Appraisal of pollution and health risks associated with coal mine contaminated soil using multimodal statistical and Fuzzy-TOPSIS approaches

Sumit Kumar[a,e], Sonali Banerjee[a], Saibal Ghosh[a], Santanu Majumder[b], Jajati Mandal[c], Pankaj Kumar Roy[d], Pradip Bhattacharyya[a]*

[a]Agricultural and Ecological Research Unit, Indian Statistical Institute, Giridih, Jharkhand-India

[b]Department of Life and Environmental Sciences, Bournemouth University (Talbot Campus), Fern Barrow, Poole BH12 5BB, UK

[c]School of Sciences, Engineering & Environment, University of Salford, UK

[d]School of Water Resource Engineering, Jadavpur University, Kolkata, West Bengal, India

[e]School of Environmental Studies, Jadavpur University, Kolkata, West Bengal, India

*Corresponding author. E-mail: pradip.bhattacharyya@gmail.com (Pradip Bhattacharyya)

Abstract

The present study assesses the concentration, probabilistic risk, source classification, and dietary risk arising from heavy metal (HMs) pollution in agricultural soils affected by coal mining in eastern part of India. Analyses of soil and rice plant indicated significantly elevated levels of HMs beyond the permissible limit in the contaminated zones (zone 1: Pb_{Soil}:108.24±72.97, Cu_{Soil}:57.26±23.91, Cd_{Soil}:8.44±2.76, Cr_{Soil}:180.05±46.90, Ni_{Soil}:70.79±25.06 mg/kg; Pb_{Grain}:0.96±0.8, Cu_{Grain}:8.6±5.1, Cd_{Grain}:0.65±0.42, Cr_{Grain}:4.78±1.89, Ni_{Grain}:11.74±4.38 mg/kg and zone 2: Pb_{Soil}:139.56±69.46, Cu_{Soil}:69.89±19.86, Cd_{Soil}:8.95±2.57, Cr_{Soil}:245.46±70.66, Ni_{Soil}:95.46±22.89 mg/kg; Pb_{Grain}:1.27±0.84, Cu_{Grain}:7.9±4.57, Cd_{Grain}:0.76±0.43, Cr_{Grain}:8.6±1.58, Ni_{Grain}:11.50±2.46 mg/kg) compared to the uncontaminated zone (zone 3). Carcinogenic and non-carcinogenic health risks were computed based on the HMs concentration in the soil and rice grain, with Pb, Cr, and Ni identified as posing a high risk to human health. Monte Carlo simulation, the solubility-free ion activity model (FIAM), and severity adjusted margin of exposure (SAMOE) were employed to predict health risk. FIAM hazard quotient (HQ) values for Ni, Cr, Cd, and Pb were > 1, indicating a significant non-carcinogenic risk. SAMOE (risk thermometer) results for contaminated zones ranged from low to moderate risk (Cr_{SAMOE}: 0.05, and Ni_{SAMOE}: 0.03). Fuzzy-TOPSIS and variable importance plots (from random forest) showed that Ni and Cr were mostly responsible for the toxicity in the rice plant, respectively. A self-organizing map for source classification revealed common origin for the studied HMs with zone 2 exhibiting the highest contamination. The positive matrix factorization model for the source apportionment identified coal mining and transportation as the predominant sources of HMs. Spatial distribution analysis indicated higher contamination near mining sites as compared to distant sampling sites. Consequently, this study will aid environmental scientists and policymakers controlling HM pollution in agricultural soils near coal mines.

Keywords: Coal mine, Free ion activity model, Monto Carlo Simulation, Pollution and Health risk, fuzzy-TOPSIS

1. Introduction

Coal, commonly referred to as the “Black Diamond," stand out as one of the most crucial minerals in the Earth’s crust. Globally, India holds 2nd and 4th positions in terms of coal production and reserve deposits. The coal mines of Giridih (Kabdibad mine and Giridih open cast) were among the first coal mines in India, covering 20 coal seams and an area of about 28.5 km^2 (11 sq mi) (Ministry of Coal, 2018). For a developing nation like India, coal mining is an ubiquitous necessity. However, mining also has negative consequences for the environment. Intensive coal extraction, consumption, and transportation result in environmental damage, leading to the contamination of heavy metals (HMs) (Pb, Ni, Cr, Cu, and Cd) in agricultural soil and evoking public health concerns (Fasinu and Orisakwe, 2013; Cortes-Ramirez et al., 2018). According to Coal India (2020), India has roughly 380 coal mines, generating massive quantities of overburden and tailings which are often stacked up close to the agricultural lands in mining regions. Furthermore, mining practices expose a significant amount of rock to weather events like rain or erosion, leaching HMs into the environment (Cortes-Ramirez et al., 2018). These mine wastes, due to their high toxicity, non-biodegradability and persistent nature pose a hazard to the ecosystem. Consequently, crops grown on agricultural land near coal mines absorb potentially toxic HMs in the plant (crop) parts (shoot, grain, and root), posing a possible threat to the food chain (Zhao and Wang, 2020; Banerjee et al., 2023). The consumption of these HMs-polluted crops from agricultural fields near mining sites poses severe health hazards, including non-carcinogenic (NCR) and carcinogenic (CR) effects (Zakir et al., 2021). Studies have already shown that long-term exposure to HMs increases the chance of various diseases like cancer in both animals and humans (Chen et al., 2022).

Multiple methods exist for estimating the degree of contamination and health risks associated with HMs in soil affected by mining wastes. Evaluating the likelihood of HMs containing the soil is pivotal, requiring an understanding of their solubility in the soil ecosystem and their transportation from agricultural soil to plant parts. Numerous researchers have employed techniques such as FIAM, SAMOE, partial dependency plots, and Fuzzy-TOPSIS have been used to estimate the risk posed by HMs-contaminated soils (Golui et al., 2017; Singh et al., 2019; Chowdhury et al., 2020; Ghosh et al.,

2023; Banerjee et al., 2023). A comprehensive analysis of the health concerns associated with HMs is required for a community living close to coal mines. The Monte Carlo simulation model (MCS) is one of the best methods for probability risk analysis. It recognizes the priorities of risks and assesses whether the risk exceeds the permissible value (Tong et al., 2019). It is possible to calculate the positive contribution of each point in the PMF model and assist in allocating the contribution of each element. It can be combined with SOM and cluster analysis to compare the quantitative attributes of different HMs (Nakagawa et al., 2020). This combined model would evaluate both environmental and human health hazards in the region of the coal mines, identify the most significant sources of pollution, and manage and regulate those sources while protecting ecosystems.

To the best of our knowledge, there is lack of substantial research on the analysis of HMs and associated health risks in coal mine impacted agricultural fields of India. This study is designed to fill this gap by examining the extent of HMs pollution from two coal mines and comparing the health risks associated. Additionally, we compared the bioavailability, plant accumulation, and effects of HMs on food chains between the two mines. Hence, the primary objectives of this research were to (a) ascertain the extent of HMs contamination in agricultural soil near two coal mining areas considering their geo-spatial pattern and (b) evaluate the relative effectiveness of the FIAM models in predicting the transport of HMs from soil to grain (rice), their exposure in diet through SAMOE and the contribution of each HMs in the rice crop through the random forest and Fuzzy-TOPSIS analysis, (c) assess both CR and NCR health risks resulting from HMs, and (d) utilize the PMF model with spatial tools to identify and allocate contamination sources of HMs.

Materials and methods

1.1. Description of study site and sampling

Giridih have two active mining sites: one is Giridih Open Cast mines (24°10'25.47"N-86°15'14.23"E), and the other is Kabdibad Mines (24°9'6.23"N-86°18'0.61"E). Human settlements and agricultural fields surround both mining sites. Topsoil (10–15 cm) from an agricultural field supposed to be affected by coal mining activities was collected from the surrounding area of two coal mining sites in

Giridih, namely Giridih Open Cast Mines (zone 1; n = 25) and Kabdibad Mines (zone 2; n = 24). An uncontaminated zone (zone 3; n = 20), at a distant location, was also selected for comparison. Sampling sites were randomly selected to cover the most extensive areas around the mines. Samples were collected first by removing the topsoil cover and digging to a depth of 10–15 cm using a shovel and spade and washed thoroughly after that to avoid overlapping samples and contamination. 1 kg of soil samples along with paddy plants were uprooted from each sampling site, and their coordinates were recorded using the GPS (Global Positioning System). The GPS coordinates for each sampling site in all three zones (zones 1, 2, and 3) are represented in Table S1. Samples (soil and paddy) were carefully stored in zipper pouches with proper labels for safe transportation and sample storage.

2.2 Preparation of sample and its analysis

The physicochemical analyses of soil samples were done following the methods outlined in Page et al. (1982). The bio-available HMs were extracted using DTPA as proposed by Lindsay and Norvell (1978). The HM content of soil (total and DTPA extractable) and plant (root, shoot, and grain) samples were estimated by Systronics make Atomic Absorption Spectrophotometer (AA S-816).*2.3 Prediction models for HMs accumulated by plant*

2.3.1 Solubility-free ion activity model (FIAM)

FIAM was used to predict the HMs in rice grain with soil pH and organic carbon as the variables. The transfer factor (TF) was calculated as the ratio of metal content in a plant $[Mplant]$ to metal ions activity in the soil pore water $[M^{n-}]$, given in Eq. (1) (Mirecki et al., 2015).

$$TF = \log \frac{[Mplant]}{[M^{n-}]} \quad (1)$$

Freundlich equation is a simple pH-dependent method to calculate free ion activity of HMs (Datta and Young, 2005) given in Eq. (2).

$$p(M^{n-}) = \{p[M_C] + k_1 + k_2\ pH\} / n_F \quad (2)$$

where (M^{n-}) was the free HMs ions activity in the soil; $M_C$was the easily displaceable soil HMs content; k_1 and $k_2$were experimental metal-specific constants; and $n_F$was the power term from the

Freundlich equation. By combining Eqs. (1) and (2), the expression for predication of HMs uptake by plants can be represented as:

$$\mathrm{p[Mplant]} = \mathrm{C} + \beta_1 \mathrm{p[M_C]} + \beta_2 \mathrm{pH} \quad (3)$$

where C = $k_1/nF - \log TF$, β1 = 1/nF, β2 = k_2/nF and C, β1 and β2 were empirical metal and plant-specific coefficients (Mandal et al., 2019). Equation (3) was parameterized by using the "SOLVER" addon facility in Microsoft Excel.

2.3.2 Risk assessment

The Hazard Quotient (HQ) was estimated to assess the health risk to humans from the consumption of rice, grown in contaminated soil. It was calculated as the average daily intake (ADI) of rice divided by their reference dose (RfD) of each HMs as outlined in IRIS (2019). The HQ was calculated by the following equation:

$$HQ = \frac{Mplant \times W}{RfD \times 68} \quad (4)$$

Where $Mplant$ is the HMs concentration in rice grain, W is the weight of rice consumed daily and RfD is the reference dose for different HMs.

2.4 Risk assessment of HMs through dietary exposure

2.4.1 Risk thermometer and severity adjusted margin of exposure (SAMOE)

Risk Thermometer is a method of risk assessment proposed by the Swedish National Food Agency (Sand et al., 2015). SAMOE of different heavy metals was calculated by the following equation:

$$\mathrm{SAMOE} = \frac{TDI}{(AFBMR \times AF \times SF \times E)} \quad (5)$$

where TDI is the provisional tolerable daily intake (PTDI), AFBMR is the non-linear relation in dose range (1/10; BMR was the benchmark response), AF is assessment factor taken as 10, SF is severity factor considered as 100, and E is the exposure factor (concentration of HMs).

2.4.2 Carcinogenic risk from dietary exposure to rice grain

The target cancer risk (TCR) is an important tool for assessing dietary risks considering the individual's lifetime exposure to carcinogenic HMs (Sengupta et al., 2021). It was calculated from the following equation:

$$TCR = \frac{E_{Fr} \times ED \times FIR \times C \times CSF}{BW \times AT} \times 10^{-3} \quad (6)$$

Where E_{Fr} is the frequency of exposure to various HMs, ED is the duration of exposure. FIR is rate of ingestion of food. C is HM content in rice grain, CSF is the oral cancer slope factor for HMs, BW is the average body weight and AT is the average carcinogenic exposure time, 10^{-3} is unit conversion factor (Antoine et al., 2017).

2.5. Fuzzy TOPSIS

Entropy was employed in the TOPSIS method to identify alternatives that were nearest to Si+ and farthest from Si-, thereby aiding in decision-making (Singh et al., 2019).

Phase 1: In order to rate the "alternatives" (rice parts) and "criteria" (metals), based on HMs contamination status. Assuming "b" viable alternatives, P = {P_1, ..., P_b}, were to be examined alongside "c" criteria. Q = {Q_1, … , Q_c}

Phase 2: The criterion ratings were assigned according to a matrix O, where z_{ij} indicates alternative P_i' value for criterion Q_j.

$$O_{b\times c} = \begin{bmatrix} z_{11} & z_{12} \cdots & z_{lc} \\ \vdots & \dots \quad z_{ij} & \vdots \\ z_{bl} & \cdots & z_{bc} \end{bmatrix} \quad (7)$$

Phase 3: The weights of the criteria were calculated based on entropy as follows:

$$r_{ij} = \frac{z_{ij}}{z_{1j}+\cdots+z_{bj}}; \forall j \in \{1, \dots, c\}, \text{ and} \quad (8)$$

$$G_j = -\frac{1}{\ln b}\sum_{i=1}^{b} r_{ij} ln r_{ij}; \forall j \in \{1, \dots, c\} \quad (9)$$

Where $0 \leq G_j \leq 1$ and indexes with higher entropy have more variation. Therefore, the criteria's weight can be estimated as follows:

$$w_j = \frac{a_j}{a_1 + \cdots + a_c} \quad (10)$$

Where, $a_j = 1 - w_j$; All the weights were aggregated into the w_{cxc} matrix.

Phase 4: Enumeration of normalized matrix as follow,

$$\overline{Y_{ij}} = \frac{Y_{ij}}{\sqrt{\sum_{i=1}^{n} Y_{ij}^2}} \quad (11)$$

Phase 5: Determination of weighted normalized matrix

$$U_{ij} = \overline{Y_{ij}} \times W_j \quad (12)$$

Phase 6: Calculate the ideal best (Si^+) and ideal worst value (Si^-)

$$S_i^+ = \left[\sum_{j=1}^{b} \left(V_{ij} - V_j^+\right)^2\right]^{0.5} \quad (13)$$

$$S_i^- = \left[\sum_{j=1}^{b} \left(V_{ij} - V_j^-\right)^2\right]^{0.5} \quad (14)$$

Phase 7: Evaluation of Performance Score and rank according by the following equation.

$$P_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (15)$$

2.6 Assessment of HM pollution through various indices

To assess the severity of pollution from the HMs pollution indices like geoaccumulation index (I_{geo}), pollution load index (PLI), degree of contamination (C_d), potential ecological risk index(R_i) were calculated according to the formula outlined by Tomlinson et al.(1980), Hakanson (1980), Zhang et al.(2007), and Zerizghi et al.(2022) respectively.

2.7 Assessment of soil to human health risk through ingestion, inhalation, and dermal contact

The presence of heavy metals (HMs) in soil can impact human health through three exposure pathways: ingestion, inhalation, and dermal contact. The accumulation of these metals in the human body has been associated with both carcinogenic and non-carcinogenic risks and hence proper assessment is required.

2.7.1 Assessment of non-carcinogenic risk

The non-carcinogenic risk was assessed as the Hazard Quotient (HQ), calculated as the ratio of ADI (Average Daily Intake) to reference dose (RfD) through different exposure pathways, as per Eq. (19) (USEPA, 1986). ADI for different exposure pathways were calculated by using Eqs. (16)-(18) (USEPA, 1989). Hazard Index (HI) was calculated from the Eq. (20) (Goumenou and Tsatsakis, 2019).

$$\mathrm{ADI_{ing}} = \mathrm{C_i} \times \frac{\mathrm{IR_{ing} \times E_{Fr} \times ED}}{\mathrm{BW \times AT}} \times 10^{-6} \quad (16)$$

$$\mathrm{ADI_{inh}} = \mathrm{C_i} \times \frac{\mathrm{IR_{inh} \times E_{Fr} \times ED}}{\mathrm{BW \times PEF \times AT}} \quad (17)$$

$$\mathrm{ADI_{derm}} = \mathrm{C_i} \times \frac{\mathrm{DAF \times SA \times SAF \times EF \times ED}}{\mathrm{BW \times AT}} \times 10^{-6} \quad (18)$$

$$\mathrm{HQ_{ing}} = \frac{\mathrm{ADI_{ing}}}{\mathrm{RfD_{ing}}}, \mathrm{HQ_{inh}} = \frac{\mathrm{ADI_{inh}}}{\mathrm{RfD_{inh}}}, \mathrm{HQ_{derm}} = \frac{\mathrm{ADI_{derm}}}{\mathrm{RfD_{derm}}} \quad (19)$$

$$\mathrm{HI} = \sum \mathrm{HQ_i} = \mathrm{HQ_{ing}} + \mathrm{HQ_{inh}} + \mathrm{HQ_{derm}} \quad (20)$$

Where $\mathrm{ADI_{ing}}$, $\mathrm{ADI_{inh}}$, and $\mathrm{ADI_{derm}}$were the average daily intake through different pathways. $\mathrm{C_i}$ is the concentration of metal in soil, $\mathrm{IR_{ing}}$ and $\mathrm{IR_{inh}}$are the rate of ingestion and the rate of inhalation, $\mathrm{E_{Fr}}$ is exposure frequency, ED is the duration of exposure, BW is body weight, PEF is particulate emission factor, DAF is dermal absorption factor, SA is skin exposure area, SAF is skin adherence factor, and AT is the average time over which the dose is averaged (Zerizghi et al., 2022).

2.7.2 Assessment of carcinogenic risk

Carcinogenic risk associated with HMs and exposure pathways from contaminated soil was calculated by Eq. (21) (Jolly et al., 2022).

$$CR = LAAD \times SF \quad (21)$$

where, LAAD (Lifetime Average Daily Dose) = ($LAAD_{ing}$+ $LAAD_{inh}$ + $LAAD_{derm}$) is the weighted average for each exposure pathway, SF denotes the slope factor of studied HMs.

2.8. Probabilistic risk assessment through Monte-Carlo simulation (MCS) model

Monte Carlo simulation was undertaken for evaluating probabilistic non-carcinogenic health risk through sensitivity analysis of studied HMs (Kalantary et al., 2022). The simulation was performed using Oracle Crystal Ball application (version11.1.1)

2.9. Pattern identification of HMs through Kohen's self-organizing map (SOM) model

SOM is an artificial neural network used for clustering, estimating, and predicting complex data. It resembles the neural network present in biological systems and is used to find the source pattern of HMs present in the affected soil. For better visualization, neurons were represented into 2-D units, called computational grid in the unsupervised learning process of the model. R-studio software was used to create the model using Kohonen package version 3.0.11 (Kohonen et al., 2001; Park et al., 2003).

2.10. Positive factorization matrix model for source apportionment of HMs

For source apportionment of the HMs the USEPA PMF 5.0, a multivariate receptor model was used. It is a reliable technique for identifying the source of metals in soils, previously used for water and air pollution (Zhang et al., 2017).

2.11. Geostatistical methods and spatial distribution maps of heavy metals

Spatial interpolation is an important technique for quality interpretation, comparative analysis, and prediction. Kriging interpolation was applied using QGIS software, and the best fitted model was selected based on nugget values (smallest) and root mean square standardized error (minimum value close to 1).

2.12 Statistical analysis and modelling

The statistical parameters such as mean, standard deviation, range, least significant difference (LSD-ANOVA), correlation plot ('corPlot' package)and fuzzy-TOPSIS were performed in R-Studio (version 1.3.1093 2.3.1). The random forest was performed in R-studio using 'randomForest' package, 'vip' package for variable importance plot and 'pdp' package for partial dependence plot.

2. Result and Discussion

3.1 Description of physio-chemical and HMs in soil

The Table S2 provides an overview of physico-chemical properties of agricultural soil contaminated with HMs, collected from the Giridih open cast mines (zone 1) and Kabdibad mines (zone 2) in Jharkhand. In this study, both zones 1 and zone 2 exhibit significantly lower pH (p =0.0005; LSD = 0.135) compared to zone 3 (uncontaminated zone). Specifically, soils of Zone 1 (6.01±0.5) were more acidic than zones 2 (6.21±0.51). The EC in the contaminated zones (zone 1 and 2) were significantly (p =0.005; LSD = 0.006) lower compared to the uncontaminated zone (zone 3). Zone 2 (Kabdibad mines) exhibited the highest OC values (2.16 ± 0.33 %) when compared to the other two zones. Carbon level in the soils near coal mining sites was high due to deposition of coal waste. A significant decline in OC content was observed in uncontaminated zone (p =0.003; LSD = 0.272) compared to contaminated zones. Since these toxic mine wastes in the mining region (zone 1 and zone 2) are not nutrient-rich but instead contain a variety of toxic HMs (Pb, Cu, Cd, Cr and Ni) both the total and bioavailable (DTPA) fraction of HMs were significantly higher in zone 2 and zone 1 than zone 3.As a result of the dumping of hazardous coal mine residues near agricultural land, Cr and Pb poisoning was prevalent in these coal mining regions (zone 1 and 2). The observations align with the findings of Xiao et al. (2020) and Zerizghi et al. (2022) where Cr, and Pb toxicity were found in coal mine area. The current situation raises concern for the well-being of humans, plants, and other living creatures.

3.2 Accumulation of HMs in paddy

The concentration of HMs in plant parts (root, shoot, and grain) collected from Zone 1, 2 and 3 has been represented in Fig. 1. The HMs concentration in the roots, shoots, and grains of zone 1 and zone

2 were significantly higher (above the WHO, 1996 prescribed limits) than zone 3. The findings of Ghosh et al. (2024) and Banerjee et al. (2023) supported the result of current study. Further, a significant positive correlation (r >0.6) was found between bioavailable (DTPA) fraction of HMs concentration and rice grains, as depicted in Fig. 2. The results were consistent with studies by Banerjee et al. (2023), which claimed that there was a direct link between hazardous HMs uptake by rice grains and their bioavailable forms.

3.3 Importance of DTPA extractable HMs on plant uptake

The availability of HMs in the plant is not governed by the total HMs present in soil, instead it is governed by their available forms (DTPA fraction). Fig. S1 illustrated the variable importance plot derived from the random forest models showcasing the relationship between DTPA extractable metal fraction and their uptake by plant (root, shoot, and grain). The %IncMSE indicated the degree of model's precision decreases if the variable is omitted.Additionally, the IncNodePurity quantified the purity of the nodes at the terminus of the tree without each variable. Identifying a split with high inter-node variation and small intra-node variance leads to higher increases in node purities.. Through this model, we observed that the rice grain and shoot parts were significantly influenced by plant bioavailable form of metals particularly in case of Cr, Pb and Cu. The metal concentrations in the contaminated soil varied significantly due to diverse waste materials deposited at mine sites over the years.

Fig. 3.demonstrates the 3-D (three dimensional) partial dependence plot from random forest algorithm representing uptake in different parts of rice plant (root, shoot, and grain) most affected by DTPA extractable HMs. The grain and shoot make substantial contribution to HMs accumulation. Concentrations of Pb and Cu in shoot and grain increased when DTPA_Pb and DTPA_Cu concentartions in soil surpassed 15 mg/kg and 7 mg/kg, respectively. In the case of Cd, the uptake in shoots and roots increased when the DTPA_Cd concentration in soil surpassed from 2.4 mg/kg in soil.

3.4 Human health risk assessment through rice grain

Rice, one of the most important cereal crops, is a major constituent in Indian diets. In Section 3.2, it can be noticed that rice grains grown in contaminated zones (1 and 2) exhibited elevated concentrations of toxic HMs. Prolonged consumption of rice from these regions may potentially endanger the health of living organisms. Therefore, in our study, we considered the concentration of HMs in the rice grain from the contaminated zones (zones 1 and 2) for calculating the health risk. This choice was made because the concentration of HMs in the rice grains from the uncontaminated zone (3) was negligible and well below the prescribed limit by WHO (1996).

3.4.1 Assessment of risk through FIAM and FIAM-HQ

pH, DTPA-extractable HMs, and OC were used to evaluate the variability of HMs in rice grain with FIAM. The prediction coefficient of FIAM, as well as plant-specific model parameters (C, β1, and β2), are shown (Table S3 and Figs. S6-7). Khaledian et al. (2017) have previously reported that pH and organic carbon (OC) of the soil are the most important factors governing the solubility of HMs in soil. This model was very useful in predicting the transfer of HMs from soil to plant without measuring the actual solubility of metals in the soil. In Zone 1, β1 was negative, and β2 was positive for all heavy metals (HMs) except β2-Cd and β2-Cr. On the other hand, in Zone 2, β1 was negative, except for β1-Cr and β1-Ni, while β2 was positive for all studied HMs compared to Zone 3 (β2-Cr, β2-Ni, and β2-Cd).. As per the results, $\beta 1$ and $\beta 2$ are negative for most of the HMs in both zones (zone 1 and zone 2), and the uptake of HMs from soil to rice grain might positively affected by OC and pH. Mandal et al. (2019) found similar trends in how metal mobility, pH, and OC are related to each other.

Using the FIAM model's hazard quotient (HQ), this study evaluated the health risk associated with HMs uptake in rice grain (Table S3). According to Raj et al. (2022), the FIAM-HQ > 0.5, rice grains was regarded as a threat to human health. The mean values of HQ-Pb, HQ-Cd, HQ-Cr, HQ-Ni for zone 1 were 1.61, 3.81, 9.39, 3.45, for zone 2, 2.14, 4.53, 9.6, 3.38 and for zone 3, 0.08, 0.15, 0.15, 0.51, respectively. These findings indicated that zone 2 has the highest health risk from consuming rice grains, with HQ values exceeding the recommended levels for Ni, Pb, Cd, and Cr except for Cu as compared with zone 1 and 3. In light of this, it might not be advisable for humans to consume rice

306 that has been produced on this agricultural soil affected by coal mine waste (zone 1 and 2). Similar
307 research by Banerjee et al. (2023) suggested that rice cultivated in metal-polluted soil was not suitable
308 for human intake.

309 *3.4.2 Dietary risk from grain intake through SAMOE and risk thermometer*

310 The presence of HMs in rice and their consumption has adverse effects on human health, which was
311 assessed in terms of target cancer risk (TCR). TCR results (Table S4) for Pb at both zones show no
312 cancer risk as their values were lower than the tolerable limit of 10^{-4} (zone 1: 2.37E-04; zone 2:
313 3.15E-04). But in the case of Ni (zone 1: 6.29E-02; zone 2: 6.162E-02), Cd (zone 1: 1.45E-03; zone 2:
314 1.72E-03), and Cr (zone 1: 1.41E-02; zone 2: 1.44E-02), TCR was much higher than the tolerable
315 limit. Similarly, the TCR values for zone 3 were below the acceptable threshold.

316 The risk thermometer and the value of SAMOE for various HMs are shown in Fig. 4 and Table S4.
317 The risk thermometer was used to evaluate the toxicity of HMs through the classification of the risk
318 scale and concern level. The results indicated that Pb, Cu, and Cd pose a low risk in zones 1 and 2,
319 while Cr and Ni pose a moderate risk. The risk level of studied HMs based on SAMOE for both zones
320 was ordered as follows: NiSAMOE>CrSAMOE>CdSAMOE>PbSAMOE>CuSAMOE. Also, zone 3
321 presents no risk for Pb, Cu, and Cd, whereas Cr and Ni represent a modest risk. As the TCR values for
322 Cd, Cr, and Ni were higher than the tolerable limit and simultaneously the NiSAMOE and CrSAMOE
323 showed moderate risk, prolonged consumption of rice from fields contaminated by coal mines (zone 1
324 and zone 2) may have a detrimental impact on the environment and living beings.

325 *3.4.3. Fuzzy TOPSIS method for risks posed by HMs*

326 The optimal alternative (metals) for heavy metal concentrations in different rice parts was
327 estimated using the fuzzy-TOPSIS multi-criteria decision-making approach. For this experiment,
328 alternatives {*A* = rice parts}were evaluated based on the criteria {Cr, Ni, Cu, Cd, Pb} for zone 1 and
329 zone 2.The ideal best and ideal worst values are shown in (TableS4), and the criteria weights of the
330 estimated concentration of heavy metals from rice parts (root, shoot and grain) were enumerated
331 through entropy technique as follows: zone 1 W_j= 0.214, 0.314, and 0.470; zone 2 W_j= 0.230, 0.281,

and 0.487; zone 3 W_j= 0.293, 0.278, and 0.427 respectively. From the Table the sequence of metal accumulation in different parts of rice, exhibiting a declining pattern among three zones (zone 1, 2 and 3) was ascertained. The order of metal accumulation follwerd the order: zone 1and 2: Ni > Cu > Cr > Cd > Pb. Conversely, in zone 3, the order was Cu > Ni > Cr > Pb > Cd. Result revealed that, for both mining sites, Ni toxicity was responsible for the highest risk in rice. This type of observation was obtained due to the deposition of toxic coal mine waste near an agricultural field. Our outcome showed resemblance with Saif-Ud-Din et al. (2022) where risk was evaluated based on accumulation of metals in different food resources.

3.5.Assessment of HM pollution through different indices

The geoaccumulation index (Igeo) evaluates the possibility of HMs accumulation in the ecosystem. In this study, Igeo values indicate extremely low to moderate pollution ($0 \geq$ Igeo ≤ 2) all over the contaminated zones, due to the presence of HMs (Pb, Cu, Cd, Cr, and Ni) (Fig. S2). The average Igeo for Cr showed a moderate level of pollution in both zone 1 and zone 2 as compared with zone 3. Further, the mean C_f values of all 3 zones were in the order of Cr > Pb > Ni > Cu > Cd, showing low to high contamination ($0 \geq C_f \leq 6$). Zone 3 was in the low-contaminated zone, but the C_f values for Cr and Pb indicated the highest contamination in zone 1 and 2. These indices aligned with the results of studies carried out by Ghosh et al. (2023) and Banerjee et al. (2023). The value of the contamination index (degree of contamination) for contaminated zones (zone 1 and zone 2) indicates a moderate degree of contamination ($0 \leq Cd \leq 6$). Igeo, and C_f together revealed contamination of Pb and Cr in the coal mine area. Several children have died from lead poisoning due to exposure to contaminated soil from mining practices in Nigeria (Mandal et al., 2022) and Senegal (WHO, 2022). The International Agency for Research on Cancer classified Cr as a group I carcinogen (Kim et al., 2015), and Núñez et al. (2016) reported cancer mortality in a population exposed to Cr-enriched soil. The ecological risk index (Ri) was the cumulative sum of the ecological risk factors of all the studies (Fig. S2). The mean ecological risk factors for each HMs were in the order of Pb>Cr>Ni>Cd>Cu and possess low ecological risk(Er^i< 30;Ri <100). Hence, the soil contamination in the coal mine area contributes to a low potential ecological risk.

3.6. Soil to human health risk assessment

HM-contaminated soil adversely affects human health through different exposure pathways such as ingestion, inhalation, and dermal contact, which might be carcinogenic or non-carcinogenic. In the current study, health risk (carcinogenic and non-carcinogenic) was calculated for contaminated zones (zones 1 and 2) and uncontaminated zone 3 using the value of total HM content (Pb, Cu, Cd, Cr, and Ni).

3.6.1 Assessment of non-carcinogenic risk

The non-carcinogenic risk was calculated in terms of the ADD, HQ, and HI for all three exposure pathways given in tables S6 and S7. In the case of ingestion, Pb had the highest hazard quotient (HQ) value for both contaminated zones (zone 1; adult: 1.86E-02; child: 1.73E-01) and (zone 2; adult: 2.41E-02; child: 2.25E-01). In contrast, for inhalation, Cr had the maximum HQ for zone 1 (adult: 5.18E-04; child: 9.19E-04) and zone 2 (adult: 7.44E-04; child: 1.32E-03). For the dermal route, Pb had the highest HQ for zone 1 (adult: 1.84E-04; child: 1.32E-03) and zone 2 (adult: 2.39E-04; child: 1.71E-03). HI represented a cumulative risk from all three exposure pathways, which is the best way to represent the non-carcinogenic risk. HI results for all studied heavy metals and sampling sites were < 1 (Table S7) for all 3 zones, suggesting low non-carcinogenic risk in the study area.

3.6.2 Assessment of carcinogenic risk using MCS

The International Agency for Research on Cancer (1990) has categorized HMs, such as Cd, Cr, and Ni, as group 1 carcinogens, and Pb as a group 2 carcinogen. The present study observed the carcinogenic risks of Pb, Cd, Cr, and Ni from contaminated soils (zones 1 and 2), considering all three modes of human body exposure, with no significant risk from zone 3 (uncontaminated site).The enrichment of Cu was studied but was not included in the cancer risk assessment due to its lower anthropogenic toxicity (SF=0). (Onyedikachi et al., 2018).HMs in all three zones had CR < 1.00E-06 (Banerjee et al., 2023), for both Cd and Pb, indicating level 1 contamination and an extremely low lifetime cancer risk (Table S8). On the other hand, the average CR value for Cr and Ni was found to be > 1.00E-06 in zone 1 (contaminated site)as compared to zone 2 and 3. Hence, both children and

adults were observed to be most likely to have a lifetime cancer risk due to Cr and Ni, as supported by the sensitivity analysis (Fig. 5a) comparing contaminated regions to uncontaminated regions. Upon comparing the cancer risk (CR) values, it was observed that children had a higher risk than their adult counterparts, indicating that children were more likely to be affected. Due to their underdeveloped organ systems and body ratios (Al Osman et al., 2019), children were at an increased risk of carcinogenic exposure. They are more susceptible because they spend more time outdoors, frequently chew non-food items, and consume food in quantities higher than their per-kilogram body weight (Ruggieri et al., 2017).Similar trends were found in a study conducted by Kabir et al. (2022). Total Cancer Risk (TCR) values represent the cumulative cancer risk of Pb, Cd, Cr, and Ni through three human exposure pathways (inhalation, ingestion, and dermal contact). TCR values were higher than the unacceptable level (10^{-6}) for zone 1 (child: 3.94E-05; adult: 3.37E-05) and zone 2 (child: 1.03E-04; adult: 4.57E-05). No anthropogenic risk was observed in zone 3 (child: 8.22E-07; adult: 1.20E-07) (Li et al., 2020), with the highest contribution from Cr. The simulation of cancer risk (Fig. 5b) for adults and children suggested that children in zone 2 were at a higher risk as compared to zone 3. Fig. 5a and 5b depict the probabilistic risk assessment values of TCR for adults and children in the contaminated zones (zones 1 and 2), supporting the results described above.

3.7 Source classification of HMs through SOM analysis

The unified distance matrix (U-matrix) and the component planes of SOM analysis were illustrated in Fig. 6 (a–c). In SOM planes, the color gradient indicates the normalized values of each variable. Similar colors show a positive correlation, while different ones show a negative correlation. The hexagon represented neurons in component planes, and the least space of the hexagon represents strong similar characteristics within the sample. (Wang et al., 2020). Fig. 6a demonstrated the neural patterns of HMs, similar neural patterns pointing towards similarities in their origin. Based on the SOM result, Ni showed a higher concentration in the upper left-to-middle portion of the matrix map. whereas Cu was found to be more concentrated in the upper right corner and lower left plane (Fig. 6a). Cr, Pb, and Cu showed a similar concentration pattern and ranked inside all HMs: Ni > Cu > Cr > Cd. HMs concentration in three different coal mine sites (zones 1, 2, and 3) showed that the neurons

in the upper middle left corner to lower middle left corner of zone 2 (Kabdibad mine) were higher than those in zone 1 (Giridih open cast mine: upper right corner to lower middle right corner) and zone 3 (uncontaminated site: lower middle to lower right) (Fig. 6c). Based on the U-matrix clustered by zone, most of the neurons belong to zone 2 (n = 16 neurons) with respect to zone 1 (n = 14 neurons) and zone 3 (n = 6 neurons) (Fig. 6b). Based on the overall SOM results, areas contaminated with HMs can be characterized by regional characteristics based on the differences in HMs emissions associated with zone 2.

3.8 Source apportionment of HMs present in affected soil by PMF

The PMF model, identified four factors (Factors 1, 2, 3, and 4) as the most significant contributors to contamination in the study region (Fig. 7 and Fig. S3). The least stable Q value hindered the evaluation of the most appropriate and best-fit factors. The determination coefficient (r^2) (Table S9) between observed value and predicted value indicates a strong correlation between HMs (Cd had the minimum r^2 value of 0.116 and Pb had the maximum r^2 of 1, and the remaining had values greater than 0.997). The results showed that for Factor 1, Ni and Cd were the primary loading elements, accounting for 59.9% and 37.3%, respectively. Ni and Cd had low Cf values, indicating a natural origin, but their higher values suggested anthropogenic contributions (Chen et al., 2015; Wang et al., 2021). It appears that Ni and Cd originated from a parent source, with Factor 1 likely linked to lithogenic enrichment resulting from displaced parent material due to open coal mining activities in the area. For Factor 2, the main contributor was Cu, accounting for 59%. Factor 2 may represent industrial activities, as Cu is widely used in industries (Wang et al., 2020). It could also be associated with agricultural activities through the application of livestock manures (Liang et al., 2017), fertilizers, and other agrochemicals (Jiang et al., 2017). Factor 3 contributed to 59.7% of the Cr. It should be noted that the mean concentration of Cr in contaminated zones exceeded its background concentration. The higher Cf value of Cr in the contaminated zones indicated anthropogenic addition through the weathering of exposed overburden accumulated during open-cast mining. The main contributors for Factor 4 were Pb and Ni, accounting for 73.6% and 35.2%, respectively. Pb is crucial for tracing vehicular emissions from leaded petrol, and although its use has been discontinued, traces

still exist in the soil. Mining operations induce high traffic on routes used to transport coal and mining wastes, allowing Pb to enter the environment through brake wear, tire friction, road surface erosion, and wasted lube oil (Yan et al., 2018; Men et al., 2019).

3.9 Geostatistical distribution of HMs

Geostatistical modeling was employed to assess the spatial distribution of HMs (specifically Cr, Ni, Cd, Pb, and Cu) within soils in two contaminated zones (zone 1 and 2) and an uncontaminated zone (zone 3). Fig.S4 and Table S10 present the characteristics of the most suitable semi-variogram model and the results from ordinary kriging, offering insights into the spatial distribution of these HMs at the sampling sites. When evaluating the semi-variogram, we observed that nugget values, indicating variability at zero distances, were positive in all zones(zone 1, 2, and 3). To determine the best-fit model, we considered the Akaike Information Criteria (AIC) for this study. Notably, the partial sill and nugget values showed the most significant differences among the data sets. For Cd, the semi-variograms exhibited a lower nugget effect in all three sample zones (zone 1, 2, and 3). Similar trends were observed for other HMs such as Cr, Cu, Pb, and Ni in zone 1 zone 3, except for Pb in zone 2. This suggests that the sampling density was adequate to capture the spatial patterns of the data (Ghosh et al., 2024).We employed a Gaussian model for Cu, Cd, and Cr in zone 1, Pb in zone 2, and Pb, Cu, and Ni in zone 3. Meanwhile, we applied a spherical model for Pb in zone 1 and Cr in zone 2. Strong spatial dependencies were observed for Cd in zone 1, and for Pb, Cu, Cd, and Ni in zone 2, as well as Cu, Cd, and Ni in zone 3.The model's suitability was demonstrated by the close Root Mean Square Error (RMSE) and Average Standard Error (ASE) values. G values for all three zones were greater than zero, indicating that spatial prediction using semi-variogram parameters outperformed assuming the mean of observed values. The semi-variogram attributes derived from the experimental data effectively depicted spatial variations (Banerjee et al., 2023).

Interpolation analysis using Kriging of the factor scores obtained from the PMF model revealed that Factor 1 was predominantly concentrated in the eastern-southern to northwestern part of zone 1, while in zone 2, it centered in the middle (Fig. S5). Factor 2 was mainly distributed in the northeastern to west-southern and eastern-southern to southwestern sections of both zone 1 and 2, particularly in

areas occupied by mines and partially used for agriculture. Consequently, high Factor 3 scores were observed in the west-southern part of both zones. An agricultural region extending from the northwestern to north-southern and eastern-southern to southwestern parts of both zones emerged as a pollution hotspot for Factor 4.

4. Conclusion

The current study has assessed HM contamination in the agricultural soil surrounding two coal mines: Giridih open cast mines (zone 1) and Kabdibad mines (zone 2) in the Giridih district of Jharkhand. The elevated levels of HMs (Pb, Cu, Cd, Cr, and Ni) were found, particularly higher concentrations of Ni, Pb, and Cr, in zone 1 and 2 compared zone 3. The concentration of HMs was also higher in the plant parts (root, shoot, and grain).Rice grains posed the greatest risk from Ni, Cr, and Pb, as determined by FIAM, SAMOE, and Fuzzy-TOPSIS analyses. TCR values for Cd, Cr, and Ni exceeded tolerable limits, with NiSAMOE and CrSAMOE indicating a moderate health risk.Consequently, prolonged consumption of rice grown from the area is not recommended. In zones 1 and 2, pollution indices indicate soil quality deterioration. The geo-statistical distribution map of HMs reveals higher contamination near mining sites compared to distant sampling sites. Analysis of the SOM component plane suggested similarities in the source and origin of HMs. PMF analysis indicated that HMs contamination in the area resulted from four different activities: lithogenic, coal mining, industrial activities, and transportation. Hence, the mining authorities should develop proper plans to mitigate HMs leaching from the dumping of toxic coal mine overburdens and tailings.

5. References

Ahmad N, Pandey P (2020). Spatio-temporal distribution, ecological risk assessment, and multivariate analysis of heavy metals in Bathinda District, Punjab, India. Water, Air, & Soil Pollution, 231(8): 1-32.

Al Osman M, Yang F, Massey I Y (2019). Exposure routes and health effects of heavy metals on children. Biometals, 32: 563-573.

Annual report, (2018). District Mineral Foundation Trust, Giridih (Jharkhand). https://cdn.s3waas.gov.in/s39b72e31dac81715466cd580a448cf823/uploads/2020/09/2020092225.pdf (Accessed 16 June 2022).

Antoine J M, Fung L A H, Grant CN (2017). Assessment of the potential health risks associated with the aluminium, arsenic, cadmium and lead content in selected fruits and vegetables grown in Jamaica. Toxicology Reports, 4: 181-187.

Antonelli A, Zizka A, Carvalho F A, Scharn R, Bacon C D, Silvestro D, Condamine F L (2018). Amazonia is the primary source of Neotropical biodiversity. Proceedings of the National Academy of Science, 115(23): 6034-6039.

Banerjee S, Ghosh S, Jha S, Kumar S, Mondal G, Sarkar D, Datta R, Mukherjee A, Bhattacharyya P (2023). Assessing pollution and health risks from chromite mine tailings contaminated soils in India by employing synergistic statistical approaches. Science of Total Environment, 880: 163228.

Brtnický M, Pecina V, Hladký J, Radziemska M, Koudelková Z, Klimánek M, Vaverková M D (2019). Assessment of phytotoxicity, environmental and health risks of historical urban park soils. Chemosphere, 220: 678-686.

Chen H, Teng Y, Lu S, Wang Y, Wang J (2015). Contamination features and health risk of soil heavy metals in China. Science of Total Environment, 512: 143-153.

Chen L, Zhou M, Wang J, Zhang Z, Duan C, Wang X, Zhao S, Bai X, Li Z, Li Z, Fang L (2022). A global meta-analysis of heavy metal (loid) s pollution in soils near copper mines: Evaluation of pollution level and probabilistic health risks. Science of Total Environment, 835: 155441.

Chowdhury N R, Das A, Joardar M, De A, Mridha D, Das R, Roychowdhury T (2020). Flow of arsenic between rice grain and water: Its interaction, accumulation, and distribution in different fractions of cooked rice. Science of Total Environment, 731: 138937.

Coal India (2020). About the company. Available at: https://archive.coalindia.in/en-us/company/aboutus.aspx (Accessed: 11 September 2022).

Cortes-Ramirez J, Naish S, Sly P D, Jagals P (2018). Mortality and morbidity in populations in the vicinity of coal mining: a systematic review. BMC Public Health, 18(1): 1-17.

Datta S P, Young S D (2005). Predicting metal uptake and risk to the human food chain from leaf vegetables grown on soils amended by long-term application of sewage sludge. Water, Air, & Soil Pollution, 163: 119-136.

Fasinu P S, Orisakwe O E (2013). Heavy metal pollution in sub-Saharan Africa and possible implications in cancer epidemiology. Asian Pacific Journal of Cancer Prevention, 14(6): 3393-3402.

Friedman J H (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Ghose M K (2004). Effect of opencast mining on soil fertility. Journal of Scientific & Industrial Research, 63: 1006-1009.

Ghosh S, Banerjee S, Prajapati J, Mandal J, Mukherjee A, Bhattacharyya P (2023). Pollution and health risk assessment of mine tailings contaminated soils in India from toxic elements with statistical approaches. Chemosphere, 138267.

Ghosh S, Mondal S, Mandal J, Mukherjee A, Bhattacharyya P (2024). Effect of metal fractions on rice grain metal uptake and biological parameters in mica mines waste contaminated soils. Journal of Environment Science, 136: 313-324.

Golui D, Mazumder D G, Sanyal S K, Datta S P, Ray P, Patra P K, Sarkar S, Bhattacharya K (2017). Safe limit of arsenic in soil in relation to dietary exposure of arsenicosis patients from Malda district, West Bengal-A case study. Ecotoxicology & Environment Safety, 144: 227-235.

Goumenou M, Tsatsakis A (2019). Proposing new approaches for the risk characterisation of single chemicals and chemical mixtures: The source related Hazard Quotient (HQS) and Hazard Index (HIS) and the adversity specific Hazard Index (HIA). Toxicology Report, 6: 632-636.

Hakanson L (1980). An ecological risk index for aquatic pollution control. A sedimentological approach. Water Research, 14(8): 975-1001.

International Agency for Research on Cancer (1990). Chromium, Nickel, and welding. IARC monographs on the Evaluation of Carcinogenic Risks to Human/World Health Organization, 49.

IRIS (2019). Integrated risk information system-database, US Environmental Protection Agency.

Jiang Y, Chao S, Liu J, Yang Y, Chen Y, Zhang A, Cao H (2017). Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China. Chemosphere, 168: 1658-1668.

Jolly Y N, Rakib M, Jahan R, Sakib M S, Shahadat M A, Rahman A, Akter S, Kabir J, Rahman M S, Begum B A, Rahman R (2022). Impact of Industrially Affected Soil on Humans: A Soil-Human and Soil-Plant-Human Exposure Assessment. Toxics, 10(7): 347.

Kabir MH, Kormoker T, Shammi RS, Tusher TR, Islam MS, Khan R, Omor MZ, Sarker ME, Yeasmin M, Idris AM, (2022). A comprehensive assessment of heavy metal contamination in road dusts along a hectic national highway of Bangladesh: Spatial distribution, sources of contamination, ecological and human health risks. Toxin Reviews, 41(3):860-79.

Kalantary RR, Barzegar G, Jorfi S (2022). Monitoring of pesticides in surface water, pesticides removal efficiency in drinking water treatment plant and potential health risk to consumers using Monte Carlo simulation in Behbahan City, Iran. Chemosphere, 286:131667.

Khaledian Y, Pereira P, Brevik E C, Pundyte N, Paliulis D (2017). The influence of organic carbon and pH on heavy metals, potassium, and magnesium levels in Lithuanian Podzols. Land Degradation & Development, 28(1): 345-354.

Kim H S, Kim Y J, Seo Y R (2015). An overview of carcinogenic heavy metal: molecular toxicity mechanism and prevention. Journal of Cancer Prevention, 20: 232.

Kohonen T (2001). Self-Organizing Maps, third edition. Springer, Berlin

Kormoker T, Proshad R, Islam M S, Tusher T R, Uddin M, Khadka S, Chandra K, Sayeed A (2022). Presence of toxic metals in rice with human health hazards in Tangail district of Bangladesh. International Journal of Environment Health Research, 32(1): 40–60.

566 Li K, Cui S, Zhang F, Hough R, Fu Q, Zhang Z, Gao S, An L (2020). Concentrations, possible
567 sources and health risk of heavy metals in multi-media environment of the Songhua River, China.
568 International Journal of Environment Research and Public Health, 17(5): 1766.

569 Liang J, Feng C, Zeng G, Gao X, Zhong M, Li X, Li X, He X, Fang Y (2017). Spatial distribution and
570 source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan,
571 China. Environmental Pollution, 225: 681-690.

572 Lindsay W L, Norvell W (1978). Development of a DTPA soil test for zinc, iron, manganese, and
573 copper. Soil Science Society of America Journal, 42(3): 421-428.

574 Mandal J, Golui D, Datta S P (2019). Assessing equilibria of organo-arsenic complexes and predicting
575 uptake of arsenic by wheat grain from organic matter amended soils. Chemosphere, 234: 419-426.

576 Masindi V, Muedi K L (2018). Environmental contamination by heavy metals. Heavy metals, 10:
577 115-132.

578 Men C, Liu R, Wang Q, Guo L, Miao Y, Shen Z (2019). Uncertainty analysis in source apportionment
579 of heavy metals in road dust based on positive matrix factorization model and geographic information
580 system. Science of Total Environment, 652: 27-39.

581 Ministry of Coal, GOI (2018). Coal Reserves in India. Ministry of Coal Government of India.
582 https://coal.gov.in/en/major-statistics/coal-reserves(Accessed 9 June 2022).

583 Mirecki N, Agic R, Sunic L, Milenkovic L, Ilic Z S (2015). Transfer factor as indicator of heavy
584 metals content in plants. Fresenius Environmental Bulletin 24(11c): 4212-4219.

585 Nakagawa K, Yu Z Q, Berndtsson R, Hosono T (2020). Temporal characteristics of groundwater
586 chemistry affected by the 2016 Kumamoto earthquake using self-organizing maps. Journal of
587 Hydrology, 582: 124519.

588 Núñez O, Fernández-Navarro P, Martín-Méndez I, Bel-Lan A, Locutura J F, López-Abente G (2016).
589 Arsenic and chromium topsoil levels and cancer mortality in Spain. Environmental Science and
590 Pollution Research, 23: 17664-17675.

591 Onyedikachi U B, Belonwu D C, Wegwu M O (2018). Human health risk assessment of heavy metals
592 in soils and commonly consumed food crops from quarry sites located at Isiagwu, Ebonyi
593 State. Ovidius University Annals of Chemistry, 29(1): 8-24.

594 Page AL, Miller RH, Keeney DR (1982). Methods of Soil Analysis. Part 2. Chemical and
595 Microbiological Properties. American Society of Agronomy. In Soil Science Society of America,
596 1159.

597 Park Y S, Céréghino R, Compin A, Lek S (2003). Applications of artificial neural networks for
598 patterning and predicting aquatic insect species richness in running waters. Ecological Modelling,
599 160(3): 265-280.

600 Rahaman S, Sinha A C, Pati R, Mukhopadhyay D (2013). Arsenic contamination: a potential hazard
601 to the affected areas of West Bengal, India. Environmental Geochemistry and Health, 35: 119-132.

602 Raj D, Kumar A, Maiti S K (2022). Health Risk Assessment of Children Exposed to the Soil
603 Containing Potentially Toxic Elements: A Case Study from Coal Mining Areas. Metals, 12(11): 1795.

604 Ruggieri F, Majorani C, Domanico F, Alimonti A (2017). Mercury in children: current state on
605 exposure through human biomonitoring studies. International journal of environmental research and
606 public health, 14(5): 519.

607 Saif-Ud-Din A S, Hussain S, Hussain J, Luqman M, Hussain J, Ali S (2022). Evaluation of heavy
608 metal contamination in indigenous fruits and associated human health risk: evidence from Fuzzy-
609 TOPSIS approach. Global Nest Journal, 24 (3): 435-444.

610 Sand S, Bjerselius R, Busk L, Eneroth H, Sanner-F'arnstrand J, Lindqvist R (2015). The risk
611 thermometer - a tool for risk comparison. Swedish National Food Agency.

612 Sengupta S, Bhattacharyya K, Mandal J, Bhattacharya P, Halder S, Pari A (2021). Deficit irrigation
613 and organic amendments can reduce dietary arsenic risk from rice: Introducing machine learning-
614 based prediction models from field data. Agriculture, Ecosystem & Environment, 319: 107516.

615 Singh K R, Dutta R, Kalamdhad A S, Kumar B (2019). Information entropy as a tool in surface water
616 quality assessment. Environmental Earth Science, 78: 1-12.

617 Taylor S R, McLennan S M (1995). The geochemical evolution of the continental crust. Rev.
618 Geophys, 33(2): 241-265.

619 Tomlinson D L, Wilson J G, Harris C R, Jeffrey D W (1980). Problems in the assessment of heavy-
620 metal levels in estuaries and the formation of a pollution
621 index. HelgoländerMeeresuntersuchungen, 33: 566-575.

622 Tong R, Cheng M, Yang X, Yang Y, Shi M (2019). Exposure levels and health damage assessment of
623 dust in a coal mine of Shanxi Province, China. Process Safety Environmental Protection, 128: 184–
624 192.

625 USEPA (1986). Guidelines for the health risk assessment of chemical mixtures. Federal
626 Register, 51(185): 34014-34025.

627 USEPA (1989). Risk assessment guidance for superfund. Volume I: human health evaluation manual
628 (part a) (p. 205). EPA/540/1-89/002.

629 USEPA (2014). EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. U.S.
630 Environment Protection Agency (Washington DC).

631 Wang L, Rinklebe J, Tack F M, Hou D (2021). A review of green remediation strategies for heavy
632 metal contaminated soil. Soil Use and Management, 37(4): 936-963.

633 Wang Y, Yang L, Kong L, Liu E, Wang L, Zhu J (2015). Spatial distribution, ecological risk
634 assessment and source identification for heavy metals in surface sediments from Dongping Lake,
635 Shandong, East China. Catena, 125: 200-205.

636 Wang Z, Xiao J, Wang L, Liang T, Guo Q, Guan Y, Rinklebe J (2020). Elucidating the differentiation
637 of soil heavy metals under different land uses with geographically weighted regression and self-
638 organizing map. Environmental Pollution, 260: 114065.

WHO (1996). World Health Organization Permissible Limits of Heavy Metals in Soil and Plants (Switzerland).

WHO (2022). Lead Poisoning. https://www.who.int/news-room/fact-sheets/detail/lead-poisoning-and-health (Accessed 28 September 2022).

Wu Y, Hoffman F O, Apostoaei A I, Kwon D, Thomas B A, Glass R, Zablotska L B (2019). Methods to account for uncertainties in exposure assessment in studies of environmental exposures. Environmental Health, 18: 1–15.

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 Total Environment, 713: 135292.

Yan G, Mao L, Liu S, Mao Y, Ye H, Huang T, Li F, Chen L (2018). Enrichment and sources of trace metals in roadside soils in Shanghai, China: A case study of two urban/rural roads. Science of Total Environment, 631: 942-950.

Zakir H M, Quadir Q F, Mollah M Z I (2021). Human health risk assessment of heavy metals through the consumption of common foodstuffs collected from two divisional cities of Bangladesh. Exposure and Health, 13(2): 253–268.

Zerizghi T, Guo Q, Tian L, Wei R, Zhao C (2022). An integrated approach to quantify ecological and human health risks of soil heavy metal contamination around coal mining area. Science of Total Environment, 814: 152653.

Zhang H B, Luo Y M, Wong M H, Zhao Q G, Zhang G L (2007). Defining the geochemical baseline: a case of Hong Kong soils. Environmental Geology, 52(5): 843-851.

Zhang H, Jiang Y, Ding M, Xie Z (2017). Level, source identification, and risk analysis of heavy metal in surface sediments from river-lake ecosystems in the Poyang Lake, China. Environmental Science and Pollution Research, 24(27): 21902-21916.

663 Zhao F J, Wang P (2020). Arsenic and cadmium accumulation in rice and mitigation strategies. Plant
664 Soil, 446(1): 1–21.
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683

684 **Figure captions**

685 **Fig. 1** Violin plot comparing the HMs (Pb, Cu, Cd, Cr, and Ni) content in grain, shoot, and root of
686 paddy sampled from coal mine contaminated zones (zone 1 and zone 2) and uncontaminated zone
687 (zone 3).

688 **Fig. 2** Interaction between total HMs and DTPA extractable HMs (Pb, Cu, Cd, Cr, and Ni) with plant
689 (root, shoot, and grain) uptake.

690 **Fig. 3** Partial dependence plot from random forest algorithm representing uptake in two plant parts
691 (among root, shoot, and grain) most affected by available heavy metals (Pb, Cu, Cd, Cr, and Ni). All
692 the values are in mg/kg.

693 **Fig. 4** Risk thermometer diagram showing risk of HMs (Pb, Cu, Cd, Cr, Ni) through the consumption
694 of rice grown on coal mine contaminated and uncontaminated soil.

695 **Fig. 5** (a) sensitivity analysis of carcinogenic risk of different HMs for child and adult populations in
696 contaminated zone, (b) predicted probability density functions of carcinogenic risk child and adult in
697 contaminated zone.

698 **Fig. 6** (a) SOM component planes of concentration of heavy metals (HMs) in coal mine affected
699 agricultural soil (b) U-Matrix cluster representing sampling zones (c) Zone-wise distribution map of
700 HMs.

701 **Fig. 7** Source allocation of HMs in coal mine contaminated soils of the study location (a) the
702 contribution percentage of each factor by PMF and (b) PMF model factor profiles of HMs in coal
703 mine contaminated soils.

704

705

706

707

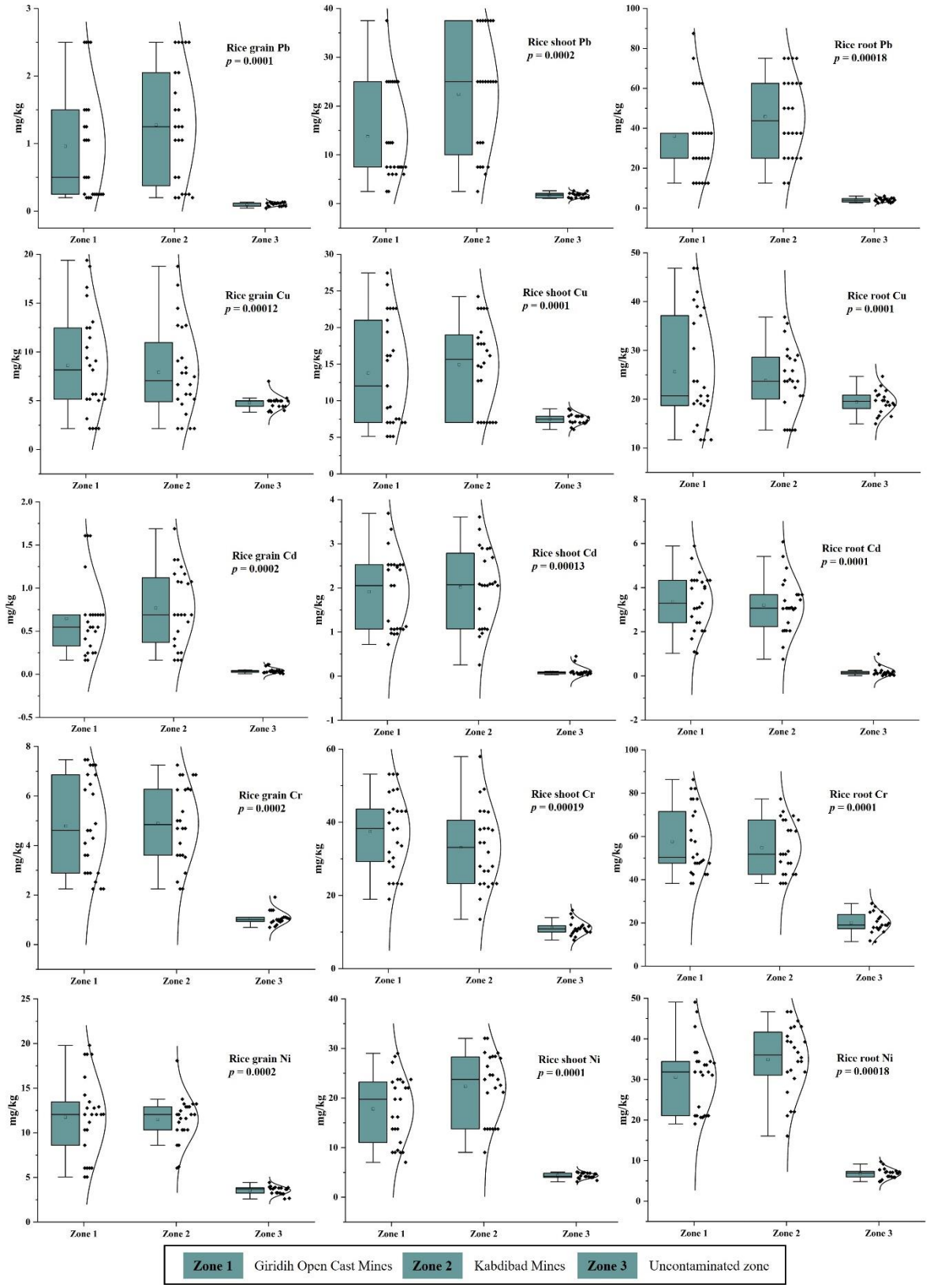
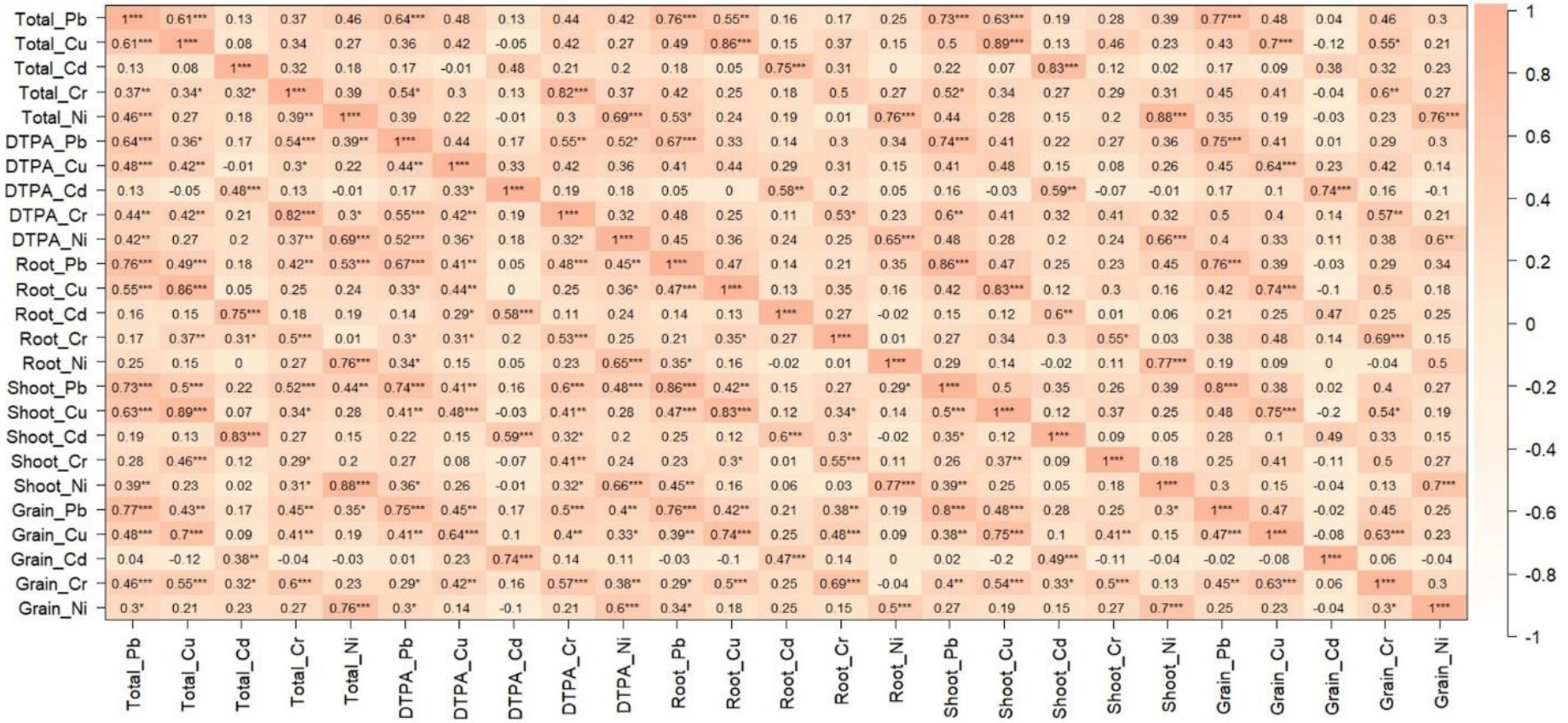


Fig. 1 Violin plot comparing the HMs (Pb, Cu, Cd, Cr, and Ni) content in grain, shoot, and root of paddy sampled from coal mine contaminated zones (zone 1 and zone 2) and uncontaminated zone (zone 3).



712

713 **Fig. 2** Interaction between total HMs and DTPA extractable HMs (Pb, Cu, Cd, Cr, and Ni) with plant (root, shoot, and grain) uptake.

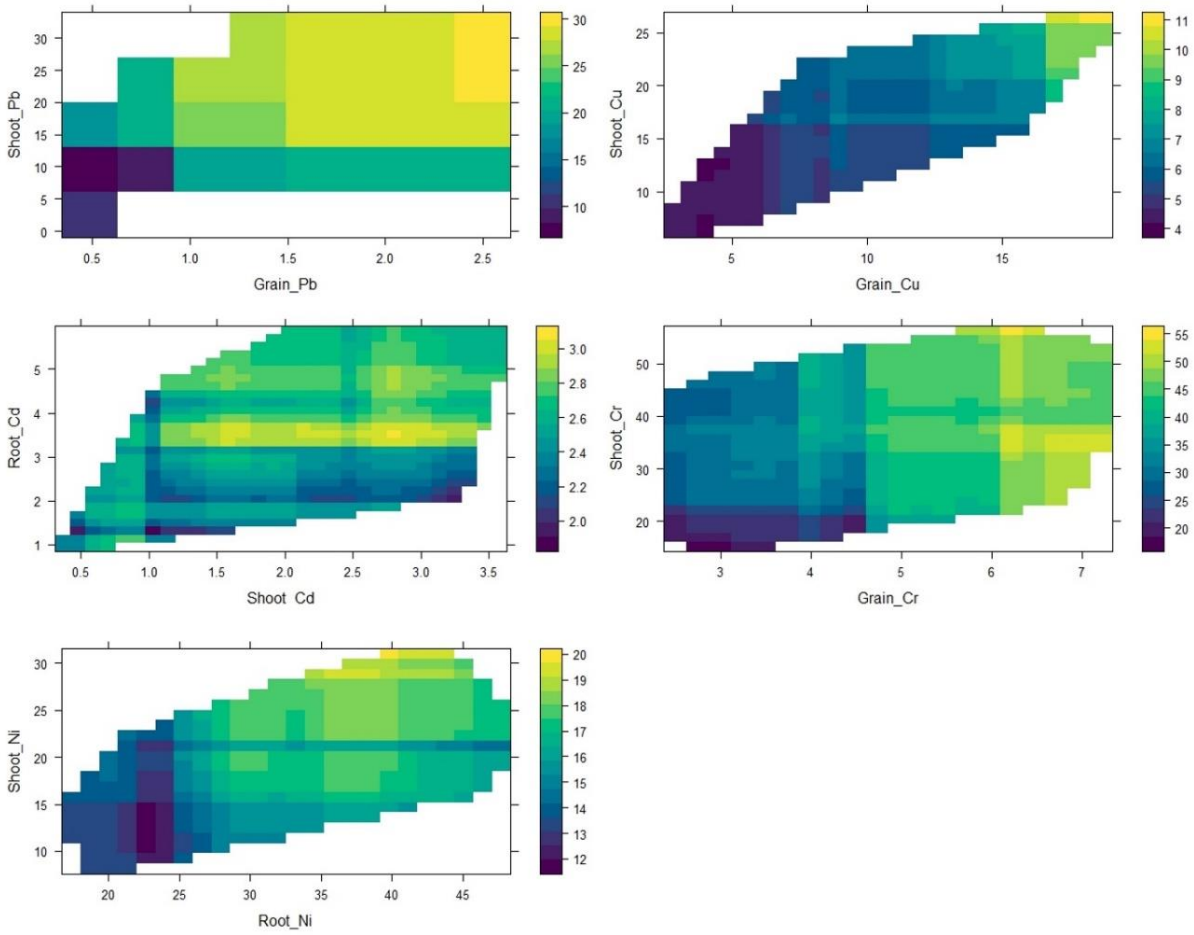


Fig. 3 Partial dependence plot from random forest algorithm representing uptake in two plant parts (among root, shoot, and grain) most affected by available heavy metals (Pb, Cu, Cd, Cr, and Ni). All the values are in mg/kg.

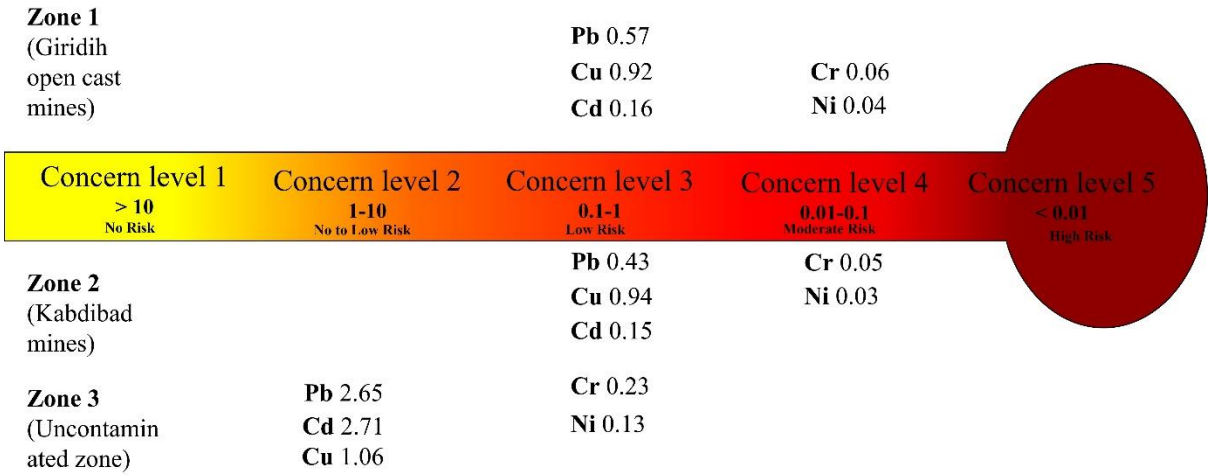


Fig. 4 Risk thermometer diagram showing risk of HMs (Pb, Cu, Cd, Cr, Ni) through the consumption of rice grown on coal mine contaminated and uncontaminated soil.

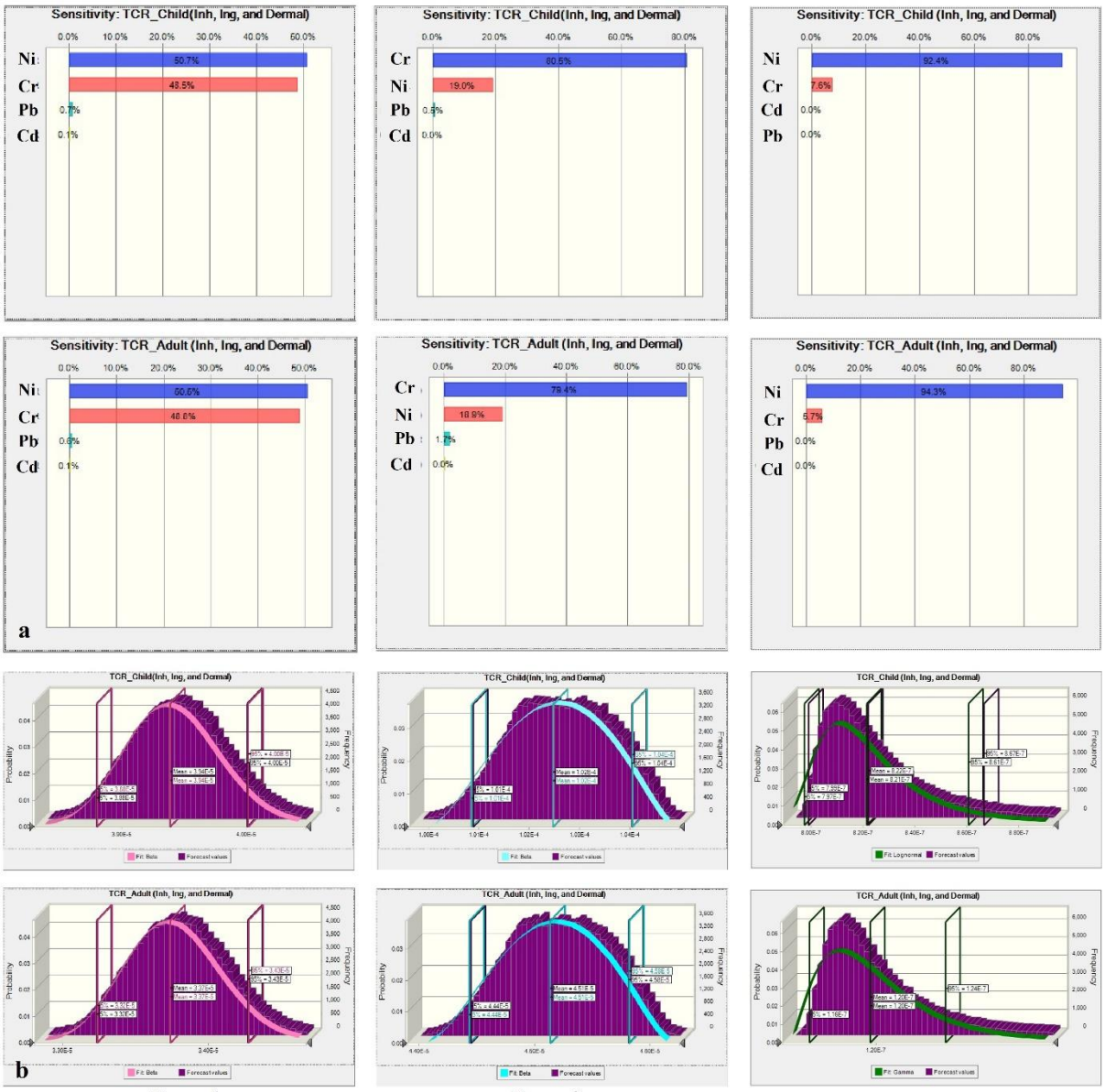
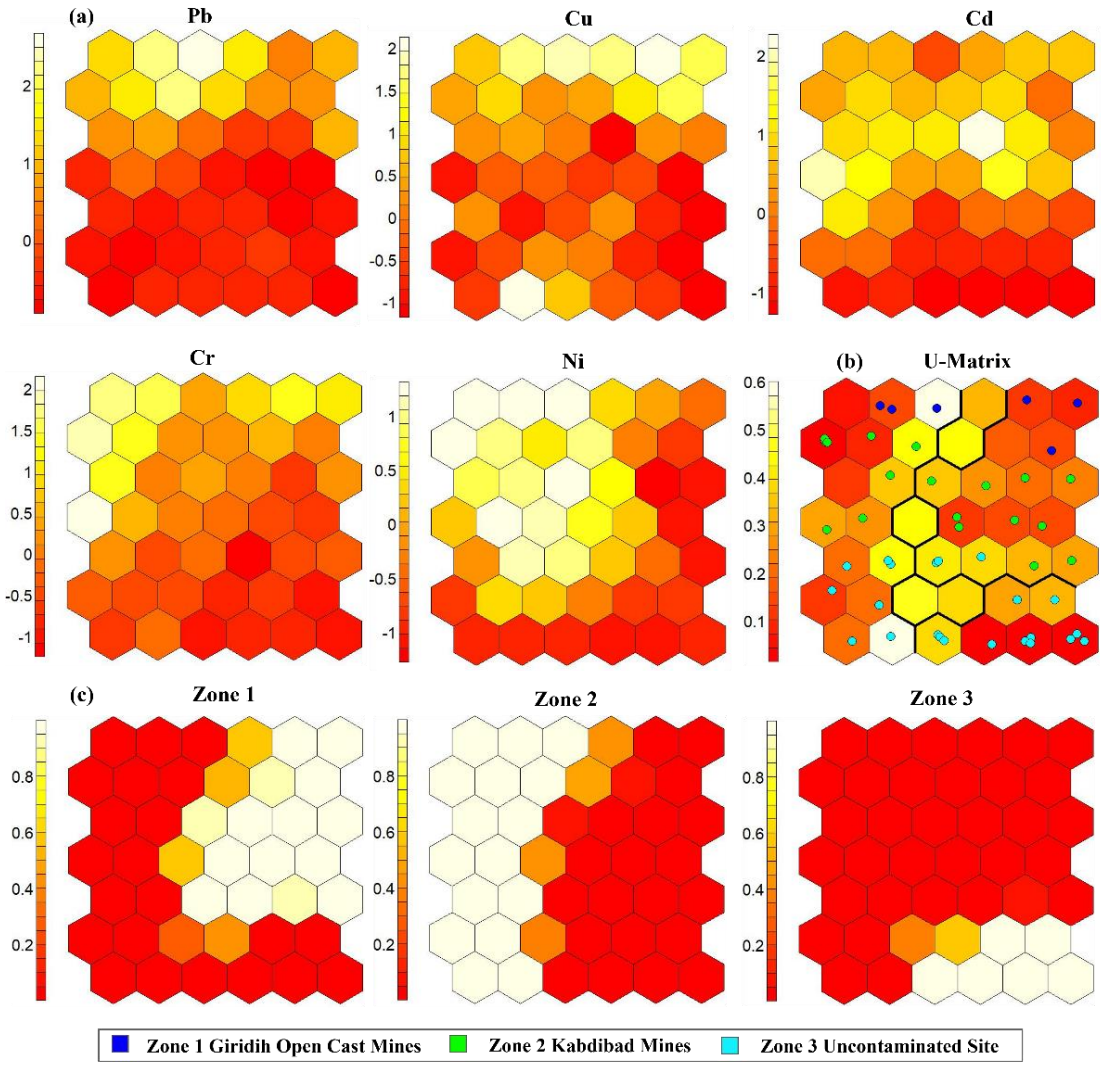


Fig. 5 (a) sensitivity analysis of carcinogenic risk of different HMs for child and adult populations in contaminated zone, (b) predicted probability density functions of carcinogenic risk child and adult in contaminated zone.



748

749 **Fig. 6** (a) SOM component planes of concentration of heavy metals (HMs) in coal mine affected
750 agricultural soil (b) U-Matrix cluster representing sampling zones (c) Zone-wise distribution map of
751 HMs.

752

753

754

755

756

757

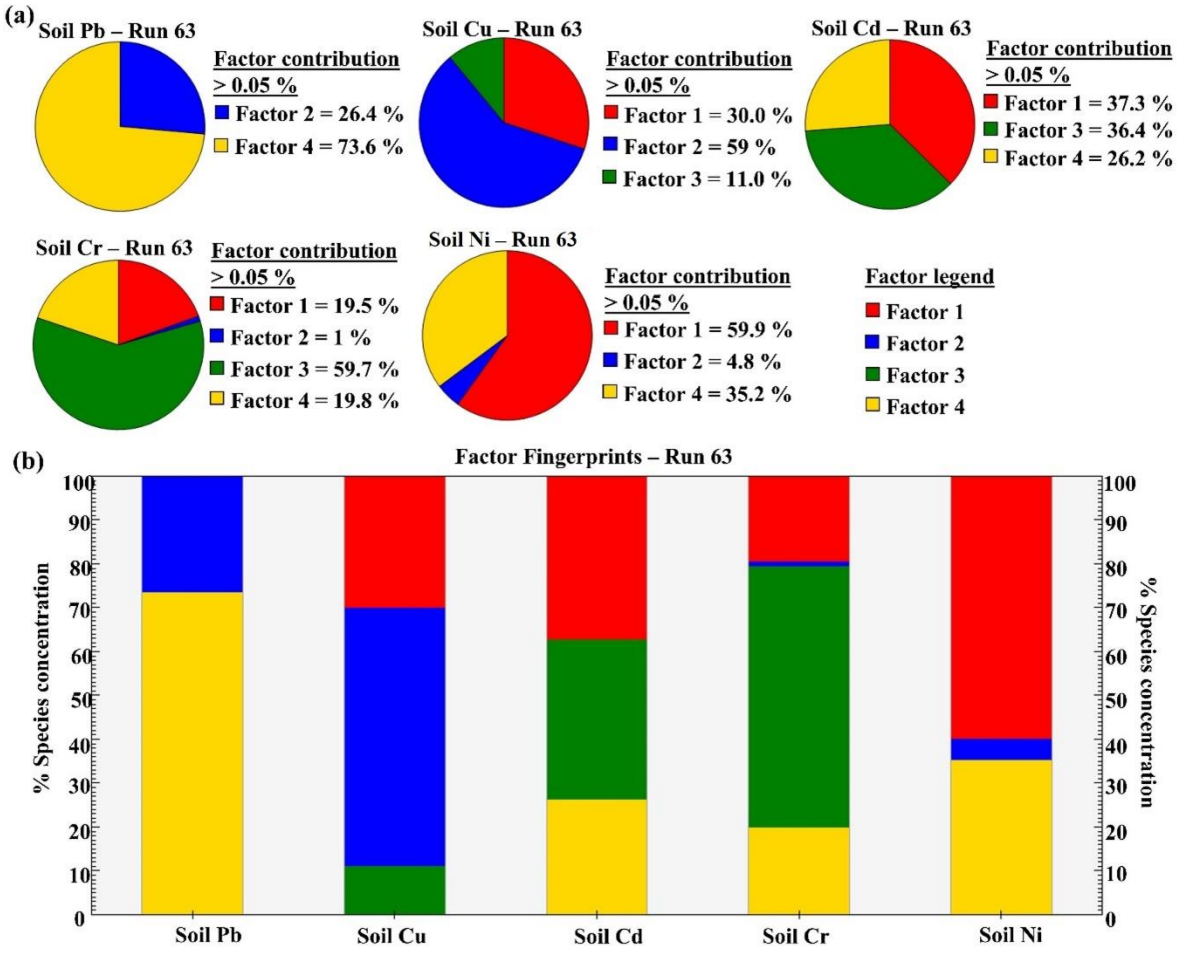


Fig. 7 Source allocation of HMs in coal mine contaminated soils of the study location (a) the contribution percentage of each factor by PMF and (b) PMF model factor profiles of HMs in coal mine contaminated soils.

Supplementary file caption:

Fig. S1 Variable importance plot from random forest algorithm representing effects of available heavy metals (Pb, Cu, Cd, Cr, and Ni) on uptake by root, shoot, and grain.

Fig. S2 Box-whisker plot comparing Geoaccumulation index (I_{geo}), contamination factor (C_f), contamination index (CI), pollution load index (PLI), and ecological risk index among the coal mine contaminated zones (zone 1 and 2) and uncontaminated zone (zone 3).

Fig. S3 Source apportionment and factor profiles of heavy metals (Pb, Cu, Cd, Cr, and Ni) from the PMF model.

Fig. S4 Spatial distribution HMs (Pb, Cu, Cd, Cr, and Ni) in coal mine contaminated zones (zone 1 and zone 2) and uncontaminated zone (zone 3).

Fig. S5 Spatial distribution of four factors derived from the PMF model for contaminated zones (zone 1 and 2) from coal mine affected soil.

Fig. S6 Comparison between observed and predicted HMs concentration resulted from FIAM analysis of rice grain grown in zone 1.

Fig. S7 Comparison between observed and predicted HMs concentration resulted from FIAM analysis of rice grain grown in zone 2.

Table S1 Coordinates of soil sample collected from contaminated zone (zone 1 and 2) and uncontaminated zone (zone 3)

Table S2 Physio-chemical, total and DTPA extractable HMs in soil collected from coal mine contaminated zones (zone 1 and zone 2) and uncontaminated zone (zone 3)

Table S3 FIAM parameters for predicting uptake of HMs by rice as a function of pH, Walkley Black organic-C and DTPA extractable PTEs, FIAM-HQ for intake of HMs through consumption of rice grains grown on coal mine contaminated soil

Table S4 Dietary risk (SAMOE and TCR) of HMs from rice grains grown on coal mine contaminated soil

Table S5 Fuzzy TOPSIS performance score values of HMs concentrations in rice plant grown in coal mine contaminated soil

Table S6 Average daily dose (ingestion, inhalation, and dermal) values on children and adults

Table S7 Non-carcinogenic risk in terms of HQ (ingestion, inhalation, and dermal) and HI on children and adults

Table S8 Carcinogenic risk (three exposure pathway) and total cancer risk values on adults and children

Table S9 Regression diagnostics test parameters recovered from the PMF (positive matrix factorization)

Table S10 Semi-variogram characteristics parameters used in geostatistical modelling of total PTEs content. RMSE – root mean square error; AIC – Akaike Information Criterion; MSE – mean square error; RMSSE – root mean square standard error; G – Goodness-of-prediction; ASE – average standard error

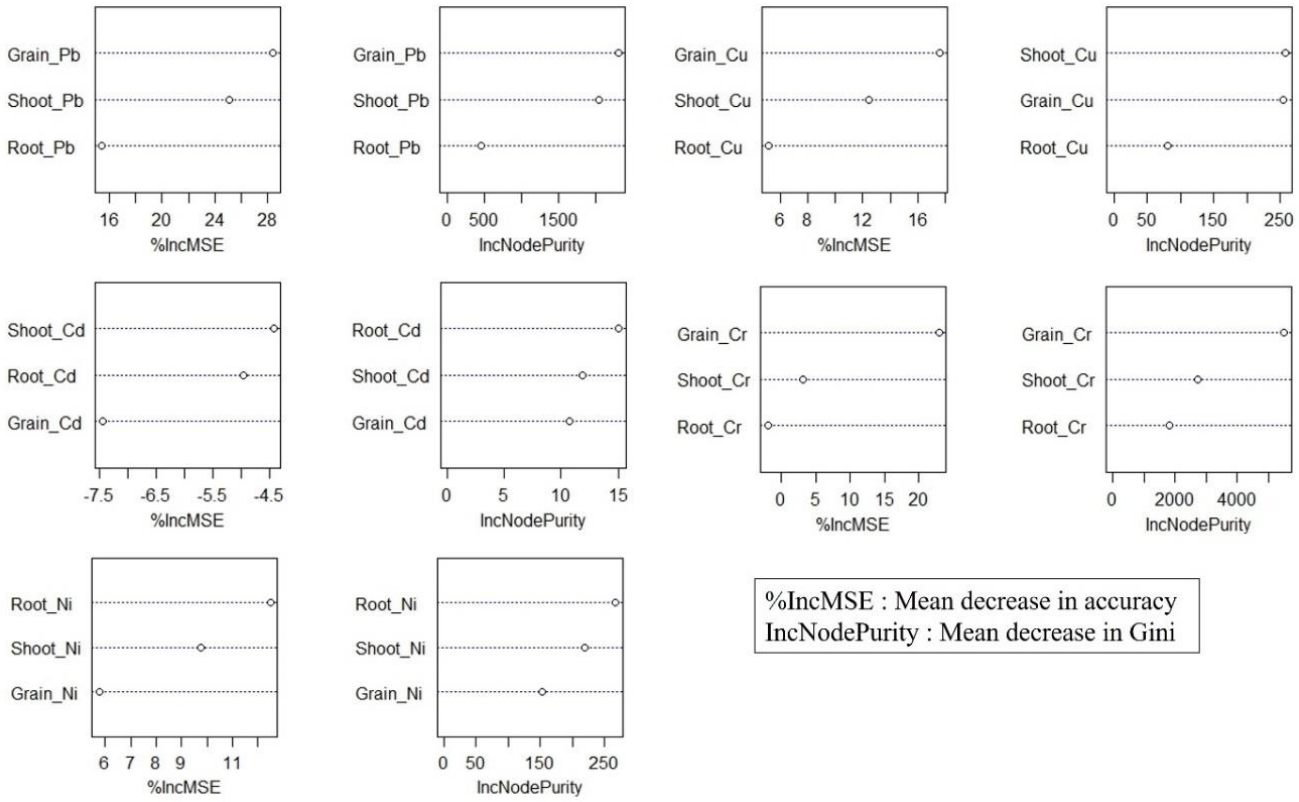
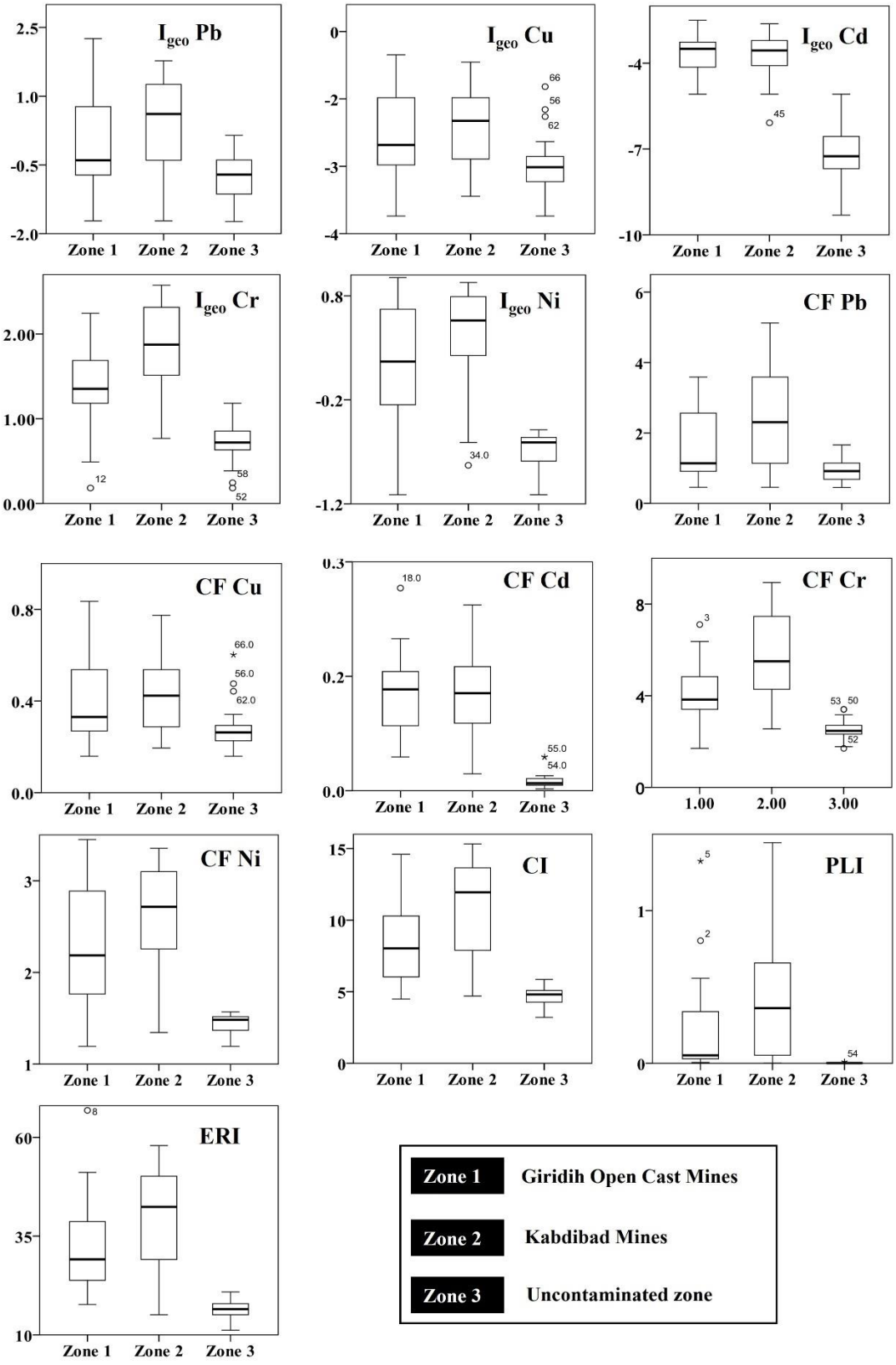


Fig. S1 Variable importance plot from random forest algorithm representing effects of available heavy metals (Pb, Cu, Cd, Cr, and Ni) on uptake by root, shoot, and grain.

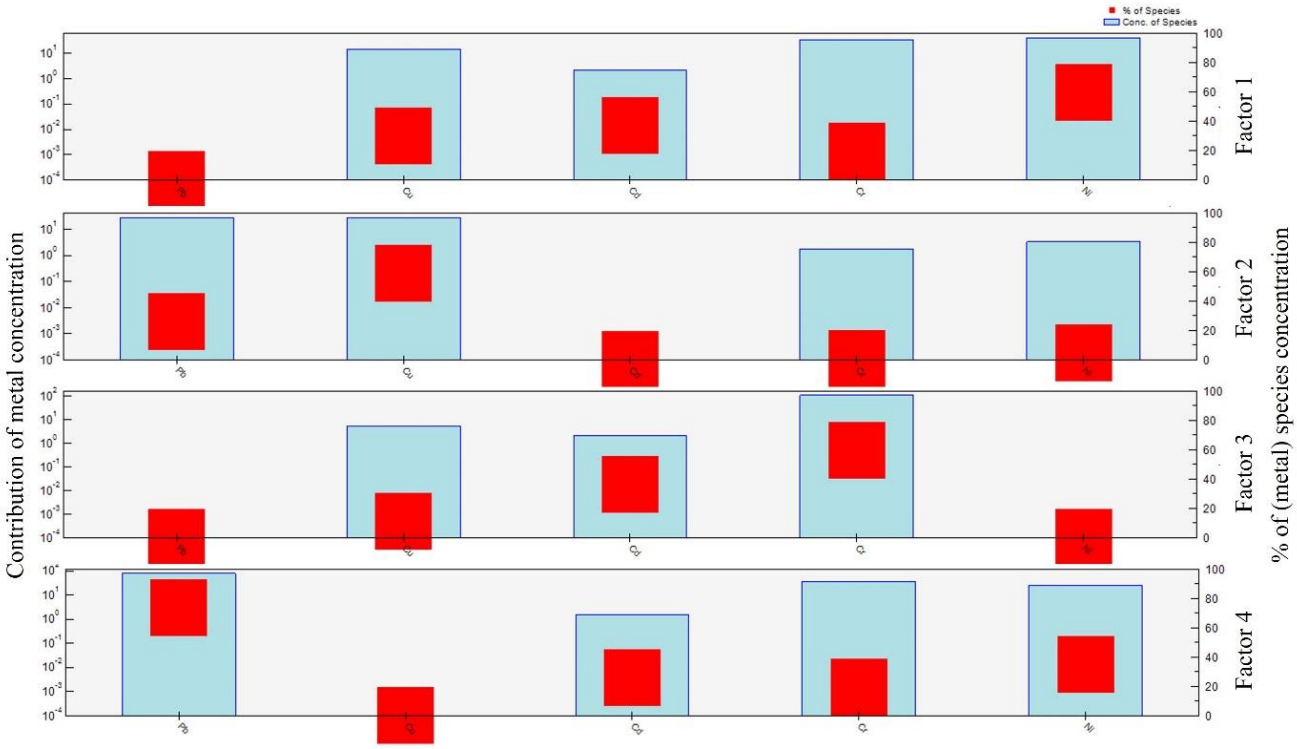


833

834 **Fig. S2** Box-whisker plot comparing Geoaccumulation index (I_{geo}), contamination factor (C_f),

835 contamination index (CI), pollution load index (PLI), and ecological risk index among the coal mine

836 contaminated zones and uncontaminated zone.

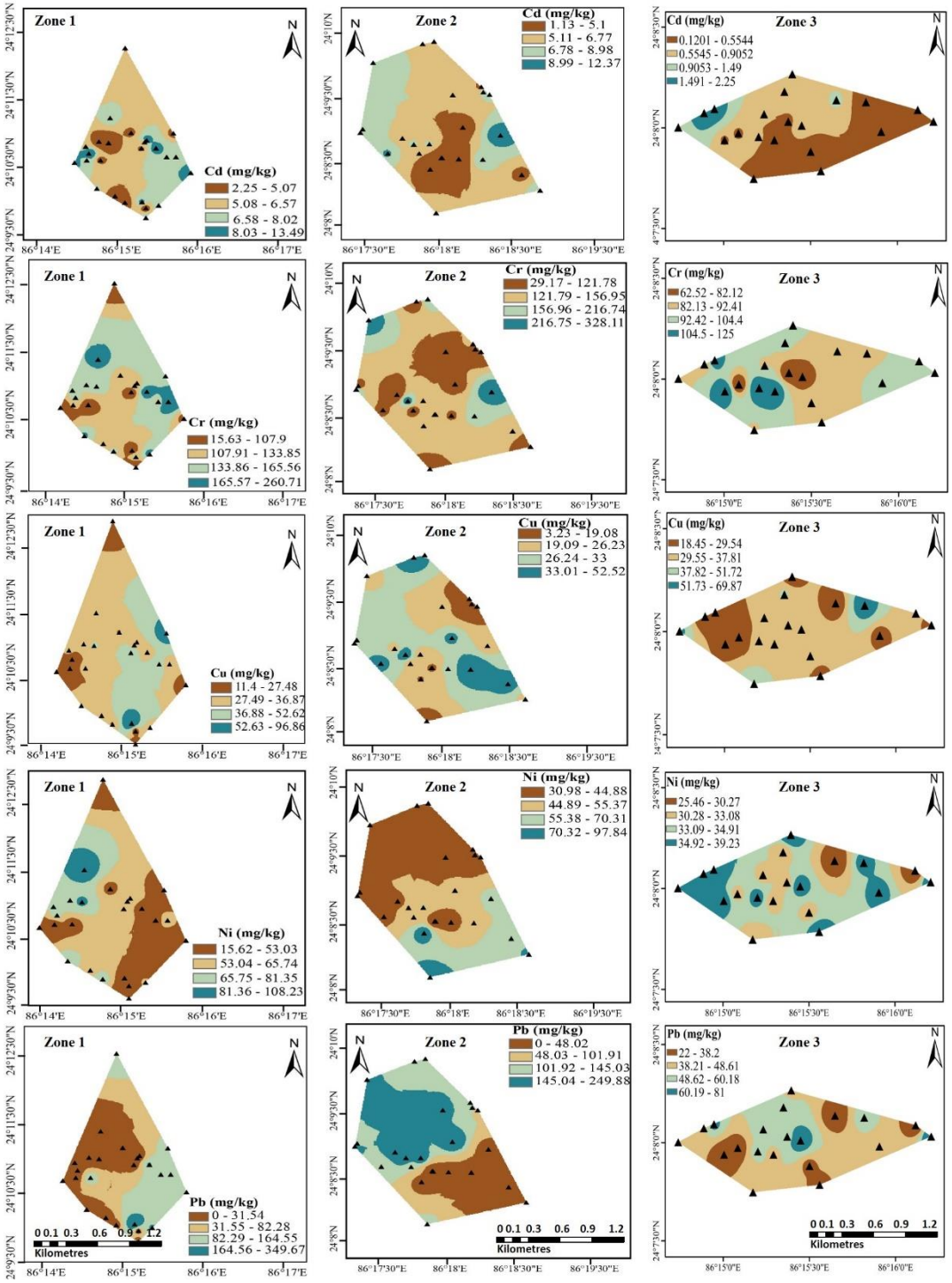


837

838 **Fig. S3** Source apportionment and factor profiles of heavy metals (Pb, Cu, Cd, Cr, and Ni) from the

839 PMF model.

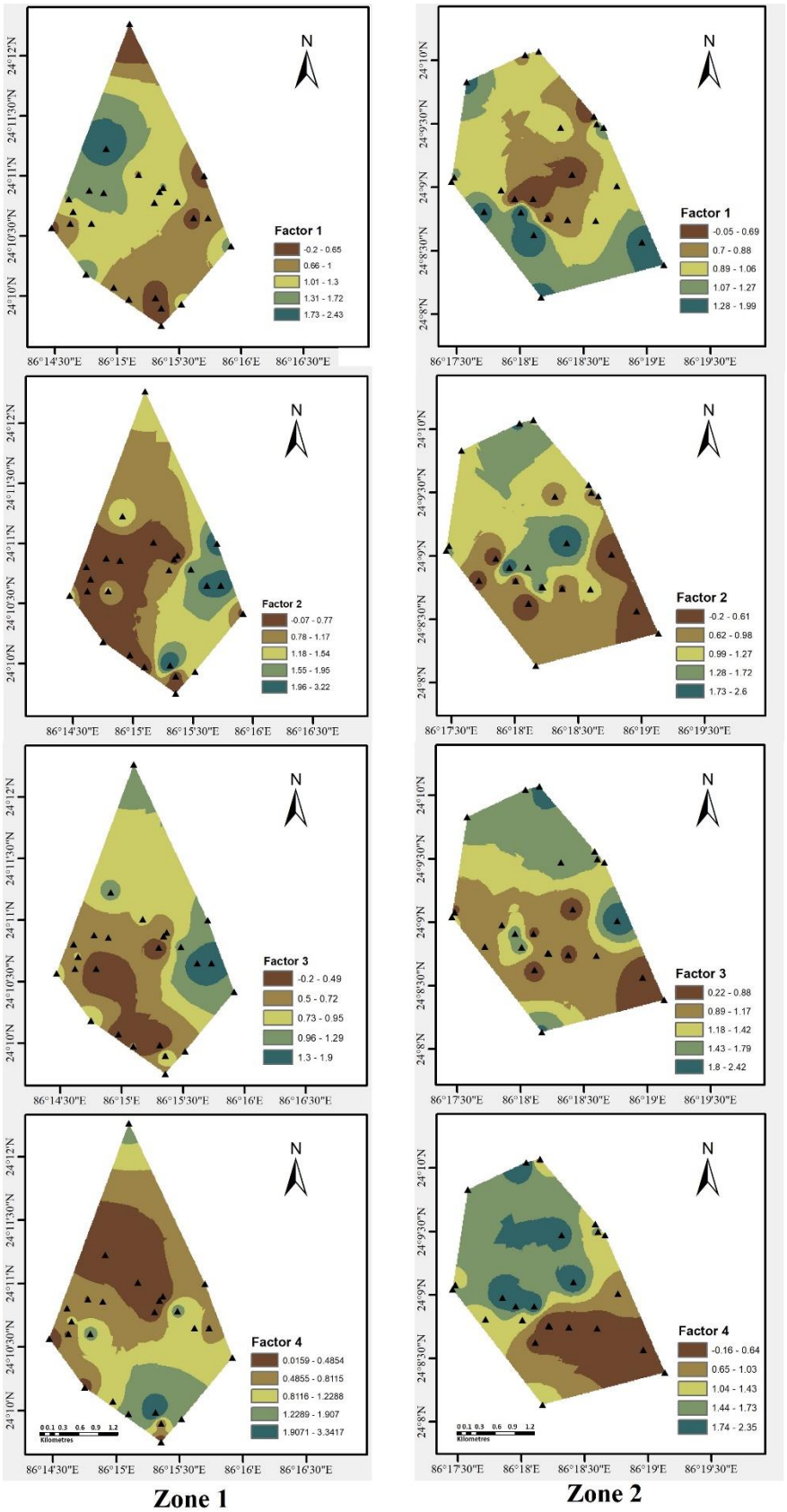
840



841

842 **Fig. S4** Spatial distribution HMs (Pb, Cu, Cd, Cr, and Ni) in coal mine contaminated zones (zone 1

843 and zone 2) and uncontaminated zone (zone 3).



844

845 **Fig. S5** Spatial distribution of four factors derived from the PMF model for contaminated zones (zone

846 1 and 2) from coal mine affected soil.

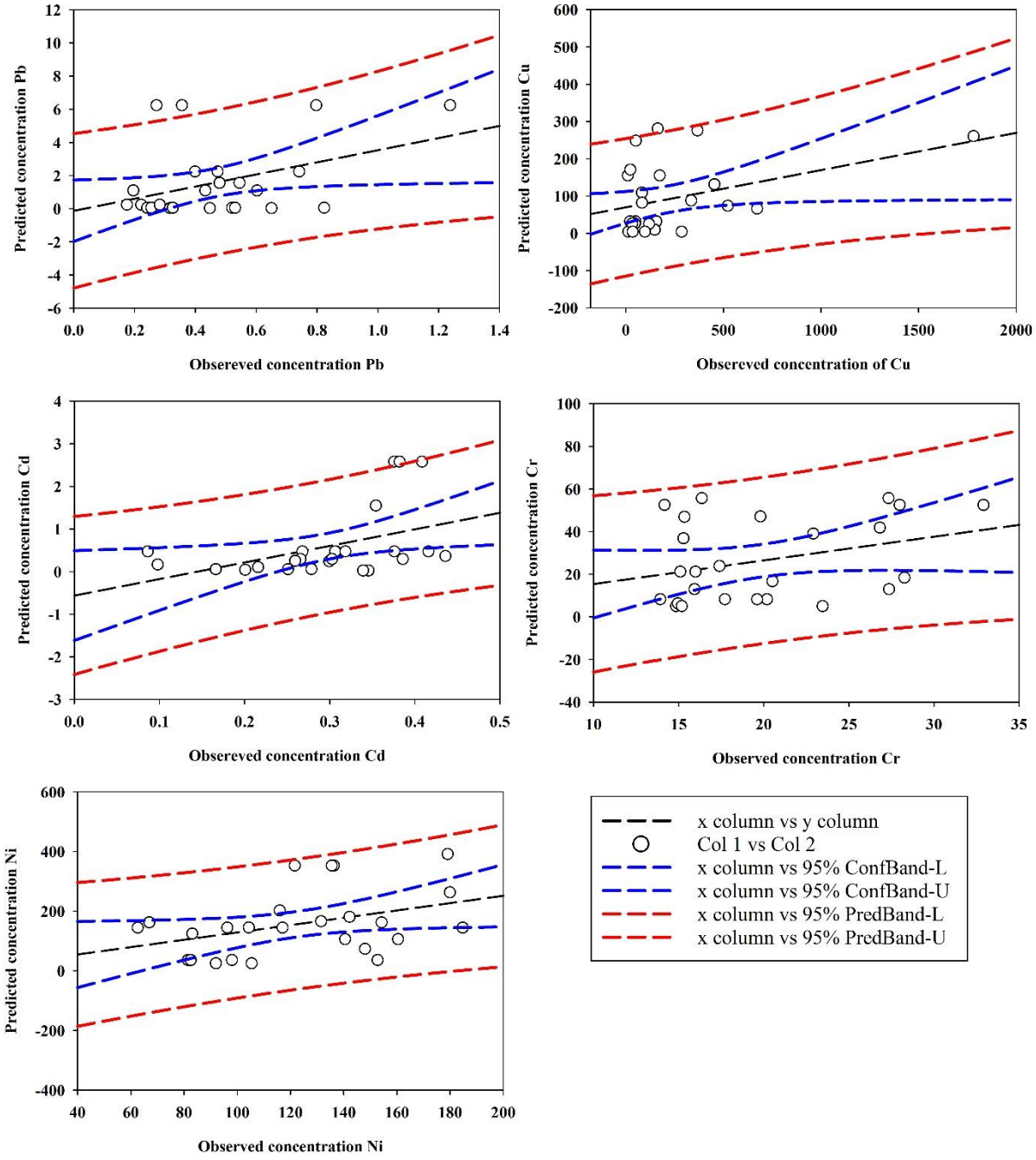
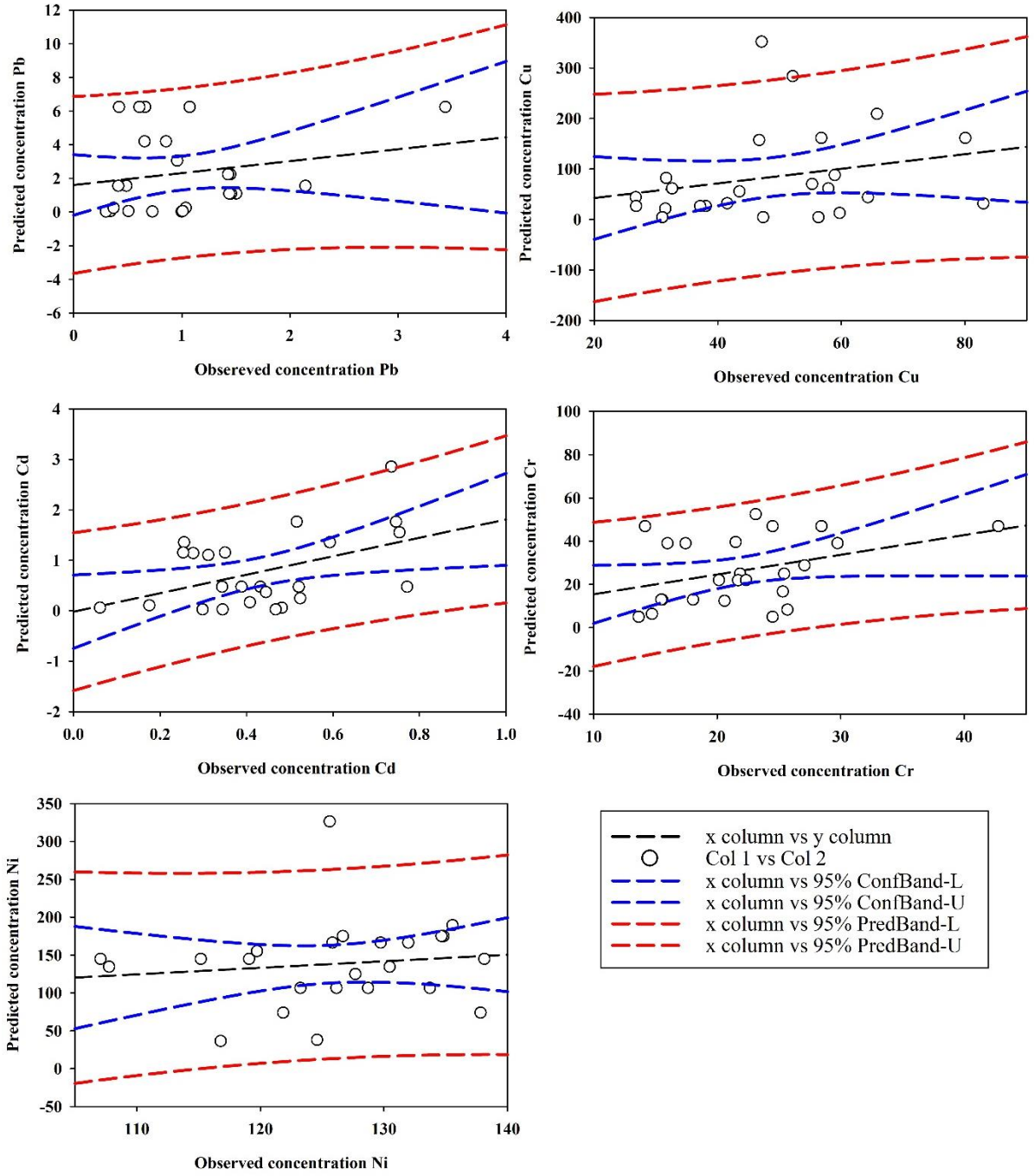


Fig. S6 Comparison between observed and predicted HMs concentration resulted from FIAM analysis of rice grain grown in zone 1.

855



856

857 **Fig. S7** Comparison between observed and predicted HMs concentration resulted from FIAM analysis

858 of rice grain grown in zone 2.

859

860

Table S1 Coordinates of soil sample collected from contaminated zone (zone 1 and 2) and uncontaminated zone (zone 3)

Zone 1 (Giridih Open Cast Mines)			**Zone 2 (Kabdibad Mines)**			**Zone 3 (Uncontaminated Zone)**		
Sample ID	Latitude	Longitude	Sample ID	Latitude	Longitude	Sample ID	Latitude	Longitude
G1	24.17361	86.265278	K1	24.16386	86.2929763	C1	24.13232	86.254853
G2	24.16556	86.258611	K2	24.14679	86.2952714	C2	24.13259	86.25334246
G3	24.1775	86.262222	K3	24.15135	86.2913227	C3	24.13289	86.25140054
G4	24.17675	86.2465793	K4	24.15834	86.3100862	C4	24.13232	86.25007553
G5	24.17972	86.258056	K5	24.15171	86.3068546	C5	24.1349	86.24910313
G6	24.20444	86.251667	K6	24.15788	86.3110131	C6	24.13456	86.24811422
G7	24.1775	86.260278	K7	24.14672	86.3001223	C7	24.13337	86.24563622
G8	24.16641	86.25517002	K8	24.13556	86.302778	C8	24.13446	86.25386367
G9	24.18333	86.261667	K9	24.15934	86.309729	C9	24.13385	86.25618069
G10	24.18168	86.25614601	K10	24.16745	86.3006423	C10	24.13352	86.25747432
G11	24.18115	86.25567968	K11	24.14851	86.3016964	C11	24.13136	86.25833673
G12	24.17963	86.25503061	K12	24.14965	86.2975128	C12	24.12976	86.25931414
G13	24.18349	86.25285998	K13	24.15786	86.3053027	C13	24.12912	86.25287476
G14	24.181	86.2482079	K14	24.16789	86.3024554	C14	24.13633	86.25581704
G15	24.1813	86.24626116	K15	24.15072	86.290963	C15	24.13778	86.25657022
G16	24.18711	86.24855584	K16	24.14849	86.2992838	C16	24.13563	86.26080468
G17	24.1801	86.24350344	K17	24.14593	86.3035563	C17	24.13546	86.26363341
G18	24.17839	86.24409366	K18	24.1459	86.3036914	C18	24.13301	86.26509951
G19	24.16969	86.24584505	K19	24.14376	86.3018148	C19	24.13386	86.27011303
G20	24.16787	86.24956686	K20	24.14573	86.3062473	C20	24.13482	86.26860667
G21	24.16622	86.25159452	K21	24.14562	86.3099559			
G22	24.16494	86.25591818	K22	24.15019	86.3127231			
G23	24.16252	86.2559185	K23	24.13981	86.318885			
G24	24.17672	86.24368916	K24	24.14276	86.3160225			
G25	24.1761	86.2411845						

Table S2 Physio-chemical, total and DTPA extractable HMs in soil collected from coal mine contaminated zones (zone 1 and zone 2) and uncontaminated zone (zone 3).

	Zone 1 (Giridih open cast mines)	**Zone 2 (Kabdibad mines)**	**Zone 3 (uncontaminated zone)**
Parameter	Mean	Mean	Mean
pH	6.01±0.5	6.21±0.51	6.93±0.3
EC (mS cm $^{-1}$)	0.03±0.02	0.04±0.02	0.05±0.02
OC (%)	1.33±0.96	2.16±1.23	0.94±0.14
Total HMs concentration (mg kg^{-1})			
Pb	108.24±72.97	139.56±69.46	46.33±17.02
Cu	57.26±23.91	69.89±19.86	33.03±12.41
Cd	8.44±2.76	8.95±2.57	0.70±0.57
Cr	180.05±46.90	245.46±70.66	93.83±17.70
Ni	70.79±25.06	95.46±22.89	34.18±3.89
DTPA-extractable HMs concentration (mg kg^{-1})			
Pb	16.46±10.60	22.56±9.74	1.49±0.34
Cu	5.67±4.23	6.5±4.48	2.92±0.39
Cd	2.43±0.91	2.67±1.06	0.08±0.05
Cr	30.37±10.50	44.35±13.56	9.38±1.78
Ni	15.67±3.64	16.72±3.73	3.44±0.36

Table S3 FIAM parameters for predicting uptake of HMs by rice as a function of pH, Walkley Black organic-C and DTPA extractable PTEs, FIAM-HQ for intake of HMs through consumption of rice grains grown on coal mine contaminated soil

	Zone 1				**Zone 2**				**Zone 3**			
	Constant	β1	β2	HQ	Constant	β1	β2	HQ	Constant	β1	β2	HQ
Pb	0.31	-0.12	0.26	1.61	1.25	-0.24	0.13	2.14	1.72	-0.11	-0.01	0.08
Cu	0.48	-0.41	0.69	0.10	-0.64	-0.12	0.24	0.09	-0.22	-0.05	-0.06	0.03
Cd	2.10	-0.21	-0.20	3.81	1.08	-0.21	0.14	4.53	0.91	-0.15	0.02	0.15
Cr	-0.16	-0.07	-0.04	9.39	-0.72	0.00	0.02	9.60	0.91	-0.15	0.02	0.15
Ni	1.66	-0.24	0.09	3.45	-1.55	0.13	0.41	3.38	-0.05	-0.09	0.04	0.51

881 **Table S4** Dietary risk (SAMOE and TCR) of HMs from rice grains grown on coal mine contaminated soil

Metal	Zone 1	Zone 2	Zone 3
SAMOE			
Pb	0.570	0.430	2.65
Cu	0.920	0.940	1.06
Cd	0.160	0.150	2.71
Cr	0.060	0.050	0.23
Ni	0.040	0.030	0.13
TCR			
Pb	2.37E-04	3.15E-04	5.13E-06
Cd	1.45E-03	1.72E-03	9.27E-05
Cr	1.41E-02	1.44E-02	6.29E-04
Ni	6.29E-02	6.16E-02	1.97E-04

882

883 **Table S5** Fuzzy TOPSIS performance score values of HMs concentrations in rice plant grown in coal mine contaminated soil

	Zone 1				Zone 2				Zone 3			
HMs	**Si+**	**Si-**	**Pi**	**Rank**	**Si+**	**Si-**	**Pi**	**Rank**	**Si+**	**Si-**	**Pi**	**Rank**
Pb	0.378637	0.100775	0.210205	5	0.360902	0.123079	0.254306	5	0.38127	0.164544	0.301465	4
Cu	0.198209	0.271195	0.577743	2	0.168593	0.260535	0.607126	2	0.207322	0.365437	0.63803	1
Cd	0.41816	0.147846	0.26121	4	0.395921	0.143283	0.265731	4	0.398914	0.200442	0.334429	5
Cr	0.259273	0.274694	0.51444	3	0.259653	0.226501	0.465903	3	0.330437	0.227994	0.408276	3
Ni	0.153611	0.364463	0.703497	1	0.108069	0.375445	0.776493	1	0.172404	0.29359	0.63003	2

884 **Table S6** Average daily dose (ingestion, inhalation, and dermal) values on children and adults

HMs	**Child**			**Adult**		
Zone 1	ADDing	ADDinh	ADDder	ADDing	ADDinh	ADDder
Pb	6.07E-04	1.69E-08	6.98E-07	6.50E-05	9.56E-09	9.76E-08
Cu	3.11E-04	8.68E-09	3.57E-07	3.33E-05	4.90E-09	5.00E-08
Cd	4.23E-05	1.18E-09	4.87E-08	4.54E-06	6.67E-10	6.81E-09
Cr	9.87E-04	2.76E-08	1.13E-06	1.06E-04	1.55E-08	1.59E-07
Ni	4.33E-04	1.21E-08	4.97E-07	4.63E-05	1.55E-08	6.96E-08
Zone 2						
Pb	7.86E-04	2.20E-08	9.04E-07	8.42E-05	1.24E-08	1.26E-07
Cu	3.28E-04	9.17E-09	3.77E-07	3.51E-05	5.17E-09	5.28E-08
Cd	4.18E-05	1.17E-09	4.80E-08	4.48E-06	6.58E-10	6.72E-09
Cr	1.42E-03	3.96E-08	1.63E-06	1.52E-04	2.23E-08	2.28E-07
Ni	4.96E-04	1.39E-08	5.71E-07	5.32E-05	2.23E-08	7.98E-08
Zone 3						
Pb	3.05E-04	8.51E-09	3.50E-07	3.26E-05	4.80E-09	4.90E-08
Cu	2.17E-04	6.07E-09	2.50E-07	2.33E-05	3.42E-09	3.49E-08
Cd	4.59E-06	1.28E-10	5.28E-09	4.92E-07	7.23E-11	7.38E-10
Cr	6.17E-04	1.72E-08	7.10E-07	6.61E-05	9.72E-09	9.93E-08
Ni	2.25E-04	6.28E-09	2.58E-07	2.41E-05	9.72E-09	3.62E-08

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

Table S7 Non-carcinogenic risk in terms of HQ (ingestion, inhalation, and dermal) and HI on children and adults

HMs	Child				Adult			
Zone 1	HQing	HQinh	HQder	HI	HQing	HQinh	HQder	HI
Pb	1.73E-01	4.84E-06	1.32E-03	1.75E-01	1.86E-02	2.73E-06	1.84E-04	1.88E-02
Cu	7.77E-03	1.93E-07	8.93E-06	7.78E-03	8.32E-04	1.09E-07	1.25E-06	8.34E-04
Cd	4.23E-02	2.08E-05	4.87E-05	4.24E-02	4.54E-03	1.17E-05	6.81E-06	4.55E-03
Cr	6.58E-04	9.19E-04	3.78E-04	1.95E-03	7.05E-05	5.18E-04	5.29E-05	6.42E-04
Ni	2.16E-02	4.83E-07	2.49E-05	2.17E-02	2.32E-03	6.22E-07	3.48E-06	2.32E-03
Zone 2								
Pb	2.25E-01	6.28E-06	1.71E-03	2.26E-01	2.41E-02	3.54E-06	2.39E-04	2.43E-02
Cu	8.20E-03	2.04E-07	9.43E-06	8.21E-03	8.79E-04	1.15E-07	1.32E-06	8.80E-04
Cd	4.18E-02	2.05E-05	4.80E-05	4.18E-02	4.48E-03	1.15E-05	6.72E-06	4.49E-03
Cr	9.44E-04	1.32E-03	5.43E-04	2.81E-03	1.01E-04	7.44E-04	7.60E-05	2.81E-03
Ni	2.48E-02	5.55E-07	2.85E-05	2.48E-02	2.66E-03	8.93E-07	3.99E-06	2.66E-03
Zone 3								
Pb	8.70E-02	2.43E-06	6.61E-04	8.77E-02	9.32E-03	1.37E-06	9.25E-05	9.42E-03
Cu	5.43E-03	1.35E-07	6.24E-06	5.44E-03	5.82E-04	7.60E-08	8.73E-07	5.83E-04
Cd	4.59E-03	2.25E-06	5.28E-06	4.60E-03	4.92E-04	1.27E-06	7.38E-07	4.94E-04
Cr	4.11E-04	5.75E-04	2.37E-04	1.22E-03	4.41E-05	3.24E-04	3.31E-05	4.01E-04
Ni	1.12E-02	2.51E-07	1.29E-05	1.12E-02	1.20E-03	3.89E-07	1.81E-06	1.21E-03

Table S8 Carcinogenic risk (three exposure pathway) and total cancer risk values on adults and children

HMs	CR		TCR	
Zone 1	Child	Adult	Child	Adult
Pb	1.09E-06	9.37E-07	3.99E-05	3.42E-05
Cd	6.91E-07	5.92E-07		
Cr	2.12E-05	1.82E-05		
Ni	1.69E-05	1.45E-05		
Zone 2				
Pb	1.42E-06	1.21E-06	5.19E-05	4.45E-05
Cd	6.82E-07	5.84E-07		
Cr	3.04E-05	2.61E-05		
Ni	1.94E-05	1.66E-05		
Zone 3				
Pb	5.5E-07	4.71E-07	2.27E-05	1.94E-05
Cd	7.49E-08	6.42E-08		
Cr	1.33E-05	1.14E-05		
Ni	8.79E-06	7.52E-06		

Table S9 Regression diagnostics test parameters recovered from the PMF (positive matrix factorization)

Species	Intercept	Slope	SE	r^2	KS Test Stat	KS Test P Value
Pb	-0.00629779	0.999997501	0.022177713	1	0.109253	0.602365
Cu	-0.009972451	1.000043539	0.043716712	0.999996	0.106952	0.629414
Cd	4.283191812	0.226657227	1.667066904	0.116359	0.117345	0.509725
Cr	0.689386979	0.996900879	0.637364726	0.999913	0.11115	0.580235
Ni	0.997373879	0.987705692	1.091000346	0.997975	0.114346	0.543454

925 **Table S10** Semi-variogram characteristics parameters used in geostatistical modelling of total PTEs
926 content. RMSE – root mean square error; AIC – Akaike Information Criterion; MSE – mean square
927 error; RMSSE – root mean square standard error; G – Goodness-of-prediction; ASE – average
928 standard error

Total HMs	Nugget	Partial sill	Spatial dependency	Model	AIC	RMSE	MSE	RMSSE	G
Zone 1									
Pb	4,339.70	1,315.52	Moderate	Spherical	198.5506	74.57028	-0.00937	1.009584	79.32624
Cu	436.2854	142.3696	Moderate	Gaussian	170.2623	23.37763	-0.01344	0.975032	98.53943
Cd	7.606961	0	Strong	Gaussian	62.53456	3.026474	-0.02697	1.05916	99.9832
Cr	191.3227	2,386.87	Moderate	Gaussian	219.2733	37.87941	0.023777	1.002513	85.32468
Ni	465.5605	136.4925	Moderate	Circular	184.4964	28.26059	0.044063	1.078711	97.41641
Zone 2									
Pb	0	6,965.03	Strong	Circular	211.5536	57.22033	0.004292	0.878583	82.82071
Cu	394.5141	0	Strong	Gaussian	172.6476	20.05908	0.035759	0.977233	98.89684
Cd	2.127732	7.44621	Strong	Gaussian	85.48904	2.563003	-0.06407	1.04124	99.98795
Cr	1491.329	3,619.90	Moderate	Spherical	232.8255	67.6139	-0.0374	1.065352	15.03446
Ni	597.982	0	Strong	Gaussian	189.2399	24.13211	-0.01125	0.928681	97.88454
Zone 3									
Pb	1231.53	4,312.53	Moderate	Gaussian	173.3306	68.50013	-0.00815	1.084135	88.76525
Cu	242.32	0	Strong	Gaussian	183.6772	27.0318	0.021652	0.988442	98.97146
Cd	2.48622	6.887	Strong	Gaussian	72.95604	2.37531	-0.01366	1.015145	99.97624
Cr	163.1984	2,532.06	Moderate	Spherical	197.3583	43.2148	-0.01875	0.976244	65.91862
Ni	687.5621	0	Strong	Gaussian	173.5423	21.98485	-0.08743	1.087153	97.98527

929
930