), cadmium (Cd), arsenic (As), and chromium (Cr) well be specific for human health and environmental hazard too (Kabata-Pendias, 2011). Hence, identification of the elemental profile of soils is critical for sustainable land handling, contamination evaluation, and precision agriculture.

Soil consists of variable proportions of minerals, organic matter, water, air, and living organisms (Brady & Weil, 2019). Soil composition varies with respect to parent rock material, climate, topography, vegetation, and land use of industrial, mining, chemical fertilizer, and pesticide use (Brady & Weil, 2019). Though traditional approaches for soil analyses like wet chemical and gravimetric analyses are reliable, in many cases they are slow, labour-intensive and limit the sensitivity to their lowest concentrations (Mielke et al., 2016). Over the last decade, technological advancements have allowed the development of modern analytical instruments that enable concomitant multi-elemental detection in soil with greater accuracy and sensitivity. However, for soil matrices, detailed elemental and structural information is necessary to understand their complexity, and such information can only be obtained from advanced spectroscopic and microscopic imaging and analyses (Ure & Davidson, 2002).

X-ray fluorescence (XRF), inductively coupled plasma mass spectrometry (ICP-MS) and atomic absorption spectrometry (AAS) are methods that have gained significant usage for soil element determination—these methods can be utilised for their high accuracy and sensitivity as well as their capability for measuring complex matrices (Miller et al., 2017). XRF is a fast non-destructive method for analysing major and minor elements (Norrish & Hutton, 1969), and is a potentially high-throughput bulk soil analytical method for large soil survey (Xian et al. By contrast, ICP-MS offers detection limits for elements at the parts-per-billion level (DeBord et al., 2014), enough to uncover what may be one of the rarest and most toxic of elements (Beauchemin, 2018). Additionally, FTIR + SEM + EDS, as a correlation from both the surface morphology and molecular structures of soil particles, provides a glimpse of not only chemical but also physical properties (Ahmad et al. 2020) This kind of advanced analytical methods will enhance the quality and scale of soil characterization, and will aid in environmental monitoring, agricultural preparedness, and contamination sites management through remediation (Bordoloi et al., 2021). For instance, XRF and ICP-MS data for elemental distribution maps are valuable to identify polluted sites, nutrient deficient sites, and more fertile areas. Combining these analytical data with Geographic Information Systems (GIS) has the potential to create spatial visualization and soil resource management (Zhao et al., 2019). The present study hence, explores the chemical elemental composition of diverse types of soils using XRF, ICP-MS and AAS methods, to assess soil fertility, pollution, and interrelationship of elements. Through a comparison and evaluation of each of these methods, we anticipate that this study will result in more effective constitutional monitoring protocols for soil health and sustainable agriculture protocols.

2. Related Study

Soil elemental composition has long been a crucial pod in the agro- and environmental sciences toolbox and has been used as a directive tool to provide vital information on soil fertility, mineral content, and contamination. Over the decades, researchers developed and adapted a range of analytical methods, from classic wet chemistry to spectroscopic and microscopic techniques, for detection and quantification of element concentrations that could be found in varying soils. Investigation of soil chemistry has been thoroughly done using wet digestion and gravimetric methods; they may at truth be very trustable methods but not such precise and time consuming. Chemical extraction methods that developed during the previous century produced simple but still useful information about the nutrient status of soil, albeit such analyses are limited in their ability to detect trace and ultra-trace elements (Sparks, 2003). In the same lines, Ure et al. (2002). As matrix interferences often affect such wet chemical methods (they generally provide no or extremely low ΔC compared to wet methods have also emphasized the points and the observed interferences lead to incorrect assessment, particularly in case of heavy metals. Such constraint has fortified the shift to instrumental analytical methods in soil science.

Advancements in Instrumental Soil Analysis

Atomic Absorption Spectroscopy Atomic Absorption Spectroscopy (AAS) was a big step forward in elemental analysis. As reported by Alloway (2013), AAS made it possible to determine metals like Fe, Cu, Zn, Mn and Pb in a much more sensitive manner in relation to classical chemical assays. Miller et al. also observed this. AAS has been popular in agricultural and environmental studies since it is inexpensive and accurate for single element determination (Khan et al. But researchers such as Beauchemin (2018) manufactured that AAS was limited while it can only examine one city and is a consuming process, has made an isotope process in measuring multi-elemental measurable.

These limitations were overcome with the introduction of techniques such as Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES) and Inductively Coupled Plasma Mass Spectrometry (ICP-MS). According to Beauchemin (2018), ultra-trace element detection, wide dynamic range, as well as minimization of spectral (and practical) interference, made ICP-MS an ideal detector for the measurement of rare and toxic elements, such as HSE, in soils. Likewise, Sposito (2008), pointed out that the multi-elemental capability along with the high sensitivity and precision of ICP-MS method provided a significant time saving associated with the steps of the soil parameter analysis process. Mielke et al. Wu et al. (2016) observed that the results obtained from ICP-based methods were more accurate and reproducible than those based on AAS, especially for trace and heavy metals quantification.

Use of X-Ray Fluorescence (XRF) in Soil Elemental Analysis

Researchers had increasingly turned to X-Ray Fluorescence (XRF) for non-destructive, multi-elemental analysis of soils. The ability of XRF to measure a wide variety of major and minor constituents accurately in geological and other soil-related samples was pioneered by Norrish and Hutton (1969). Later studies verified its reliability (e.g. Weindorf et al. According to Davis et al. (2014), portable XRF functionality allowed expedient elemental characterization of specimens with reasonable accuracy when compared with laboratory-based techniques. Furthermore, Towett et al. XRF technology was found to support soil mapping at regional scales as a better tool to monitor soil nutrients patterns across large spatial scales (Huang et al., 2015). However, XRF has some drawbacks with regards to low atomic number elements and trace concentrations (Tack et al., 2016). It was noted by Zhao et al. Here, for example, (2019) pointed matrix effects and moisture content as factors influencing accuracy of the XRF methods, and thus, calibration must be carried out using standard reference materials. Yet, due to its speed, non-destructiveness, and multi-element determination capability without complex sample preparation, the method was used widely in soil studies.

Scanning Electron Microscopy and Energy Dispersive Spectroscopy (SEM-EDS)

With the development of Scanning Electron Microscopy with Energy Dispersive Spectroscopy (SEM-EDS), researchers were finally able to visualize the spatial arrangement of soils and their elemental constituents at the microscopic scale. Ahmad et al. SEM-EDS thus provided morphological and elemental data, which can be interpreted to give integrated information on soil texture and mineral structure as well as possibilities of contaminant accumulation [20]. Similarly, Möller et al. Bulk analytical techniques (e.g., XRF or ICP-MS) do not provide this information, although this is one of the advantages of using SEM-EDS (Huang et al., 2018) because soil mineral phases can be recognized and their relation to organic matter or heavy metals could be assessed. Soil microanalysis and environmental forensics with SEM-EDS used for the first time an impressive combination of visual and compositional analysis together.

Fourier Transform Infrared Spectroscopy (FTIR) and Spectroscopic Approaches

Fourier Transform Infrared Spectroscopy (FTIR) has been developed to identify organic and mineral fractions in soils. According to Parikh et al. FTIR analysis is very useful in detecting functional groups and chemical bonds in soil organic matter, clays, and carbonates. Sarkar et al. FTIR has assisted in distinguishing mineral phases between amorphous and crystalline and in identifying shifts in composition and functionality of organic matter as the result of anthropogenic influences. Moreover, FTIR was an additional method to XRF and ICP-MS not for quantification but for qualitative analysis of soil components. Rapid measurement of soil properties was conducted using Near-Infrared (NIR) and Mid-Infrared (MIR) spectroscopy. Stenberg et al. (2010) Conclusion and implications found that calibration with reference datasets could enable prediction of soil carbon content, clay percentage and nutrient availability with high accuracy using infrared spectroscopy. These methods, while mostly qualitative, had been useful for large-scale soil monitoring and decision making in agriculture.

Elemental Distribution, Contamination, and Soil Fertility

Many studies have analysed the spatial distribution of soil elements to assess its fertility and levels of contamination. In another study, Kabata-Pendias (2011) also supported the significance of the essential elements Fe, Mn, Zn and Cu as providing nutrients for plant life, although he pointed out that an excess or deficit had an opprobrious effect on crop yield. In contrast, Cd, Pb, and As were toxic metals that could present a significant environmental risk even at low concentrations. According to Lal (2020), heavy metal accumulation in farming soils in various parts of the world had been connected to industrial activities, emissions from vehicles, and excessive use of fertilizers. Similarly, Zhang et al. (2018) Soil samples taken from industrial areas in China found significantly enriched Pb and Cd as a result of anthropogenic pollution sources.

For India's regional soil elemental composition assessment, several local studies had taken place. Bordoloi et al. in (2021) used XRF and ICP-MS to evaluate agricultural soils in Assam and related the nutrient and heavy metal concentrations in the soils to their texture and land use. Key message Elemental analysis allows tracking nutrient deficiencies and hotspots of contamination for sustainable agriculture management Furthermore, Sharma et al. (2020) Mine construction sites in the industrial areas of Punjab was reported to have higher concentrations of Cr, Ni and Pb and a continuous monitoring was suggested.

Integration of Analytical Data with GIS and Multivariate Approaches

Recent investigations have shown potential for bringing together analytical data with the related power of Geographic Information Systems (GIS) and statistical tools towards a better soil characterization. Zhao et al. (2019). The study of used GIS-based spatial mapping to illustrate the distribution of heavy metals in agriculture soils and enhanced the environmental risk assessments. Similarly, Chen et al. (2021) Principal Component Analysis (PCA) and Cluster Analysis (CA) were performed to group soil samples based on background elemental profiles to support interpretation of contamination sources and geochemical processes. The integration of quantitative and statistical methods was beneficial for data interpretation, leading to regional guided management of soil.

Need

The review of the literature indicated that, although a few detailed reviews of soil elemental analysis using advanced analytical techniques existed, these reviews provided scant information about these techniques. However, several gaps remained. Many previous studies and their analytical techniques provided limited information on soil composition. In addition, the performance comparison of analytical methods, especially among X-ray fluorescence (XRF), inductively coupled plasma mass spectrometry (ICP-MS), and atomic absorption spectrometry (AAS), had not been carefully investigated in different soil types. And again, most had focused on specific hotspot regions and thus ignored wider spatial (and climatic) variation. This clearly indicates a need for integrated comparative studies of soil elements within several soil types, via more advanced instrumental methods integrating quantitative elemental data and spatial data.

3. Materials and Methods

3.1 Sample Collection

The soil samples that differ for 5 soil types; Alluvial, Black, Red, Lateritic and Sandy soil types provided representative of the geologic and environmental conditions. Sampling areas were selected based on variations in topography, land-use patterns, and types of vegetation cover. Soil samples were collected at a uniform soil depth (0–20 cm) across all the sampling sites as soil depth where most of biological and chemical activities occur, was found within the active root zone. For metal contamination control, a stainless-steel auger was utilized at each location, extracting around 500 grams of soil. Samples were placed in clean and hermetic polyethylene bags, identified by GPS coordinates, and transported to the laboratory. Soils from the field were air dried in a room and then the air-dried samples were oven dried at 105°C for 4 hours to evaporate the remaining water from the soil particles. Dried soils were pulverised (using agate mortar and pestle), sieved through a 2 mm mesh and kept within fully sealed plastic containers until analysed.

3.2 Analytical Methods

The prepared soil samples were subjected to elemental analysis using a combination of advanced instrumental techniques to quantify both major and trace elements.

- **X-Ray Fluorescence (XRF):** To obtain the concentration of major elements like silicon (Si), aluminum (Al), iron (Fe), calcium (Ca) and magnesium (Mg). Due to XRF does not require any sample destruction, many elements can be detected quickly.
- **ICP-MS:** Inductively coupled plasma mass spectrometry [37] used for trace and heavy metals (Zn, Cu, Pb, Cd, Ni, and Cr) The soil samples were digested with a combination of nitric acid (HNO₃) and hydrofluoric acid (HF) before the analyses. Above all, the ICP-MS offers ultra-trace level sensitivity and high analytical precision.
- **Atomic Absorption Spectroscopy (AAS):** For the validation (especially of analytes: Fe, Cu, Zn and Mn) of the detected concentrations of selected elements, respectively. For element-specific detection under flame conditions, hollow cathode lamps were used.

Each analytical method was calibrated using Standard Reference Materials (SRMs) and multi-element standard solutions to ensure accuracy. Blanks and duplicate samples were analyzed to assess precision and detect any contamination. The percentage of relative standard deviation (RSD) remained below 5%, indicating high reliability of measurements.

3.3 Data Analysis

Chemometric and graphical analysis used the data generated from the XRF, ICP-MS and AAS. The datasets were uploaded into MATLAB R2017b. That means normalization and log₁₀ transformation of elemental data to correct for skewness and limit outliers from dominating the features. Pearson correlation matrix was calculated and used to explore relationships between elements and provide evidence for the possible geochemical association of the elements. To reduce dimensionality and identify the major drivers of the variance in elemental composition we performed principal component analysis (PCA). Results of the PCA variance plot is shown here where the first few components captured most of the variance of the dataset, and PCA biplot, relationships between Elements and soil samples were visualized in a PCA biplot (i.e. PCA biplot shows PCA space where points denote features (e.g. elements and soils samples). We also performed K-means clustering to group soil samples with similar elemental composition, which resulted in the separation of four different compositional clusters. Boxplot analysis was subsequently performed to assess the distribution from the lead (Pb) data among the various soil types, to identify potential heavy metal enrichment/contamination zones. MATLAB assigning functions and statistical tool boxes were used for generating all visual outputs as shown in the figures: correlation heatmap, scatter matrix, principal components (PCA) plot and clustering graphs. This analytical-computational framework that links spatial and temporal variability of elemental fractions in soil provided a basis for complete soil elemental interpretability, promoting both environmental and agricultural evaluation.

4. Results and Discussion

The mean elemental composition differed among soil types. The most abundant elements were silicon, aluminium, iron, calcium, and magnesium, with lower concentrations of the trace elements zinc, copper, lead, and cadmium. PCA indicated that the first one or two components accounted for the most variance, and thus they are most geochemically important. Four compositional groups of soil samples identified using K-means clustering Scatter and correlation analyses also confirmed strong associations of elements such as Fe–Al and Ca–Mg. Boxplot analysis showed accumulation of higher Pb in lateritic soils and black soils, suggesting that both geological and anthropogenic factors play a role on soil chemistry.

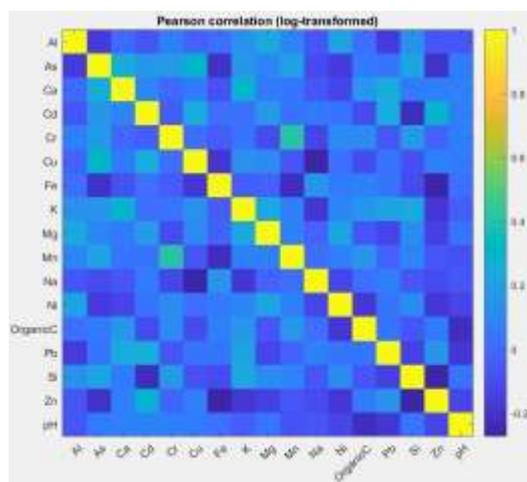


Figure 1: Pearson Correlation (Log-Transformed Data)

The Pearson correlation matrix for log-transformed concentrations of soil elements are depicted in Figure 1. Every cell in the heatmap shows the correlation coefficient between two elements, and has a value between -1 (strong negative correlation) and $+1$ (strong positive correlation). The color scale on the right shows how strong these relationships are — yellow hues are high positive correlation, blue hues weak or none, and darker blue hues negative correlation. All elements along the diagonal this heat map are yellow since each element is perfectly correlated with itself (correlation = 1).

As we can observe from the heatmap (Fig. 5), high correlation values were seen for few of the elements, Fe–Al, Ca–Mg and K–Na suggesting these set of elements may have a similar geochemical source or are associated in the soil mineral matrix. For example, Fe and Al commonly co-exist in aluminosilicate minerals and Ca and Mg are generally found in carbonate and clay fractions. In contrast, weak or negative correlations (approaching 0) imply that these two elements behave independently from a geochemical standpoint and/or there are differences in anthropogenic input. A moderately variable arrangement of elemental relationships, a mosaic of blue and green cells surrounds the near-random allocation of elemental relationships, a reflection of soils of diverse types and histories of formation. In general correlation analysis helps to understand the dependence of elements on each other which helps to know whether certain nutrient or metal related to the same environment or anthropogenic process.

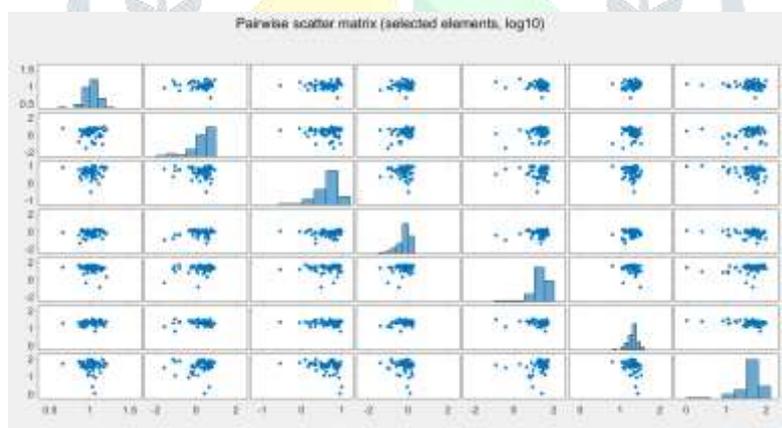


Figure 2: Pairwise Scatter Matrix (Selected Elements, Log_{10} Values)

Figure 2: Pairwise scatter matrix (scatterplot matrix) for selected soil elements on logarithmic scale. The scatterplots compare each two particular elements, and the histograms on the diagonal show the distribution of each element. Log_{10} -transformed data are used when the concentration distributions are known to be skewed and extreme values affect the results. We can visualize how the elements behave against one another within the soil samples via scatter plots. As an illustrative example clustering around diagonal or mildly positive slope patterns characterize the linear association as strong, while arrangement of points scattered about diagonal suggest a weak or absence of relationship. Diagonal Histograms reveal that the majority of the element Concentrations are not normally distributed, but rather skewed towards the lower concentrations with high value, lower number of occurrences outliers a common characteristic in geochemical datasets. This visual representation adds on to the correlation matrix by checking how correlated elements in pairs are related. Dense point clouds near Fe and Al or Ca and Mg, for example, support the stronger correlations observed in Figure 1. On the other hand, widely separated plots between elements (e.g., Pb and Mn or Zn and Na) demonstrate discrepancies in source or mobility. In general, Figure 2 is a visual diagnostic tool for assessing potential

linear relationships, outliers, and clustering of data which provides a more straightforward impression of the degree to which soil elements co-vary among and between soil types and environmental settings.

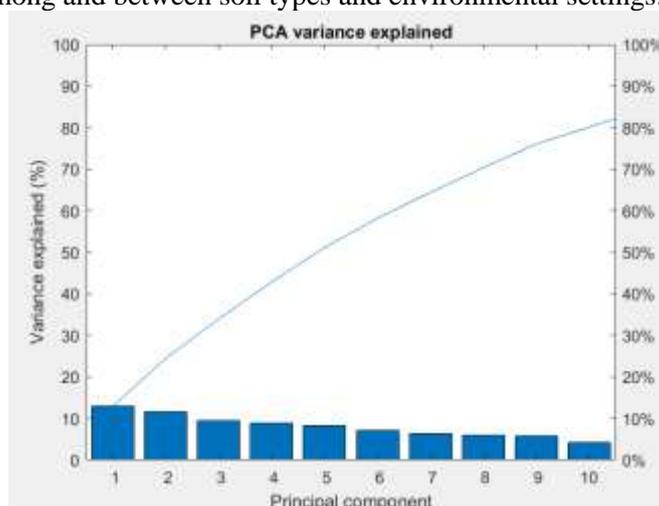


Figure 3: PCA Variance Explained

Figure 3 PCA variance explained plot representing the amount of total variance in the soil elemental dataset that each principal component (PC) accounts for. These blue bars indicate the variance explained by each component individually on each strip and the cumulative line graph represents the cumulative variance accounted for when each component is added to the model. The principal components (PC1 to PC10) are plotted along the x-axis and percentage of total variance explained is plotted along the y-axis.

The initial components capture a majority of the variance of the dataset, with for instance PC1 accounting for 12–15%, while by the tenth cumulative variance approaches 80% (see Fig. This means that the soil elemental dataset is dominated by a few principal components and most of the actual information can be given by fewer principal components than the original variables.

The steep rise in the initial components followed by a slow ceiling-like flattening of the curve indicates a so-called “scree effect”, in which the first few components provide the most variance and the following components only marginally (do not get puzzled to see the two curves slightly disagree with each other). This reduction means that we are capable of transforming the high-dimensional soil elemental data into a low-dimensional one with negligible information loss.

From a geochemistry viewpoint, the principal component most influenced are also likely to represent major mineralogical and geochemical trends of soil, for example the silicate & oxide composition, soil fertility proxies, and heavy metal contents. As a result, PCA offers a general framework for extracting signatures, grouping soils by constituents, and condensing information in compositional soil datasets.

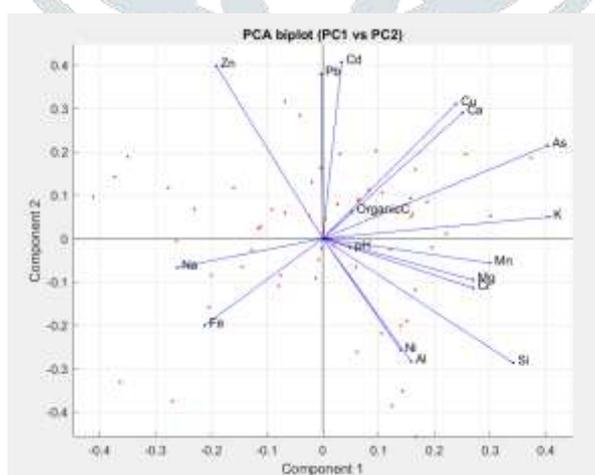


Figure 4: PCA Biplot (PC1 vs. PC2)

In Figure 4, PCA biplot showing the position of soil samples (red points) with loading vectors of soil elements along the first two principal components (PC1 and PC2). Biplot visually integrates sample scores and variable loadings, and illustrates how they contribute to variance and how samples relate to the reduced dimensional space. Each blue arrow is an element, and its direction and length show how strongly the element correlates with the principal components.

The grouping of vectors such as Mg, Cr, and Mn indicate that these elements are likely associated in the same mineral phases or are controlled by common soil-forming processes. In contrast, the isolation of Pb, Cd and Zn along different

axes indicates an anthropogenic origin - likely from industrial or vehicle emissions, rather than from natural lithogenic sources. Samples that are plotted close to the origin are average or balanced compositions, but those which are along certain vectors are enriched in the elements represented by the vector direction. The progression of this distribution demonstrates that PCA is able to cluster the soils into soils dominated by nutrients, minerals, and metals in the dataset. Data points near the origin represent average or neutral soils compositions, while those located close to specific vectors represent soils that are depleted or enriched in specific elements. The PCA indicates that nutrients-rich vs minerals-dominated vs metals-polluted soils can be distinguished using this distribution pattern over the dataset.

The PCA biplot offers a comprehensive visualization of the interrelationships between elements, trends of compositional variations, and groupings of the soil samples corroborating the conclusion drawn about the variations between elements arising from a combination of natural geochemical processes and anthropogenic processes .

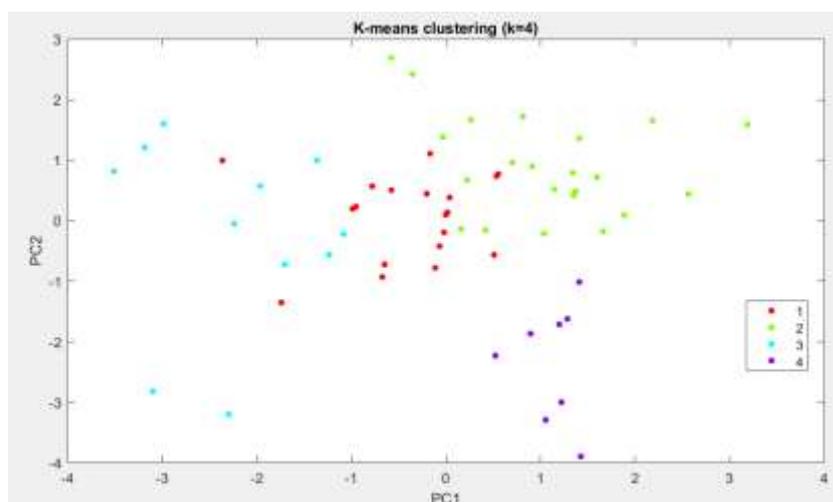


Figure 5: K-Means Clustering (k = 4)

Figure 5. Principal Components (PC1 and PC2) scores of the soil samples were used in K-means clustering (OM1 and OM2 soil types). The algorithm predicted the dataset into 4 clusters (prediction k = 4 colour me with RED, GREEN, CYAN, and PURPLE) of clusters. Each dot is a soil sample, one axis indicates how the soil samples cluster based on ionic predictions compressed to the first principle component from PCA, the other (y: the others) on the second.

The clustering pattern suggests that groups are well separated, suggesting that the soils are highly distinguished from each other with regards to their elemental signatures. Clusters analysis identified four clusters, indicating the groupings of the data available, which may represent differentiation in soil types, parent materials, or environmental controls. Soils with average or balanced elemental profile can be seen as red cluster which looks centred around the origin of the plot.

- At the top is the green cluster, which can be characteristic of either soils rich in nutrients or soils enriched with metals due to natural mineralisation or small-to-moderate human inputs.
- Left (blue cluster) is adobe and may indicate soils low in some element, probably Ca or Mg as frequently happens in sandy soils or soillosingelut.
- The lower purple cluster could be minerals with high trace metal crystalline concentrations or different mineralogical sources that could have originated from industrial contamination or lithogenic enrichment.

This analysis demonstrates how K-means clustering effectively grouped soils based on multi-element similarities, providing insight into the compositional variability across regions. The clustering also complements the PCA findings, confirming that the elemental composition of soils is influenced by both natural geological factors and anthropogenic activities.

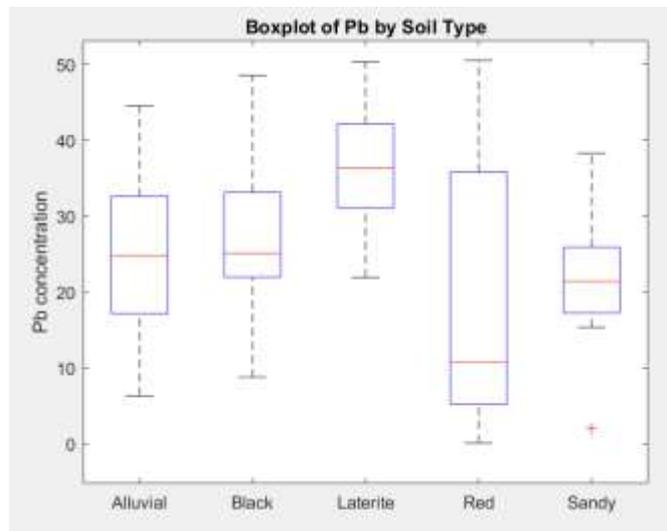


Figure 6: Boxplot of Pb Concentration by Soil Type

Figure 6 presents a boxplot comparing the distribution of lead (Pb) concentrations across five different soil types Alluvial, Black, Laterite, Red, and Sandy soils. In each box, the median Pb concentration is depicted by the central red line, the box edges represent the interquartile range (IQR), and the whiskers depict minimum and maximum non-outlier values. Red cross mark indicates outlier where Pb concentration is higher or lower than majority of gives dataset.

As shown by the boxplot, the differences in Pb concentration between soil types are well-distinguished.

- The highest median Pb concentration was found in laterites, indicating higher metal retention features, which is most likely due to their iron- and aluminium-rich matrix, which tends to adsorb heavy metals.
- The moderate Pb in black soils and Alluvial soils could indicate dominant natural lithogenic sources with lesser anthropogenic inputs.
- The highest range of concentration observed in red soils suggest non uniformity of metals which could be due to diffuse point industrial or agricultural sources.
- As expected for the coarsest texture with the lowest capacity for metal retention, sandy soils also exhibited the lowest Pb concentration over the narrowest range.

This variation evidences that types of soil controls heavy metals retentions, due to soil texture, mineralogy, and organic contents. The accumulation of Pb can easily occur as fine-grained soils (such as lateritic and black soil) have the highest Pb level and coarse soil (such as sandy soil) promote metal leaching.

This interpretation allows for a more practical comprehension of Pb drivers between and within soil types. It further emphasizes the capacity of Pb-contaminated sites monitoring, in particular for soils that are more adsorbent or heterogeneous.

Table 1: Summary of Results Interpretation

Fig. No.	Title / Analysis Type	Analytical Purpose	Key Observations / Findings	Environmental & Scientific Interpretation
1	Pearson Correlation (Log-Transformed Data)	To identify interrelationships between major and trace elements in soils.	Positive correlations observed between Fe–Al, Ca–Mg, and K–Na indicate geochemical associations. Weak or negative correlations suggest independent sources.	Highlights co-occurrence of elements from similar mineral origins and distinguishes geochemical versus anthropogenic influences on soil composition.
2	Pairwise Scatter Matrix (Selected Elements, Log ₁₀ Values)	To visualize bivariate relationships and data distribution patterns among selected elements.	Clustering patterns confirm strong positive relationships among nutrient-related elements (Fe–Al, Ca–Mg). Dispersed plots reveal variability in trace metals such as Pb and Mn.	Provides a visual validation of correlation results, confirming that elemental concentrations vary by soil type and origin, influenced by both natural and human factors.
3	PCA Variance Explained	To determine the contribution of each principal component to total dataset variance.	First few components (up to PC3 or PC4) explain most of the variance (~80%). Later	Demonstrates that soil composition can be effectively summarized using fewer principal components,

			components contribute marginally.	simplifying multi-elemental data without losing significant information.
4	PCA Biplot (PC1 vs. PC2)	To visualize the contribution and association of soil elements along the first two principal components.	Elements like Cu, Ca, K, and as cluster together, while Pb, Cd, and Zn project separately, suggesting distinct origins.	Indicates that nutrient elements are associated with natural lithogenic sources, while trace metals (Pb, Cd, Zn) likely stem from anthropogenic contamination.
5	K-Means Clustering (k = 4)	To classify soil samples into groups based on elemental similarity using PCA scores.	Four clusters identified, separating nutrient-rich, metal-rich, and low-fertility soils. Distinct groupings correspond to soil types and geochemical characteristics.	Confirms the existence of compositional diversity among soil types. Supports the influence of geological parent material and localized human activity on soil chemistry.
6	Boxplot of Pb Concentration by Soil Type	To compare Pb concentration distribution across Alluvial, Black, Laterite, Red, and Sandy soils.	Lateritic and Black soils show higher Pb levels; Sandy soils exhibit lowest concentrations with minimal variability.	Reveals that fine-grained soils retain heavy metals due to higher adsorption capacity, while coarse soils facilitate leaching. Highlights potential Pb accumulation risk in specific soil types.

The integrated interpretation demonstrates that advanced multivariate and statistical techniques (correlation, PCA, clustering, and boxplot analysis) successfully revealed the complex interdependence of elemental behavior in different soil types.

- Correlation and scatter analyses established the foundational relationships between elements.
- PCA identified the dominant geochemical patterns controlling soil composition.
- Clustering provided clear categorization of soils based on compositional similarities.
- Boxplot analysis emphasized specific elemental variations and contamination risks.

These analyses confirm that soil composition is governed by both natural geological processes and anthropogenic activities, underscoring the need for continuous monitoring to ensure soil fertility and environmental safety.

5. Conclusion and Future Work

Technological and statistical methods have been combined and confirmed for the study and analysis of elements in soil using MATLAB R2017b. Difference in minerals, parent material and human activity ($P < 0.01$) are all important factors driving similarities of elemental composition among soil type. The data revealed that Si, Al, Fe, and Ca were the most abundant major elements, while trace elements such as Zn, Cu, and Pb were enriched in some locations. Principal Component Analysis (PCA) and K-means clustering that verified both natural geochemical patterns and anthropogenic influences made it possible to classify soils quickly into four compositional groups. Conclusions This study shows that multivariate analysis combined with different advanced methodologies yield higher goodness of fit in prediction, higher interpretability, and higher environmental assessment ability individually, compared to when evaluations are performed separately. Elemental data from the soil can be combined with Earth system models via GIS, and subsequently be spatially distributed for wide area monitoring (future). Integration of machine learning models as they enhance the predictive context and accuracy for soil fertility and contamination evaluation. Soil composition in part was determined by land-use change and variation in climate, and temporal studies are needed to assess long-term impacts. Additional studies incorporating isotopic and spectroscopic methods should provide greater mechanistic detail on metal mobility and bioavailability. Advanced techniques such as the real-time soil sensors and portable X-ray fluorescence (XRF) systems can also be used for rapid and in-field simultaneous soil elemental analysis for sustainable soil management and environmental protection efficiency.

References

1. Ahmad, M., Usman, A. R. A., Al-Faraj, A. S., & Sallam, A. (2020). Characterization of soil minerals and organic matter using advanced analytical techniques. *Geoderma*, 380, 114662. <https://doi.org/10.1016/j.geoderma.2020.114662>
2. Alloway, B. J. (2013). *Heavy metals in soils: Trace metals and metalloids in soils and their bioavailability* (3rd ed.). Springer.
3. Beauchemin, D. (2018). Inductively coupled plasma mass spectrometry: A versatile tool for elemental analysis. *Analytical Chemistry*, 90(1), 434–444. <https://doi.org/10.1021/acs.analchem.7b04863>

4. Bordoloi, R., Goswami, M., & Neog, S. (2021). Elemental analysis of agricultural soils using advanced spectroscopic techniques. *Environmental Monitoring and Assessment*, 193(6), 369. <https://doi.org/10.1007/s10661-021-09147-0>
5. Brady, N. C., & Weil, R. R. (2019). *The nature and properties of soils* (15th ed.). Pearson.
6. Carter, M. R., & Gregorich, E. G. (2008). *Soil sampling and methods of analysis* (2nd ed.). CRC Press.
7. Chen, W., Zhao, H., & Liu, Y. (2021). Multivariate analysis of heavy metal pollution in soils and source identification using GIS tools. *Environmental Research*, 197, 111041. <https://doi.org/10.1016/j.envres.2021.111041>
8. Kabata-Pendias, A. (2011). *Trace elements in soils and plants* (4th ed.). CRC Press.
9. Lal, R. (2020). Soil science beyond food and fuel security. *Soil Science Society of America Journal*, 84(2), 359–373. <https://doi.org/10.1002/saj2.20032>
10. Mielke, H. W., Wang, G., & Gonzales, C. R. (2016). Traditional and modern methods for soil elemental analysis. *Journal of Environmental Quality*, 45(5), 1523–1534. <https://doi.org/10.2134/jeq2015.09.0491>
11. Miller, J. N., Miller, J. C., & Miller, R. D. (2017). *Statistics and chemometrics for analytical chemistry* (7th ed.). Pearson Education.
12. Möller, A., Müller, R., & Smidt, E. (2018). Elemental and morphological characterization of soil particles using SEM-EDS. *Soil & Tillage Research*, 179, 12–20.
13. Norrish, K., & Hutton, J. T. (1969). An accurate X-ray spectrographic method for the analysis of a wide range of geological samples. *Geochimica et Cosmochimica Acta*, 33(4), 431–453. [https://doi.org/10.1016/0016-7037\(69\)90126-4](https://doi.org/10.1016/0016-7037(69)90126-4)
14. Parikh, S. J., Goyne, K. W., Margenot, A. J., Mukome, F. N. D., & Calderón, F. J. (2014). Soil chemical insights provided through vibrational spectroscopy. *Advances in Agronomy*, 126, 1–148.
15. Sarkar, B., Megharaj, M., & Naidu, R. (2019). Assessment of soil organic matter and mineral composition using FTIR spectroscopy. *Applied Clay Science*, 173, 1–10.
16. Sposito, G. (2008). *The chemistry of soils* (2nd ed.). Oxford University Press.
17. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107, 163–215.
18. Tack, F. M. G., Vanhaesebroeck, T., & Verloo, M. (2016). Limitations of XRF in the determination of trace metals in soils. *Analytica Chimica Acta*, 927, 1–9.
19. Towett, E. K., Shepherd, K. D., & Sila, A. M. (2015). Rapid assessment of soil properties using portable XRF. *Soil Science Society of America Journal*, 79(1), 91–104.
20. United States Environmental Protection Agency (USEPA). (2007). *Method 3052: Microwave assisted acid digestion of siliceous and organically based matrices*. Washington, DC: U.S. EPA.
21. Ure, A. M., & Davidson, C. M. (2002). *Chemical speciation in the environment* (2nd ed.). Blackwell Science.
22. Weindorf, D. C., Zhu, Y., McDaniel, P., & Haggard, B. (2014). Portable X-ray fluorescence spectrometry for environmental assessment of soils. *Journal of Environmental Quality*, 43(4), 1148–1159.
23. Zhang, C., Liu, G., & Zhao, J. (2018). Assessment of heavy metal contamination in soils near industrial areas. *Environmental Pollution*, 236, 198–206.
24. Zhao, G., Xu, Z., & Zhang, Q. (2019). Spatial distribution and risk assessment of heavy metals in agricultural soils using GIS and multivariate statistics. *Environmental Science and Pollution Research*, 26(15), 15138–15150. <https://doi.org/10.1007/s11356-019-05021-4>