



OPEN

Inclusion of key soil parameters in the modified contamination factor (MCF) model as a tool for assessing heavy metal pollution in agricultural soils

Mohssen Elbagory¹, Nagwa EL-Khateeb², Sahar El-Nahrawy³, Alaa El-Dein Omara³, Ibrahim Mohamed⁴, Mahmoud El-Sharkawy⁵, Amal Zayed⁶, Madhumita Goal⁷, Deep Gupta⁸, Piyush Kumar⁹, Pankaj Kumar^{10,11} & Ivan Širić¹²✉

This study aimed to evaluate the applicability of a Modified Contamination Factor (MCF) model to assess heavy metal (HM) pollution in rural, urban, and industrial agricultural soils of Haridwar, India. Traditional Contamination Factor (CF) indices, though widely employed, fail to consider intrinsic soil buffering capacity, potentially misrepresenting ecological risk. To address this limitation, an MCF model was developed by integrating soil pH, organic matter (OM; %), and cation exchange capacity (CEC; cmol/kg) through a principal component analysis (PCA)-derived weight assignment method. Soil samples were collected and analyzed for nine HMs (Cd, Cr, Cu, Co, Fe, Mn, Ni, Pb, Zn) along with key soil properties including pH, organic matter (OM), and cation exchange capacity (CEC). Results revealed significant declines in soil pH (7.10 in rural to 6.30 in industrial), organic matter (2.50% to 1.20%), and cation exchange capacity (18.20 to 12.30 cmol/kg) from rural to industrial areas, reflecting progressive soil degradation. Concurrently, HM concentrations showed a significant increase, with Pb (79.5 mg/kg; permissible limit 85 mg/kg, USEPA) and Zn (188.3 mg/kg; permissible limit 300 mg/kg, WHO/FAO) approaching or exceeding guideline values in industrial zones. The MCF model yielded refined contamination estimates by incorporating adjustment factors (f), thereby magnifying contamination in low-retention soils and suppressing overestimation in resilient soils. Validation parameters ($R^2 = 0.9729$, RMSE = 0.190, MAE = 0.152, NSE = 0.939) also demonstrated high agreement between MCF and traditional CF values, supporting model strength. The MCF model improves upon the traditional CF by accounting for variations in soil properties, offering a more accurate and ecologically meaningful assessment of contamination risk.

Keywords Modified contamination factor, HMs, Soil buffering capacity, Ecological risk assessment, Haridwar agriculture

¹Health Specialties, Basic Sciences and Applications Unit, Applied College, King Khalid University, 61421 Mohayil, Asir Abha, Saudi Arabia. ²Botany Department, (Agricultural Microbiology), Faculty of Agriculture, Kafrelsheikh University, Kafr El-Sheikh 33516, Egypt. ³Soil Microbiology Research Department, Soils, Water, and Environment Research Institute (SWERI), Agriculture Research Center (ARC), Giza 12112, Egypt. ⁴Department of Soil and Water Sciences, Faculty of Agriculture, Benha University, Moshtohor, Toukh 3736, Egypt. ⁵Soil and Water Department, Faculty of Agriculture, Tanta University, Tanta 31527, Egypt. ⁶Educational Psychology Department, Faculty of Education, Kafrelsheikh University, Kafr El-Sheikh 33516, Egypt. ⁷Department of Environmental Science, Graphic Era (Deemed to Be University), Dehradun 248002, India. ⁸College of Smart Agriculture and College of Allied Sciences, COER University, Roorkee 247667, India. ⁹Department of Zoology, School of Science, Vivek University, Bijnor 246701, India. ¹⁰Research and Development Division, Society for AgroEnvironmental Sustainability, Dehradun 248007, India. ¹¹School of Environmental Studies, Maa Shakumbhari University, Punwarka, Saharanpur 247120, India. ¹²University of Zagreb, Faculty of Agriculture, Svetosimunska 25, 10000 Zagreb, Croatia. ✉email: isiric@agr.hr

The rapid industrialization, urban expansion, and unsustainable agricultural practices have resulted in a substantial contamination of heavy metals (HMs) into terrestrial ecosystems, particularly agricultural lands¹, HM pollution in soils is a growing global concern, posing serious threats to food security, ecosystem health, and human well-being². Chronic exposure to HMs is linked with severe health consequences in humans, including nephrotoxicity, neurotoxicity, cardiovascular dysfunction, carcinogenesis, and impaired cognitive development in children³. HMs such as Pb and Cd bioaccumulate through the food chain and can cause long-term damage to vital organs even at low exposure levels. These health risks show the importance of accurate soil pollution assessment to safeguard public health. In India, agricultural soils surrounding rapidly developing regions are increasingly exposed to HMs through atmospheric deposition, industrial effluents, sewage sludge, and agrochemical overuse. Globally, nearly 30% of land is already degraded, affecting more than 3.2 billion people. Soil erosion alone accounts for an estimated 35 Pg/yr of soil loss from croplands worldwide⁴. In India, soil erosion rates average 16.4 tons/ha/year, and approximately 120 million hectares (~33%) of the country's land area are classified as degraded due to erosion, salinity, vegetation loss, and other processes⁵. These figures highlight the urgent need to monitor soil contamination and degradation, particularly in agricultural landscapes influenced by rapid industrialization and urban expansion⁶. These metals can persist in soil for decades, gradually entering the food chain via crop uptake^{7,8}. The excessive accumulation of HMs in agricultural soils affects soil biological activity, alters essential nutrient cycling, and significantly reduces crop productivity and safety³. Current evidence indicates high levels of metal contamination in soils adjoining industrial corridors, urban centers, and even rural agricultural landscapes, reflecting a complex gradient of anthropogenic influence⁹.

Soil pollution in and around Haridwar city has become a pressing environmental concern due to rapid urbanization and industrial expansion, particularly following the establishment of the industrial estate in Haridwar¹⁰. In the adjoining region, Singh et al.¹¹ conducted a contamination-factor (CF) and geo-accumulation index (Igeo) study of bed sediments of the River Ghaghara, a major tributary of the Ganga, finding that CF values varied seasonally and that certain metals such as Cd and Pb showed anthropogenic enrichment ($CF > 1$) in locations downstream of urban areas. Another recent work by Kumar an Anshumali¹² on upland and riparian soils in the Ganga River basin also used CF, enrichment factor (EF), and ecological risk indices, showing significantly higher CF and risk in areas with intensive agriculture and close to river banks. The increasing anthropogenic pressure from industrial waste, vehicular emissions, domestic waste, and unregulated use of agrochemicals has led to the accumulation of various HMs in agricultural soils¹³. Different anthropogenic sources contribute distinct HMs to soils. For instance, vehicular emissions are a major contributor of Pb, Zn, and Cd due to fuel combustion, lubricants, and tire wear^{14,15}. Industrial activities such as electroplating, smelting, and chemical manufacturing release Cr, Ni, Cu, and Cd, while domestic and municipal wastes often contribute Zn, Pb, and Fe¹⁶. The excessive use of agrochemicals, including phosphate fertilizers and pesticides, is another significant source of Cd, As, and Pb accumulation in agricultural soils. These source-specific contributions highlight the complex nature of HM inputs in the Haridwar region. These contaminants pose significant risks to soil health, crop productivity, and food safety. Haridwar, being a densely populated religious and cultural hub, experiences rapid land conversion and high volumes of municipal waste generation. Inadequately managed waste, open dumping, and leachate formation contribute to the leaching of HMs such as Pb, Cd, and Zn into soils, while untreated wastewater used for irrigation further increases pollutant load. In addition, particulate emissions from dense traffic and biomass burning in urban areas are deposited on agricultural fields through atmospheric fallout, collectively contributing to the mobilization and accumulation of HMs in nearby soils¹⁷. Industrial discharges from processing, electroplating, and chemical manufacturing facilities are often inadequately treated before entering surface water and soil systems, leading to long-term contamination. The industrial estate in Haridwar hosts a diverse range of industries, including pharmaceuticals, textiles, electroplating and metal finishing units, food processing, plastics, and chemical manufacturing facilities. Electroplating and metal-based industries are major contributors of Cr, Ni, Zn, and Cu; textile industries release dyes and auxiliaries containing Pb, Cd, and Zn; while pharmaceutical and chemical plants discharge effluents rich in organic and inorganic residues, including HMs. Together, these industrial activities explain the elevated concentrations of Pb, Zn, Cr, and Cd observed in soils adjacent to the estate. Furthermore, atmospheric deposition of particulate matter from traffic and factories also contributes to the diffuse input of metals into soils¹⁸. These processes disrupt the soil's physicochemical properties, ultimately affecting its fertility and buffering capacity¹⁹. Continuous exposure to HM concentrations can alter microbial communities, impair nutrient cycling, and lead to bioaccumulation in crops, posing ecological and human health risks²⁰. Despite regulatory frameworks, the lack of regular monitoring and integrated assessment tools limits our understanding of pollution dynamics in the region. Hence, evaluating contamination using modified indices that incorporate local soil characteristics is essential for accurate risk assessment and sustainable land management.

Quantitative assessment of HM pollution in soils is crucial for environmental monitoring, agricultural safety, and policymaking²¹. Among the tools developed for this purpose, the contamination factor (CF) model has been widely used due to its simplicity and capacity to indicate the degree of contamination relative to background levels²². Traditional CF is calculated as the ratio of measured HM concentration to a regional background reference²³. While CF is effective as a preliminary screening method, this approach has a fundamental limitation—it overlooks the intrinsic buffering capacity of soils. Soils are heterogeneous matrices with complex physicochemical properties, including pH, organic matter (OM), and cation exchange capacity (CEC), which strongly influence metal mobility, bioavailability, and toxicity^{24,25}. Soil pH regulates the solubility and speciation of HMs; under acidic conditions, HMs such as Cd, Pb, and Zn remain in more soluble ionic forms, thereby increasing their mobility and uptake by plants, while alkaline pH promotes precipitation and adsorption, reducing bioavailability²⁶. OM acts as both a sink and a source of HMs: it immobilizes them through chelation and complexation with functional groups, but its decomposition under certain conditions can release bound HM back into the soil solution²⁷. Similarly, cation exchange capacity (CEC) shows the soil's ability to

adsorb and retain positively charged HM ions on exchange sites; soils with high CEC buffer metal toxicity by immobilizing them, whereas soils with low CEC allow greater mobility and leaching²⁸. Together, these three parameters determine the extent to which HMs are retained, mobilized, or transferred into the food chain, thereby influencing their ecological risk. For example, alkaline soils with high OM and CEC tend to immobilize metals through adsorption, complexation, and ion exchange processes, thereby reducing their ecological risk^{28,29}. In contrast, acidic or OM-deficient soils with low retention potential allow higher metal solubility and plant uptake^{30,31}. Therefore, relying solely on concentration-based indices like traditional CF can lead to misleading interpretations of pollution severity, especially when comparing heterogeneous land uses or soil types. This shortcoming suggests the need for more integrative contamination assessment tools that factor in soil-specific retention characteristics.

To address this gap, a modified contamination factor (MCF) model that accounts for both contaminant load and soil retention capacity can be developed. The proposed method in this study refines traditional CF by incorporating three critical soil parameters—pH, OM, and CEC, into a correction factor, derived from weighted normalization based on principal component analysis (PCA). This approach adjusts the contamination assessment by scaling the CF according to the soil's inherent ability to retain HMs, thus providing a more ecologically meaningful assessment. The adjustment factor is constructed as the inverse of a weighted composite index of the normalized soil parameters, where higher values indicate weaker buffering and thus amplify contamination risk. The use of PCA ensures objectivity in assigning weights, reflecting the relative importance of each parameter in influencing metal behavior. By multiplying the traditional CF by this adjustment factor, the MCF model differentiates pollution impacts under variable soil conditions, reducing the contamination signal in high-retention soils and magnifying it in vulnerable ones. This study introduces a novel PCA-weighted Modified Contamination Factor (MCF) model that integrates soil physicochemical characteristics (pH, OM, CEC) into contamination assessment. Unlike traditional concentration-based indices, the MCF provides an ecologically meaningful risk evaluation by accounting for soil buffering capacity. The approach enhances accuracy in pollution classification, reduces misinterpretation of risk, and offers a transferable framework for adaptive soil monitoring and sustainable agricultural land management. Based on the above rationale, this study hypothesized that integrating soil physicochemical properties (pH, OM, and CEC) into the contamination factor framework will provide a more accurate and ecologically meaningful assessment of HM pollution than the traditional CF alone. Specifically, we expected that the MCF would (1) reduce contamination scores in soils with high retention capacity (high OM and CEC, neutral pH), and (2) amplify contamination scores in soils with low buffering capacity (acidic, low OM and CEC), thereby offering improved discrimination of pollution severity across rural, urban, and industrial zones.

Considering the above, this study aims to evaluate the applicability of the MCF model across three contrasting agricultural zones in Haridwar, Uttarakhand: rural (traditional village farming systems), urban (peri-urban cultivation in city limits), and industrial (fields adjacent to the SIDCUL industrial estate). Each region represents a distinct anthropogenic pressure gradient, with differing land use histories, pollution sources, and soil characteristics. The objectives of this study are: (1) to determine the extent and variability of HM contamination in agricultural soils of Haridwar; (2) to demonstrate the advantage of the MCF model over the traditional CF in providing a more realistic representation of pollution severity; and (3) to quantify the influence of key soil properties (pH, OM, CEC) on metal retention.

Materials and methods

Study area

The study was conducted in three distinct agricultural regions of Haridwar district, Uttarakhand, India, based on varying anthropogenic influences. The rural area included non-impacted agricultural fields surrounding villages near Bahadrbabad (29.9523° N, 78.1480° E), characterized by no urban or industrial intrusion. The urban area comprised cultivated fields within the city limits of Haridwar (29.9457° N, 78.1632° E), exposed to urban pressures. The industrial area is focused on agricultural lands adjacent to the high-density zone (29.9294° N, 78.1231° E), known for extensive industrial activities. Figure 1 shows the detailed map of the study area depicting the regions considered in this study. The three sampling zones were located within a spatial gradient across Haridwar district. The rural sites near Bahadrbabad were situated approximately 7 km from the Haridwar city limits, while the urban agricultural fields were distributed within and around the city. The industrial sampling sites were concentrated adjacent to the industrial estate, lying about 15 km from the urban fields and 12 km from the rural sites. This spatial separation allowed clear representation of rural, peri-urban, and industrial influences along the land-use gradient.

Study design and sample collection

A total of 30 composite soil samples were collected during September 2024, comprising 10 representative samples from each land-use category (rural, urban, and industrial). Each composite was prepared by pooling five subsamples collected within an individual agricultural field, ensuring adequate representation of spatial variability. In each category, a total of 10 agricultural fields were selected. From each field, five subsamples were collected at the topsoil layer (0–15 cm) using a sterile stainless-steel auger, following a random sampling pattern to capture within-field variability. The subsamples were then homogenized and pooled to form one representative composite sample per field. This strategy ensured adequate representation of soil heterogeneity across each land-use category. The samples were stored in polyethylene bags, labeled clearly with site codes, and transported in cool boxes to the laboratory within 24 h. All sampling tools were pre-cleaned with 10% nitric acid and rinsed with deionized water before and after each use to prevent cross-contamination.

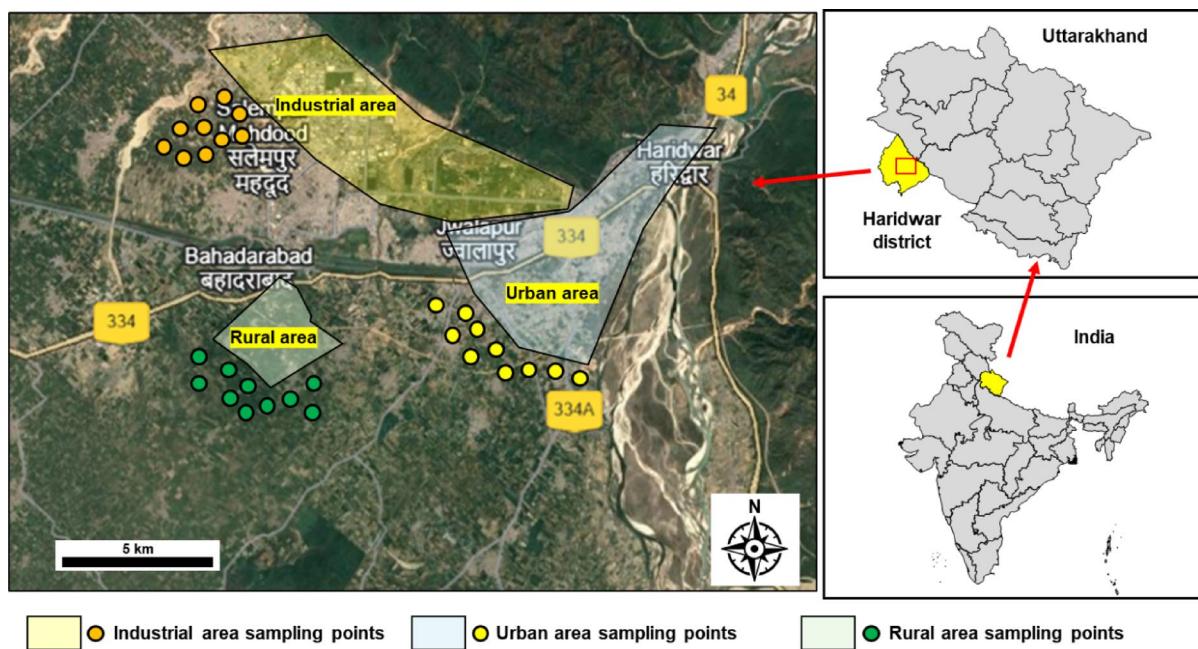


Fig. 1. Location map showing urban, industrial, and rural areas of Haridwar district, Uttarakhand, India. Base satellite imagery from Google Earth Pro (version 7.3.6; Imagery © 2024 Airbus, Map data © Google). Map annotations prepared by the authors using MS Power Point 365 and Powe User (<https://www.powerusersoftware.com/>).

Analytical methods

Each composite sample was air-dried at room temperature, gently crushed, and sieved through a 2 mm mesh. The ≤ 2 mm soil fraction was used for HM analysis, in accordance with USEPA Method 3051A, as it represents the standard soil fraction for environmental monitoring and plant-soil interaction studies. Soil pH was measured in a 1:2.5 soil-to-water suspension using a digital pH meter (ESICO International, India). Organic matter (OM) was determined by the Walkley–Black dichromate oxidation method³². Cation exchange capacity (CEC) was quantified using the ammonium acetate saturation method and expressed in cmol/kg³³. Total concentrations of cadmium (Cd), chromium (Cr), copper (Cu), cobalt (Co), iron (Fe), manganese (Mn), nickel (Ni), lead (Pb), and zinc (Zn) were determined after microwave-assisted acid digestion (HNO_3 –HCl mixture) following USEPA Method 3051A. The digested extracts were analyzed using Inductively Coupled Plasma–Optical Emission Spectrometry (ICP-OES, Model: Thermo Scientific iCAP 7400). Quality assurance and quality control (QA/QC) procedures were followed using certified reference materials (CRM, NIST SRM 2711a), method blanks, and sample duplicates. Recovery rates for all metals ranged between 94–105%, as verified using certified reference material (NIST SRM 2711a), method blanks, and sample duplicates. The detailed QA/QC results are provided in Supplementary Table S1 to support analytical reliability and data traceability³⁴.

Model development and verification

The detailed steps adopted for the development of the MCF model are provided in Fig. 2.

Calculation of traditional contamination factors

The CF was initially calculated using the traditional method by dividing the measured concentration of the metal in the soil by its corresponding background concentration³⁵. The measured value represents the specific metal concentration (in mg/kg), while the background value (B_i) represents the natural baseline concentration of each HM in uncontaminated soils, as reported in standard references. These background values are used to differentiate anthropogenic enrichment from geogenic levels. The baseline values adopted in this study were: Cd = 0.5 mg/kg, Cr = 40 mg/kg, Cu = 20 mg/kg, Co = 10 mg/kg, Fe = 38,000 mg/kg, Mn = 500 mg/kg, Ni = 30 mg/kg, Pb = 25 mg/kg, and Zn = 70 mg/kg³⁶. The traditional CF of each metal in the soil was computed using the following model (Eq. 1):

$$CF = C_i / B_i \quad (1)$$

where C_i : concentration of the i -th metal in the sample; and B_i : background (reference) concentration of the i -th metal as per³⁶. The CF quantifies enrichment of a given metal relative to its regional background concentration and therefore indicates the degree of anthropogenic input independent of regulatory thresholds. CF categorization ($CF < 1$: low; $1 \leq CF < 3$: moderate; $3 \leq CF < 6$: considerable; $CF \geq 6$: very high) is an enrichment-based classification that complements concentration-based guideline values. Because total concentration alone does not account for soil-specific factors that control mobility and bioavailability, we developed the Modified

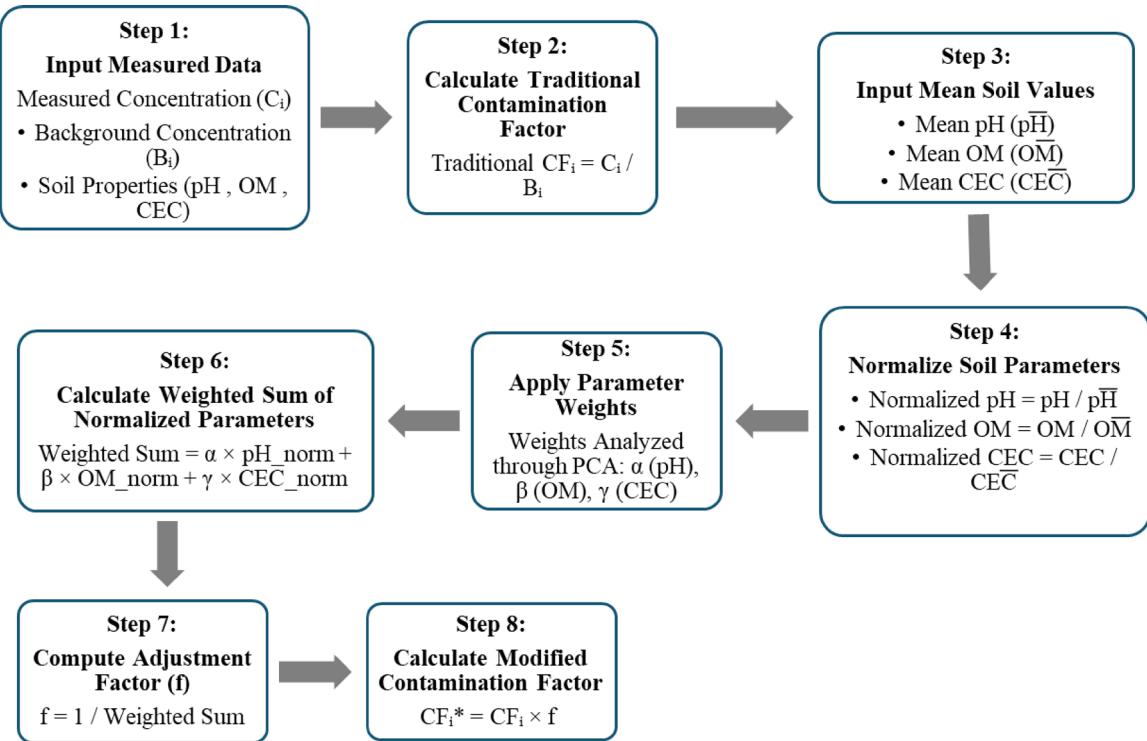


Fig. 2. Step-by-step flowchart for calculating modified contamination factors (MCF) incorporating soil retention properties.

Contamination Factor (MCF) to adjust CF by an empirically derived soil-retention factor (f) based on pH, OM and CEC. CF therefore signals enrichment relative to geogenic background, while MCF adjusts that enrichment for soil buffering capacity to better reflect potential ecological risk.

Soil parameter normalization

To account for the soil's retention capacity, three key soil physicochemical properties were selected: soil pH, organic matter content (OM, %), and cation exchange capacity (CEC). These parameters were normalized by dividing each site-specific value by the corresponding mean value across the study area (Eq. 2).

$$X_{norm,i} = \frac{x_i}{\frac{1}{n} \sum_{j=1}^n \bar{x}} \quad (2)$$

where x_i is the measured value of the soil parameter for the i -th observation, while \bar{x} is the mean value of the respective parameter over n observations. This normalization process allows comparison across different treatments and standardizes the influence of each parameter. Normalized values (X_{norm}) show how much more or less capable the local soil is at retaining metals compared to the average condition. For normalization of soil parameters (pH, OM, CEC), mean normalization was applied. This method was selected because it reduces the influence of extreme values, maintains all values as positive for multiplicative adjustment, and ensures comparability among parameters. Maximum normalization was avoided because it can disproportionately weight single outliers, while standard (z-score) normalization introduces negative values that complicate integration into multiplicative indices. Mean normalization thus provides a robust and interpretable scaling method suitable for incorporation into the MCF framework.

PCA-based weight calculation

Principal Component Analysis (PCA) was employed to objectively assign weights to soil properties (pH, OM, CEC) in the calculation of the soil retention factor (f). PCA reduces data dimensionality while maximizing explained variance, thereby identifying the relative importance of variables without arbitrary assumptions. The loadings of each soil property on the principal components were used to determine their weights in the model. This approach has been widely applied in soil contamination studies to derive variable importance and assign weights in composite indices³⁷. By adopting PCA, the weighting process becomes data-driven and reduces subjectivity compared to equal-weighting schemes. Principal component analysis (PCA) was performed to derive weights for soil parameters (pH, OM, CEC). Components were selected using the Kaiser criterion (eigenvalues > 1) and by ensuring that cumulative variance explained exceeded 70%. These derived weights (α for pH, β for OM, γ for CEC) were used to construct a weighted sum of the normalized parameters, representing the

soil's composite retention capacity. Weights were assigned to each soil property based on its relative importance in influencing metal retention. Then, a weighted sum of the normalized parameters was computed.

Let L_{pH} , L_{OM} , and L_{CEC} be PCA loadings for pH, OM, and CEC from PC1. The raw loadings were normalized to ensure their sum equals 1, yielding weights as given in Eq. (3):

$$\alpha = \frac{L_{pH}}{L_{pH} + L_{OM} + L_{CEC}}, \beta = \frac{L_{OM}}{L_{pH} + L_{OM} + L_{CEC}}, \gamma = \frac{L_{CEC}}{L_{pH} + L_{OM} + L_{CEC}} \quad (3)$$

These weights were then applied to normalized soil parameters to compute a composite soil retention index using Eq. (4):

$$\text{Weighted Sum} = \alpha \cdot pH + \beta \cdot OM + \gamma \cdot CEC \quad (4)$$

This weighted sum represents the integrated influence of soil properties on HM retention and is later used to calculate the adjustment factors (f):

Calculation of adjustment factors and MCF model

An adjustment factor (f) was computed as the inverse of the weighted sum of normalized soil parameters (Eq. 5):

$$f(pH, OM, CEC) = \frac{1}{\alpha \cdot pH + \beta \cdot OM + \gamma \cdot CEC} \quad (5)$$

where: α , β , γ are empirically determined weights (unit-normalized). Finally, the traditional CF was multiplied by the adjustment factor to obtain the MCF. This approach refines the contamination assessment by incorporating soil buffering potential, leading to a more ecologically relevant evaluation of pollution risk. The MCF was then calculated as (Eq. 6):

$$MCF = \frac{C_i}{B_i} \cdot f(pH, OM, CEC) \quad (6)$$

Here, defining the adjustment function $f(\cdot)$, we seek a multiplicative correction factor f that decreases MCF in high retention soils and increases MCF in mobile conditions. This form shows that higher pH, OM, and CEC reduce the mobility and bioavailability of metals due to increased adsorption and complexation in soils. If soil has high pH, OM, and CEC, then: $MCF < CF \rightarrow$ pollution impact is mitigated, and if soil has low pH, OM, and CEC, then: $MCF > CF \rightarrow$ pollution impact is magnified.

To facilitate interpretation, the same classification scheme used for traditional CF was applied to the MCF values, with adjustment for soil retention properties. Accordingly, $MCF < 1$ indicates low contamination, $1 \leq MCF < 3$ indicates moderate contamination, $3 \leq MCF < 6$ indicates considerable contamination, and $MCF \geq 6$ indicates very high contamination^{27,37}. These thresholds allow direct comparison of risk categories between CF and MCF while accounting for the soil's buffering capacity.

Validation parameters

The dataset was randomly split into three parts: 70% for training, 15% for validation, and 15% for testing. This partitioning ensures a sufficient number of samples for model calibration while retaining independent data for evaluation of predictive performance. To validate the MCF model, several validation parameters were adopted. Out of them, the coefficient of determination (R^2) quantifies the proportion of variance in traditional contamination factor (CF) values that are explained by the MCF model. Higher R^2 values (closer to 1) indicate stronger agreement and predictive accuracy between CF and MCF (Eq. 1).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Here, y_i denotes the observed (measured) value for the i -th sample, while \hat{y}_i represents the corresponding model-predicted value.

Mean Squared Error (MSE) measures the average of the squared differences between traditional and MCF values³⁸. It identifies larger errors more precisely, making it a sensitive indicator of model performance and overall deviation (Eq. 8).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

Root Mean Squared Error (RMSE) is the square root of MSE and provides an error estimate in the same units as the original values³⁸. It reflects the typical magnitude of prediction error, with lower values indicating a better fit (Eq. 9).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

Mean Absolute Error (MAE) represents the average absolute difference between predicted MCF and observed CF values (Eq. 10). It offers an intuitive measure of model accuracy, less influenced by extreme outliers compared to RMSE³⁸.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

Model Efficiency (ME) or Nash–Sutcliffe Efficiency (NSE) evaluates the predictive power of the model relative to the mean of the observed data³⁹. A value close to 1 indicates high model reliability, whereas values near or below zero suggest poor predictive performance (Eq. 11).

$$\text{ME} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

Mean bias calculates the average directional deviation between the MCF and CF values. A negative bias implies systematic underestimation by the model, while a positive value indicates overestimation (Eq. 12).

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (12)$$

Software and statistics

Data arrangement, preliminary processing, and graphical representation were performed using Excel (Microsoft Office 2021). Statistical analyses were performed to evaluate differences among land-use categories and to explore multivariate patterns in soil properties and metal concentrations. Normality of datasets was first assessed using the Shapiro–Wilk test ($p < 0.05$). For normally distributed parameters (pH, OM, CEC), one-way analysis of variance (ANOVA) was applied, followed by Tukey's Honest Significant Difference (HSD) test for pairwise comparisons. For datasets not conforming to normal distribution (HM concentrations), the non-parametric Kruskal–Wallis test was employed. To explore interrelationships among metals and sampling sites, multivariate analyses were conducted. Hierarchical Cluster Analysis (HCA), based on Ward's linkage and Euclidean distance, was used to classify metals and sites according to similarity in contamination profiles. PCA was applied to reduce dimensionality, identify key sources of variation, and assign weights to soil properties (pH, OM, CEC) in the MCF model. Data for PCA and HCA were standardized (z -scores) prior to analysis to ensure comparability among variables. These statistical approaches together provided both inferential testing of group differences and pattern recognition for source apportionment. The cluster analysis was also performed to evaluate similarities among sampling areas and soil properties. OriginPro 2022 was employed to derive factor loadings and weights for the MCF model based on principal component analysis (PCA). A significant level of $p < 0.05$ was adopted throughout the statistical interpretation.

Results and discussion

Properties of soil in the study area

The soil properties across the rural, urban, and industrial zones of Haridwar showed statistically significant variations, indicating the influence of land use on soil quality and contamination (Table 1). The mean soil pH declined from 7.10 in rural areas to 6.30 in industrial fields ($p = 0.021$). This progressive acidification is likely

Property	Safe limit	Sampling area			K–W test (p -value)
		Rural	Urban	Industrial	
pH	6.5–8.5	7.10 ± 0.12 a	6.80 ± 0.15 b	6.30 ± 0.20 c	0.021*
OM (%)	> 0.75 [#]	2.50 ± 0.30 a	1.90 ± 0.25 b	1.20 ± 0.18 c	0.034*
CEC (cmol/kg)	> 10	18.20 ± 1.75 a	15.70 ± 1.60 b	12.30 ± 1.45 c	0.049*
Cd (mg/kg)	< 0.8	0.12 ± 0.03 a	0.29 ± 0.05 b	0.67 ± 0.08 c	0.003*
Cr (mg/kg)	< 64	12.50 ± 1.90 a	33.80 ± 3.40 b	58.70 ± 4.20 c	0.001*
Cu (mg/kg)	< 36	15.20 ± 1.35 a	24.10 ± 2.10 b	34.60 ± 2.75 c	0.006*
Co (mg/kg)	< 20	4.10 ± 0.55 a	8.60 ± 0.90 b	14.20 ± 1.10 c	0.008*
Fe (mg/kg)	–	5123.00 ± 320.00 a	7981.00 ± 470.00 b	$10,435.00 \pm 610.00$ c	0.017*
Mn (mg/kg)	< 1500	476.00 ± 35.00 a	623.00 ± 40.00 b	899.00 ± 50.00 c	0.040*
Ni (mg/kg)	< 35	11.30 ± 1.10 a	24.60 ± 2.00 b	32.80 ± 2.50 c	0.004*
Pb (mg/kg)	< 85	9.60 ± 1.00 a	43.10 ± 3.50 b	79.50 ± 5.00 c	0.002*
Zn (mg/kg)	< 300	35.20 ± 3.00 a	87.40 ± 5.50 b	188.30 ± 8.70 c	0.005*

Table 1. Physicochemical and HM properties of soil in three different zones of Haridwar district, India. Values are mean followed by standard deviation of ten samples; source of safe limits: derived from^{40–42}; [#]Adopted from ICAR⁴³; *Kruskal–Wallis test (K–W test) indicates significant difference between sites at $p < 0.05$; The safe pH range of 6.5–8.5 was adopted from WHO and BIS guidelines, which define this interval as optimal for sustaining soil fertility, nutrient availability, and crop productivity.

linked to industrial emissions and the accumulation of acidic byproducts in nearby agricultural soils. Also, mean OM content decreased from 2.50% in rural soils to 1.20% in industrial sites ($p=0.034$), which may be attributed to reduced organic inputs, microbial suppression, and higher disturbance in urbanized environments. Similarly, the mean CEC dropped from 18.20 cmol/kg in rural areas to 12.30 cmol/kg in industrial soil ($p=0.049$), indicating a loss of soil fertility and buffering capacity. On the other hand, HM concentrations exhibited a clear increasing trend from rural to industrial areas. For instance, Cd rose from 0.12 mg/kg in rural areas to 0.67 mg/kg in industrial soils ($p=0.003$), approaching the threshold of concern. Cr increased nearly fivefold across zones, from 12.5 mg/kg (rural) to 58.7 mg/kg (industrial; $p=0.001$), while Cu and Pb concentrations increased more than twofold and eightfold, respectively, with industrial Pb levels reaching 79.5 mg/kg ($p=0.002$). Zn showed the highest enrichment, rising from 35.2 mg/kg in rural soils to 188.3 mg/kg in industrial fields ($p=0.005$). The high levels of Co, Ni, Mn, and Fe in industrial zones further showed the cumulative impact of metal deposition from manufacturing, waste disposal, and traffic emissions. Thus, all parameters showed significant differences between land-use types ($p<0.05$), with industrial soils showing the highest contamination and lowest fertility indicators. The HCA heatmap (Fig. 3) demonstrates that soil samples cluster according to land-use categories. Industrial soils form a distinct group characterized by elevated Fe, Mn, Zn, and other HMs, reflecting strong anthropogenic influence. In contrast, rural soils cluster separately and are associated with higher buffering properties such as pH, OM, and CEC. Urban soils occupy an intermediate position, showing moderate enrichment of metals along with reduced soil quality compared to rural soils. This clustering showed that both HM contamination and soil physicochemical properties are strongly differentiated by land use. This indicates a strong anthropogenic influence on soil contamination in urbanized and industrial regions.

OM and CEC are key regulators of HM mobility in soils. High OM content reduces mobility by binding metals through complexation, chelation, and adsorption onto functional groups such as carboxyl and phenolic moieties⁴⁴. This immobilization decreases immediate bioavailability, although decomposition of OM under acidic or anaerobic conditions can remobilize bound metals. Similarly, CEC represents the density of negatively charged exchange sites (primarily clay minerals and organic colloids) that attract and retain cationic metals such as Pb^{2+} , Cd^{2+} , and Zn^{2+} ⁴⁵. Soils with higher CEC thus act as stronger buffers, reducing leaching losses and limiting plant uptake, whereas soils with low CEC exhibit greater metal solubility and mobility⁷. In the present study, the progressive decline in OM and CEC from rural to industrial soils explains the elevated bioavailability and contamination risk of metals observed in the industrial zone, despite comparable total concentrations. These findings confirm that soil buffering capacity strongly mediates the ecological risk posed by HMs.

The observed declines in soil pH, OM, and CEC from rural to industrial zones in Haridwar align closely with findings from recent regional studies, reinforcing the influence of land use on soil degradation and contamination. Bharti et al.⁴⁶ reported that industrial activities in the Ranipur Rao watershed reduced OM (0.48–0.75%) compared to cleaner forest soils (0.70–0.84%) and elevated HMs, particularly Cd, Cr, Cu, Zn, and Pb. Also, Bhardwaj et al.⁴⁷ found significantly higher HM and electrical conductivity values in industrial soils around Haridwar. A broader meta-analysis by Padbhushan et al.⁴⁸ observed similar organic carbon losses and lower buffer capacities under intensified land use in India, comparable to the drop in our CEC values. Additionally, urban–rural interface studies in China by Liao et al.⁴⁹ demonstrated progressive pH acidification and HM enrichment from rural to urban zones, supporting trends observed in our study. Thus, these studies corroborate our conclusion that industrial land use significantly degrades soil fertility parameters while elevating HM contamination in Haridwar soils.

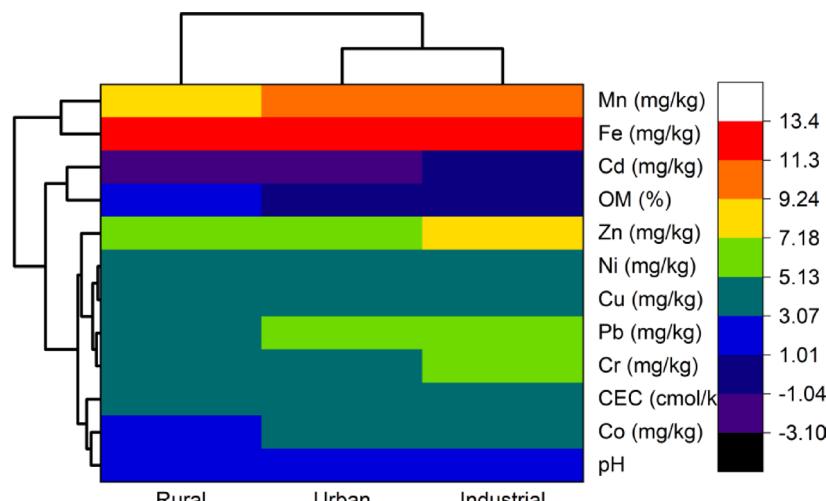


Fig. 3. Hierarchical cluster analysis (HCA) of soil physicochemical and HM properties across rural, urban, and industrial zones in Haridwar.

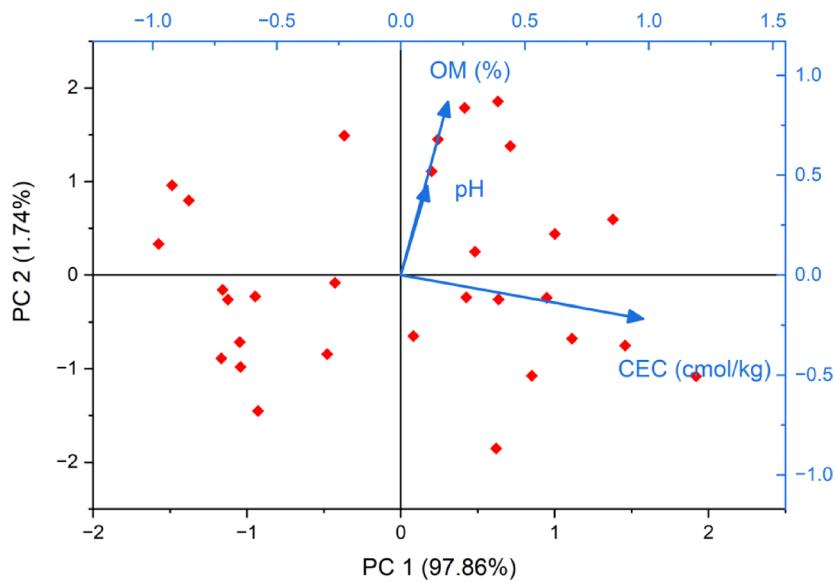


Fig. 4. Principal component analysis biplot for extracted eigenvectors from PC1 (covariance of 97.86%) for weight assignment to soil properties.

Parameter	Sampling area	Heavy metals									
		Cd	Cr	Cu	Co	Fe	Mn	Ni	Pb	Zn	
Measured (C_i)	Rural	0.13	12.16	15.19	3.91	5166.65	460.92	10.77	9.75	35.63	
	Urban	0.29	34.86	22.74	8.91	7820.07	633.22	25.30	42.66	88.96	
	Industrial	0.71	59.35	33.33	14.34	10,415.80	913.87	31.83	81.86	187.60	
Background reference (B_i)	–	0.30	40.00	20.00	10.00	38,000.00	500.00	30.00	25.00	70.00	
Traditional CF	Rural	0.43	0.30	0.76	0.39	0.14	0.92	0.36	0.39	0.51	
	Urban	0.96	0.87	1.14	0.89	0.21	1.27	0.84	1.71	1.27	
	Industrial	2.37	1.48	1.67	1.43	0.27	1.83	1.06	3.27	2.68	
Adjustment factor (f)	Rural	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	
	Urban	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	
	Industrial	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	
Modified CF	Rural	0.29	0.20	0.51	0.26	0.09	0.62	0.24	0.26	0.34	
	Urban	0.72	0.65	0.85	0.66	0.15	0.94	0.63	1.27	0.95	
	Industrial	2.42	1.52	1.70	1.47	0.28	1.87	1.08	3.35	2.74	

Table 2. Traditional and modified contamination factors of heavy metals (HMs) by sampling area (measured concentrations are expressed in mg/kg dry weight). Background values for HMs were derived from^{41,42,50}, along with international standards provided by the FAO and USEPA. CF = contamination factor (measured concentration / background). CF classification: $CF < 1$ = low; $1 \leq CF < 3$ = moderate; $3 \leq CF < 6$ = considerable; $CF \geq 6$ = very high. MCF = CF \times f (adjusted for soil retention; f derived from pH, OM, CEC). Regulatory permissible limits (Table 1) are concentration-based and differ conceptually from enrichment-based CF categories.

Results of traditional versus modified contamination factor

Figure 4 presents a PCA biplot used to derive weighting coefficients for soil properties in the MCF model. In this, PC1 explains 97.86% of the total variance, indicating strong dimensionality reduction. CEC shows the highest positive loading along PC1, followed by pH and OM, suggesting their dominant contribution to spatial variability in soil quality. The vector lengths and orientations showed the relative influence of each parameter. These eigenvectors were used to assign differential weights to each property, ensuring that soil characteristics influencing contaminant mobility and retention were appropriately integrated into the MCF calculation. The assigned weights, *i.e.*, 0.97 for CEC, 0.19 for OM, and 0.10 for pH, indicate that CEC contributed most significantly to the variability in soil quality and was thus given the highest importance in the MCF model. The comparison between traditional and MCF across the three land-use categories is shown in Table 2. Herein, traditional CF values showed increasing contamination trends from rural to industrial areas, with the industrial zone exhibiting the highest CFs for most metals viz., Pb (3.27), Zn (2.68), and Cd (2.37) exceeded the threshold for moderate to

considerable contamination ($CF > 1$), indicating substantial anthropogenic inputs. Urban soils showed moderate contamination levels for Pb (1.71) and Zn (1.27), suggesting the influence of vehicular emissions, urban runoff, and possible use of contaminated compost or sludge. In contrast, rural soils exhibited CFs well below 1 for all metals, reflecting baseline or geogenic concentrations with minimal anthropogenic influence.

In rural regions of Haridwar, agriculture is the predominant land use, with wheat, rice, and sugarcane being the main crops. Farming practices in these areas rely on conventional inputs, including urea, diammonium phosphate (DAP), and pesticides, but at lower intensities compared to peri-urban and industrial-adjacent zones. The absence of industrial effluent irrigation or direct waste disposal reduces the contamination load. Nevertheless, repeated use of phosphate fertilizers is a known source of Cd and Pb accumulation in soils, which may explain the trace but detectable levels of these metals in rural samples. Thus, while rural areas remain the least polluted in comparison to urban and industrial sites, agricultural inputs still contribute to low-level background enrichment. On the other hand, the MCF, which incorporated adjustment factors based on soil properties ($f=0.67$ for rural, 0.74 for urban, 1.02 for industrial), provided a more differentiated risk profile. This adjustment appropriately downscaled the rural and urban contamination estimates while amplifying industrial site risk. For instance, the MCF for Pb increased slightly from 3.27 to 3.35 in the industrial zone, reinforcing its role as a priority pollutant. On the other hand, in rural areas, Pb's MCF decreased to 0.26, aligning more accurately with its negligible pollution level. Similar patterns were observed for Zn, Cd, and Mn, where industrial zones consistently presented values above 1.5 after modification, denoting considerable contamination. Overall, the MCF proved effective in refining the risk characterization by integrating soil properties into contamination assessment. It provided a more realistic interpretation than traditional CF alone, especially for soils with heterogeneous properties and pollutant profiles.

It should be noted that CF values represent enrichment relative to regional background concentrations and are not identical to regulatory or health-based permissible limits. Thus, $CF > 1$ indicates anthropogenic enrichment even where the absolute metal concentration may still be below a jurisdictional guideline. For example, several HM in the industrial zone showed $CF > 1$ (Table 2) indicating enrichment relative to background, while their measured concentrations remained at or below some international guideline values listed in Table 1. The MCF provides additional context i.e., by increasing or decreasing CF values according to soil retention capacity, MCF helps identify cases where enrichment is likely to translate to higher mobility and ecological exposure (low CEC soils), and conversely where enrichment is buffered by strong retention (high CEC, high OM).

Although pH, OM, and CEC are key determinants of soil buffering capacity, we recognize that other parameters such as clay content, carbonate levels, mineralogy, and redox potential (Eh) also play important roles in regulating HM behavior. Their exclusion in this study was due to practical and analytical limitations, but integrating them into future iterations of the MCF model would further improve accuracy and transferability. This is the first study that utilizes the PCA-based weight assignment for refining the CF model. However, previous studies have used PCA and other tools to assess pollution levels. Out of them, Bashir et al.⁵¹ applied contamination and enrichment indices to assess agricultural soils near industrial clusters in Guntur, revealing that PCA explained over 78% of variance and highlighted differential parameter weights. Similarly, Tiwari et al.⁵² used PCA loading adjustments in Punjab's soil contamination assessments, identifying key HMs and parameter weights.

Validation and verification of the modified contamination factor model

The plot between traditional and MCF values showed strong predictive alignment as presented in Fig. 5. The linear regression reveals a high correlation coefficient ($r=0.9863$) and an $R^2=0.9729$, indicating that over 97% of the variance in MCF values can be explained by the traditional CF values. The regression equation ($y=1.0589x-0.1886$) implies a near-unity slope and minimal intercept bias, confirming that the MCF model retains the magnitude and directionality of contamination trends with only a slight offset. Validation statistics in Table 3 further support model adequacy. The coefficient of determination from observed vs. predicted values closely aligns with the graphical fit and reinforces the model's ability to capture the underlying data. The RMSE (0.190) and MAE (0.152) indicate low average deviations between traditional and MCF, while the MSE (0.036) confirms the absence of large residuals. The model bias (-0.124) suggests a minor underestimation tendency, but the deviation is small and does not compromise predictive reliability. The model efficiency, calculated as 0.939, mirrors the R^2 value and confirms that the MCF performs substantially better than using the means of traditional CFs alone. The high efficiency is especially significant given that the MCF model incorporates adjustment factors and weightings to account for soil properties. Thus, these results validate the MCF as a statistically sound and context-sensitive alternative to the traditional method, offering improved accuracy in pollution quantification without introducing significant distortion or variability.

The validation results align well with recent multivariate soil assessment frameworks. For instance, Agyeman et al.⁵³ using CF-PMF (contamination factor-positive matrix factorization), reported improved R^2 and reduced RMSE compared to traditional methods in Czech soils. Similarly, SD-PCA (spatial distribution – principal component analysis) applied in China demonstrated > 90% variance capture and robust source attribution. Environmental chemical surveys have leveraged PCA-coupled regression approaches, yielding $R^2 > 0.95$ and low RMSE in heavy-metal predictions⁵⁴.

Conclusion

The present investigation demonstrates that integrating soil physicochemical characteristics such as pH, OM, and CEC into the contamination assessment process significantly improves the accuracy of ecological risk evaluation in metal-impacted agricultural soils. The MCF model, developed using principal component-based weighting and normalization, effectively differentiates pollution severity across land-use gradients in Haridwar. The industrial zone exhibited elevated contamination levels and reduced soil buffering capacity, indicating an

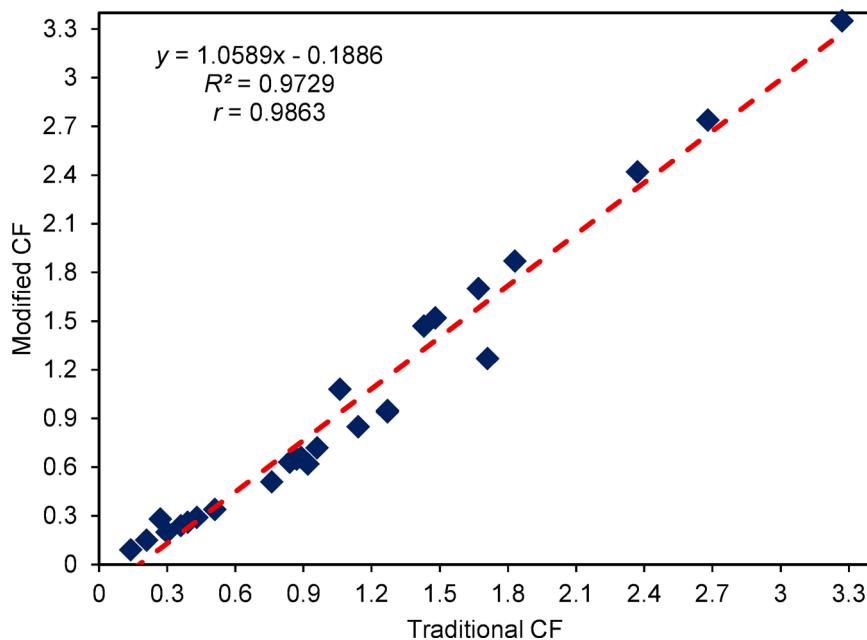


Fig. 5. Traditional versus modified contamination factor plots for linear fitness and correlation.

Validation parameter	Value
R^2	0.972
RMSE	0.190
MSE	0.036
MAE	0.152
Model bias	-0.124
Model efficiency	0.939

Table 3. Statistical validation parameters for the performance of the modified contamination factor model. R^2 : Coefficient of determination, RMSE: Root mean squared error, MSE: Mean squared error, MAE: Mean absolute error, Model bias: Average deviation between traditional and modified CF values, Model efficiency: Nash–Sutcliffe model efficiency coefficient.

intensified ecological risk. In contrast, rural soils, despite minor trace metal presence, retained their functional integrity due to higher pH, OM, and CEC. The MCF model can mitigate misclassification of pollution risk by scaling traditional CF values based on retention potential, as evidenced by lower MCFs in resilient soils and increased values in vulnerable conditions. Validation statistics confirmed strong model reliability, with high predictive accuracy and minimal bias. This study shows the necessity of adopting context-sensitive indices like MCF for environmental monitoring, particularly in heterogeneous agro-ecosystems subjected to multiple stressors. While the model provides a robust framework, its effectiveness across broader agroecological zones warrants further evaluation. Although this study quantified total HM concentrations using USEPA Method 3051A, it did not directly assess bioavailable metal fractions (e.g., DTPA- or EDTA-extractable forms). Instead, bioavailability was indirectly inferred through soil properties such as pH, OM, and CEC, which are known to regulate metal solubility and retention.

The present study used 10 composite samples per land-use category due to budget and logistical constraints. While compositing subsamples improves field representativeness and the balanced design supports comparative analyses across land uses, the sample size and single-season sampling limit the ability to generalize results temporally and to detect small effect sizes. Seasonal changes in rainfall, irrigation, redox status, and agricultural activities can influence HM mobility, pH, OM, and CEC; therefore, multi-season monitoring with increased replication is recommended to quantify temporal variability and improve risk assessment. Future studies should incorporate direct measurement of bioavailable forms to strengthen ecological risk assessments and provide a more complete picture of soil contamination dynamics. While this study validates the MCF against traditional CF, benchmarking with other widely used indices such as the Pollution Load Index (PLI), Geo-accumulation Index (Igeo), and Ecological Risk Index (RI) could further confirm its robustness across diverse soil conditions, which we recommend as a direction for further studies. Future studies should incorporate temporal monitoring and explore the role of additional soil attributes such as redox potential and clay content.

Data availability

Data will be made available on reasonable request to the corresponding author.

Received: 17 July 2025; Accepted: 31 October 2025

Published online: 02 December 2025

References

- Hou, D. et al. Metal contamination and bioremediation of agricultural soils for food safety and sustainability. *Nat. Rev. Earth Environ.* **1**, 366–381 (2020).
- Angon, P. B. et al. Sources, effects and present perspectives of heavy metals contamination: Soil, plants and human food chain. *Helijon* **10** (2024).
- Alengebawy, A., Abdelkhalak, S. T., Qureshi, S. R. & Wang, M. Q. Heavy metals and pesticides toxicity in agricultural soil and plants: Ecological risks and human health implications. *Toxics* **9**, 1–34 (2021).
- UNESCO. Worsening land degradation impacts 3.2 billion people worldwide. UNESCO. (2018). At <https://www.unesco.org/en/articles/worsening-land-degradation-impacts-32-billion-people-worldwide>
- Bhattacharyya, R. et al. Soil degradation in India: Challenges and potential solutions. *Sustainability (Switzerland)* **7**, 3528–3570 (2015).
- Jayakumar, M., Surendran, U., Raja, P., Kumar, A. & Senapathi, V. A review of heavy metals accumulation pathways, sources and management in soils. *Arab. J. Geosci.* **14**, 2156 (2021).
- Uchimiya, M. et al. Chemical speciation, plant uptake, and toxicity of heavy metals in agricultural soils. *J. Agric. Food Chem.* **68**, 12856–12869 (2020).
- Lallawmzuali, G. et al. Assessment of the heavy metal contaminations of roadside soil in Aizawl Mizoram (India): An in-depth analysis utilising advanced scientific methodologies. *Asian J. Water Environ. Pollut.* **21**(5), 37. <https://doi.org/10.3233/AJW240058> (2024).
- Li, Y. et al. Accumulation, fractionation and health risk assessment of fluoride and heavy metals in soil-crop systems in northwest China. *Sci. Total Environ.* **663**, 307–314 (2019).
- Bhutiani, R., Kulkarni, D. B., Khanna, D. R., Tyagi, V. & Ahamed, F. Spatial and seasonal variations in particulate matter and gaseous pollutants around integrated industrial estate (IIE), SIDCUL, Haridwar: a case study. *Environ. Dev. Sustain.* **23**, 15619–15638 (2021).
- Singh, H., Pandey, R., Singh, S. K. & Shukla, D. N. Assessment of heavy metal contamination in the sediment of the River Ghaghara, a major tributary of the River Ganga in Northern India. *Appl. Water Sci.* **7**, 4133–4149 (2017).
- Kumar, S. & Anshumali. Heavy metal pollution and health risk assessment in upland and riparian soils of the Ganga River basin. *Discov. Soil* **2** (2025).
- Thakur, R. K. & Kumar, V. in *Springer Proceedings in Earth and Environmental Sciences* 1–17 (Springer International Publishing, 2022). https://doi.org/10.1007/978-3-031-05335-1_1
- Gupta, V. Vehicle-generated heavy metal pollution in an urban environment and its distribution into various environmental components. *Environ. Concerns Sustain. Dev.* **113**–127 (2020). https://doi.org/10.1007/978-981-13-5889-0_5
- Singh, J. et al. Spatio-temporal analysis of air quality in high traffic density zones of Haridwar City, India. *Environ. Monit. Assess.* **197** (2025).
- Lee, J. C. & Pandey, B. D. Bio-processing of solid wastes and secondary resources for metal extraction—A review. *Waste Manag.* **32**, 3–18 (2012).
- Kumar, S., Shwetank & Jain, K. A multi-temporal landsat data analysis for land-use/land-cover change in Haridwar Region using remote sensing techniques. *Procedia Comput. Sci.* **171**, 1184–1193 (2020).
- Arora, T. et al. Pollution load assessment and potential environmental risks of composite industrial effluents discharged from SIIDCUL Integrated Industrial Estate, Haridwar (Uttarakhand), India. *J. Environ. Biol.* **38**, 205–216 (2017).
- Biswas, B. et al. The fate of chemical pollutants with soil properties and processes in the climate change paradigm—A review. *Soil Syst.* **2**, 1–20 (2018).
- Popoola, O. J., Ogundele, O. D., Ladapo, E. A. & Senbore, S. in *Soil Microbiome in Green Technology Sustainability* 351–375 (Springer Nature Switzerland, 2024). https://doi.org/10.1007/978-3-031-71844-1_15
- Zhou, H. et al. Identification and hazard analysis of heavy metal sources in agricultural soils in ancient mining areas: A quantitative method based on the receptor model and risk assessment. *J. Hazard Mater.* **445**, 130528 (2023).
- Zinkutė, R., Taraskevičius, R., Jankauskaitė, M. & Stankevičius, Ž. Methodological alternatives for calculation of enrichment factors used for assessment of topsoil contamination. *J. Soils Sediments* **17**, 440–452 (2017).
- Milinovic, J. et al. Multivariate analysis applied to X-ray fluorescence to assess soil contamination pathways: Case studies of mass magnetic susceptibility in soils near abandoned coal and W/Sn mines. *Environ. Geochim. Health* **46**, 202 (2024).
- Caporale, A. G. & Violante, A. Chemical processes affecting the mobility of heavy metals and metalloids in soil environments. *Curr. Pollut. Rep.* **2**, 15–27 (2016).
- Zhong, R. et al. Freeze–thaw–induced regulation of petroleum hydrocarbon adsorption in cold-region soils: Role of organic matter dynamics. *Water Res.* **287**, 124495. <https://doi.org/10.1016/j.watres.2025.124495> (2025).
- Kumpiene, J., Lagerkvist, A. & Maurice, C. Stabilization of As, Cr, Cu, Pb and Zn in soil using amendments—A review. *Waste Manag.* **28**, 215–225 (2008).
- Bao, Y. et al. Interactions between organic matter and Fe (hydr)oxides and their influences on immobilization and remobilization of metal(loid)s: A review. *Crit. Rev. Environ. Sci. Technol.* **52**, 4016–4037 (2022).
- Lwin, C. S., Seo, B. H., Kim, H. U., Owens, G. & Kim, K. R. Application of soil amendments to contaminated soils for heavy metal immobilization and improved soil quality—A critical review. *Soil Sci. Plant Nutr.* **64**, 156–167 (2018).
- Lu, H. et al. Alkaline amendments improve the health of soils degraded by metal contamination and acidification: Crop performance and soil bacterial community responses. *Chemosphere* **257**, 127309 (2020).
- Zeng, F. et al. The influence of pH and organic matter content in paddy soil on heavy metal availability and their uptake by rice plants. *Environ. Pollut.* **159**, 84–91 (2011).
- Clarholm, M. & Skjellberg, U. Translocation of metals by trees and fungi regulates pH, soil organic matter turnover and nitrogen availability in acidic forest soils. *Soil Biol. Biochem.* **63**, 142–153 (2013).
- Walkley, A. & Black, I. A. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **37**, 29–38 (1934).
- Ma, D. et al. Long-term effects of conventional cultivation on soil cation exchange capacity and base saturation in an arid desert region. *Sci. Total Environ.* **949**, 175075 (2024).
- Tibane, L. V. & Mamba, D. Environmental risk assessment, principal component analysis, tracking the source of toxic heavy metals of solid gold mine waste tailings, South Africa. *Environ. Forensics* **25**, 254–270 (2024).
- Li, C. et al. Soil physiochemical properties and landscape patterns control trace metal contamination at the urban-rural interface in southern China. *Environ. Pollut.* **250**, 537–545 (2019).
- Kabata-Pendias, A. *Trace Elements in Soils and Plants*. CRC Press (2011).

37. Sarala Thambavani, D. & Prathipa, V. Quantitative assessment of soil metal pollution with principal component analysis, geo accumulation index and enrichment index. *Asian J. Environ. Sci.* **7**, 125–134 (2012).
38. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, 1–24 (2021).
39. Lin, F., Chen, X. & Yao, H. Evaluating the use of Nash–Sutcliffe efficiency coefficient in goodness-of-fit measures for daily runoff simulation with SWAT. *J. Hydrol. Eng.* **22**, 5017023 (2017).
40. WHO. *Guidance document for WHO monographers and reviewers evaluating contaminants in food and feed (replacing 2001 guidance)*. WHO (2001).
41. Bureau of Indian Standards. *IS 14684:1999—Determination of nitrogen and nitrogenous compounds in soil by Kjeldahl method*. BIS (1999).
42. Kabata-Pendias, A. *Trace elements in soils and plants (3rd ed.)*. CRC Press. (2001).
43. Indian Council of Agricultural Research (ICAR). *Handbook of agriculture (6th ed.)*. New Delhi: Directorate of Knowledge Management in Agriculture. (2015).
44. Parvin, A. *et al.* Chemical speciation and potential mobility of heavy metals in organic matter amended soil. *Appl. Environ. Soil. Sci.* **2022** (2022).
45. Sajidu, S. M. I., Persson, I., Masamba, W. R. L., Henry, E. M. T. & Kayambazinthu, D. Removal of Cd²⁺, Cr³⁺, Cu²⁺, Hg²⁺, Pb²⁺ and Zn²⁺ cations and AsO₄³⁻ anions from aqueous solutions by mixed clay from Tundulu in Malawi and characterisation of the clay. *Water SA* **32**, 519–526 (2006).
46. Bharti, M., Kamboj, N. & Kamboj, V. Effect of untreated wastewater on soil quality: A case study in Ranipur Rao watershed in Haridwar region (Uttarakhand), India. *Adv. Environ. Pollut. Manag. Wastew. Impacts Treat. Technol.* **145**, 145–157 (2020).
47. Bhardwaj, S. *et al.* Assessment of the soil quality of Haridwar Uttarakhand India: A comparative study. *Environ. Conserv.* **21**, 155–164 (2020).
48. Padbhushan, R. *et al.* Impact of land-use changes on soil properties and carbon pools in India: A meta-analysis. *Front. Environ. Sci.* **9**, 794866 (2022).
49. Liao, R. *et al.* Three-dimensional assessment of heavy metal contamination in soil affected by urbanization at the urban-rural interface of Chengdu. *Emerg. Contam.* **11**, 100482 (2025).
50. Alloway, B. J. *Heavy Metals in Soils (2nd ed.)*. Blackie Academic & Professional (1995).
51. Bashir, Z., Raj, D. & Selvasebian, R. Source profiling, pollution and health risk assessment of heavy metals in agricultural soils around an industrial cluster using PCA and GIS-assisted PMF. *Environ. Monit. Assess.* **197**, 711 (2025).
52. Tiwari, M., Rathod, T., Bhangare, R., Sahu, S. K. & Pulhani, V. Comprehensive elemental and carbon profiling of Punjab's soils: Distribution patterns, pollution assessment, and source attribution. *Environ. Geochem. Health* **47**, 1–19 (2025).
53. Agyeman, P. C. *et al.* Health risk assessment and the application of CF-PMF: A pollution assessment-based receptor model in an urban soil. *J. Soils Sediments* **21**, 3117–3136 (2021).
54. Agyeman, P. C., John, K., Kebonye, N. M., Borůvka, L. & Vašát, R. Combination of enrichment factor and positive matrix factorization in the estimation of potentially toxic element source distribution in agricultural soil. *Environ. Geochem. Health* **45**, 2359–2385 (2023).

Acknowledgements

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/104/46.

Author contributions

Mohssen Elbagory: Conceptualization, Funding acquisition, Methodology, Project administration, Validation, Writing—original draft; Nagwa EL-Khateeb: Data curation, Methodology, Software, Validation, Visualization, Writing—review & editing; Sahar El-Nahrawy: Data curation, Software, Validation, Visualization, Writing—review & editing; Alaa El-Dein Omara: Data curation, Software, Validation, Visualization, Writing—review & editing; Ibrahim Mohamed: Data curation, Software, Validation, Visualization, Writing—review & editing; Mahmoud El-Sharkawy: Data curation, Software, Validation, Visualization, Writing—review & editing; Amal Zayed: Data curation, Software, Validation, Visualization, Writing—review & editing; Madhumita Goalal: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Writing—original draft; Deep Gupta: Formal analysis, Investigation, Methodology, Resources, Writing—original draft; Piyush Kumar: Formal analysis, Investigation, Methodology, Resources, Writing—original draft; Pankaj Kumar: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Writing—original draft; Ivan Širić: Conceptualization, Methodology, Software, Supervision, Validation, Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding

The Deanship of Research and Graduate Studies at King Khalid University funding this work Through Large Research Project under Grant Number RGP2/104/46.

Declarations

Ethics approval and consent to participate

This study did not involve any animal or human participants, so ethical approval was not required. All authors agreed to participate in the study.

Consent for publication

All authors agreed to publish in the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-27110-w>.

Correspondence and requests for materials should be addressed to I.Š.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025