MethodsX 14 (2025) 103180

Contents lists available at ScienceDirect

MethodsX

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

# A novel method to determine background concentrations and spatial distributions of heavy metals in soil at large scale using machine learning coupled with remote sensing-terrain attributes ☆

Magboul M. Sulieman [a,b,c,*], Fuat Kaya [d], Abdullah S. Al-Farraj [a], Eric C. Brevik [e]

[a] Department of Soil Science, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, 11451, Riyadh, Saudi Arabia
[b] Department of Soil and Environment Sciences, Faculty of Agriculture, University of Khartoum, P.O. Box 32, 13314, Khartoum North, Shambat, Sudan
[c] Department of Physical Geography, University of Göttingen, 37077, Göttingen, Germany
[d] Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Isparta University of Applied Sciences, 32260, Isparta, Türkiye
[e] School of Agricultural Sciences and School of Earth Systems and Sustainability, Southern Illinois University, Carbondale, IL, USA

ARTICLE INFO

*Method name:*
Establish soil heavy metals background concentrations and spatial variability using machine learning algorithms coupled with remote sensing and digital elevation model derivatives

*Keywords:*
Environmental pollution-monitoring studies
Environmental variables
Landsat 8/9-OLI
Random forest
Stepwise multiple linear regression

ABSTRACT

Soil heavy metals are among the most hazardous materials in the environment. Their harmful effects can extend to surrounding systems (air, plants, water), and given the appropriate conditions may ultimately have negative effects on human health. Thus, preventing pollution and protecting pristine soils and preindustrial areas from human activities that lead to the concentration of heavy metals (HMs) is a priority. Here, a novel methodology was proposed to establish background concentrations of eight soil HMs, cobalt (Co), chromium (Cr), copper (Cu), iron (Fe), manganese (Mn), nickel (Ni), lead (Pb), and zinc (Zn), and digitally map their spatial distributions in an area (i.e., harrats region) that has not yet been impacted by industrial activity. The proposed methodology combined measurements of the target HMs and fifty-two environmental covariates (ECOVs) derived from 2017 to 2021 Landsat 8/9 OLI and Shuttle Radar Topography Mission (SRTM)-derived terrain attributes. Random forest and stepwise multiple linear regression models were further used to digitally map the studied HMs. The methodology is important for any future environmental pollution/monitoring studies in the area and can be applied in other similar environments. Machine learning algorithms show great ability to use available environmental variables and investigate the relationships between the factors influencing HMs accumulation under a given soil environment. The proposed methodology was effective for describing HMs spatial variability in the environments investigated.

- The proposed method is a novel way to predict soil HMs and their spatial distribution over large areas.
- Remote sensing/digital elevation models (DEMs)-derived ECOVs are useful for predicting and digitally mapping soil HMs, thus important for future environmental monitoring studies.

☆ **Related research article:** Sulieman, M. M., Kaya, F., Keshavarzi, A., Hussein, A, M., Al-Farraj, A. S., Brevik, E., C. (2024). Spatial variability of some heavy metals in arid harrats soils: Combining machine learning algorithms and synthetic indexes based-multitemporal Landsat 8/9 to establish background levels. Catena 234, 107,579. https://doi.org/10.1016/j.catena.2023.107579.

* Corresponding author.
*E-mail addresses:* 434108505@student.ksu.edu.sa, magboul@uofk.edu, magboul.siddig@uni-goettingen.de (M.M. Sulieman), fuatkaya@isparta.edu.tr (F. Kaya), sfarraj@ksu.edu.sa (A.S. Al-Farraj), eric.brevik@siu.edu (E.C. Brevik).

https://doi.org/10.1016/j.mex.2025.103180
Received 5 December 2023; Accepted 20 January 2025
Available online 21 January 2025

- Explainable algorithms (i.e., RF and SMLR) are able to utilize ECOVs for HMs prediction and to establish background concentrations over large areas.

Therefore, the combination of machine learning and RS/DEMs-based ECOVs is crucial to overcome the disadvantages of HMs determination via conventional methods.

Specifications table

| | |
|---|---|
| Subject area: | Environmental Science |
| More specific subject area: | Heavy metals in soils |
| Name of your method: | Establish soil heavy metals background concentrations and spatial variability using machine learning algorithms coupled with remote sensing and digital elevation model derivatives. |
| Name and reference of original method: | Spatial variability of some heavy metals in arid harrats soils: Combining machine learning algorithms and synthetic indexes based-multitemporal Landsat 8/9 to establish background levels. Catena 234(2024) 107,579. https://doi.org/10.1016/j.catena.2023.107579. |
| Resource availability: | The data is available with this article, the published related article, and their supplementary materials. |

## Method details

### *Study area*

The proposed method was applied in three harrats regions (i.e., Harrat Khaybar, Harrat Ithnayn, Harrat Kurma) located in the west central portion of Saudi Arabia (Fig. 1). The harrats region has not been affected by industrial activities. The soils are typically clean with little to no anthropogenic activities (i.e., agriculture, industry, extraction) that may cause heavy metals (HMs) accumulation in soils.

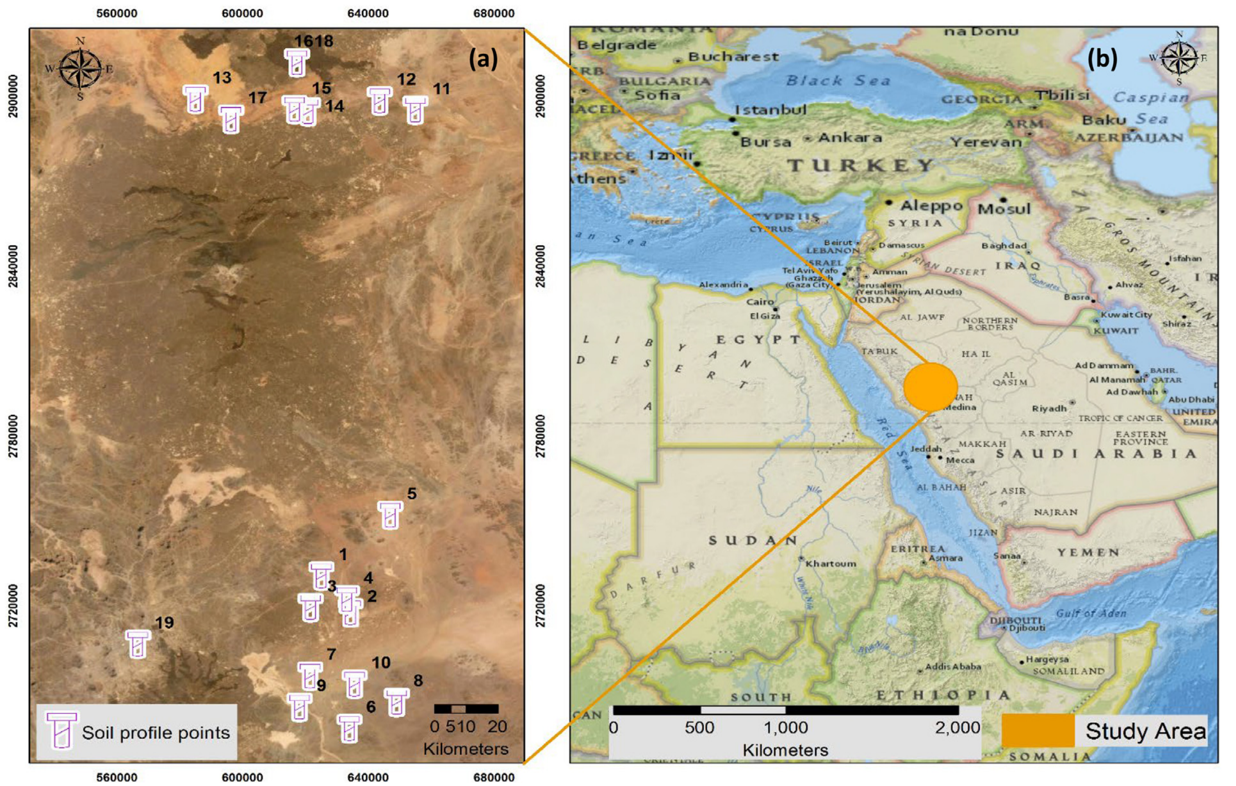


**Fig. 1.** Maps showing sampling locations at three studied harrats areas (a) and the geographical location of the harrats region (orange circle) inside the Arabian Shield, Saudi Arabia (b).

*Soil sampling, preparation and heavy metals analysis*

A total of 19 surface soil samples were collected from 19 locations at three harrats (their sites are shown in Fig. 1). A sampling strategy based on landform types was followed to select the studied profiles, and soil samples were taken from the top 0–30 cm for each genetic horizon of all profiles during the summer season. Soil samples were air dried at ambient laboratory temperature (≈25 °C), crushed, and processed via a 2 mm sieve (Haver & Boecker Germany, Inc.) to separate soil from non-soil materials (e.g., gravel, rock fragments). The fine soil particles that passed through the 0.25 mm sieve from each subsample were used because they would be easy to fully digest for HMs analysis. Afterward, the chosen subsamples were digested in a microwave (MARS, CEM Corporation, USA) according to the USEPA 3051 method [1]. All samples were prepared and analyzed at the Soil Sciences Department laboratories, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia. Cobalt (Co), chromium (Cr), copper (Cu), lead (Pb), nickel (Ni), zinc (Zn), manganese (Mn), and iron (Fe) were determined using the USEPA 3051 method [1] analyzed via inductively coupled plasma-optical emission spectrometry (ICP-OES; PerkinElmer Inc. Optima 4300 DV, USA). Quality assurance was carefully followed during sample preparation and analysis. At each specific process (i.e., digestion, analyses) three replicates were utilized for each sample. For ICP-OES anomalous HMs readings, identification, and elimination, a $Q_{test}$ with 95 % confidence level was used [2,3]. The precision of the procedure was calculated as relative standard deviation at ≤ 5 %. The ICP-OES detection limits were < 0.1 μg $L^{-1}$ for Fe and 1 μg $L^{-1}$ for Co, Cu, Cr, Pb, Mn, Ni, and Zn. Any HM values that were below the detection limits were excluded from further modeling. Three different soil standard references (i.e., Till-1, Till-2, and Till-4) were used to perform quality control for the HMs analyzed. The recovery percentage of the HMs was calculated using Eq. (1) [4–6] and shown in Fig. 2:

$$Recovery\% = \frac{Rsample}{Rstandard} * 100 \tag{1}$$

where $R_{sample}$ is the HMs concentration in a specified sample extracted by solution (mg $kg^{-1}$) and $R_{standard}$ is the concentration of HMs in the standard reference soil "Till" (mg $kg^{-1}$).

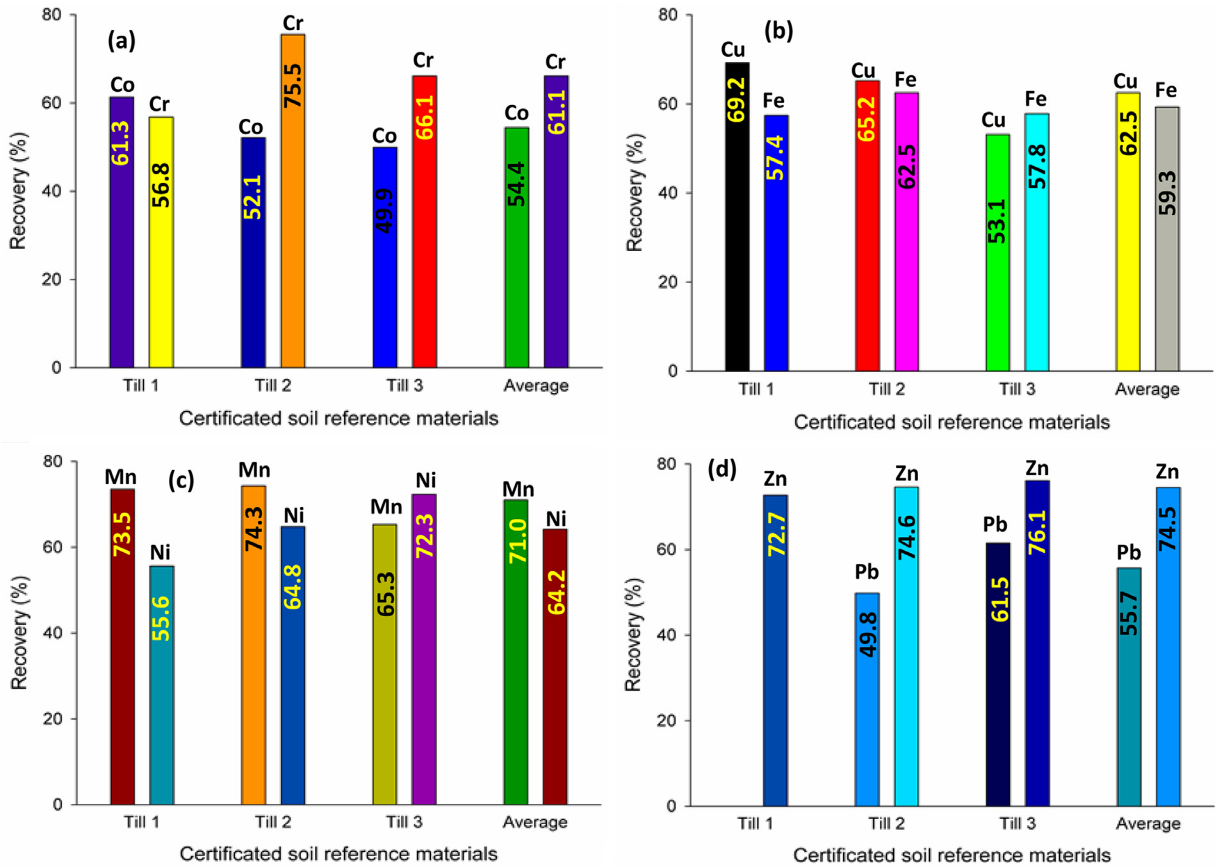


**Fig. 2.** Recovery of the eight heavy metal contents and their averages in the three certificated reference materials (Till 1, Till 2 and Till 4) digested by the EPA 3051 method for: (a) cobalt (Co) and chromium (Cr); (b) copper (Cu) and iron (Fe); (c) manganese (Mn) and nickel (Ni); and (d) lead (Pb) and zinc (Zn). Till 1 does not contain Pb and therefore is not shown in this figure.

Basic descriptive statistics (e.g., mean, standard deviation, standard error, coefficient of variation (CV), skewness, kurtosis, etc.)) were performed for the studied HMs. Additionally, Spearman correlation tests were run for HMs and different ECOVs. All statistical analysis was performed utilizing R program.

*Remote sensing and digital elevation model (DEM) variables*

This study utilized fifty-two environmental variables that were extracted from remote sensing (RS) datasets and a digital elevation model (DEM). All variables from the raster were calculated for the four time steps from 2017 to 2021 to a 30 m spatial resolution. Landsat 8 and 9 time-series data was used to extract RS-based indices using time-series to reduce mapping uncertainty, taking into account the soil sampling from November 2018. Landsat 8 surface reflectance (SR) images from November 2017, 2018, 2020, and 2021 were provided by Earth Explorer (EE-https://earthexplorer.usgs.gov/). To gauge the degree of uncertainty inherent in the machine learning models (here RF and SMLR) across these multiple runs, the standard deviation (SD) values were computed and considered as a representative uncertainty measure [7,8]. The scaling process for SR was conducted from collection 2 Landsat Level-2 prior use by applying a 0.0000275 scale factor with an extra −0.2 offset per pixel (i.e. bands 1 to 7) to RS outputs. This is a data scripts technique that can be done using manual calculations that is found in many GIS programs and select other software programs. The bands scale processing was achieved using Raster Calculator from Map Algebra, which can be found in Spatial Analyst Tools as an option for Arctoolbox available in ArcMap-ArcGIS 10.8 [9]. Seven indices were calculated that related to soil moisture (i.e., Normalized Difference Moisture Index; NDMI), vegetation (i.e., Normalized Difference Vegetation Index; NDVI), and bare soil (i.e., Clay Normalized Ratio; CLNR). The 52 variables included 35 RS data and seventeen DEM-based topography attributes (including the 1st and 2nd DEM derivatives with a resolution of 30 m x 30 m). The DEM was obtained from Shuttle Radar Topography Mission (SRTM) 1 arc-second DEM (http://earthexplorer.usgs.gov), accessed on February 15, 2022 [10]. System for Automated Geoscientific Analysis (SAGA GIS software) version 8.1.3 was used to obtain topography attributes [11]. These attributes (e.g., elevation, slope and terrain ruggedness index) are commonly used as terrain representations in digital soil mapping studies. The Landsat8/9 satellite bands information properties (i.e., pixel sizes and wavelengths) [12] are presented in Fig. 3. The complete list of 52 environmental variables (17 DEM and 35 RS) used in this study are shown in Fig. 4.

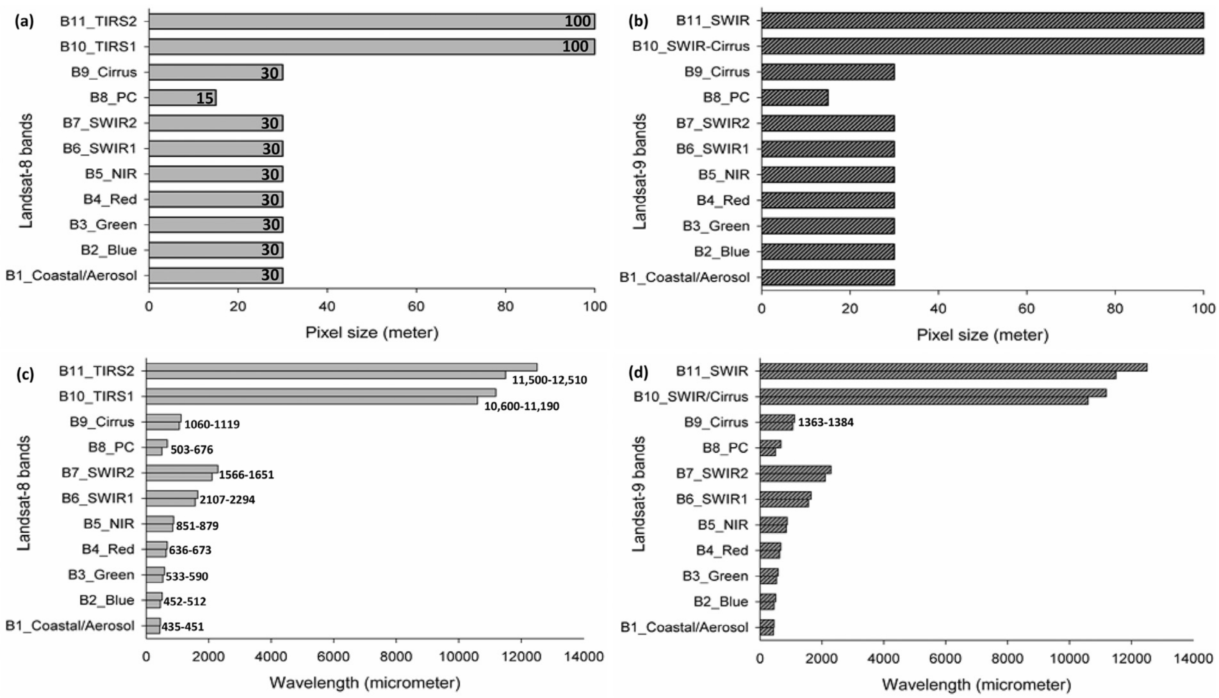


**Fig. 3.** Landsat 8/9 bands pixel size (a, b) and bands wavelength range (c, d). Abbreviations: NIR = near infrared; SWIR = short wavelength infrared; PC = panchromatic; TIRS = thermal infrared sensor. Band pixel sizes and wavelengths are the same for both Landsat 8 and 9 and if different are given.

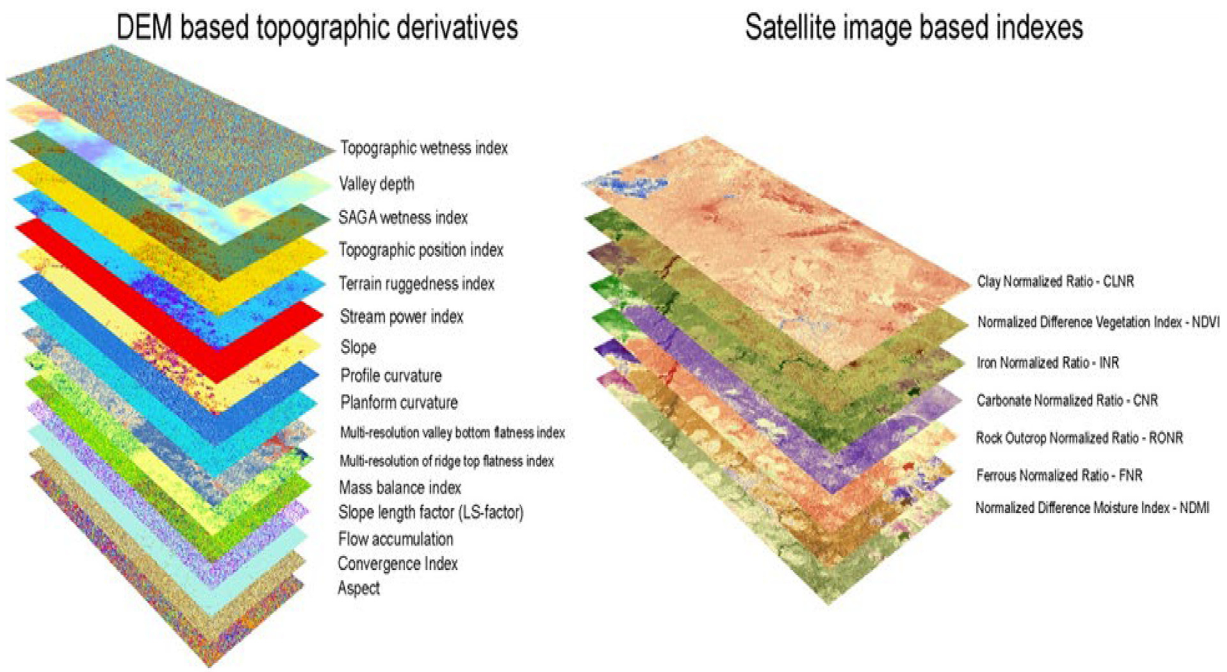


**Fig. 4.** The fifty-two environmental covariates (17 DEM-based topographic derivatives including elevation, which is the same as the DEM, and 35 RS-based satellite image indices (seven variables each one with five index statistics including minimum, maximum, median, mean, and standard deviation)) used to predict soil heavy metal contents in the study area.

*Data pre-processing and variable selection*

First, the 52 RS-DEM attributes were stacked in one layer and then the values of the 19 studied soil sample locations were extracted. These processes were performed via the R "raster" package [13]. Second, a multicollinearity test was generated using Spearman correlation, confirmed when the correlation was >0.7, and avoided use of the attribute if multicollinearity was indicated. Third, the Spearman correlation between index statistics (minimum, maximum, median, mean, and standard deviation) and the eight studied soil HMs was conducted prior to establishing spatial predictive models of soil HMs using the environmental attributes [14]. The average values over the four-year time-series had stronger relationships to HMs than the other statistic values. Fourth, the modeling process was continued using the average index values. Fifth, principal component analysis (PCA) was applied to select ECOVs that should be used in the final modelling. PCA was used because it linearly transformed the primary environmental variables dataset into a new set of significantly smaller, unrelated environmental variables with the ability to provide the greatest possible amount of information in the primary dataset. A small set of uncorrelated variables is much easier to understand and use in further analysis than a larger set of correlated variables. In this selection, one of the ECOVs with a Spearman correlation coefficient >0.7 was chosen, the one with the highest correlation with HMs [15]. These variables were different for each soil heavy metal. Fig. 5 illustrates a detailed flowchart that depicts the work process followed in this study.

*Machine learning-based stepwise multiple linear regression (SMLR) and random forest (RF) model building*

The SMLR and RF models were applied to construct estimation models that combined soil HM content and environmental variables [16]. These two models were used due to their simplicity and explainability. The SMLR model uses multiple explanatory environmental variables to predict the output of the soil HMs that were targeted in this study. The model tries to illustrate the spatial distribution of a dependent variable through a linear correlation between the soil and ECOVs that were used as explanatory variables and HMs as the target variables as given in Eq. (2):

$$y = \alpha + \sum_{i=1}^{n} b_i * x_i \pm \varepsilon \tag{2}$$

where y is the dependent variable (soil HMs), $x_i$ are independent variables (spectral and terrain variable values), n is the number of variables, $b_i$ are the partial regression coefficients, $\alpha$ is the intercept, and $\varepsilon$ is the standard error of the estimate. A function named stepwise linear selection that included selections of forward and backward was applied [17]. The model type is "lmStepAIC", available in the caret package in the R software package [18]. Residuals of SMLR were modelled using the "hist", "qqnorm", and "qqline" functions in the R Core Environment [18]. The RF algorithm utilizes ensemble learning by combining multiple decision or

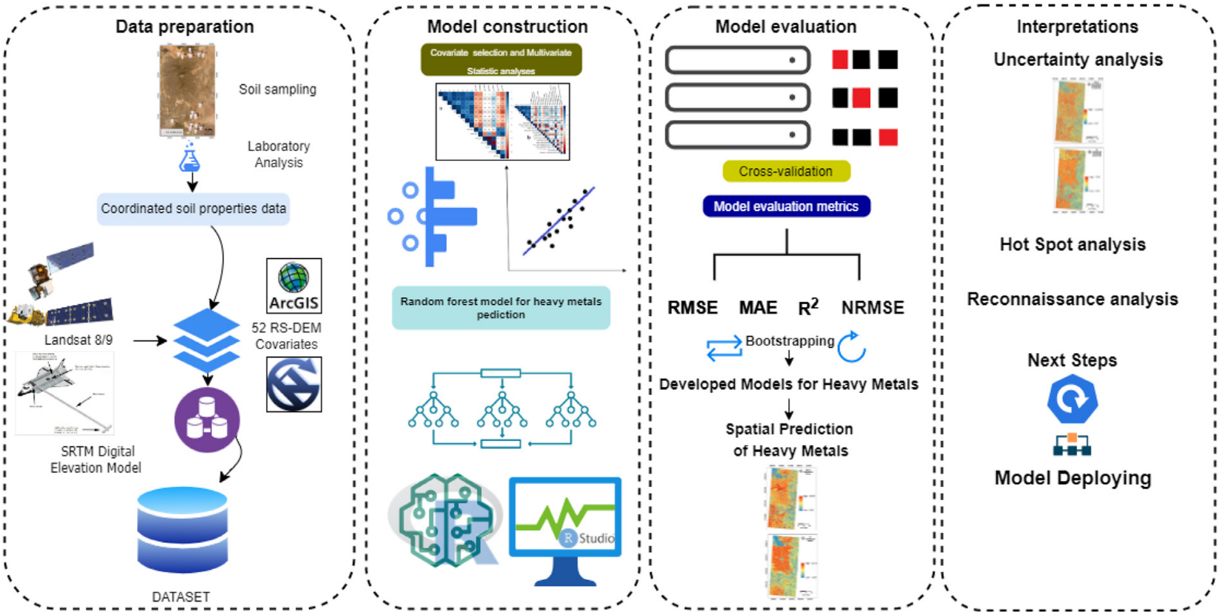


**Fig. 5.** The proposed methodological framework for the spatial prediction and background concentrations of soil heavy metals in the study area.

regression trees. Using bagging techniques as described by Breiman [19], RF parallelizes dataset partitioning into homogeneous tree subsets. Each tree is generated with random subsets of training data and features that from predictive models. The final predictions are the average of all trees [20,21]. The modelling process was carried out for each HM using the “train” function in the “caret” package in the R Core Environment software, version 4.2.1 [17].

*Model performance assessment*

The model performance on the full dataset was assessed using 5-fold cross-validation with 3 repetitions. The two models were further assessed using four different indicators: coefficient of determination ($R^2$) (Eq. (3)), coefficient of determination adjustment (adj-$R^2$) (Eq. (4)), mean absolute error (MAE) (Eq. (5)), root mean square error (RMSE) (Eq. (6)), and normalized root mean square error (NRMSE) (Eq. (7)).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(\Delta_i - M_i\right)^2}{\sum_{i=1}^{n} \left(\nexists_i - M_i\right)^2} \tag{3}$$

$$\text{Adjusted } R^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} \tag{4}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N} \left(Mi - \Delta_i\right)}{N} \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(M_i - \Delta_i\right)^2} \tag{6}$$

$$\text{NRMSE} = 100 \times \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(M_i - \Delta_i\right)^2}}{maxmin\left(M_i\right)} \tag{7}$$

*Models validation*

The residuals histogram and qqplots graphics obtained by the SMLR model for the eight HMs are shown in Fig. 6. The performance of the prediction models (SMLR and RF) was validated using root mean square error (RMSE), mean $R^2$, and normalized root mean square error (NRMSE) based on the predicted and measured heavy metal values. Fig. 7 shows detailed information about these metrics for both tested models. Briefly, in general, the SMLR outperformed the RF model for the studied HMs. Fig. 8 gives the Adj. $R^2$ and MAEs values for the eight heavy metals obtained by the SMLR model. The SMLR regression model used in this study passed the 0.05 significance test in terms of determining the environmental variables affecting the distribution of the studied soil HMs in the harrats region (Fig. 8a). Based on the results of the studied models, they were suitable for this study. The models established with remote sensing variables (i.e., Normalized Difference Moisture Index, Carbonate Normalized Ratio, Iron Normalized Ratio, Rock Outcrop Normalized Ratio) and terrain attributes (i.e., Mass balance index, Valley depth, Profile curvature, Convergence Index, Topographic wetness index, and Multi-resolution of ridge top flatness index) were most effective for predicting HMs concentrations. The environmental variables used in the final SMLR models are shown in Fig. 9.

In this method article, a novel methodology frame-work was proposed for estimation of the content and spatial distribution of soil HMs over relatively large areas that have not had major previous anthropogenic activity. The proposed integrated method combined 35 Landsat 8/9-OLI-based remote sensing variables and 17 Shuttle Radar Topography Mission (SRTM)-based terrain attributes coupled with stepwise multiple linear regression (SMLR) and random forest (RF) algorithms. The proposed method was effective for determining the eight studied soil HMs and describing their spatial distributions in the harrats region and can be easily applied to other similar geo-environmental settings.

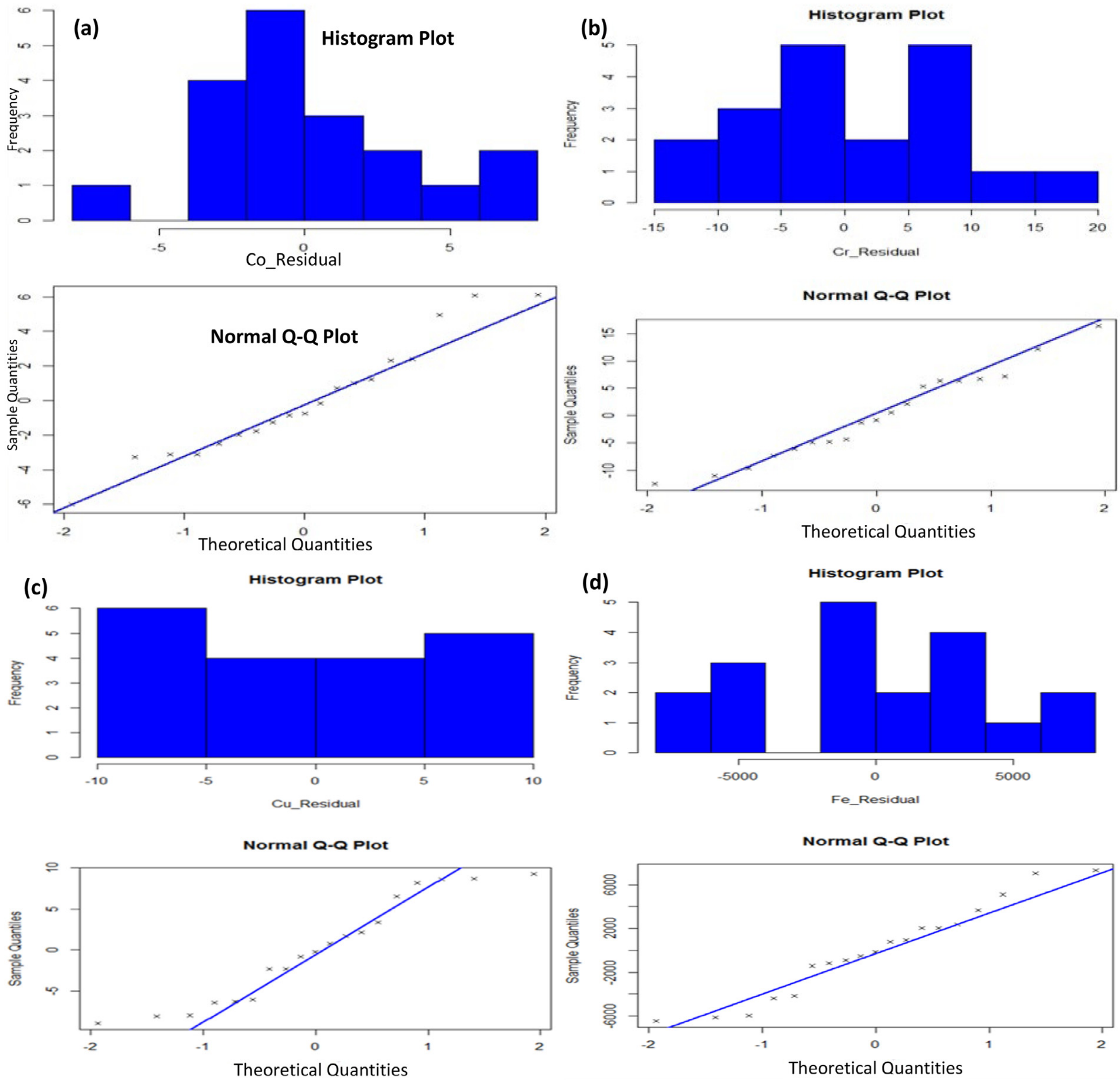


**Fig. 6.** Stepwise multiple linear regression (SMLR) model residuals histogram and qqplots graphics for: (a) Co, (b) Cr, (c) Cu, and (d) Fe. Stepwise multiple linear regression (SMLR) model residuals histogram and qqplots graphics for: (a) Mn, (b) Ni, (c) Pb, and (d) Zn.

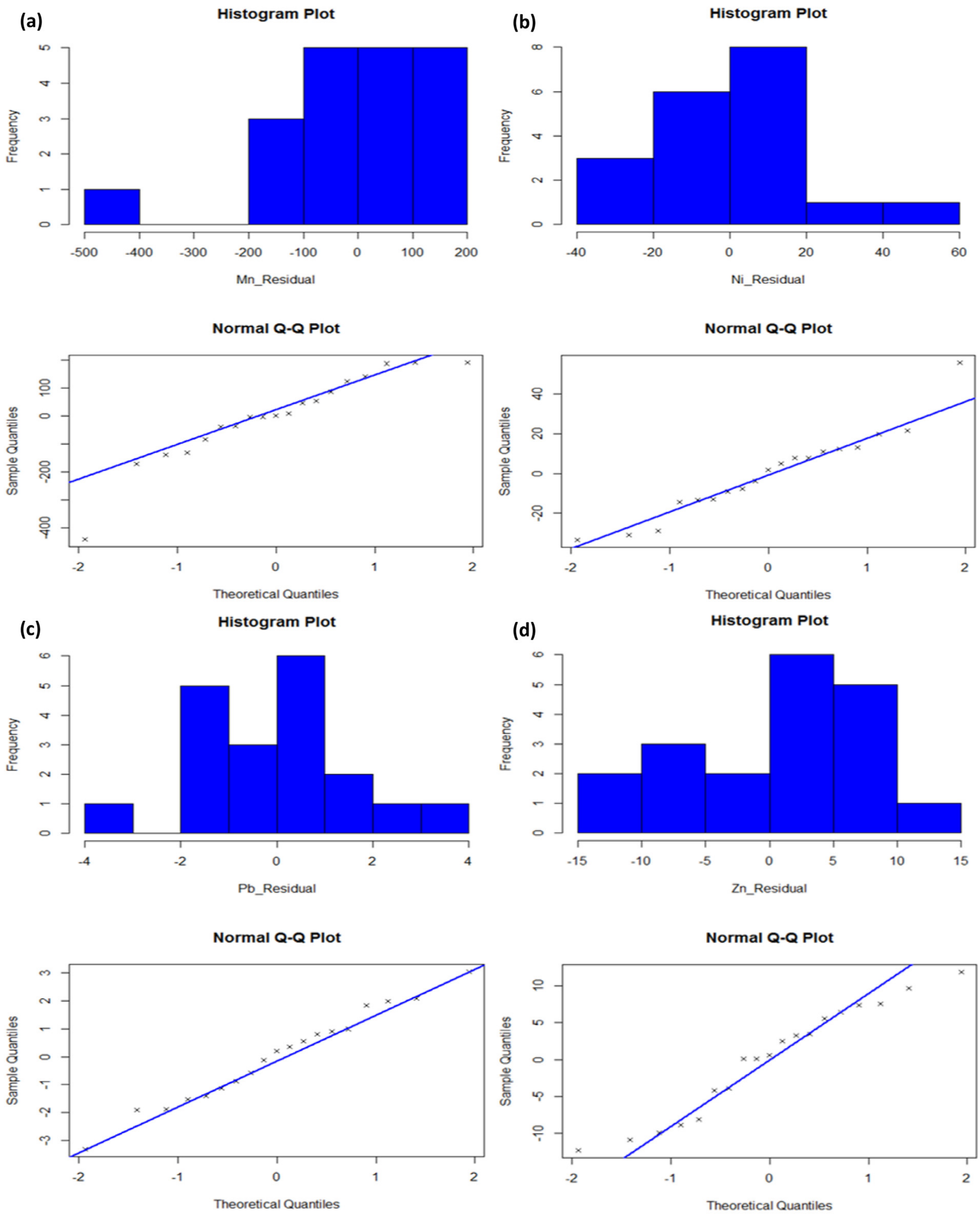


**Fig. 6.** Continued

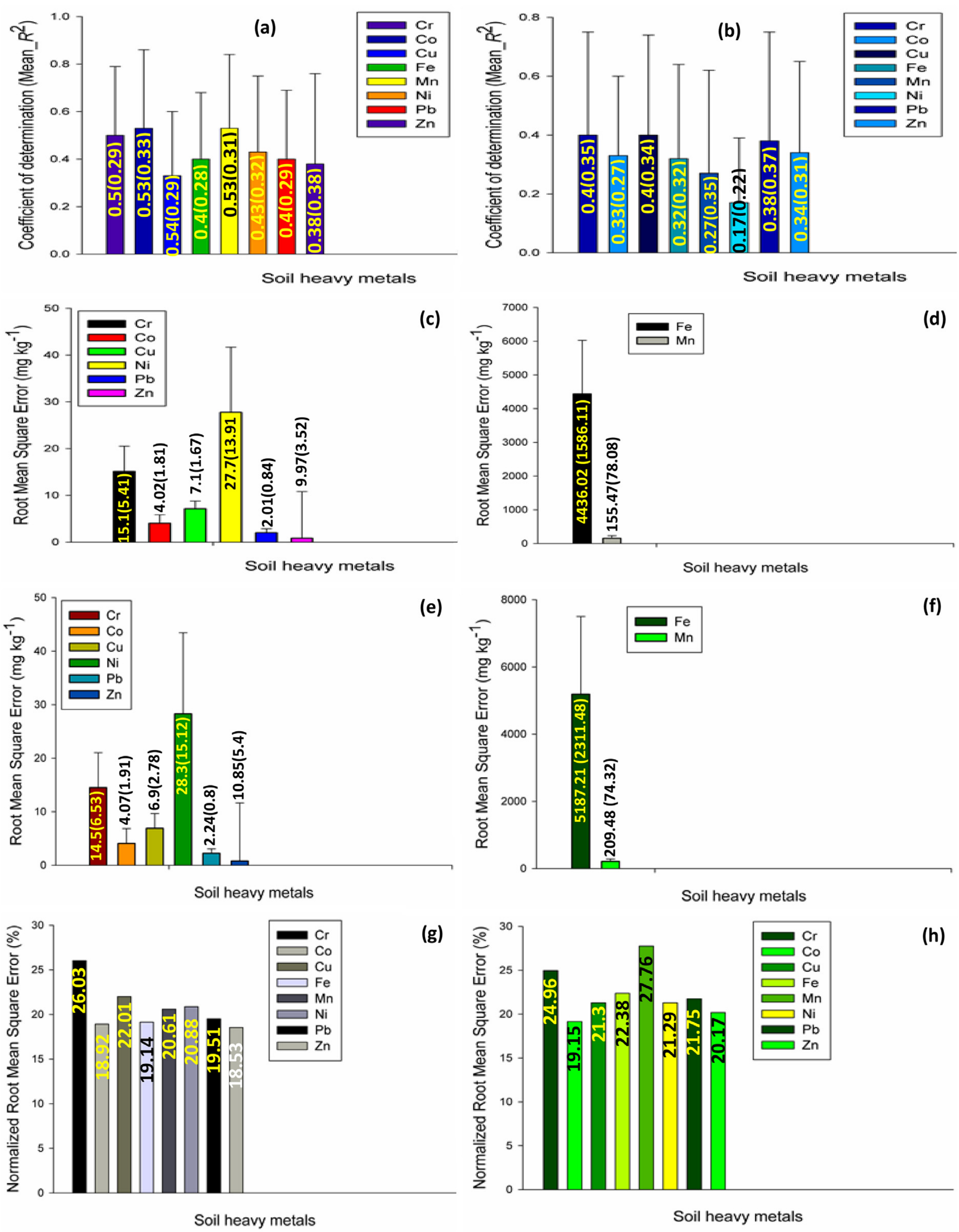


**Fig. 7.** The performance metrics (i.e., coefficient of determination ($R^2$), root mean square error (RMSE), and normalized root mean square error (NRMSE)) of the different models used for predicted soil HMs in the study area: (a), (c), (d), (g) obtained by stepwise multiple linear regression (SMLR); (b), (e), (f), (h) by random forest (RF). $R^2$ and RMSE are given as mean values with standard deviations.

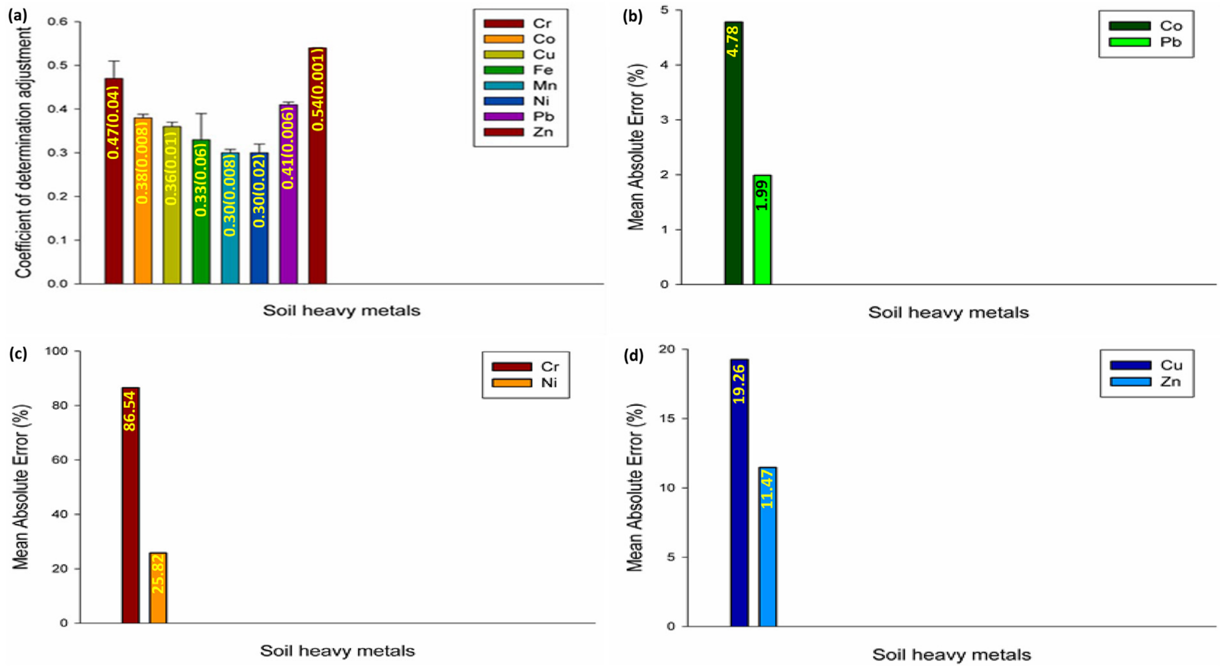


**Fig. 8.** Coefficient of determination adjustment (Adj. R2) with p-values for the eight soil HMs (a) and mean absolute errors (MAEs) for some HMs obtained by the SMLR model. MAE for: (b) Co and Pb, (c) Cr and Ni, and (d) Cu and Zn.

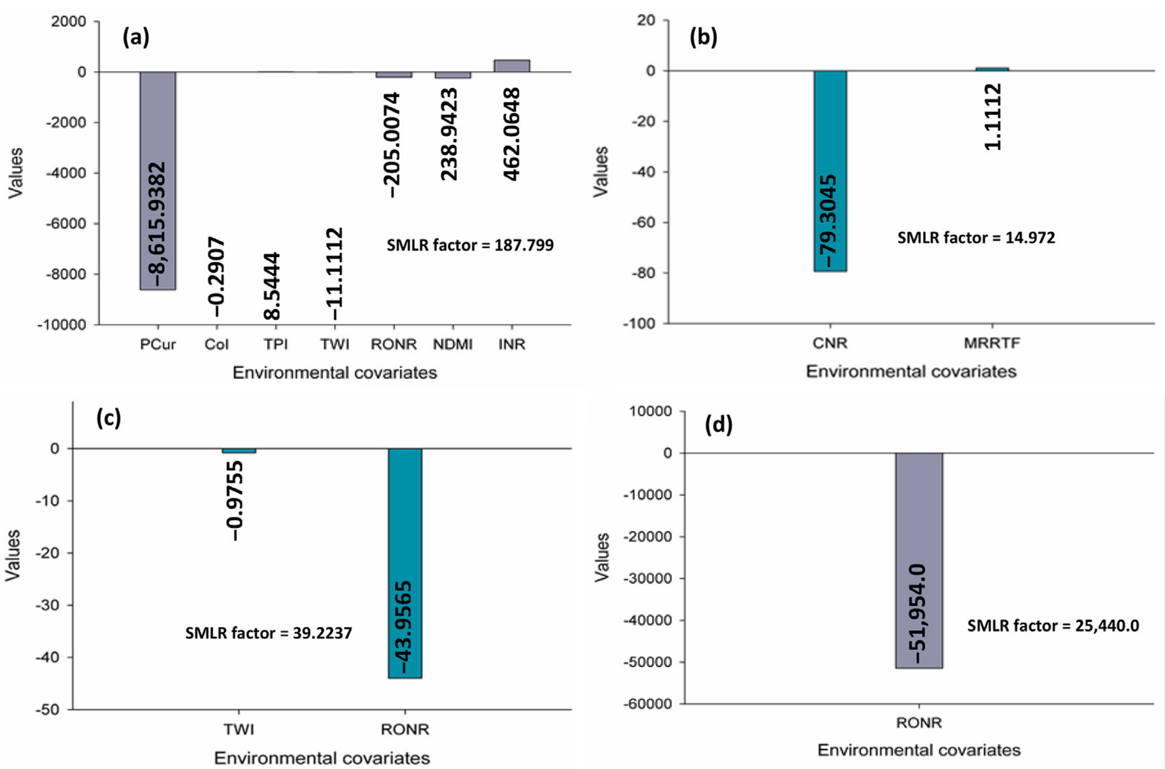


**Fig. 9.** Final environmental variables from the SMLR model for: (a) Co, (b) Cr, (c) Cu, (d) Fe, (e) Mn, (f) Ni, (g) Pb, and (h) Zn. Abbreviations: PCur = profile curvature; CoI = convergence index; TPI = topographic position index; TWI = topographic wetness index; RONR = rock outcrop normalized ratio; NDMI = normalized difference moisture index; INR = iron normalized ratio; CNR = MRRTF = multi-resolution of ridge top flatness index; carbonate normalized ratio; MBI = mass balance index.

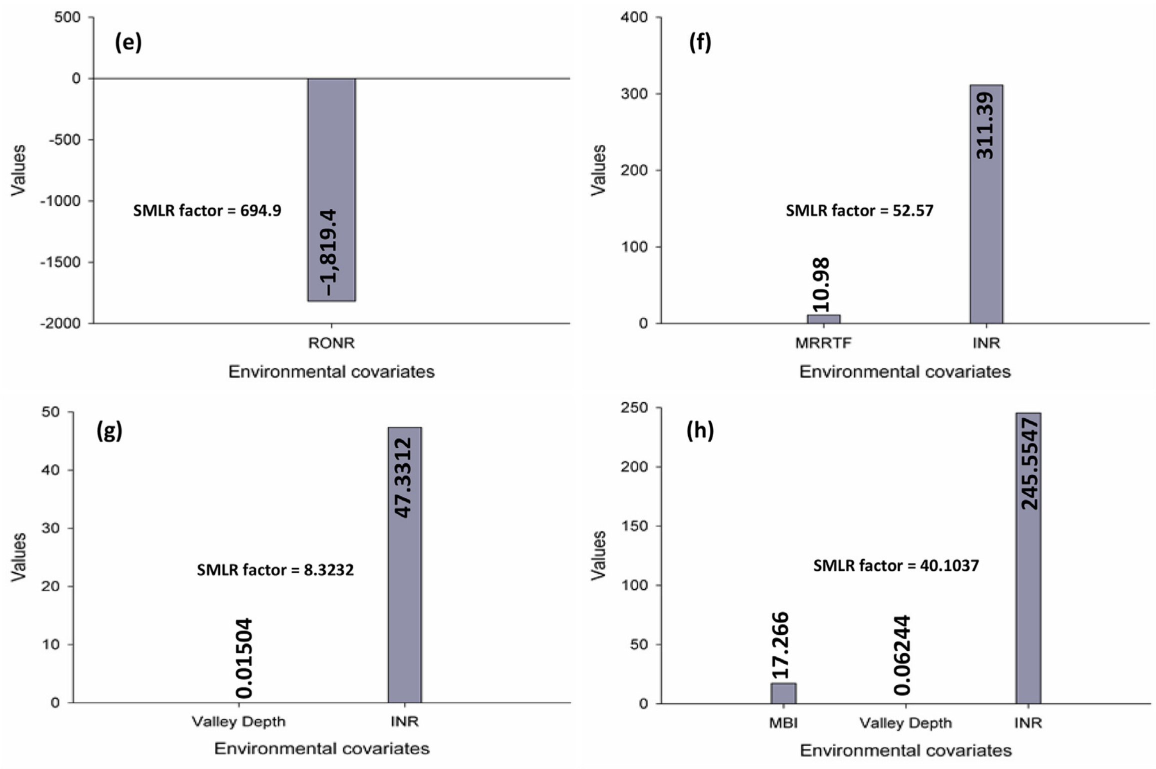


**Fig. 9.** Continued

## Ethics statements

This method does not include work with human subjects, animal experiments, or data gathered from social media platforms.

## Declaration of competing interest

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

## CRediT authorship contribution statement

**Magboul M. Sulieman:** Conceptualization, Methodology, Data curation, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Fuat Kaya:** Methodology, Data curation, Visualization, Investigation, Validation, Writing – review & editing. **Abdullah S. Al-Farraj:** Supervision, Writing – review & editing. **Eric C. Brevik:** Writing – original draft, Writing – review & editing.

## Data availability

Data will be made available on request.

## Acknowledgments

The authors extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for supporting this work through the College of Food and Agricultural Sciences Research Center (Grant ID 434108505). The first author thanks the Alexander von Humboldt (AvH) Foundation for awarding him the 2023 Georg-Forester Postdoctoral Fellowship (Grant No. 1237101). We thank the Editor-in-Chief (Prof. Dr. Damià Barceló) for considering our manuscript. We also acknowledge the anonymous reviewers for providing their constructive comments and suggestions that improved the final version of the manuscript.

## References

[1] USEPA, (2007). Method 3051A (SW-846): Microwave Assisted Acid Digestion of seDiments, Sludges, Soils, and Oils. Revision 1. Washington, DC.

[2] M. Xu, Z. Gao, Q. Wei, G. Chen, D. Tang, Label-free hairpin DNA-scaffolded silver nanoclusters for fluorescent detection of Hg2+ using exonuclease III-assisted target recycling amplification, Biosensor Bioelectron. 79 (2016) 411–415, doi:10.1016/j.bios.2015.12.081.
[3] Z. Qiu, J. Shu, G. Jin, M. Xu, Q. Wei, G. Chen, D. Tang, Invertase-labeling gold-dendrimer for in situ amplified detection mercury (II) with glucometer readout and thymine–Hg2+–thymine coordination chemistry, Biosensor. Bioelectron. 77 (2016) 681–686, doi:10.1016/j.bios.2015.10.044.
[4] M.M. Sulieman, F. Kaya, A. Keshavarzi, A.M. Hussein, A.S. Al-Farraj, E.C. Brevik, Spatial variability of some heavy metals in arid harrats soils: combining machine learning algorithms and synthetic indexes based-multitemporal Landsat 8/9 to establish background levels, Catena (Amst) 234 (2024) 107579, doi:10.1016/j.catena.2023.107579.
[5] Y. Gao, J. Tang, Q. Zhou, Z. Yu, D. Wu, D. Tang, Excited-state intramolecular proton transfer-driven photon-gating for photoelectrochemical sensing of CO-releasing molecule-3, Anal. Chem. 96 (12) (2024) 5014–5021, doi:10.1021/acs.analchem.4c00324.
[6] Y. Gao, Z. Yu, L. Huang, Y. Zeng, X. Liu, D. Tang, Photoinduced electron transfer modulated photoelectric signal: toward an organic small molecule-based photoelectrochemical platform for formaldehyde detection, Anal. Chem. 95 (23) (2023) 9130–9137, doi:10.1021/acs.analchem.3c01690.
[7] C. Luo, X. Zhang, X. Meng, H. Zhu, C. Ni, M. Chen, H. Liu, Regional mapping of soil organic matter content using multitemporal synthetic Landsat 8 images in Google Earth Engine, Catena (Amst) 209 (2022) 105842, doi:10.1016/j.catena.2021.105842.
[8] M. Zeraatpisheh, Y. Garosi, H.R. Owliaie, S. Ayoubi, R. Taghizadeh-Mehrjardi, T. Scholten, M. Xu, Improving the spatial prediction of soil organic carbon using environmental covariates selection: a comparison of a group of environmental covariates, Catena (Amst) 208 (2022) 105723, doi:10.1016/j.catena.2021.105723.
[9] ESRIArcGIS Desktop: Release 10.8.1, Environmental Systems Research Institute, Redlands, CA, USA, 2022.
[10] NASA-SRTM. (2013). Shuttle radar topography mission 1 arc-Second Global https://doi.org/10.5066/F7PR7TFT.
[11] O. Conrad, B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, J. Böhner, System for automated geoscientific analyses (SAGA) v. 2.1.4, Geosci. Model Dev. 8 (2015) 1991–2007, doi:10.5194/gmd-8-1991.
[12] K. Sayler, K. Zanter, Landsat 8 Collection 2 (C2) Level 2 Science Product (L2SP) Guide LSDS-1619 Version 2.0, EROS, Sioux Falls, South Dakota, 2020.
[13] R.J. Hijmans, raster: geographic Data Analysis and Modeling, R package version (2020) 3.4-5 https://CRAN.R-project.org/package=raster3 .
[14] A.M.J.-C. Wadoux, B. Minasny, A.B. McBratney, Machine learning for digital soil mapping: applications, challenges and suggested solutions, Earth. Sci. Rev. 210 (2020) 103359, doi:10.1016/j.earscirev.2020.103359.
[15] B. Droz, S. Payraudeau, J.A. Rodríguez Martín, G. Tóth, P. Panagos, L. Montanarella, P. Borrelli, G. Imfeld, Copper content and export in European vineyard soils influenced by climate and soil properties, Environ. Sci. Technol 55 (2021) 7327–7334, doi:10.1021/acs.est.0c02093.
[16] Hengl, T., & MacMillan, R.A. (2019). Predictive soil mapping with R. OpenGeoHub foundation. www.soilmapper.org.
[17] M. Kuhn, caret: classification and regression training, R package version (2020) 6.0-86 https://CRAN.R-project.org/package=caret .
[18] R Core TeamR: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[19] L. Breiman, Random forests, Mach. Learn. (2001) 5–32 pages, doi:10.1023/A:1010933404324.
[20] F. Kaya, G. Mishra, R. Francaviglia, A. Keshavarzi, in: Combining Digital Covariates and Machine Learning Models to Predict the Spatial Variation of Soil Cation Exchange Capacity, 12, Land, 2023, p. 819, doi:10.3390/land12040819.
[21] G. Mishra, M.M. Sulieman, F. Kaya, R. Francaviglia, A. Keshavarzi, E. Bakhshandeh, M. Loum, A. Jangir, I. Ahmed, A. Elmobarak, A. Basher, D. Rawat, Machine learning for cation exchange capacity prediction in different land uses, Catena (Amst) (2022), doi:10.1016/j.catena.2022.106404.