

STATISTICAL ASSESSMENT OF HEAVY METAL CONTAMINATION IN INDUSTRIAL SOILS USING POLLUTION INDICES AND MULTIVARIATE ANALYSIS

Dr. James Wilson^{1*}, Dr. Elena García², Dr. Thomas Fischer³

¹*Department of Environmental Statistics, Example University, Manchester, United Kingdom*

²*Institute of Environmental Sciences, Example University, Madrid, Spain*

³*Department of Environmental Chemistry and Data Analysis, Example Institute, Hamburg, Germany*

**Corresponding Author: Dr. James Wilson*

Abstract

The total concentrations of metals in industrial soils may not adequately reflect enrichment, mobility, bioavailability, or multivariate contamination patterns, thereby necessitating an integrated statistical assessment. This study assessed Cd, Cr, Mn, Ni, Pb, and Zn using 100 soil observations obtained from 50 sampling locations at two depths (0–30 and 30–60 cm). The Wilcoxon signed-rank test, bioavailable fractions, contamination factor, geo-accumulation index, Pollution Load Index, degree of contamination, Spearman correlation, principal component analysis, hierarchical cluster analysis, and Hotelling's T² test were among the descriptive statistics and statistical analyses. Cd showed the highest contamination factor at both depths, with mean relative enrichment of approximately 3.47 and positive geo-accumulation values at all sampling locations. Pb also exhibited notable enrichment, whereas mean PLI values remained below 1; however, 14% of surface and 10% of subsurface sites exceeded this threshold, indicating localized contamination hotspots. Zn and Pb showed the highest relative bioavailable fractions. Contamination indices and several metals were positively associated with cation-exchange capacity and soil organic matter, highlighting the influence of soil physicochemical characteristics on metal accumulation. PCA showed strong contributions of Cr, Ni, Pb, Cd, and CEC to the dominant component, while HCA identified close Cr–Ni and Pb–Zn associations. Hotelling's T² analysis further demonstrated a significant overall difference in heavy-metal profiles between soil depths. These findings demonstrate that combining pollution indices with multivariate statistical techniques provides a comprehensive framework for characterizing contamination intensity, vertical variability, inter-metal relationships, and site-specific contamination hotspots in industrial soils.

Keyword: - heavy metals; industrial soils; pollution indices; principal component analysis; hierarchical cluster analysis

1. INTRODUCTION

Soils contaminated with heavy metals are becoming a significant environmental issue since potentially toxic elements can remain in soils for extended periods, bioaccumulate in the terrestrial environment and become part of exposure pathways for both the environment and humans. Metal levels can be significantly increased above background concentrations due to industrialization, mining, smelting, fossil-fuel processing, waste disposal and intensive agricultural activities. Metals are not biodegradable like other organic pollutants, and can persist in soil matrices for long periods, causing long-term environmental management problems (Hou et al., 2020). The industrial zones are also regarded as a significant area of risk as the continuous emission, deposition of waste, fallout from the atmosphere and accidental spillages may add more than one metal to the surrounding soils (Verma et al., 2021).

The most often tested pollutants in industrial and peri-industrial soils are cadmium (Cd), chromium (Cr), manganese (Mn), nickel (Ni), lead (Pb), and zinc (Zn). They also have environmental relevance due to their mobile, bioavailable, physicochemical, and soil property interactions, as well as total concentration. Research around smelting and industrial sites have shown significant enrichment and complex fractionation patterns of Pb and Zn related to historic industrial processes (Adnan et al., 2025). Likewise, it has been noted that in southern India there are significant differences between the distribution of heavy metals at various depths with the soil, and therefore, a single depth sampling is not enough (Mitran et al., 2024).

A typical pollution assessment approach is to use quantitative pollution indices to assess soil contamination by comparing the measured concentration with a reference or background concentration. The contamination factor (CF), the geo-accumulation index (I_{geo}), the pollution load index (PLI) and the degree of contamination (DOC) are indicators that give complementary information on the enrichment of individual metals and cumulative contamination. With the complexity of multi-element pollution, there is an increasing application of integrated index-based approaches (Imoni & Jonathan, 2025). In recent years, studies on industrial soils have also highlighted the value of using multi-index approaches with native or geochemical reference concentrations for better interpretation of contamination (Tarighat et al., 2025).

However, the concentration of total metals does not always indicate the mobility of the metal in the environment and/or potential biological exposure. The bioavailable or extractable fractions can potentially give further insights into the available or mobilizable part of the metal pool. The retention and release of metals can be greatly affected by soil physicochemical properties such as pH, organic matter, electrical conductivity, calcium carbonate content, and cation-exchange capacity (Xu et al., 2022). In recent years, too, there has been research that accounts for bioaccessibility and availability, which has indicated that metal speciation and metal availability can significantly affect the interpretation of exposure when compared to those based on bulk concentrations only (Li et al., 2024; Zhang et al., 2024).

Industrial-soil data typically include several metals and soil properties that can be correlated, making multivariate statistical methods advantageous for detecting underlying structures which would not be apparent from univariate comparisons. While hierarchical cluster analysis (HCA) can be used to categorize metals based on similarities in their distributional profiles, principal component analysis (PCA) can be used to reduce the dimension and identify groups of factors that explain common variance. In recent contamination studies, these methods have been successfully used to interpret interrelationships between the associated metals, soil properties and potential source related patterns (Bashir et al., 2025; Islam et al., 2025). Furthermore, multivariate methods have been used in industrial and agricultural environments for source interpretation and enhancing the statistical description of complicated contamination profiles (Saraswat et al., 2023).

Another major aspect of heavy-metal contamination is the spatial and site-specific heterogeneity. There may also be instances where the average pollution index is relatively low, but the pollution index at a particular sampling location is quite high. Local hotspot areas can be found in conjunction with regional average contamination levels, as demonstrated by various studies that have employed pollution indices and multivariate analyses (Lin et al., 2024). Some similar results have been observed in industrial estates where both multivariate and ecological-risk approaches revealed significant differences in sampling site and metals (Karthikayini et al., 2025).

Based on this, the present study comprehensively statistically assesses the level of heavy metals contamination in industrial soils by applying concentration distribution, depth-wise comparison, bioavailable fraction, integrated pollution index, correlation analysis, PCA, HCA and multivariate comparison of metal profiles. The analysis considers two soil depths and Cd, Cr, Mn, Ni, Pb, and Zn relationships with the most important physicochemical soil properties. The study combines traditional pollution indexes to multivariate statistical techniques, to get a more complete pollution characterization of intensity, depth, association and sampling sites with high concentrations.

2. MATERIALS AND METHODS

2.1 Study Data and Variables

The dataset was obtained from an industrial area, with soil physicochemical properties and heavy metal concentrations (Stefanou, 2026) and analysed. The data was collected from 50 soil sampling points at a depth of 0-30cm and 30-60cm, making a total of 100 soil observations. Total concentrations of cadmium (Cd), chromium (Cr), manganese (Mn), nickel (Ni), lead (Pb) and zinc (Zn) were available in mg kg⁻¹ for each depth. The available metal concentrations were added for comparison. The soil physicochemical properties analysed were pH, soil organic matter (SOM), electrical conductivity (EC), calcium carbonate (CaCO₃) and cation-exchange capacity (CEC).

2.2 Descriptive and Depth-Wise Comparison of Heavy Metals

For every heavy metal at every soil depth, descriptive statistics (mean, standard deviation, median, minimum, and

maximum concentrations) were calculated. Mean concentrations were used to compare the overall distribution of metals in both surface and subsurface soils.

The two depth readings were collected at the same locations, and statistical comparisons were made on a paired basis between the depths to assess depth-wise differences in total metal concentrations. To determine if the concentrations of metals were significantly different between the 0–30 cm and 30–60 cm soil layers, the Wilcoxon signed-rank test was performed. A p value of less than 0.05 was statistically significant. The percentage difference between the two depths was determined as:

$$\text{Difference (\%)} = \frac{C_{\text{surface}} - C_{\text{subsurface}}}{C_{\text{subsurface}}} \times 100$$

where C_{surface} and $C_{\text{subsurface}}$ represent the concentrations measured at 0–30 cm and 30–60 cm, respectively.

2.3 Bioavailable Fraction of Heavy Metals

The relative bioavailability of each metal was estimated as the concentration available relative to the total concentration:

$$\text{Bioavailable Fraction (\%)} = \frac{C_{\text{available}}}{C_{\text{total}}} \times 100$$

Separate averages, standard deviations, medians, minimums and maximums were computed for the bioavailable fractions of each metal at each soil depth. These values were used to decide the percentage of the total soil metal pool that was potentially available to the plant.

2.4 Heavy Metal Contamination Indices

The Contamination Factor (CF), Geo-accumulation Index (I_{geo}), Pollution Load Index (PLI), and Degree of Contamination were used to assess the contamination of heavy metals

The contamination factor for each metal was calculated as:

$$CF_i = \frac{C_i}{B_i}$$

where C_i is the measured concentration of metal i and B_i is the corresponding geochemical background concentration. The background concentrations applied in the analysis were Cd = 0.3, Cr = 90, Mn = 850, Ni = 68, Pb = 20, and Zn = 95 mg kg⁻¹.

The geo-accumulation index was determined by the following formula:

$$I_{\text{geo}} = \log_2 \left(\frac{C_i}{1.5B_i} \right)$$

where the factor 1.5 accounts for possible natural variability in background concentrations.

The degree of contamination was calculated as the sum of the contamination factors of all six metals:

$$C_d = \sum_{i=1}^n CF_i$$

The Pollution Load Index was calculated from the geometric mean of the individual contamination factors:

$$PLI = (CF_1 \times CF_2 \times \dots \times CF_n)^{1/n}$$

where $n=6$ represented the six metals included in the analysis. PLI values were summarized separately for the two soil depths, and the proportion of sampling locations with $PLI > 1$ was determined.

2.5 Correlation Analysis

Spearman's rank correlation coefficient (r_s) was used to study the relationships between soil physicochemical properties and total heavy metal contents in surface layers (0–30 cm). Correlation analysis was done between the total Cd, Cr, Mn, Ni, Pb, Zn concentrations and the measured soil properties such as pH, SOM, EC, CaCO₃, and CEC. A correlation matrix was used to summarize the resulting correlation coefficients, in order to determine the direction and relative strength of association of soil characteristics with accumulation of metals.

2.6 Principal Component Analysis

The multivariate analysis of the structure of the total measure of heavy metals and the physicochemical properties of soils was analyzed using principal component analysis (PCA). The variables measured were total concentrations of Cd, Cr, Mn, Ni, Pb and Zn, pH, EC, CaCO₃, soil organic matter (SOM) and cation-exchange capacity (CEC). All variables were z-scored prior to PCA as they were measured on different scales.

Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were used to assess the adequacy of the data for PCA. The standardized data set was subjected to Principal components analysis and the percentage of total variance accounted for by each principal component was determined. The loadings of the components were explored to determine the relative importance of individual metals and soil properties in each component extracted. Several components were retained for interpretation depending on the amount of variance in the data they explain, and in this case, the first three principal components were retained. To understand the association between the variables and the distribution of the observations at both depths, a PCA biplot was drawn.

2.7 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) was performed to identify groups of heavy metals exhibiting similar concentration

profiles across the combined surface and subsurface observations. Total concentrations of Cd, Cr, Mn, Ni, Pb, and Zn were analysed across the 100 soil observations.

The Spearman rank correlation coefficient was used to determine similarity between metals and the dissimilarity measure was defined as:

$$D_{ij} = 1 - r_{s,ij}$$

where D_{ij} represents the dissimilarity between metals i and j , and $r_{s,ij}$ represents their Spearman correlation coefficient.

The average-linkage method was next used for agglomerative hierarchical clustering. This cluster structure was displayed in a dendrogram. The dendrogram was analysed to see if metals were highly similar in their distribution profiles and to find the main metal groupings.

2.8 Multivariate Comparison of Heavy Metal Profiles Between Soil Depths

A paired multivariate comparison using Hotelling's T^2 test was performed to determine whether the overall heavy-metal composition differed between the two matched soil depths. The multivariate response consisted of total concentrations of Cd, Cr, Mn, Ni, Pb and Zn measured at 0–30 cm and 30–60 cm at the same 50 sampling points.

The differences between surface and subsurface concentrations were determined for each metal for each sampling point. The significance of the vector of mean paired differences across the six metals was tested using the Hotelling's T^2 statistic. The T^2 statistic was transformed into the appropriate F statistic, and statistical significance was determined using a p value of < 0.05 .

Standardized paired effect sizes were computed as:

$$d_z = \frac{\bar{D}}{SD_D}$$

where $\bar{D}$ is the mean paired difference and SD_D is the standard deviation of the paired differences.

2.9 Relationships Between Pollution Indices and Soil Physicochemical Properties

A Spearman correlation analysis was used to examine the associations between integrated soil contamination and physico-chemical characteristics. The analysis was carried out in the two soil depths 0–30 cm and 30–60 cm.

Two integrated pollution measures have been considered: Pollution Load Index (PLI) and degree of contamination. All of the indices were correlated with the pH, EC, CaCO_3 , SOM, and CEC. Each relationship was assessed using Spearman correlation coefficient (r_s) to determine the direction and level of the relationship. The value of $p < 0.05$ was used as the criterion for statistical significance.

2.10 Identification of High-Contamination Sampling Points

The degree of contamination and calculated Pollution Load Index (PLI) were used to determine site-specific contamination at each of the 50 sampling sites for both soil depths. Sampling points with $\text{PLI} > 1$ were identified as locations where the integrated contamination level exceeded the adopted geochemical background condition represented by the index.

The number and the percentage of the sampling points above these limits were calculated separately for the 0–30 cm and 30–60 cm layers. The sampling points that were found to be above the threshold at both depths were also flagged to find locations where the combined contamination was found to be elevated across the depth sampled. PLI-values at the sampling points were plotted against sampling-point number, using $\text{PLI} = 1$ as reference.

2.11 Data Presentation

Grouped bar charts were used to show the concentrations of the heavy metals at the two soil depths. The concentrations of the metals varied greatly, therefore the total metal concentration was plotted on a Log scale. Bioavailable fractions and contamination factors were graphically shown with regard to metal and depth and relationships between soil characteristics and heavy metal contents were presented by a Spearman correlation matrix.

A PCA biplot and HCA dendrogram were used to present the multivariate patterns. A sampling-point profile of PLI values was shown to represent the site-specific variation in integrated contamination for the basis of a reference line ($\text{PLI} = 1$). Descriptive tables, principal component loadings, correlation coefficients, multivariate test statistics and site-specific pollution-index values were used to summarize statistical results.

2.12 Ethical Consideration

This research used only an existing secondary dataset and did not engage in direct research on human subjects or experimental animals. There was no personally identifiable or confidential information included in the dataset or analysis. Data were used only for scientific analysis and original data measurements were kept intact without changes.

3. RESULTS

3.1 Heavy Metal Concentrations and Vertical Distribution

There were significant differences in the levels of the six metals measured and the pattern of spatial variation in industrial soils (Table 1). In the 0–30 cm layer, Mn was the quantitatively dominant metal, with a mean concentration of $500.31 \pm 354.65 \text{ mg kg}^{-1}$, followed by Zn ($66.55 \pm 71.02 \text{ mg kg}^{-1}$), Pb ($34.67 \pm 9.45 \text{ mg kg}^{-1}$), Ni ($21.25 \pm 4.92 \text{ mg kg}^{-1}$), Cr ($20.58 \pm 5.68 \text{ mg kg}^{-1}$), and Cd ($1.04 \pm 0.25 \text{ mg kg}^{-1}$). The overall concentration hierarchy was similar at 30–60 cm with Mn averaging $486.09 \pm 292.94 \text{ mg kg}^{-1}$ and Cd averaging $1.04 \pm 0.29 \text{ mg kg}^{-1}$.

Table 1. Descriptive statistics of total heavy metal concentrations at the two soil depths (mg kg⁻¹)

Depth	Metal	Mean $\pm$ SD	Median	Minimum	Maximum
0–30 cm	Cd	1.04 $\pm$ 0.25	1.05	0.61	1.82
	Cr	20.58 $\pm$ 5.68	19.58	12.34	36.42
	Mn	500.31 $\pm$ 354.65	397.35	253.49	2688.31
	Ni	21.25 $\pm$ 4.92	20.36	14.35	38.82
	Pb	34.67 $\pm$ 9.45	32.14	19.97	56.06
	Zn	66.55 $\pm$ 71.02	52.98	26.97	531.46
30–60 cm	Cd	1.04 $\pm$ 0.29	1.04	0.47	1.82
	Cr	20.90 $\pm$ 5.50	20.72	9.29	36.85
	Mn	486.09 $\pm$ 292.94	415.22	164.70	1916.62
	Ni	22.69 $\pm$ 6.76	22.15	10.67	47.93
	Pb	33.06 $\pm$ 10.09	31.44	13.78	57.18
	Zn	49.10 $\pm$ 20.56	43.99	22.39	148.97

The large ranges of Mn and Zn showed significant spatial variation. Surface Zn ranged from 26.97 to 531.46 mg kg⁻¹, while Mn ranged from 253.49 to 2688.31 mg kg⁻¹. As can be seen in Figure 1, the depth-specific concentrations of the metals are shown on a logarithmic scale and most metals had similar mean concentrations for both layers, with a more distinct surface enrichment for Zn.

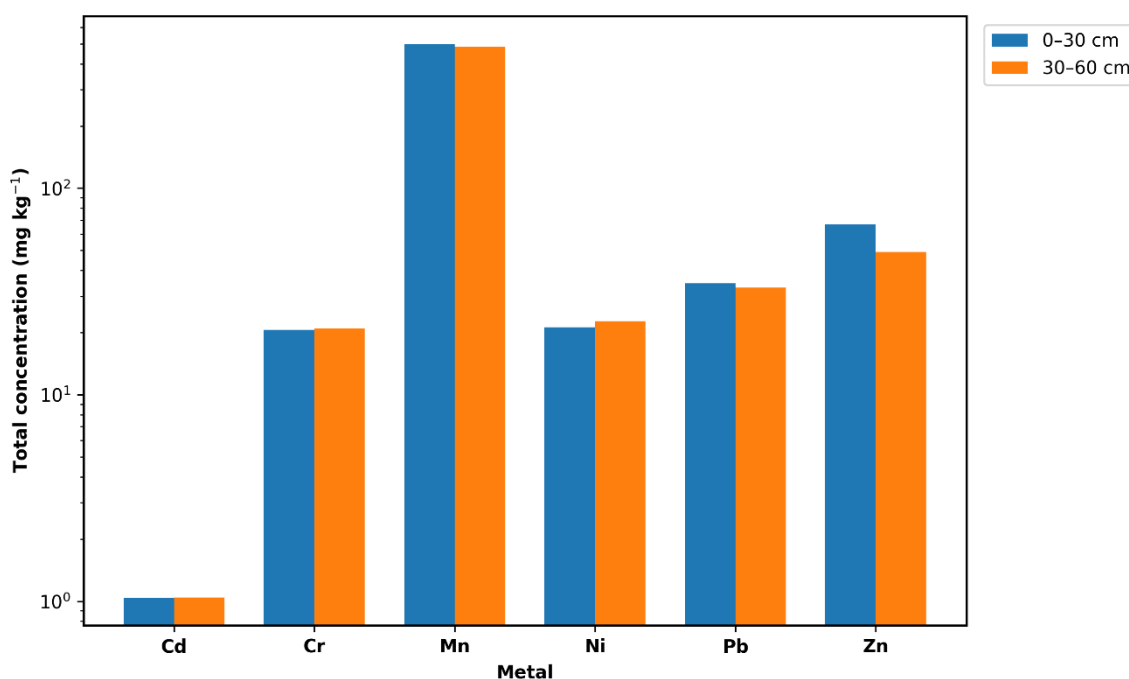


Figure 1. Mean total concentrations of metals in surface (0–30 cm) and subsurface (30–60 cm) industrial soils, presented on a logarithmic scale

Further depth comparisons in pairs indicated that there were no significant vertical differences at the 5% level for the elements Cd, Cr, Mn, Ni, and Pb (Table 2). The Wilcoxon signed-rank test indicated that there was a clear distributional difference in Zn in the matched soil layers ($p < 0.001$), in contrast. The mean concentration of the surface layer of Zn is about 1.35 times the mean concentration of the subsurface layer. The results of the non-parametric test showed that there was a tendency towards higher concentrations of Zn in the surface layer, although the paired t-test did not achieve the $p < 0.05$ threshold due to the high level of extreme samples in the surface layer and a strongly right skewed surface distribution.

Table 2. Paired comparison of total metal concentrations between the 0–30 and 30–60 cm soil layers

Metal	Surface mean	Subsurface mean	Difference (%)	Wilcoxon W	p-value
Cd	1.040	1.042	−0.2	563	0.472
Cr	20.583	20.903	−1.5	526	0.287
Mn	500.312	486.088	+2.9	625	0.909
Ni	21.248	22.695	−6.4	454	0.077
Pb	34.665	33.058	+4.9	498	0.181
Zn	66.551	49.102	+35.5	238	<0.001

Note: Percentage difference = $[(\text{surface} - \text{subsurface})/\text{subsurface}] \times 100$.

3.2 Bioavailable Metal Fractions

The fractions of the total metal pool available varied by a large factor between elements (Table 3 and Figure 2). In surface soil, Zn had the highest mean available fraction (9.23%), followed by Pb (8.69%), Cd (5.85%), Mn (2.56%), Ni (2.22%), and Cr (0.05%). The same trend was found in the deeper layer, where Zn and Pb were the metals with highest relative available fractions.

Table 3. Bioavailable fractions of heavy metals relative to their corresponding total concentrations

Depth	Metal	Mean (%)	SD	Median (%)	Range (%)
0–30 cm	Cd	5.85	2.66	5.66	1.43–12.12
	Cr	0.05	0.02	0.04	0.01–0.15
	Mn	2.56	1.25	2.37	0.43–6.99
	Ni	2.22	1.05	1.97	0.79–6.76
	Pb	8.69	2.32	8.28	4.74–16.36
	Zn	9.23	3.37	8.87	2.56–19.55
30–60 cm	Cd	4.04	2.55	3.71	0.60–11.66
	Cr	0.06	0.06	0.04	0.02–0.40
	Mn	2.97	1.32	2.64	0.79–7.36
	Ni	2.04	1.52	1.54	0.73–9.03
	Pb	8.44	2.69	8.01	4.27–14.01
	Zn	8.97	4.36	8.00	2.63–20.96

However, the relative availability of Cr was very low, as illustrated in Figure 2, although its measurable total concentration was quite high. On the other hand, the available fractions of Zn and Pb were significantly higher. The available fraction of Cd was also higher in the surface layer than in the subsurface layer with a mean of 5.85% versus 4.04% respectively. The results showed that total concentration was not directly related to the potentially mobile and/or available fraction of metal.

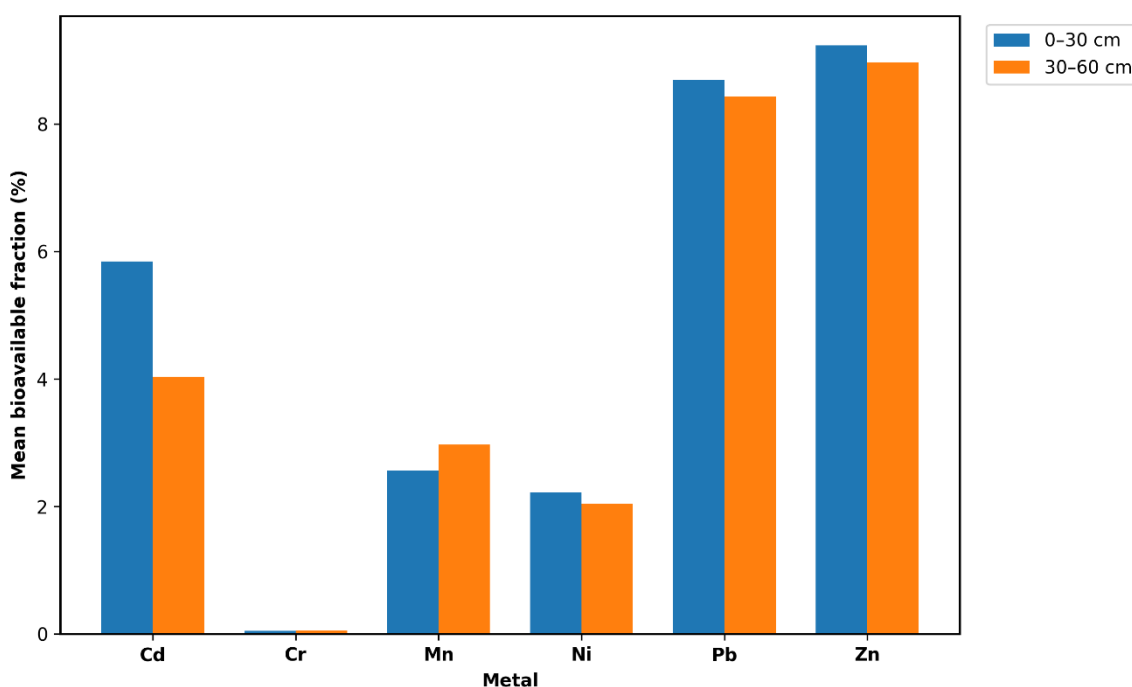


Figure 2. Mean bioavailable fraction of each heavy metal in the 0–30 and 30–60 cm soil layers, expressed as percentage of the corresponding total concentration

There were significant variations in contamination factors among the different metals (Table 4). Cd was the most enriched, with mean CF values of 3.47 at both depths. Pb also exceeded the adopted background level, with mean CF values of 1.73 and 1.65 in surface and subsurface soils, respectively. By contrast, the mean CF values for Cr, Mn, Ni and Zn were below 1.0 at both depths (Figure 3).

3.3 Integrated Heavy Metal Contamination Indices

There were significant variations in contamination factors among the different metals (Table 4). Cd was most enriched, with mean CF values of 3.47 at both depths. This exceeded background concentrations as well with mean concentrations of 1.73 and 1.65 for surface and subsurface soils, respectively for Pb. By contrast, the mean CF values for Cr, Mn, Ni and Zn were both lower than 1.0 at both depths (Figure 3).

Table 4. Contamination factor and geo-accumulation index of metals at the two soil depths

Depth	Metal	Mean CF	CF range	Mean Igeo	Maximum Igeo	Samples with Igeo > 0 (%)
0–30 cm	Cd	3.47	2.04–6.07	1.17	2.02	100
	Cr	0.23	0.14–0.40	–2.76	–1.89	0
	Mn	0.59	0.30–3.16	–1.50	1.08	2
	Ni	0.31	0.21–0.57	–2.30	–1.39	0
	Pb	1.73	1.00–2.80	0.16	0.90	64
	Zn	0.70	0.28–5.59	–1.36	1.90	4
30–60 cm	Cd	3.47	1.58–6.06	1.15	2.01	100

	Cr	0.23	0.10–0.41	–2.74	–1.87	0
	Mn	0.57	0.19–2.25	–1.54	0.59	4
	Ni	0.33	0.16–0.70	–2.23	–1.09	0
	Pb	1.65	0.69–2.86	0.07	0.93	60
	Zn	0.52	0.24–1.57	–1.63	0.06	2

The geo-accumulation values confirmed the significance of Cd. At 0–30 cm, the mean I_{geo} values were 1.17 and at 30–60 cm, 1.15 and all 100 observations gave a positive Cd I_{geo} values. Of the surface samples, 64% had positive I_{geo} values and of the subsurface samples, 60% had positive I_{geo} values. All Cr and Ni observations were negative I_{geo} values, however.

The integrated PLI averaged 0.72 for surface soil, and 0.70 for subsurface soil. However, despite the overall mean value being below unity, PLI exceeded 1.0 for a high percentage of surface sites (14%) and a high percentage of subsurface sites (10%). PLI ranged from 0.46 to 1.13 at 0–30 cm and from 0.31 to 1.17 at 30–60 cm. The mean degree of contamination (ΣCF) was found to be 7.03 and 6.78 respectively.

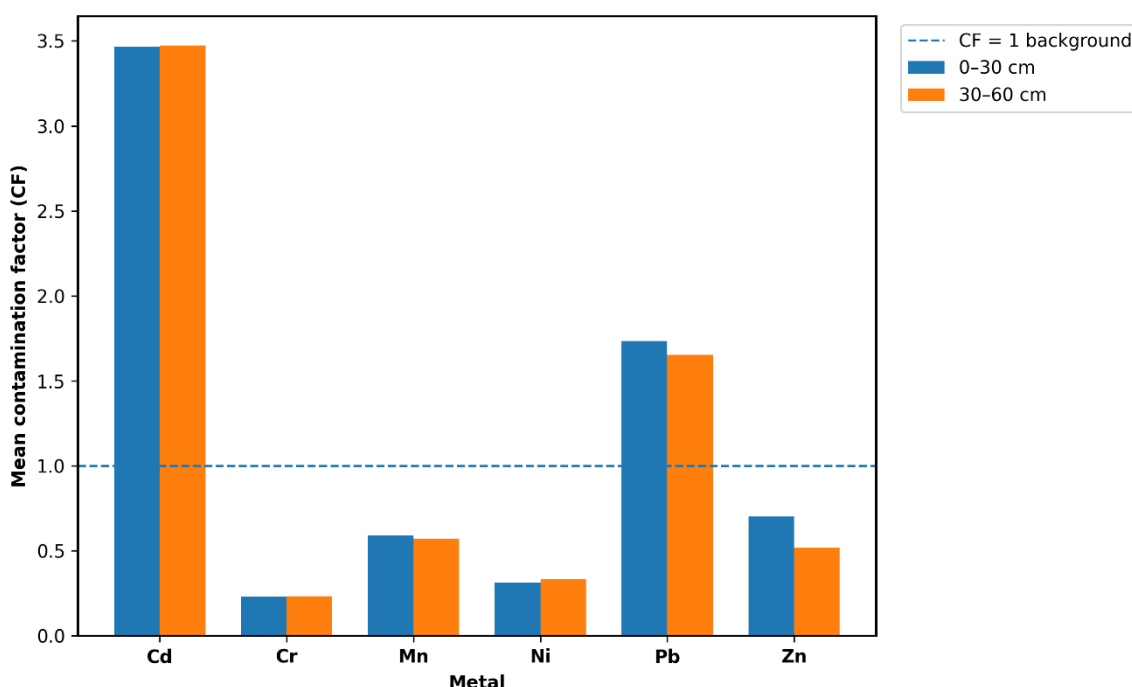


Figure 3. Mean contamination factors (CF) of the six heavy metals at both soil depths, with CF = 1 indicating equivalence to the adopted geochemical background concentration

3.4 Relationships Between Soil Properties and Metal Accumulation

Several relationships were found between surface-soil physicochemical properties and concentrations of total metals using Spearman correlation analysis (Figure 4). Cation-exchange capacity (CEC) showed the strongest positive associations with Cr ($r_s = 0.83$), Ni ($r_s = 0.80$), Pb ($r_s = 0.73$), Zn ($r_s = 0.56$), Mn ($r_s = 0.55$), and Cd ($r_s = 0.49$). The consistently positive coefficients suggested that accumulation of metals in soils increased with increasing exchange capacity.

Soil organic matter (SOM) also showed positive relationships with Zn ($r_s = 0.49$), Mn ($r_s = 0.46$), and Pb ($r_s = 0.41$). There was a positive relationship between electrical conductivity and all six metals, but the coefficient values were relatively small. CaCO_3 had weak negative correlations with Cr, Mn, Ni, Pb and Zn, however. There were only weak correlations between metal concentrations and soil pH.

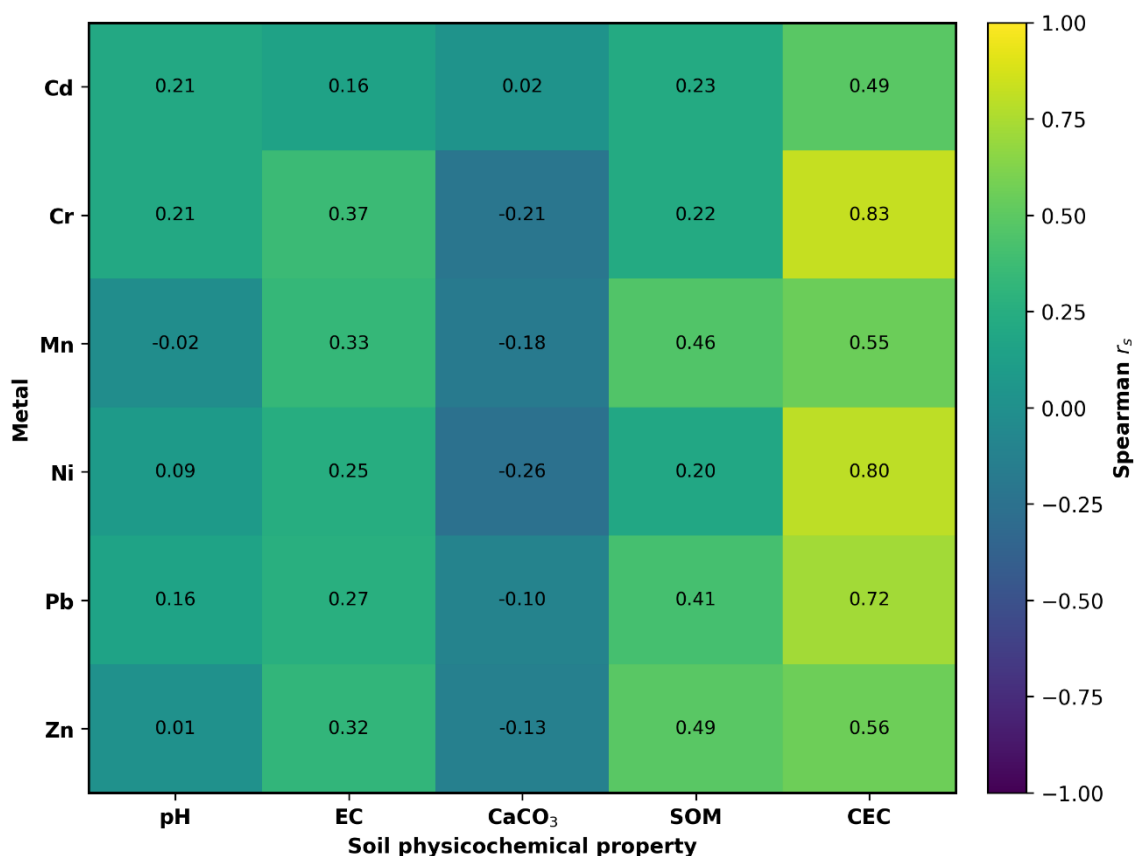


Figure 4. Spearman correlation matrix showing relationships between total heavy metal concentrations and physicochemical properties of surface soils (0–30 cm)

3.5 Principal Component Analysis of Heavy Metals and Soil Properties

Principal component analysis (PCA) was used to study the two soil depths simultaneously, using standardized concentrations of Cd, Cr, Mn, Ni, Pb and Zn together with pH, electrical conductivity (EC), CaCO₃, soil organic matter (SOM) and cation-exchange capacity (CEC). A Kaiser–Meyer–Olkin value of 0.734 supported the suitability of the dataset for multivariate reduction. Bartlett's test of sphericity was significant, $\chi^2(55) = 510.83$, $p < 0.001$, confirming sufficient correlation among variables for PCA.

The first three principal components accounted for the larger part (63.29%) of the total variance (Table 5). PC1 accounted for 39.51%, PC2 for 12.28%, and PC3 for 11.50% of the variance. PC1 was characterized by strong positive loadings for Pb (0.881), Cr (0.878), Ni (0.868), CEC (0.839), and Cd (0.753). SOM also had a positive loading on PC1 (0.576). This part, therefore, was seen as a wide range of simultaneous metal deposition correlated with increased soil exchange capacity.

PC2 was primarily characterized by CaCO₃ (0.739) and pH (0.721), whereas metal loadings on this component were comparatively weak. PC3 was characterized mainly by Zn (0.667), followed by CaCO₃ (0.407) and Mn (0.399), while EC exhibited a negative loading of -0.484 . Thus, Zn displayed a partly distinct multivariate pattern from Cr, Ni, Pb, and Cd, which were primarily associated with PC1.

Table 5. Principal component loadings for heavy metal concentrations and soil physicochemical properties

Variable	PC1	PC2	PC3
Cd	0.753	0.293	−0.030
Cr	0.878	0.007	−0.269
Mn	0.444	−0.309	0.399
Ni	0.868	−0.134	−0.208
Pb	0.881	0.022	0.050
Zn	0.305	−0.233	0.667
pH	0.401	0.721	0.027
EC	0.144	−0.037	−0.484
CaCO ₃	−0.126	0.739	0.407
SOM	0.576	−0.205	0.377
CEC	0.839	0.007	−0.110
Variance explained (%)	39.51	12.28	11.50
Cumulative variance (%)	39.51	51.79	63.29

The PCA biplot (Figure 5) visually verified the high multivariate correlation between Cr, Ni, Pb, Cd and CEC and the more independent orientation of Zn towards PC3. There was also significant overlap between the PCA score space in the surface and subsurface soil layers, suggesting that the principal multivariate gradients were similar for the two soil layers, not completely distinct.

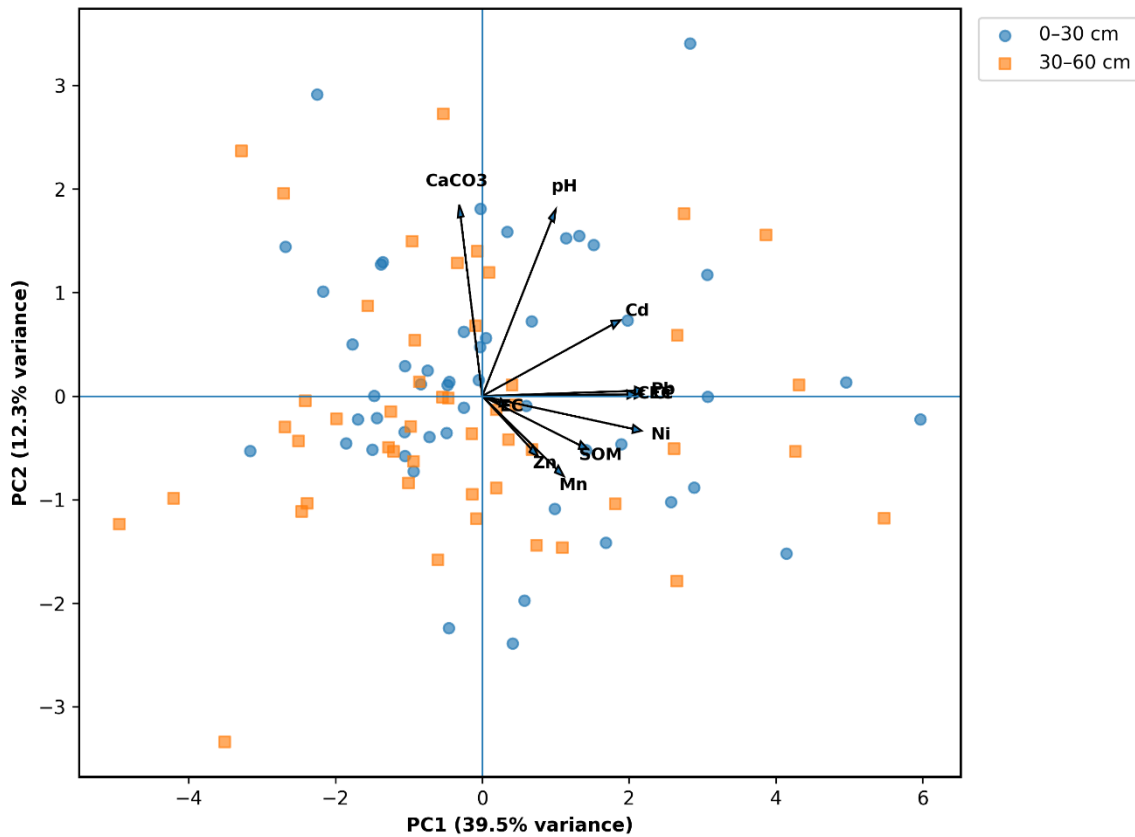


Figure 5. PCA biplot of standardized heavy metal concentrations and physicochemical properties across surface and subsurface industrial soils

3.6 Hierarchical Cluster Analysis of Heavy Metals

Hierarchical cluster analysis was performed on the six total heavy metal concentrations to examine similarities in their profiles across the 100 soil observations. The clustering pattern corroborated the patterns shown in the PCA and correlation analyses. The strongest association was between Cr and Ni ($r_s = 0.910$), and these metals formed the first cluster at the lowest linkage distance.

A second group of association was found between Pb and Zn with $r_s = 0.759$. Since then, it has been clustered with Cr and Ni, which has positive correlation with Ni ($r_s = 0.775$) and Cr ($r_s = 0.671$) (Table 6). The separation of metals into the three-cluster solution was: Cluster 1: Pb–Zn, Cluster 2: Cr–Ni–Mn and Cluster 3: Cd. Cd was only added to the other metals at the highest linkage distance, suggesting that the pattern of distribution of Cd was relatively different from that of the other elements.

Table 6. Spearman correlation coefficients among total heavy metal concentrations used in hierarchical clustering

Metal	Cd	Cr	Mn	Ni	Pb	Zn
Cd	1.000	0.507	0.535	0.627	0.649	0.316
Cr	0.507	1.000	0.671	0.910	0.715	0.600
Mn	0.535	0.671	1.000	0.775	0.696	0.638
Ni	0.627	0.910	0.775	1.000	0.750	0.534
Pb	0.649	0.715	0.696	0.750	1.000	0.759
Zn	0.316	0.600	0.638	0.534	0.759	1.000

This clustering structure is displayed in the dendrogram in figure 6, which shows that Cd was the most distinct from the other metals.

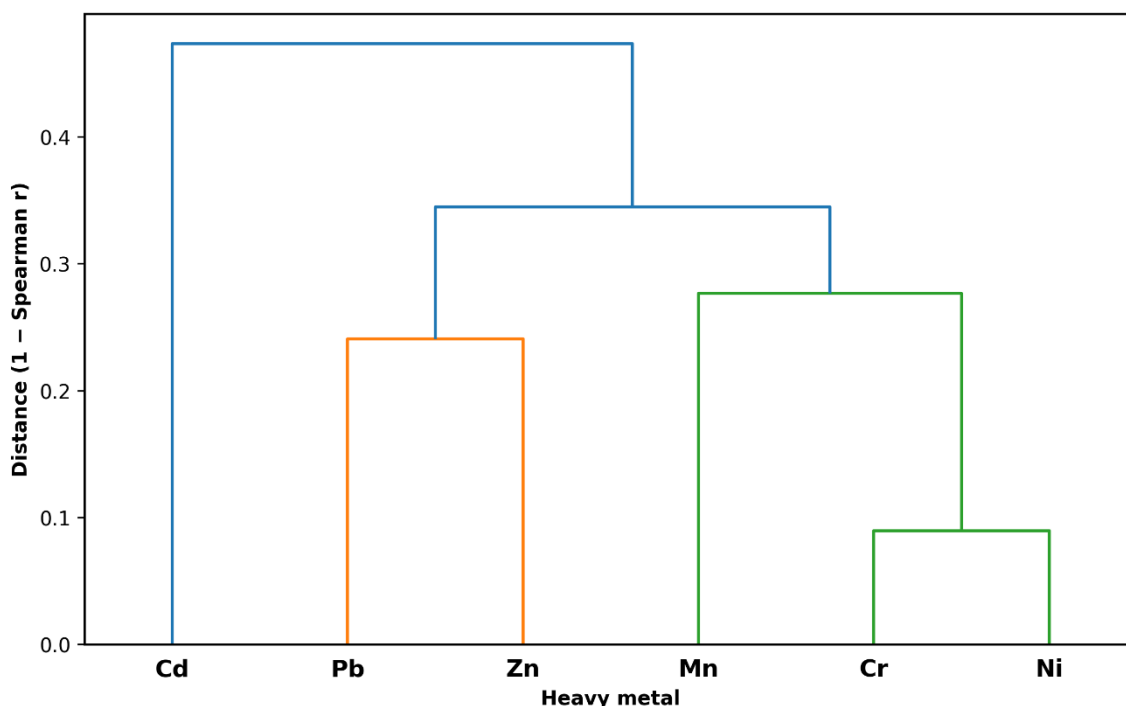


Figure 6. Hierarchical clustering dendrogram of Cd, Cr, Mn, Ni, Pb, and Zn based on 1 – r, distance and average linkage

3.7 Multivariate Comparison of Heavy Metal Profiles Between Soil Depths

The concentration profiles of the six metals were compared using a paired multivariate test for matched soil samples from the 0–30 cm and 30–60 cm layers. Hotelling's T^2 analysis showed a statistically significant difference in the overall heavy-metal profile between the two depths, $T^2 = 17.33$, $F(6, 44) = 2.59$, $p = 0.031$.

The differences in depth means were different for the various metals (Table 7). On average, the surface concentrations of Mn, Pb and Zn were 14.22, 1.61 and 17.45 mg kg^{-1} higher, respectively. Subsurface concentrations were slightly elevated for Cd, Cr and Ni, however. The greatest standardized paired difference was for Zn ($d_z = 0.265$) followed by Ni ($d_z = -0.251$) and Pb ($d_z = 0.209$). These effect magnitudes were small, reflecting the fact that statistically significant effects due to the multivariate depth effect was due to the combination in the six metals, and not due to large differences in all the individual metals.

Table 7. Paired multivariate comparison of total heavy metal concentrations between soil depths

Metal	Mean difference: surface – subsurface (mg kg^{-1})	Standardized paired difference (d_z)
Cd	–0.002	–0.008
Cr	–0.320	–0.071
Mn	14.224	0.082
Ni	–1.447	–0.251
Pb	1.607	0.209
Zn	17.449	0.265

Overall multivariate test: Hotelling's $T^2 = 17.33$; $F(6, 44) = 2.59$; $p = 0.031$.

This multivariate finding is in support of the earlier reported univariate depth differences, which indicated the greatest depth difference for Zn.

3.8 Relationships Between Integrated Pollution Indices and Soil Properties

Spearman correlation analysis was used to examine relationships between integrated pollution indices and soil physicochemical characteristics separately at each depth.

Spearman correlation analysis was used to analyse the relationships of PLI and degree of contamination with soil physicochemical characteristics separately by depth (Table 8). In the 0–30 cm layer, PLI was strongly and positively associated with CEC ($r_s = 0.736$, $p < 0.001$) and moderately associated with SOM ($r_s = 0.482$, $p < 0.001$). A statistically significant positive correlation was also found between PLI and EC, although this relationship was weaker ($r_s = 0.371$, $p = 0.008$). No significant associations were observed between surface PLI and either pH or CaCO_3 .

The surface-soil degree of contamination similarly showed positive relationships with CEC ($r_s = 0.612$, $p < 0.001$), SOM ($r_s = 0.426$, $p = 0.002$), and EC ($r_s = 0.284$, $p = 0.046$). In subsurface soil, PLI was again strongly associated with CEC ($r_s = 0.669$, $p < 0.001$) and SOM ($r_s = 0.550$, $p < 0.001$). Degree of contamination showed significant positive correlations with CEC ($r_s = 0.570$, $p < 0.001$), SOM ($r_s = 0.459$, $p < 0.001$), and pH ($r_s = 0.374$, $p = 0.008$). CaCO_3 was

not significantly correlated with either contamination index at either depth.

Table 8. Spearman correlations between pollution indices and soil physicochemical properties

Depth	Index	pH	EC	CaCO ₃	SOM	CEC
0–30 cm	PLI	0.090	0.371	−0.126	0.482	0.736
	Degree of contamination	0.147	0.284	−0.025	0.426	0.612
30–60 cm	PLI	0.264	0.223	−0.033	0.550	0.669
	Degree of contamination	0.374	0.146	0.035	0.459	0.570

Overall, CEC represented the soil property most consistently associated with integrated contamination levels at both depths, followed by SOM.

3.9 Identification of High-Contamination Sampling Points

The site-level PLI values were studied to find the sampling points where there were combination effects due to heavy metals. The surface layer had 7 of 50 sampling points (14%) with PLI values above 1.0 while the subsurface layer had 5 of 50 sampling points (10%).

At 0–30 cm, the highest PLI was recorded at sampling point 43 (PLI = 1.127), followed by points 26 (1.039), 38 (1.033), 28 (1.029), 2 (1.025), 37 (1.022), and 14 (1.002). In the 30–60 cm layer, the highest PLI occurred at sampling point 26 (PLI = 1.166), followed by points 43 (1.110), 37 (1.067), 14 (1.037), and 28 (1.024) (Table 9).

Persistent combined contamination was observed at both soil depths, with PLI > 1 at five sampling points: 14, 26, 28, 37 and 43. Sampling point 37 recorded the highest surface degree of contamination (10.887), whereas sampling point 26 exhibited the highest subsurface degree of contamination (11.406).

Table 9. Sampling points with Pollution Load Index greater than 1

Depth	Sampling point	PLI	Degree of contamination
0–30 cm	43	1.127	10.053
	26	1.039	10.076
	38	1.033	10.436
	28	1.029	9.506
	2	1.025	10.073
	37	1.022	10.887
	14	1.002	9.710
30–60 cm	26	1.166	11.406
	43	1.110	10.147
	37	1.067	9.106
	14	1.037	9.806
	28	1.024	10.762

The PLI is also shown to be varying at the site level (figure 7), with few sites having a PLI greater than 1. Elevated PLI was found at the same five sites at both depths, suggesting that the greatest contamination did not occur just at the surface.

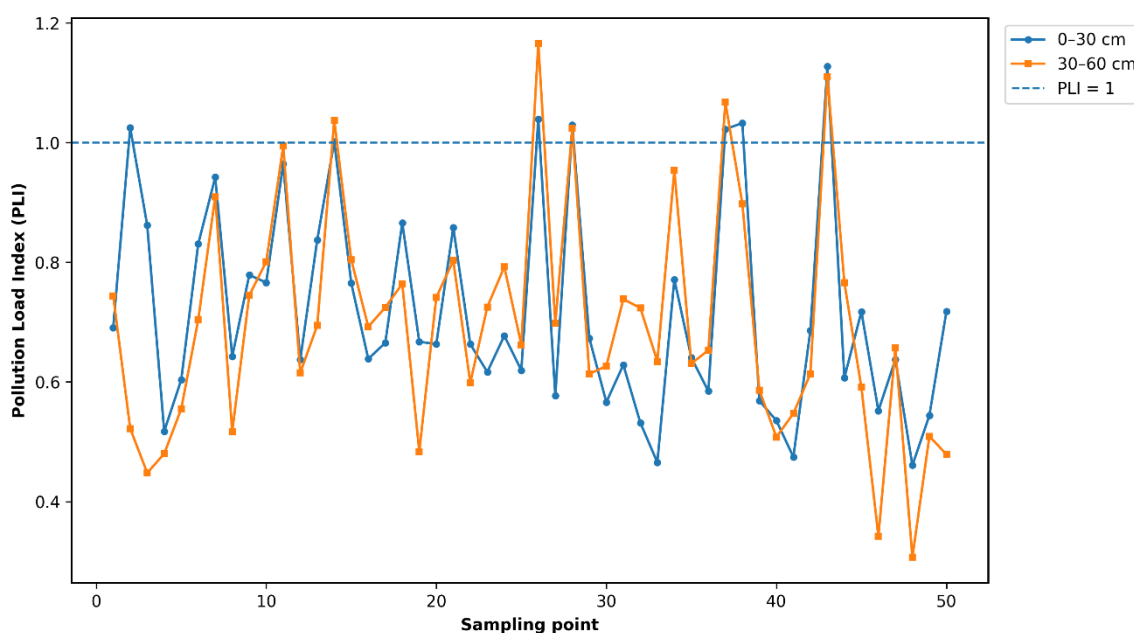


Figure 7. Pollution Load Index values across the 50 sampling points at 0–30 cm and 30–60 cm soil depths, with the reference line indicating PLI = 1

4. DISCUSSION

The present study showed substantial heterogeneity in heavy-metal contamination among elements, soil depths and sampling sites. Although Mn exhibited the highest absolute concentration, contamination-index analysis identified Cd as the most enriched element relative to the adopted geochemical background, followed by Pb. A similar trend has been observed in soils affected by industry and mining, where relatively low concentrations of metals can show high enrichment when evaluated against background levels (Gogoi et al., 2024). Industrial contamination can also result from cumulative anthropogenic inputs superimposed on natural geochemical variation (Gan et al., 2022).

Cd contamination factor (CF) means of around 3.47 at both depths were higher than that of Cr, Mn, Ni and Zn. Cd has been reported as a significant contaminant in soils as a result of industrial activities due to its mobility, persistency and low natural background concentration (Liu et al., 2025). Similar studies have also identified Cd as a primary source of contamination and concern for potential exposure (Xia et al., 2025).

The mean PLI values were not uniformly elevated, with mean values at both depths below 1. However, when PLI was calculated at 14% of surface and 10% of subsurface locations, a few hotspots were found that would not have been identified by using the mean values. The studies have also indicated that the severity of contamination within a heterogeneous industrial setting can be underestimated by aggregated statistics (Tao et al., 2026). Site-specific evaluation is thus relevant in the presence of spatially non-uniform contamination from emissions and waste deposition as well as from historical industrial activity (Molla et al., 2026).

The bioavailable-fraction analysis added information to the total concentration. Similarly, Ahmed et al. (2025) showed that the ability to study total concentration, transfer behavior and bioavailability in a complementary manner contributes to the understanding of metal contamination. Bioavailability is significant in interpreting environmental risk (Mensah et al., 2021) as only a portion of the total metal in the soil is bioavailable for biological uptake and/or direct exposure.

The majority of metals were found in similar concentrations when compared by depth, while Zn was found to be enriched on the surface and the six-metal composite profile significantly differed between depth. Soils with surface horizons that are more contaminated than deeper horizons have been found to have depth-dependent contamination (Kang et al., 2020). The high pollution index values at some stations at both depths, however, suggest contamination was not only superficial, but may be attributed to persistent pollution and vertical redistribution (Kumar & Anshumali, 2025).

The accumulation may also be influenced by adsorption and complexation by SOM. Multivariate contamination assessment also has similar relationships between soil characteristics and heavy metal distribution (Shukla et al., 2024). These interactions are important in understanding the differences in contamination intensity that can be measured between locations that are close together.

These relations were further elucidated using multivariate analyses. Cr, Ni, Pb, Cd, CEC were grouped together strongly by PCA on the first component, while Zn displayed a more distinct multivariate pattern. Close Cr–Ni and Pb–Zn associations were also found by HCA, and Cd was relatively separate. Such patterns may suggest similar geochemical activity or, at least partially, similar anthropogenic activity, but do not automatically mean that the source of contamination is the same. PCA and cluster-based methods have been proven to be effective in identifying dominant contamination gradients and metal associations in complex soil systems (Hendawy et al., 2025; Sharma et al., 2026).

Pollution indices must be interpreted as not being a direct measure of human-health risk from a risk perspective. High analytical values for CF, Igeo or PLI are an indicator of enrichment but not of exposure or toxicity. For formal health-risk assessment, pathways of exposure, intakes and toxicological reference parameters are all necessary (Birch et al.,

2024; Kong et al., 2021). Karimian et al. (2021) have demonstrated the differences in non-carcinogenic and carcinogenic risks of metals and receptor groups based on the previous soil investigations. The large-scale evidence also indicates that the soils contaminated by industry can be important contributors for potentially toxic element exposure (Xiao et al., 2020). Thus, the present integrated statistical assessment gives a comprehensive assessment of contamination, and exposure-based calculations are needed to assess direct human-health effects.

5. CONCLUSION

In this study, an integrated statistical analysis of heavy metal contamination in industrial soils was carried out using descriptive analysis, pollution indices, bioavailability assessment analysis, correlation analysis, principal component analysis, hierarchical clustering and multivariate depth comparison. The relative enrichment levels were highest for Cd, with consistently high contamination factor values and geo-accumulation index values of all soils at both depths. There was also significant enrichment of Pb and some hot spots of both Zn and Mn at certain sites. While the mean Pollution Load Index values were below unity, however, a percentage of sites surpassed the limit, indicating that the pollution level of the area can be moderate but still be highly contaminated at certain locations.

It also revealed that the concentrations of Zn, Pb and Cd in bioavailable fractions were more environmentally accessible than those of several other metals. The soil properties were significantly related to the contamination indices and metal concentrations, especially cation exchange capacity and soil organic matter. Strong associations among Cr, Ni, Pb and Cd were found in multivariate analyses while Zn showed a relatively distinct association. Vertical variation in the combined metal profile was further substantiated by the significant multivariate difference between soil depths. In general, the results indicate that indices of pollution and multivariate statistics give a more complete description of industrial-soil contamination than do single-concentration measures of contamination alone. To quantify direct human-health implications a calculation of future exposure-based risk would be necessary.

REFERENCES

1. Adnan, M., Zhao, P., Xiao, B., Ali, M. U., & Xiao, P. (2025). Heavy metal pollution and source analysis of soils around abandoned Pb/Zn smelting sites: Environmental risks and fractionation analysis. *Environmental Technology & Innovation*, 38, 104084.
2. Ahmed, M., Aslam, A. A., Aftab, F., Sanaullah, M., Hussain, R., Eiman, E., ... & Zargar, S. (2025). Multivariate analysis of potentially toxic metal contents in soil and vegetables: Enrichment, bioconcentration, translocation from soil to vegetables, and assessment of human health toxicity. *Food and Chemical Toxicology*, 200, 115413.
3. Bashir, Z., Raj, D., & Selvasembian, R. (2025). Source profiling, pollution and health risk assessment of heavy metals in agricultural soils around an industrial cluster using PCA and GIS-assisted PMF. *Environmental Monitoring and Assessment*, 197(7), 711.
4. Birch, G., Wang, X., & Liu, E. (2024). Human health risk assessment of metal-contaminated soils in Sydney estuary catchment (Australia). *Environmental Geochemistry and Health*, 46(4), 125.
5. Gan, L., Wang, J., Xie, M., & Yang, B. (2022). Ecological risk and health risk analysis of soil potentially toxic elements from oil production plants in central China. *Scientific Reports*, 12(1), 17077.
6. Gogoi, N., Sarma, A., Choudhury, M., Samanta, P., Faizan, M., & Sadhak, S. (2024). Soil heavy metal pollution and ecological risk assessment in disturbed and undisturbed soil of Morigaon, Assam. *Discover Environment*, 2(1), 13.
7. Hendawy, E., Belal, A. A., Rebouh, N. Y., Shokr, M. S., Sheta, A. E. A. S., Saleh, A. M., & Abou-Hadid, A. F. (2025). Health risk assessment of heavy metals in arid soils of the Nile Delta, Egypt using GIS and multivariate statistical analyses. *Plos one*, 20(11), e0335394.
8. Hou, D., O'Connor, D., Igalavithana, A. D., Alessi, D. S., Luo, J., Tsang, D. C., ... & Ok, Y. S. (2020). Metal contamination and bioremediation of agricultural soils for food safety and sustainability. *Nature Reviews Earth & Environment*, 1(7), 366-381.
9. Imoni, O., & Jonathan, E. L. (2025). Data-driven assessment of soil heavy metal contamination in Joinkrama, Rivers State, Nigeria using pollution indices and multivariate analytics. *Journal of Materials Science Research and Reviews*, 8(3), 628-643.
10. Islam, M. S., Hassan, F. U., Toriman, M. E., Ahmad, R., Bashir, M. A., Rehim, A., ... & Halim, S. B. A. (2025). Spatial assessment and ecological risk evaluation of soil heavy metal contamination using multivariate statistical techniques. *Catena*, 261, 109550.
11. Kang, Z., Wang, S., Qin, J., Wu, R., & Li, H. (2020). Pollution characteristics and ecological risk assessment of heavy metals in paddy fields of Fujian province, China. *Scientific Reports*, 10(1), 12244.
12. Karimian, S., Shekoohiyan, S., & Moussavi, G. (2021). Health and ecological risk assessment and simulation of heavy metal-contaminated soil of Tehran landfill. *RSC advances*, 11(14), 8080-8095.
13. Karthikayini, S., Chandrasekaran, A., Manjunatha, & Bennal, A. S. (2025). Ecological and human health risk assessment of heavy metal contamination in soils of major industrial estates of Tamil Nadu, India. *International Journal of Environmental Analytical Chemistry*, 105(16), 4509-4528.
14. Kong, F., Chen, Y., Huang, L., Yang, Z., & Zhu, K. (2021). Human health risk visualization of potentially toxic elements in farmland soil: A combined method of source and probability. *Ecotoxicology and Environmental Safety*, 211, 111922.
15. Kumar, S., & Anshumali. (2025). Heavy metal pollution and health risk assessment in upland and riparian soils of the Ganga River basin. *Discover Soil*, 2(1), 33.

16. Li, G., Chi, H., Hou, Y., Williams, P. N., Liu, Z., & Cai, C. (2024). Accurate bioaccessibility assessment of soil heavy metals by combining their speciation and in vitro model with human gut microbiota. *Environmental Sciences Europe*, 36(1), 209.
17. Lin, M., Liu, X., Deng, H., & Lin, Z. (2024). Quantitative analysis of heavy metal pollution sources, influencing factors, and delineation of pollution risk areas in industrial soil: A combined application of receptor models and geodetector. *ACS ES&T Engineering*, 4(3), 562-571.
18. Liu, Z., Mo, L., Liang, J., Shi, H., Yao, J., & Lun, X. (2025). Heavy Metal Pollution and Health-Ecological Risk Assessment in Agricultural Soils: A Case Study from the Yellow River Bend Industrial Parks. *Toxics*, 13(10), 834.
19. Mensah, A. K., Marschner, B., Antoniadis, V., Stemn, E., Shaheen, S. M., & Rinklebe, J. (2021). Human health risk via soil ingestion of potentially toxic elements and remediation potential of native plants near an abandoned mine spoil in Ghana. *Science of the Total Environment*, 798, 149272.
20. Mitran, T., Gunnam, J. R. S., Gourigari, S., & Kandrika, S. (2024). Assessment of depth wise distribution, enrichment, contamination, ecological risk and sources of soil heavy metals over an Industrial area in Southern India. *Journal of Geochemical Exploration*, 257, 107379.
21. Molla, A., Emmanouil, C., Lolos, A., Charvalas, G., & Kyzas, G. Z. (2026). Quantitative Health Risk Evaluation of Soil Contamination by Potentially Toxic Elements near an Industrial Area in Central Greece. *Environmental Processes*, 13(2), 31.
22. Saraswat, A., Ram, S., Raza, M. B., Islam, S., Sharma, S., Omeka, M. E., ... & Golui, D. (2023). Potentially toxic metals contamination, health risk, and source apportionment in the agricultural soils around industrial areas, Firozabad, Uttar Pradesh, India: a multivariate statistical approach. *Environmental Monitoring and Assessment*, 195(7), 863.
23. Sharma, L. D., Bhupenchandra, I., Singh, E. J., Dutta, S., Pandey, Y., Chongtham, S. K., ... & Sahoo, M. R. (2026). Multivariate source apportionment and integrated indexical assessment of heavy metal contamination in northeast Indian wetland sediments and soils. *Scientific Reports*.
24. Shukla, A. K., Pradhan, M., & Tiwari, O. N. (2024). Heavy Metal Contamination of Surface Sediments-Soil Adjoining the Largest Copper Mine Waste Dump in Central India Using Multivariate Pattern Recognition Techniques and Geo-Statistical Mapping. *Nature Environment & Pollution Technology*, 23(3), 1363.
25. Stefanou, S. (2026). Soil physicochemical properties and heavy metal concentrations in soils of the Industrial Zone of Thessaloniki city, Sindos – Greece (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/wxc3wg7dzs.1>
26. Tao, L., Ran, Y., Zhang, Z., Pan, Y., Guo, Y., Zhou, F., ... & Jiao, W. (2026). Heavy Metal Pollution, Source Apportionment and Human Health Risk Assessment: A Case Study of Soil in Bohai New Area Industrial Park. *Soil and Sediment Contamination: An International Journal*, 1-17.
27. Tarighat, H., Baghdadi, M., & Nasrabadi, T. (2025). Integrating native reference values with fractionation and multi-index analysis for heavy metal risk assessment in industrial soils of the South Pars Economic Energy Zone, Iran. *Environmental Pollution*, 127543.
28. Verma, F., Singh, S., Dhaliwal, S. S., Kumar, V., Kumar, R., Singh, J., & Parkash, C. (2021). Appraisal of pollution of potentially toxic elements in different soils collected around the industrial area. *Heliyon*, 7(10).
29. Xia, L., Zhang, Q., Zhang, Y., & Jiang, B. (2025). Source-specific ecological and health risks of potentially toxic elements in an agricultural industrialization town, Shandong Province, China. *Environmental Monitoring and Assessment*, 197(12), 1313.
30. Xiao, X., Zhang, J., Wang, H., Han, X., Ma, J., Ma, Y., & Luan, H. (2020). Distribution and health risk assessment of potentially toxic elements in soils around coal industrial areas: A global meta-analysis. *Science of the Total Environment*, 713, 135292.
31. Xu, D., Shen, Z., Dou, C., Dou, Z., Li, Y., Gao, Y., & Sun, Q. (2022). Effects of soil properties on heavy metal bioavailability and accumulation in crop grains under different farmland use patterns. *Scientific Reports*, 12(1), 9211.
32. Zhang, J., Wang, X., Li, J., Luo, J., Wang, X., Ai, S., ... & Liu, Z. (2024). Bioavailability (BA)-based risk assessment of soil heavy metals in provinces of China through the predictive BA-models. *Journal of Hazardous Materials*, 465, 133327.