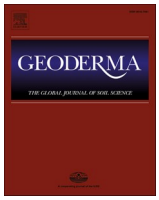
Geoderma 453 (2025) 117138

Contents lists available at ScienceDirect

Geoderma

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



Comprehensive spatial mapping of metals and metalloids in the Peruvian Mantaro Valley using advanced geospatial data Integration

Samuel Pizarro [a,*], Narcisa G. Pricope [b], Jesús Vera [a], Juancarlos Cruz [c], Sphyros Lastra [c], Richard Solórzano-Acosta [c], Patricia Verástegui Martínez [a]

[a] Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), Carretera Saños Grande - Hualahoyo Km 8 Santa Ana, Huancayo, Junin 12002, Peru
[b] Department of Geosciences, Mississippi State University, Starkville 39579, USA
[c] Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru

ARTICLE INFO

Handling Editor: B. Minasny

Keywords:
Random forest
Soil mapping
Google earth engine
Machine learning
Cloud computing

ABSTRACT

The quality and safety of soil are crucial for ensuring social and economic development and providing contaminant-free food. The availability and quality of soil data, particularly for multiple metals and metalloids, are often insufficient for comprehensive analysis. Soil formation and the distribution of metals are shaped by various factors such as geology, climate, topography, and human activities, making accurate modeling highly challenging. Additionally, agricultural intensification, urban expansion, road construction, and mining activities frequently result in soil pollution, posing serious risks to ecosystems and human health. This study aims to integrate diverse geospatial datasets with machine learning for high resolution soil contamination mapping (10 m spatial resolution) in a major agricultural region of Peruvian highlands. This study mapped 25 elements (Ca, Mg, Sr, Ba, Be, K, Na, As, Sb, Se, Tl, Cd, Zn, Al, Pb, Hg, Cr, Ni, Cu, Mo, Ag, Fe, Co, Mn, V) in the Peruvian Mantaro Valley using a training dataset of 109 topsoil samples combined with various geospatial datasets (remote sensing, climate, topography, soil data, and distance). The model provided satisfactory results in predicting the spatial distribution of the selected elements, with R^2 values ranging from 0.6 to 0.9 for most elements. Edaphic, climate, and topographic covariates were the most significant predictors, particularly for croplands near rivers, whereas spectral variables were less important. The results reveal As, Pb, and Cd concentrations significantly above permissible limits, highlighting urgent health risks. These findings suggest that it is feasible to identify polluted soils and improve regulations based on widely available geospatial datasets with minimal training data. The study contributes to the development of models to assess the impact of pollutants on environmental and human health in the short-to-medium term, emphasizing the need for further research on the translocation of toxic metals into food crops and the implications for public health.

1. Introduction

Soils, alongside with air and water, are crucial for agriculture, forming the base for most food production and contributing 98.8 % of daily calorie intake worldwide (Kopittke et al., 2019). Beyond food production, soil is pivotal in supporting terrestrial ecosystems, offering critical ecosystem services such as nutrient provision, water filtration, and contaminant removal (Brevik et al., 2020). Healthy soil regulates water flow, supports diverse organisms, and acts as a carbon sink, helping to mitigate the effects of climate change and global environmental changes (Tahat et al., 2020). Healthy soils contribute significantly to agricultural productivity and sustainability. Implementing sustainable soil management practices can lead to cost savings by reducing the need for chemical inputs and improving crop yields (Tegtmeier and Duffy, 2004). Conversely, the degradation of soil quality can lead to significant ecological and economic consequences, including reduced agricultural productivity and increased greenhouse gas emissions that can result in substantial economic losses due to reduced

* Corresponding author.
E-mail addresses: samuel.pizarro@untrm.edu.pe (S. Pizarro), npricope@research.msstate.edu (N.G. Pricope), jvera@lamolina.edu.pe (J. Vera), jcruz@inia.gob.pe (J. Cruz), slastrapaucar@gmail.com (S. Lastra), investigacion_labsaf@inia.gob.pe (R. Solórzano-Acosta), patymarve@gmail.com (P.V. Martínez).

https://doi.org/10.1016/j.geoderma.2024.117138
Received 27 May 2024; Received in revised form 6 December 2024; Accepted 7 December 2024
Available online 12 December 2024

productivity and increased costs associated with soil restoration (Lal, 2020).Fig. 1..

Soils have considerable effects on human health, both directly and indirectly as well as positively and negatively (Brevik and Sauer, 2015; Steffan et al., 2018). Human nutritional deficiencies often stem from soil nutrient limitations, and while agricultural practices have improved crop yields, there has been a concerning decline in the nutritional quality of crops, particularly in essential minerals such as calcium, iron, and zinc (Silver et al., 2021). This phenomenon, known as "hidden hunger," affects billions of people worldwide and underscores the importance of soil health in ensuring food security and nutrition (Graham et al., 2007). Additionally, soils can influence human health through their ability to filter and decompose contaminants. However, contaminated soils can pose significant health risks by transferring harmful elements and pathogens to crops and water sources, ultimately affecting food safety and quality (Steffan et al., 2018). Thus, the quality and safety of soil resources is essential for ensuring social and economic development (Chen et al., 2016).

Agricultural intensification, the continued expansion of urban areas, the construction of roads, and mining activities worldwide can lead to soil degradation and contamination through a wide range of mechanisms. Factors such as pollutant translocation through water and air, incorporation via fertilization, and distribution by erosion and infiltration processes – which are affected by topography, soil characteristics and climate –contribute to soil pollution. This pollution, in turn, leads to increased environmental risks to ecosystem functions and human health (Silver et al., 2021;Wu et al., 2023). Metallic elements such as Manganese (Mn), Zinc (Zn), Copper (Cu), Iron (Fe), Molybdenum (Mo), Nickel (Ni), Magnesium (Mg), Calcium (Ca) or Boron (B), when present in soils at relatively low concentrations, can enhance the growth and development of plants and the human body, but concentrations above optimal levels can negatively affect plant development (Rashid et al., 2023). Other elements, such as Cadmium (Cd), Lead (Pb), Chromium (Cr) and Arsenic (As) are considered toxic and detrimental for plant and human health when detectable in soils at any levels (Brevik et al., 2020; Steffan et al., 2018).

Recent advancements in soil management include the use of remote sensing technologies to monitor soil moisture levels, soil sensors that provide real-time data on soil properties, and data analytics platforms that integrate various data sources to enhance soil health management (Minasny and McBratney, 2016; Viscarra Rossel et al., 2016). With the increasing availability of spatial datasets from satellites and models, it is now possible to use advanced machine learning techniques to develop spatial prediction frameworks for soil classification, fertility assessment, and heavy metal contamination.

The digital soil mapping framework leverages multiple sources and scales of spatial data, enabling more complex and accurate analyses that have shown promising results in various studies. (Chen et al., 2016; Liu et al., 2023; Taghizadeh-Mehrjardi et al., 2019). The random forest (RF) (Breiman, 2001), a supervised machine-learning algorithm typically used for regression problems, has been widely used in various applications for soil mapping, effectively handling multiple classes of data (Lachaud et al., 2023; Moradpour et al., 2023; Omondi and Boitt, 2020; Pizarro et al., 2023; Shi et al., 2021; Thomas et al., 2023). RF combines multiple decision trees to predict values through a voting system. This method makes no assumptions about data distribution and can handle both categorical and continuous variables effectively. RF is known for its robust nonlinear data mining and generalization capabilities (Gholizadeh et al., 2018). Utilizing complex models and large volumes of spatial data poses significant challenges, necessitating the use of cloud computing platforms like Google Earth Engine (GEE) (Gorelick et al., 2017) to reduce time processing and resources while ensuring accurate results.

Similar to many agricultural valleys and regions where extractive activities predominate worldwide, in the Peruvian Mantaro Valley certain areas have been identified as having polluted soils with high

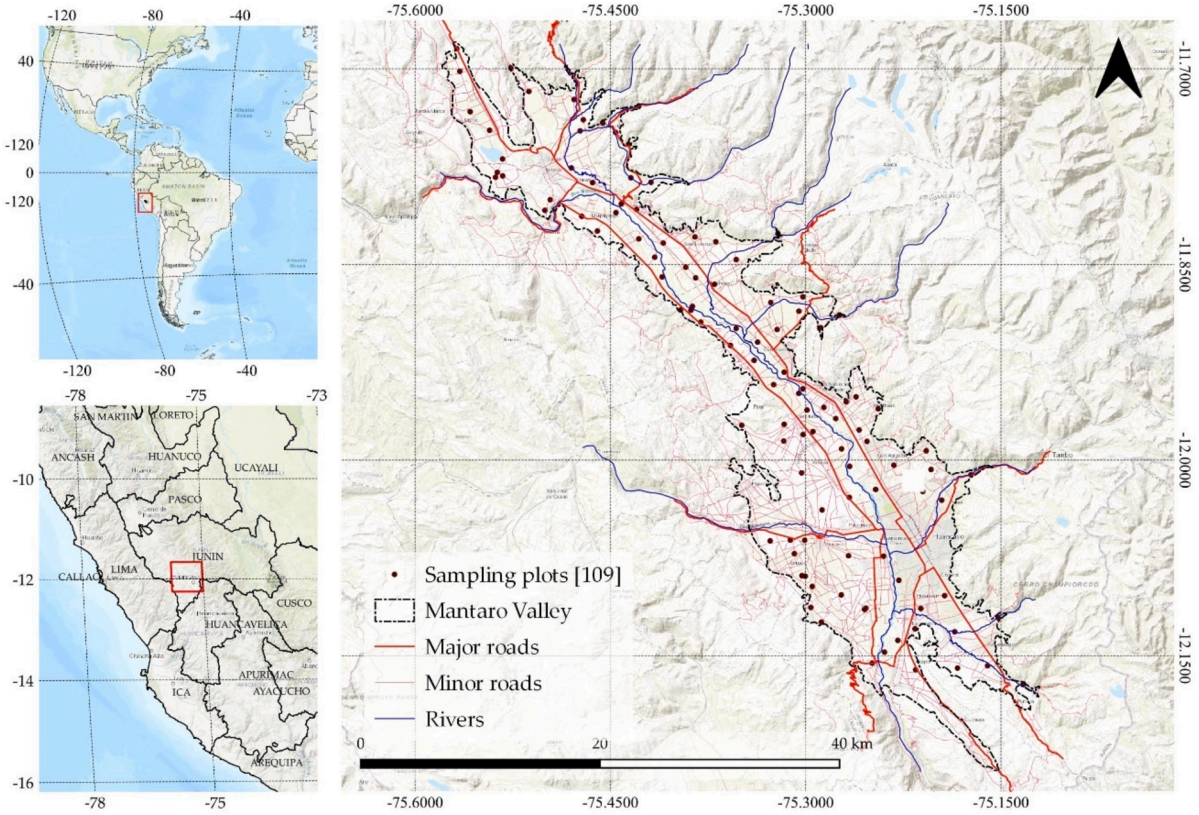


Fig. 1. Location of the study area, Mantaro Valley, Junin (Peru).

concentrations of Cd, Pb, Zn, As and Cu, particularly in regions irrigated from the Mantaro River (Custodio et al., 2021; Munive et al., 2020). In such areas, it is essential to understand the spatial distribution of these elements in soils so as to identify polluted areas, enhance planning and rehabilitation efforts and improve regulations. This is particularly important in developing countries, where effective remediation strategies are needed to mitigate the environmental and health risks associated with contaminated soils.

This study aimed to develop a framework for digital mapping of soil elements, focusing on 25 key elements: calcium (Ca), magnesium (Mg), strontium (Sr), barium (Ba), beryllium (Be), potassium (K), sodium (Na), arsenic (As), antimony (Sb), selenium (Se), thallium (Tl), cadmium (Cd), zinc (Zn), aluminum (Al), lead (Pb), mercury (Hg), chromium (Cr), nickel (Ni), copper (Cu), molybdenum (Mo), silver (Ag), iron (Fe), cobalt (Co), manganese (Mn), and vanadium (V). To achieve this, various geospatial datasets, including remote sensing, climate, topography, soil data, and distance, were integrated to establish spatial estimation models for these elements using a dataset consisting of 109 soil samples. Understanding the spatial distribution of soil contaminants is crucial because it allows for the identification of pollution hotspots that require targeted remediation efforts, thereby protecting both environmental and human health (Steffan et al., 2018). Moreover, it informs sustainable agricultural practices by identifying areas where soil quality needs improvement, ensuring safe food production and enhancing crop yields (Graham et al., 2007; Lal, 2020). This knowledge is essential for policymakers to develop effective regulations and strategies for soil management, ultimately contributing to long-term ecological and economic stability (Chen et al., 2016).

In this context, the aims of this work are to: (i) generate digital maps of soil elements using the Google Earth Engine (GEE) platform, incorporating various geospatial covariates to map the spatial distribution of 25 different soil elements, (ii) assess and compare the accuracy of multiple Random Forest models in predicting element values based on spatial covariates; and (iii) identify the most important covariates for predicting soil element content, especially for metals and metalloids.

2. Methodology

2.1. Study site

The Mantaro Valley (MV) is located in the Peruvian central highlands, between 12.2377S and 11.6793 latitude South and 75.5792 W and 75.1202 longitude West, with an altitudinal gradient between 3150 to 3750 m.a.s.L. The study region is composed of four provinces: Chupaca, Concepción, Huancayo and Jauja, within 57 districts in total, along 53 km between the urban concentrations of Jauja and Huancayo City, and a width ranging from 4 to 21 km in places and flanked by Cordillera Occidental on the west and the Cordillera Central on the east.

The MV has the largest proportion of agricultural land in the Andes, with almost 43,000 ha of croplands, typically planted with potato, wheat, corn, onion, garlic, leafy vegetables and livestock. The sowing is recognized as the “big growing season” which starts in September and October and harvesting in February and March. The “small growing season” takes place from May to August under irrigation, but it represents only 10 % of the croplands (Fujimoto et al., 2004). The climate is characterized by periods of rain between October and March, and a dry season between April and September, with an average annual precipitation of 650 mm/year. The mean temperature ranges from 4 to 18 °C, with the lowest temperatures between May and August, and frost events between July and August (Instituto Geofísico del Perú, 2005).

2.2. Methodological framework

The methodological framework employed in this study is presented in Fig. 2, and described in more detail in the following four methods subsections below.Fig. 3..

2.2.1. Land cover analysis

To determine the extent of land cover for the sampling date, we generated a supervised map based on a Sentinel-2 imagery collection from July to August 2023, corresponding to the land preparation season. We used the Land Cover classification approach proposed by Pizarro

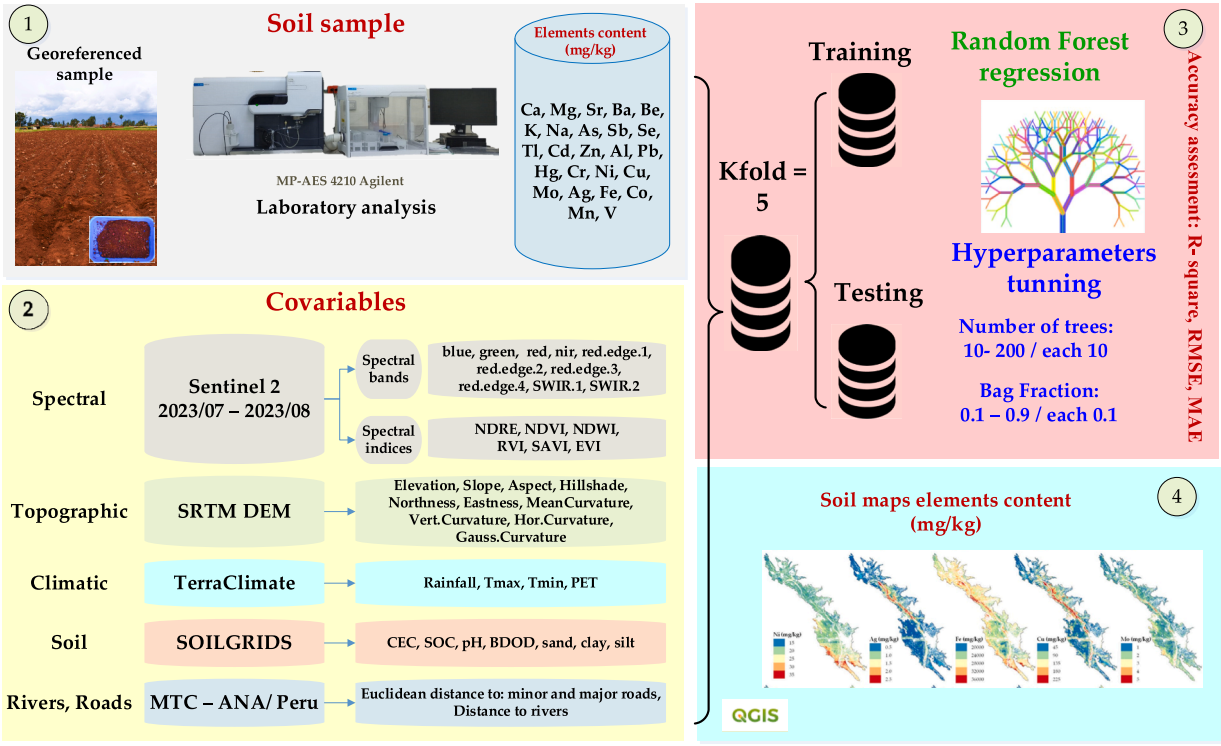


Fig. 2. Representation of the methodological framework used in this study.

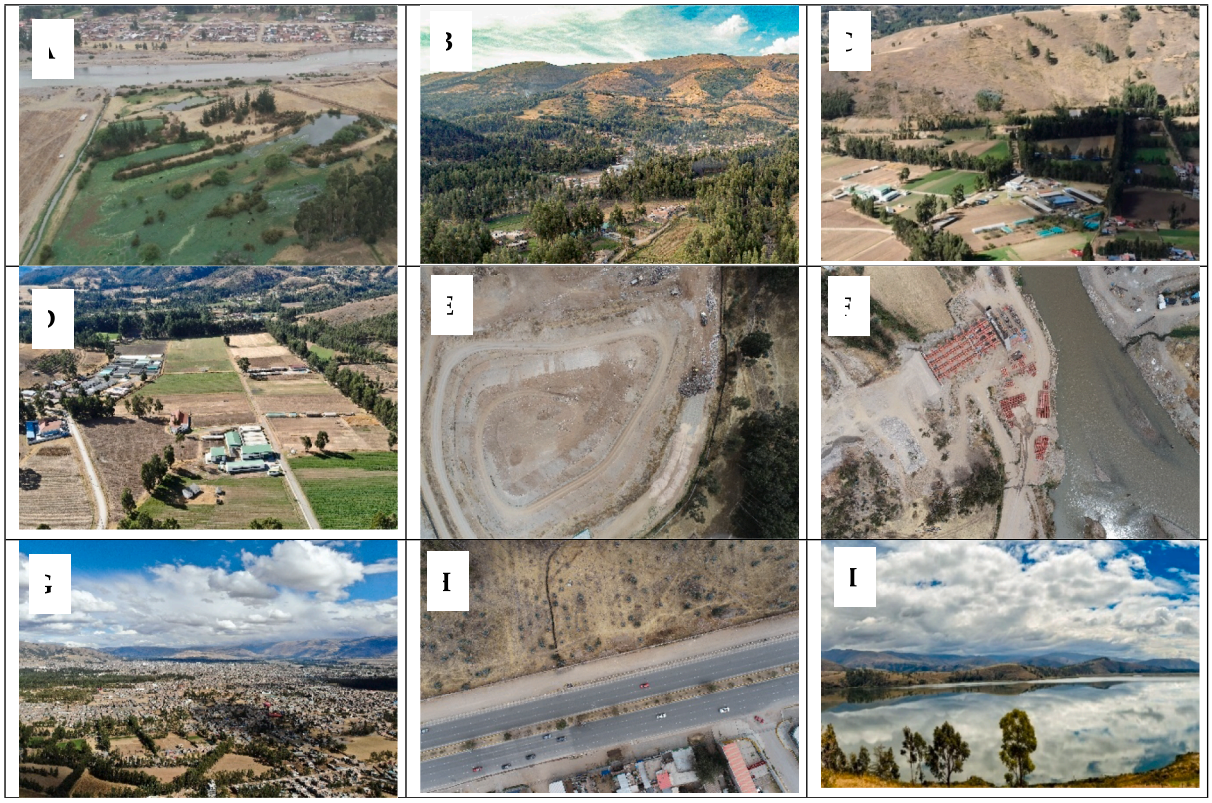


Fig. 3. Depiction of land cover classes adopted in the LULC product and found in Mantaro Valley. (A) Wetlands. (B) Forest. (C) Shrubland. (D) Cropland. (E) Bareland. (F) Sand bar. (G) Urban (built up). (H) Roads. (I) Water body.

et al. (Pizarro et al., 2022) that combines spectral bands and topographic covariates derived from a digital elevation model (DEM). The land cover training set was composed of nine land cover classes and completed with other reference sources such as the 2015 National Vegetation Map (Ministerio del Ambiente (MINAM), 2015) into several classes of interest: croplands, urban zones, bare lands, shrublands, wetlands, forest, water bodies and roads. Finally, we included a Sand Bar class near rivers and floodplains using a Google Earth imagery (pan-sharpened QuickBird, GeoEye, and WorldView-2 imagery), and visual interpretation (Fig. 2).

We used a set of randomly selected set consisting of 9000 pixel samples (1000 for each class), derived 80 % from field data, and 20 % from Google Earth imagery and the 2015 National Vegetation Map. The samples were split into a validation dataset in a 30/70 % proportion using a stratified random sampling method for each land cover class, ensuring independen.

2.2.2. ce between the training and validation data

The accuracy of the classifier was evaluated with the overall accuracy (OA), kappa coefficient (K), producer accuracy (PA), and user accuracy (UA), where K indicates the degree of agreement between the ground-truth data and the predicted values, while the PA measures how well a pixel has been classified and includes the error of omission (the proportion of observed features on the ground that are erroneously excluded from a class). The UA measures the reliability of the map, representing how well the map portrays what is on the ground and it includes the error of commission which refers to pixels erroneously included in a class (Congalton, 1991).

2.2.3. Field sampling of chemical and physical soil parameters

European soil-sampling guidelines (Theocharopoulos et al., 2001) were followed to conduct a thorough sampling of Mantaro Valley croplands. A grid with a 2.5-kilometer interval was used, yielding 109 soil samples at a density of one point per 390 ha. Each soil sample consisted of four subsamples collected at a depth of 30 cm between July and August 2023. The subsamples were spaced 20 m apart in a cross pattern and georeferenced using a Leica Zeno 5 sub-metric GPS.

Twenty-five elements were analyzed to assess metal and metalloid concentrations in the soil samples. These elements included Ca, Mg, Sr, Ba, Be, K, Na, As, Sb, Se, Tl, Cd, Zn, Al, Pb, Hg, Cr, Ni, Cu, Mo, Ag, Fe, Co, Mn, and V. The analysis was conducted at the Soil, Water, and Foliar Laboratory (LABSAF) of the Santa Ana Agricultural Station.

After being dried for 48 h at 40 °C, the soil samples were ground and sieved in a 2 mm sieve. Then 1 g of each dry soil sample was taken and digested in a 50 ml centrifugal tube with 0.5 ml of concentrated HNO_3 acid (69–70 %, v/v) and 0.5 ml of hydrogen peroxide (H_2O_2). Next, the solution was filtered and the Cd, Pb, Mn and Zn concentrations were measured using MP-AES. The analytical standards were prepared in a matrix of 1 % HNO_3 and 1 % HCl. Arsenic (As) was measured by pre-reduction of arsenic with potassium iodide (KI) and then by hydride generation. The element concentrations are expressed in mg / kg.

Two separate methodologies were used for the analysis of the respective elements: EPA Method 3050A (U.S. EPA, 1992) and EPA Method 6020B (SW-846) (U.S. EPA, 2014). EPA Method 3050A involves the acid digestion of sediments, sludge, and soils, which was used to digest each of the soil samples. EPA Method 6020B employs Inductively Coupled Plasma – Mass Spectrometry (ICP-MS) for the analysis of elemental content. Both methodologies and the supporting equipment can detect the necessary limits required for the purposes of this analysis.

2.2.4. Acquisition and processing of environmental covariates

To develop the spatial soil elements distribution models, we utilized five sets of covariates, shown in Table 1.

A. Spectral variables

As mentioned above, we used the Google Earth Engine platform to build a Sentinel-2 multispectral composite mosaic by a median from the

Table 1
All environmental covariates and their abbreviations and sources.

Data type	Covariate	Abbreviation	Source
Spectral	Sentinel-2 reflectance bands	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12,	blue, green, red, red edge 1, red edge 2, red edge 3, nir, red edge 4, SWIR 1, SWIR 2
	Normalized difference vegetation index	NDVI	(NIR − RED)/ (NIR + RED)
	Normalized difference water index	NDWI	(GREEN − NIR)/ (GREEN + NIR)
	Soil-adjusted vegetation index	SAVI	(NIR − RED)/ (NIR + RED + 1)(1 + 0.6)
	Enhanced Vegetation Index	EVI	2.5×(NIR-RED)/ (NIR + 6 × RED + 7.5 × BLUE + 1)
	Ratio Vegetation Index	RVI	NIR/GREEN
Topographic	Elevation	Elevation	https://srtm.csi.cgiar.org
	Slope	Slope	Calculated from Elevation
	Aspect	Aspect	Calculated from Elevation
	Eastness	Easting	Calculated from Elevation
	Northness	Northing	Calculated from Elevation
	Gaussian curvature	Gauss. Curvature	Calculated from Elevation
	Horizontal curvature	Hor.Curvature	Calculated from Elevation
	Vertical curvature	Vert.Curvature	Calculated from Elevation
	Mean Curvature	Mean. Curvature	Calculated from Elevation
Climatic	Reference evapotranspiration (ASCE Penman-Montieth)	PET	https://www.climatologylab.org/terraclimate.html
	Minimum temperature	Temp. min	https://www.climatologylab.org/terraclimate.html
	Maximum temperature	Temp. max	https://www.climatologylab.org/terraclimate.html
	Precipitation	Precipitation	https://www.climatologylab.org/terraclimate.html
Soil	Proportion of clay particles (< 0.002 mm) in the fine earth fraction	clay	https://www.isric.org/explore/soilgrids
	Proportion of sand particles (> 0.05 mm) in the fine earth fraction	sand	https://www.isric.org/explore/soilgrids
	Proportion of silt particles (≥ 0.002 mm and ≤ 0.05 mm) in the fine earth fraction	silt	https://www.isric.org/explore/soilgrids
	Soil pH	pH	https://www.isric.org/explore/soilgrids
	Cation exchange capacity of the soil	CEC	https://www.isric.org/explore/soilgrids
	Bulk density of the fine earth fraction	BDOD	https://www.isric.org/explore/soilgrids
	Soil organic carbon content in the fine earth fraction	SOC	https://www.isric.org/explore/soilgrids
	Total nitrogen (N)	Nitrogen	https://www.isric.org/explore/soilgrids
Distance	Distance to major roads	D_MjR	https://portal.mtc.gob.pe/estadisticas/descarga.html
	Distance to minor roads	D_mnR	https://portal.mtc.gob.pe/estadisticas/descarga.html
	Distance to rivers	D_Riv	https://www.geoidep.gob.pe/autoridad-nacional-del-agua-ana

image collection "COPERNICUS/S2_SR". We applied a cloud filter to all the scenes and performed pan sharpening for bands B5, B6, B7, B8A, B11 and B12 to achieve a 10-m resolution for bands B2, B3, B4 and B8. This process was conducted for each scene from July to August 2023, covering the entire study area with a collection of 16 images.

Additionally, based on previous research, we computed spectral indices including the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), the Normalized Difference Water Index (NDWI) (McFeeters, 1996), the Soil-Adjusted Vegetation Index (SAVI) (Qi et al., 1994), the Enhanced Vegetation Index (EVI) (Huete et al., 2002), the Ratio Vegetation Index (RVI) (Pearson and Miller, 1972) and the Red-edge Normalized Difference Vegetation Index (NDRE) (Gitelson and Merzlyak, 1994). This range of spectral indices has been shown to present considerable advantages in detecting and enhancing predictions of heavy metal presence in soils (Omondi and Boitt, 2020; Peng et al., 2021). Table 1 includes not only the complete list of variables included, as well as the formulas used to calculate them.

B. Topographic variables

Topography affects the redistribution of elements in soil (Wu et al., 2023b), therefore the effect of topographic factors on soil composition is critical to consider. Topographic variables were derived from the Shuttle Radar Topographic Mission (SRTM) digital elevation model and resampled to 10-m resolution. Using the TAGEE package implemented in GEE (Safanelli et al., 2020), we extracted covariates related to terrain that include slope, aspect, eastness, northness, Gaussian curvature, horizontal curvature, vertical curvature and mean curvature. These variables are important as they influence soil formation processes, erosion patterns, water drainage, and nutrient distribution, which are crucial for accurate digital soil mapping (Liu et al., 2023).

C. Climatic variables

Climate has an influence on geochemical processes in soils and facilitates the soil-to-plant transfer of various elements (Cornu et al., 2016). As such, we included the following climatic variables: precipitation, maximum temperature (Temp.min), minimum temperature (Temp.max), and reference evapotranspiration (PET). These variables help in understanding the environmental conditions that affect soil properties (leaching, nutrient availability, microbial activity, or water retention) and element redistribution. The data were obtained from the TerraClimate repository in GEE (Abatzoglou et al., 2018). TerraClimate provides monthly gridded climate and climatic data at a spatial resolution of 0.5° from 1958 to 2023. The climate data was processed to average PET, Temp.max and Temp.min, average annual precipitation for the last 30 years.

D. Soil variables

Gridded soil data were obtained from the SoilGrids database, published by the International Soil Reference and Information Centre (https://www.isric.org/explore/soilgrids). SoilGrids is a system for global digital soil mapping based on machine learning and trained with multiple covariates and soil profile databases obtained from the WoSIS (Poggio et al., 2021). To construct the soil variable set for our model, we selected the proportions of clay, sand, and silt, as well as soil pH, the

cation exchange capacity (CEC), the bulk density of the fine earth fraction (BD), the soil organic carbon content (SOC) and the total soil nitrogen content. All soil covariates were resampled to a 10-m grid resolution using bilinear interpolation in the GEE environment. The selected soil variables are critically important as they collectively influence soil texture, fertility, structure, and overall health. These properties affect key soil functions, including water retention, nutrient availability, microbial activity, root growth, filtering, buffering, and transforming of organic and inorganic pollutants, minimizing the leaching of contaminants into the groundwater (Hengl et al., 2014; McBratney et al., 2003; Poggio et al., 2021; Sarkar et al., 2021).

E. Distance variable

The major and minor roads inventory was obtained from the Peruvian Transport and Communication Ministry (https://portal.mtc.gob.pe/estadisticas/descarga.html), and the regional rivers inventory was downloaded from the Peruvian National Water Administration (https://www.geoidep.gob.pe/autoridad-nacional-del-agua-ana). To constr uct distance variable set, we used the Euclidean distance mapping approach (Danielsson, 1980). These distance variables are crucial because proximity to roads and rivers can significantly influence the distribution of heavy metals in soils due to factors like vehicular emissions, runoff, and erosion. These variables has been employed in other studies focused on heavy metal soil mapping (Zhang et al., 2020).

For inclusion in the final prediction model, all covariates were resampled to the resolution of 10 m using the nearest neighbor interpolation method.

2.2.5. Covariates selection

Considering the large number of available covariables (40 layers), in order to reduce redundancy between covariates, and obtain a more computationally efficient model, we implemented a de-correlation analysis, selecting only covariate layers that had a pairwise correlation coefficient below 0.85 with all other covariates. For each pair of covariates **correlated** above this threshold, only the first one in alphabetical order was selected for inclusion in the modelling phase. This step reduced the number of initial covariates to 24.

2.2.6. Selection of samples and optimal parameter of RF

We used machine learning Random Forest model (Breiman, 2001) to construct multiple models, and applied to each element mapped. RF makes no assumptions about data distribution and can handle both categorical and continuous variables simultaneously. It is also known for its robust nonlinear data mining capabilities and strong generalization performance (Gholizadeh et al., 2018).

The operation of the RF algorithm on the GEE platform requires providing six parameters. In this study, the number of trees (NT) from 50 to 500 at 50-tree intervals, the variables per split (VPS) known as *mtry*, from 1 to 24 at 1-variable intervals, and the bag fraction (BF) from 0.1 to 1 at 0.2 fraction intervals were examined to identify optimal parameters of RF for elements prediction. As a result, a total of 1,200 combinations (NT = 10, VPS = 24, and BF = 5) were generated.

Soil samples were divided into 5-fold CV 20 times, and before each repetition the dataset was randomly shuffled, and new folds were generated to increase the robustness of the prediction. The model was trained with 4 folds and evaluated in the validation fold (fold 5), meaning that final was computed from 100 models for each combination of hyperparameters.

Then, optimal values of three primary parameters: NT, VPS, and BF, were selected based on the best regression metrics. Meanwhile, the other three parameters, maximum nodes, minimum leaf population, and seed were set up using the default values with values of null, 1, and 0, respectively.

2.2.7. Model validation and accuracy assessment

To evaluate the performance of the models developed, we conducted accuracy assessments. The coefficient of determination (R^2), the Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were used to compare the accuracy of different models (Chai and Draxler, 2014; Legates and McCabe, 1999; Willmott and Matsuura, 2005). More specifically, the R^2 was used to measure the variation between the measured and predicted soil parameters evaluated; the RMSE was used to assess the magnitude of error between the measurements and the predicted soil parameters and MAE was used to measure the average magnitude of errors in a set of predictions, without considering their direction. MAE and RMSE express the average prediction error in units of the variable of interest. The closer R^2 is to 1, and the closer RMSE and MAE are to 0, the better the model fit is.

Multiple decision trees in each model help determine the importance of different variables. This importance is measured using two methods: mean decrease accuracy (MDA) and mean decrease Gini (MDG). MDA calculates the accuracy of each variable by looking at how much the model's accuracy changes when that variable is removed. This is done using the out-of-bag error rate, which is the error made by the model on data it hasn't seen before. Gini impurity measures how "pure" a set of data is. A high Gini impurity means that the data is mixed up, and a low Gini impurity means that the data is mostly in one category (Nguyen et al., 2020). Variable importance of the environmental covariates used in optimized RF models was extracted from the model properties into GEE. The variable importance determines the contribution of each predictor variable to the general regression model (Ishwaran and Kogalur, 2010).

2.2.8. Uncertainty

To quantify the uncertainty between the results of predicted maps we calculated the coefficient of variation between the k-fold maps generated for the selected best model for each element, to reveal areas with strongly differing values.

3. Results

3.1. Supervised land cover classification results

In this study, accuracy assessment was performed using error matrix, resulting in an overall classification accuracy of 89.7 % and kappa coefficient of 0.852. The classification revealed that croplands cover the most extensive area (42,563.46 ha) including annual croplands (66.4 %), and permanent pastures. Urban zones, which include both urban and rural areas, are most concentrated in the south and comprise 12.3 % of the area. Roads accounted for 3.1 % and include both urban and rural zones. Forests, consisting of eucalyptus, pine and native wood species covered 7.6 %, while shrublands, distributed around the valley in piedmont areas, made up 3.9 %. Water bodies, including lakes and rivers, accounted for 1.5 %, while sandbars are distributed along rivers, covering 1.8 %. Finally, wetlands, located around lakes and rivers, comprised 1.7 % of the total area of the valley included in this analysis. This analysis provided a baseline reference to consider only the cropland areas sampled for the elements mapped over the period examined (Fig. 4).

3.1.1. Summary statistics descriptive of the measured elements in soil

The statistical information for the analyzed soil elements is shown in Table 2. The most abundant elements were iron (Fe), calcium (Ca), and aluminium (Al) with concentrations exceeding 19,000 mg/kg. Conversely, the least abundant elements were beryllium (Be), mercury (Hg), silver (Ag), thallium (Tl) and selenium (Se) with concentrations below 1 mg/kg. Alkali metals show moderate variability and has low correlation between them, however, has moderate correlation with Alkali-earth metals (Fig. S1). Transition Metals and Post-transition (alkali) metals showed moderate variability and had low correlation between them but exhibited moderate correlation with alkali-earth metals (Fig. S1).

Transition metals and post-transition metals were positively

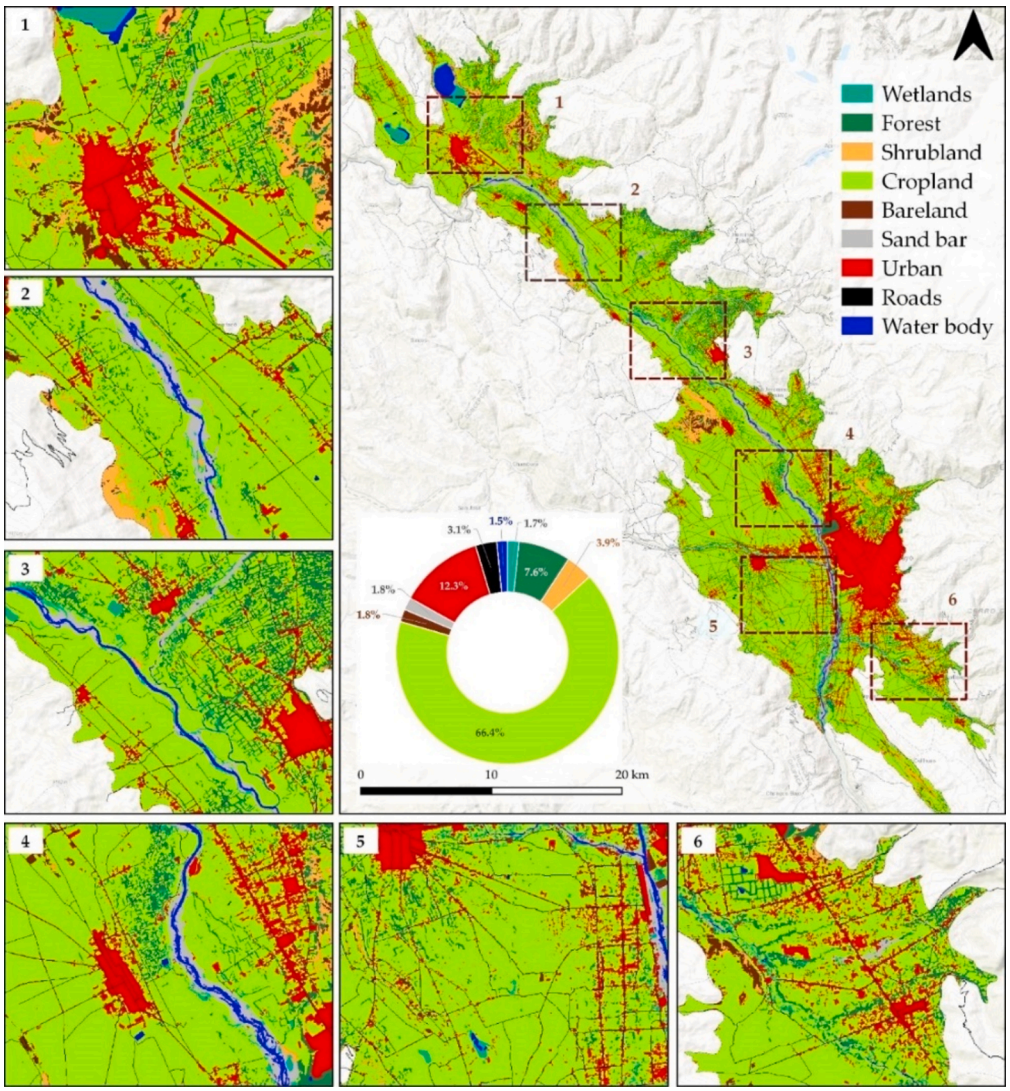


Fig. 4. Land cover map of Mantaro Valley, Peru, based on a supervised classification algorithm for the 2022 – 2023 period.

correlated, with moderate to high correlation, particularly among Ag, Pb, Zn, and Cu. Metalloids demonstrated a higher correlation with transition metals. Cobalt (Co), iron (Fe), chromium (Cr), beryllium (Be), nickel (Ni), aluminium (Al), potassium (K), sodium (Na), barium (Ba), and vanadium (V) had a coefficient of variation (CV) of less than 50 %. Magnesium (Mg), manganese (Mn), and thallium (Tl) had a CV between 5 % and 100 %, while the other elements had a CV greater than 100 %. Both CV and standard deviation (Std) indicate the degree of variability, with CV being independent of the number of dimensions. The majority of soil elements analyzed showed high variability, reflecting the significant variation of soil-forming factors in the study region.

3.1.2. Correlation analysis between model predictors and mapped soil elements

Out of the 24 environmental covariates chosen to model the soil content of 25 elements, the Pearson's correlation coefficients (r) were calculated among these to detect multicollinearity among the input covariates using the corrplot library (Wei and Simko, 2017) in the R environment (R Core Team, 2021) (Fig. 5). Most variables exhibited low, but significant correlation with the analyzed elements. Nevertheless, there were noticeable negative correlations between mineral content and elevation, as well as climatic covariates. Negative correlations between mineral content and elevation may indicate that certain minerals are more concentrated in lower-lying areas due to processes like runoff and sediment deposition, while climatic factors such as temperature and precipitation can influence soil chemical properties and mineral content through processes like weathering, leaching, and organic matter decomposition. Overall, multiple covariates showed correlations with the elements of interest, suggesting that incorporating these covariates as predictors could improve the model's performance to varying extents.

3.1.3. Analysis of modeling results

The assessment of the best RF hyper-parameters selected for each element analyzed, based on environmental covariates, is presented in Table 3. The VPS parameter variable for each element is always greater than the square root of the number of covariates which is the default value provided for RF in GEE.

There is no clear correspondence between NT and VPS, but the VPS are usually smaller when the model uses less than 100 trees. All final models show better performance when trained with a Bag Fraction of 0.8.

The chosen models exhibited satisfactory predictive capabilities for all the analyzed elements, with marked gains observed when incorporating more trees and VPS for certain elements. We selected the best models for each element to generate spatial distribution maps based on these results.

Table 2
Descriptive statistics of all elements analyzed (mg/kg).

Type	Element	Mean	Median	Min	Max	Std	C.V (%)
Alkali metals	**Na**	190.121	159.950	65.960	422.420	87.412	45.98
	K	2,067.770	1,813.580	956.790	5,279.980	910.816	44.05
Alkali-earth metals	**Ba**	159.418	138.370	56.750	485.570	77.360	48.53
	Be	0.944	0.920	0.220	2.110	0.359	38.05
	Ca	22,987.902	8,041.850	863.520	159,196.410	28,872.454	125.60
	Mg	4,743.025	4,208.620	815.540	14,072.900	2,882.404	60.77
	Sr	64.807	35.760	5.400	479.720	81.436	125.66
Transition Metals	**Ag**	0.628	0.100	0.002	9.200	1.527	243.26
	Co	10.245	10.210	3.630	17.520	2.360	23.04
	Cr	21.589	19.630	10.400	52.160	7.700	35.67
	Cu	83.737	30.340	13.010	980.760	165.675	197.85
	Fe	27,444.919	25,916.790	11,643.700	71,509.240	9,036.092	32.92
	Mn	917.072	738.810	67.270	5,258.920	725.168	79.07
	Mo	1.689	1.100	0.360	12.210	1.768	104.64
	Ni	20.394	18.490	9.990	68.020	8.099	39.71
	V	45.027	40.750	17.720	168.560	22.031	48.93
Post-transition metals	**Al**	19,064.720	16,594.570	8,375.640	46,054.440	7,632.829	40.04
	Cd	1.896	0.620	0.040	21.800	3.541	186.70
	Hg	0.749	0.240	0.001	9.570	1.524	203.40
	Pb	146.317	39.360	10.670	1,674.760	321.796	219.93
	Tl	0.488	0.360	0.040	3.300	0.474	97.06
	Zn	691.053	132.350	26.190	7,638.580	1,546.541	223.79
Metalloid	**As**	61.939	31.800	10.380	477.440	93.159	150.40
	Sb	15.417	1.300	0.460	217.960	45.303	293.86
	Se	0.478	0.001	0.002	4.680	0.832	174.08

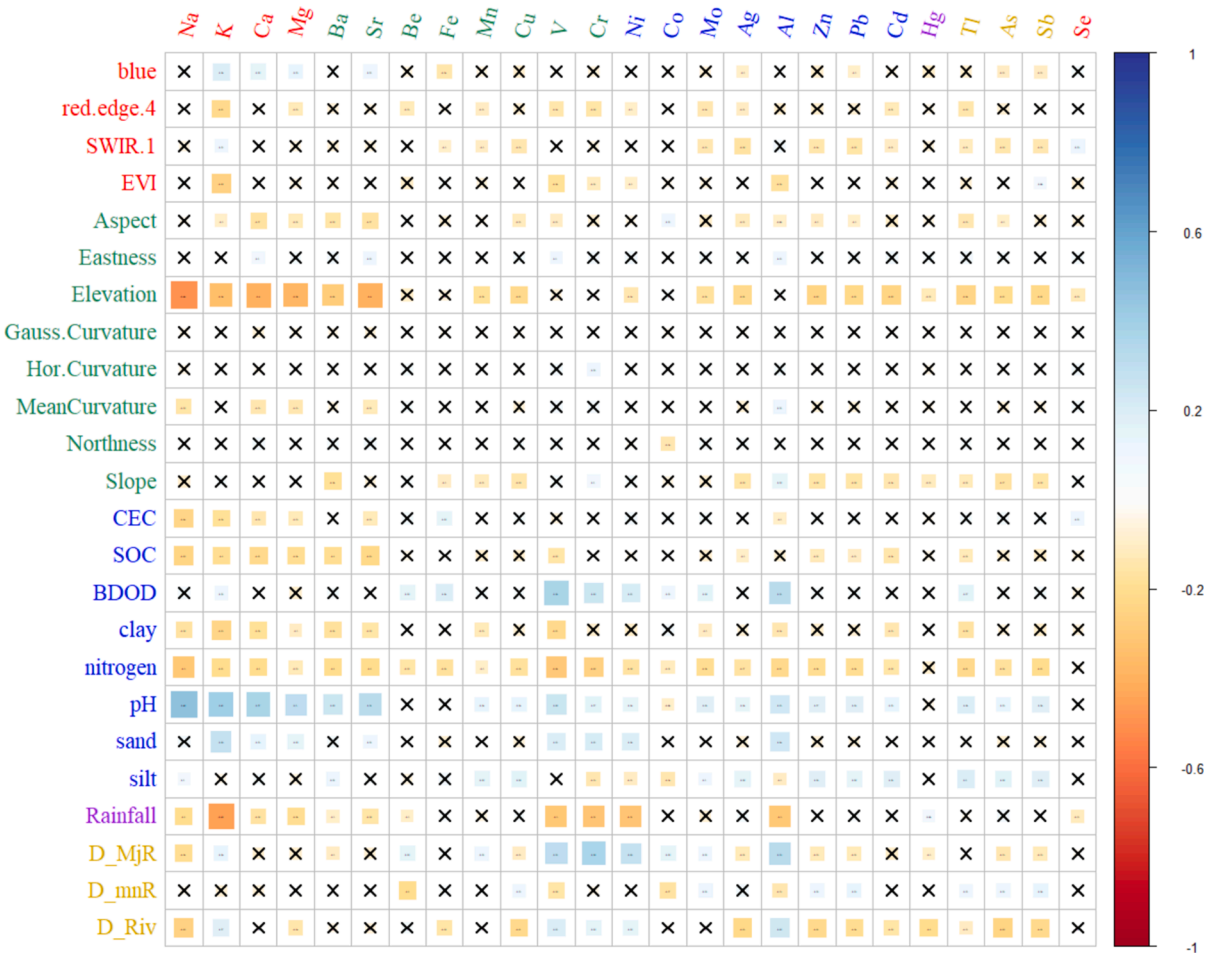


Fig. 5. Correlation coefficients between measured soil elements and model predictors. r—Pearson's correlation coefficient, significant at 5% probability; X—non-significant.

Table 3
Hyper-parameters and regression metrics for each element.

Element	NT	VPS	BF	R-square	RMSE	MAE
Na	100	20	0.8	0.9472	19.1113	11.9315
K	450	13	0.8	0.9456	207.2953	138.0432
Ba	100	9	0.8	0.8787	25.9149	17.0112
Be	250	15	0.8	0.8867	0.1208	0.0845
Ca	150	8	0.8	0.8735	9,969.6157	5,747.9678
Mg	500	12	0.8	0.8896	938.8697	536.7802
Sr	150	8	0.8	0.8485	30.1830	14.9180
Ag	200	22	0.8	0.8771	0.5005	0.2003
Co	150	23	0.8	0.8894	0.7735	0.4795
Cr	400	11	0.8	0.9196	2.1190	1.3852
Cu	200	19	0.8	0.8266	63.6495	25.7100
Fe	50	6	0.8	0.8601	3,264.4926	2,007.7215
Mn	450	23	0.8	0.8424	271.7074	125.4945
Mo	50	8	0.8	0.8391	0.7076	0.3450
Ni	250	16	0.8	0.9432	1.8683	1.3186
V	500	9	0.8	0.9220	6.0616	3.6318
Al	500	17	0.8	0.9664	1,380.8397	990.9776
Cd	400	22	0.8	0.8366	1.3683	0.5076
Hg	50	17	0.8	0.8157	0.6273	0.2903
Pb	200	12	0.8	0.8535	115.0437	48.9777
Tl	50	11	0.8	0.8401	0.1804	0.0860
Zn	150	11	0.8	0.8548	548.3749	245.3823
As	50	15	0.8	0.8583	33.0037	14.2915
Sb	200	13	0.8	0.8605	16.2511	6.1983
Se	150	16	0.8	0.7337	0.4138	0.2382

3.1.4. Prediction results and the relative importance of the predictors

After selecting the best regression models, we created maps of the spatial quantitative distribution of each element (Fig. 6) and correspondent uncertainty map (Figure S2). The 25 maps generated by the selected models for each element show a gradient of element content, with differentiation between the north and south regions of Mantaro Vallery. Additionally, elements such as Na, Ca, Mg, Sr, Ba, Tl, Cd, Zn, Pb, Hg, Ag, Cu, As and Sb, are concentrated along rivers.

3.1.5. Importance of individual covariates

We calculated the relative importance of 24 covariates (note that the importance value has been converted to percentage) for the best-selected models by element and grouped by the type of elements (Figs. 7 - 8). For alkali and alkali − earth metals (Na, K, Mg, Ba, Ca, Be and Sr), the most important covariates were pH, elevation, precipitation and distance to rivers, additionally associated with nitrogen and SOC.

The most important covariates in transition metals (Cr, V, Ni, Ag, Fe, Cu, Mo, Co and Mn) were distance to major roads, precipitation, soil bulk density and distance to major roads, as well elevation, sand and nitrogen content. In post-transition metals (Al, Tl, Cd, Zn, Pb and Hg), the distance to rivers, precipitation, bulk density, and elevation were the most important covariates, especially for Cd, sand and silt content were important.

For metalloids, nitrogen content, distance to rivers and major roads, pH and elevation, were the most important covariates. It is also worth noting that topography-derived covariates (slope, aspect, curvatures) and spectral-band derived covariates (multispectral bands and derived indices), have medium and low importance in all cases. Overall, elevation is consistently a key predictor for all five element groups.

4. Discussion

In this research paper, we present the first-ever maps of alkali, alkali-earth, transition metals, post-transition metals, and metalloids for the croplands of Mantaro Valley one of the most agriculturally productive areas in the Peruvian Andes. We found a positive correlation between some elements, especially heavy metals, indicating an aggregation degree of these elements, similar to the results obtained by Liu et al. (2023) (Liu et al., 2023) and Zhou et al. (2021) (Zhou et al., 2021). These correlations suggest that elements such as Pb, Zn, and Cu may originate from common anthropogenic sources such as mining and industrial activities, consistent with findings from other regions affected by similar activities (Liu et al., 2023; Zhou et al., 2021). These correlations indicate that local industrial activities, mining operations, and agricultural practices may significantly contribute to the observed heavy metal concentrations (Nouri et al., 2009).

High variability of elements across different areas, from north-to-south and near rivers, was observed for all elements analyzed. This can be explained by the complex interaction of variables in soil formation, despite the Mantaro Valley being composed mostly of debris and transported material (Martínez, 1978). Topography affects soil formation through erosion, runoff and infiltration processes, influencing chemical and physical properties and resulting in spatial variability (Wu et al., 2023b). Zgłobicki (Zgłobicki, 2013) found that in areas with high slope exposed to erosion, the concentration of elements in soils was low and more concentrated at the foot of slopes or bottoms of depressions. Climate plays an important role in pedogenetic processes, with events like heavy rainfall transporting elements from abandoned mining dams, geological deposits, biological degradation of organic matter, atmospheric deposition and accelerating industrial and domestic discharge (Custodio et al., 2020). Additionally, soil-forming processes like weathering and organic matter decomposition, coupled with human activities such as agriculture and industrial waste disposal, significantly influence the spatial distribution of these elements. Combined with the natural topography and climatic influences, these processes create a highly heterogeneous distribution of soil elements (Weil and Brady, 2017).

Distance to rivers had high relative importance in mapping the sampled soil elements, although Pearson correlation coefficients showed moderate to low relationships with some elements. This indicates that water used for irrigating croplands can transport elements like heavy metals and other pollutants (Chira et al., 2022), a product of human activities, leading to metal accumulation in areas close to rivers. Rivers can act as conduits for the redistribution of contaminants, facilitating their spread through irrigation practices, flooding events, and sediment transport, which explains the observed elevated concentrations of elements like Cd, Pb, and Zn near riverbanks. Rivers, therefore, play a critical role in the hydrological cycle that redistributes contaminants from both natural and anthropogenic sources across the landscape.

Similarly, distance to roads had medium importance in the mapping models as the roads adjacent to croplands in the Mantaro Valley are typically low traffic. However, it is recognized that heavy metals like Pb often accumulate in surface soils near roads due to oil combustion, which can lead to accumulation in plants growing in those soils (Wang et al., 2020). The accumulation of heavy metals such as Pb in surface soils near roads can also be attributed to vehicular emissions and road maintenance practices, even in areas with low traffic. These contaminants can persist in the environment and affect soil quality and plant health over time. Additionally, other anthropogenic activities, such as agricultural practices and changes in land use management, can contribute to the accumulation of elements, especially heavy metals like Cu found in fungicides (Mouazen et al., 2021). This highlights the need for continuous monitoring and regulation of emissions from vehicles and other transportation sources to mitigate soil contamination.

Other soil elements such as soil organic carbon (SOC), sand, lime and clay content are linked with metals and metalloids (Thomas et al., 2023). The addition of manure can immobilize some elements by forming complexes with metal ions, thus changing the pH and affecting the accumulation of metals in soils (De Temmerman et al., 2003). Soil organic matter, clay content, and other components play crucial roles in metals' adsorption and desorption processes, influencing their mobility and bioavailability in the soil. Organic amendments like manure can enhance metal immobilization by increasing soil pH and forming stable complexes with metal ions. Furthermore, these findings underscore the potential benefits of using organic amendments to manage soil health and reduce metal mobility in contaminated soils (Bolan et al., 2014).

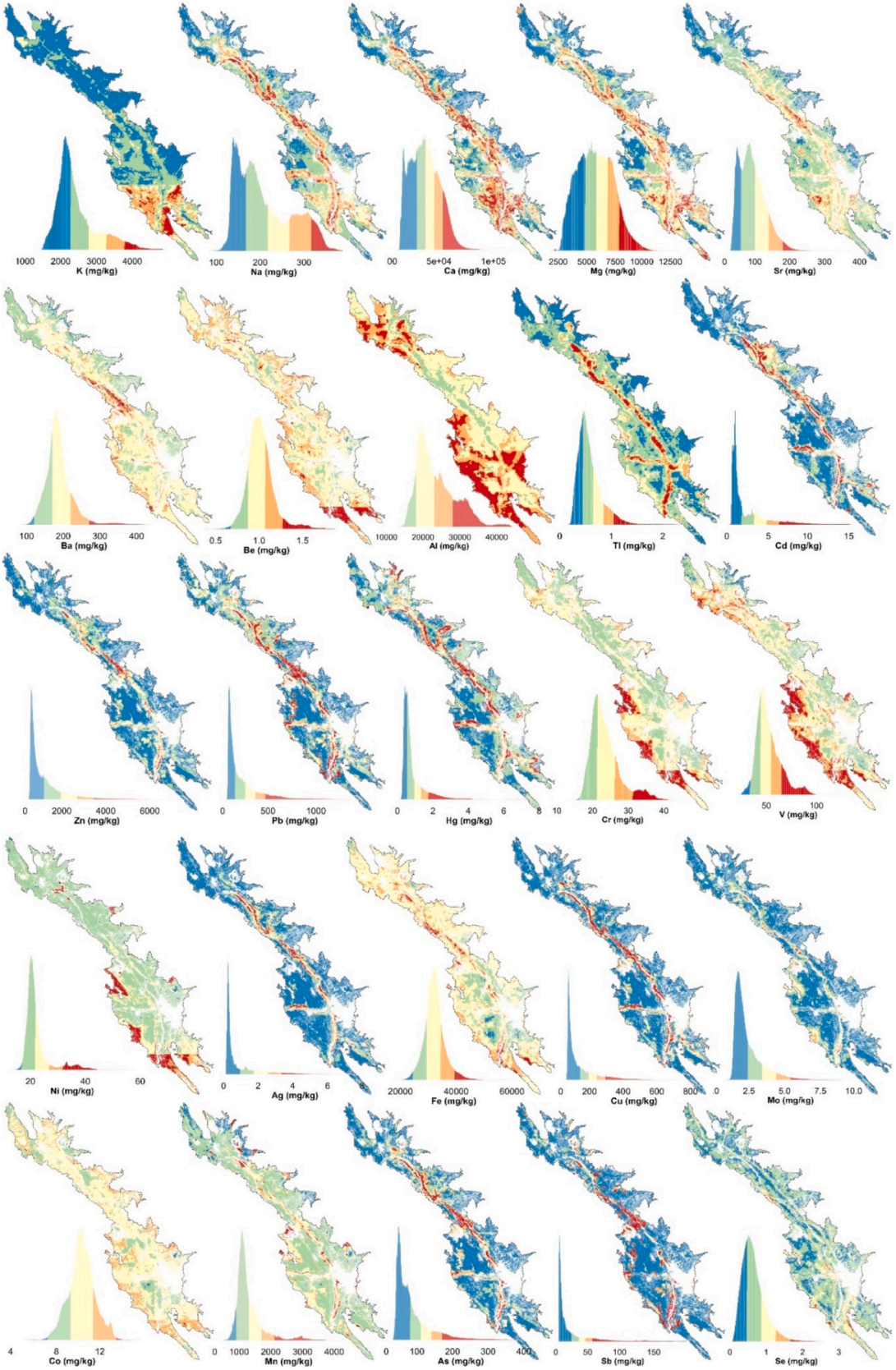


Fig. 6. Spatial distribution maps of selected elements.

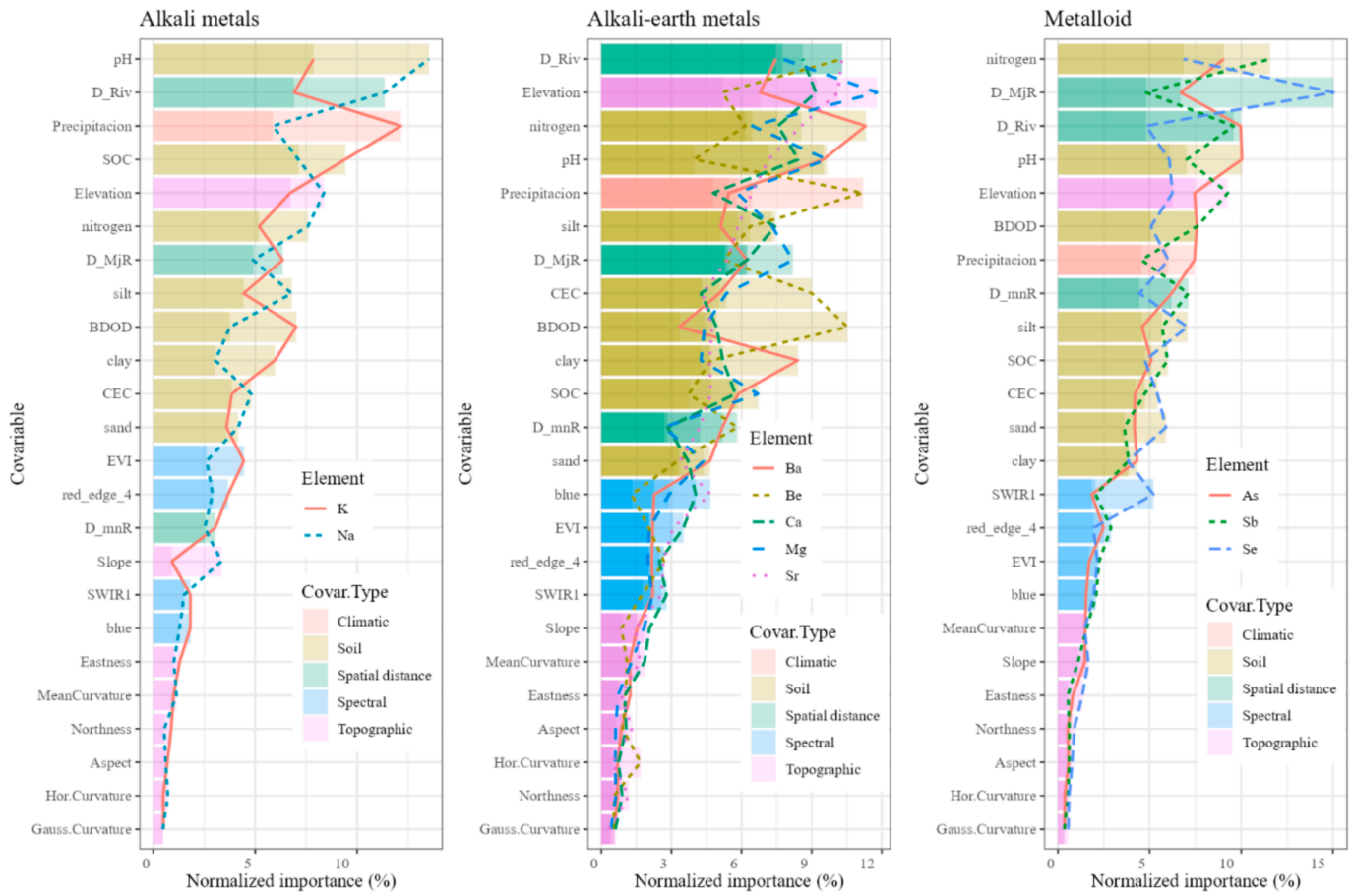


Fig. 7. Relative importance of 41 environmental covariables on the alkali metals (A), alkali-earth metals (B), metalloid (C).

Contrary to our expectations, spectral covariates had relatively low importance because these variables are more sensitive to vegetation growth status, making other variables more important in predicting element content. Spectral data, while useful for monitoring vegetation health and often indicative of soil conditions indirectly, may not capture the subsurface variability and element concentrations effectively. Therefore, direct soil measurements and topographic data provide more accurate insights into soil element distributions. This result indicates that metals and metalloids concentration are highly affected by topography and human activities (Peng et al., 2016) and that while remote sensing is valuable for large-scale monitoring, it should be supplemented with ground-based measurements for accurate soil assessments (Pizarro et al., 2023).

According to the Peruvian environmental standard for agricultural soils (MINAM, 2017), the maximum permissible limit for arsenic (As) is 50 mg/kg, while in this study we report areas with concentrations up to 477.44 mg/kg. For lead (Pb), the maximum permissible limit is 50 mg/kg, and we found areas with concentrations reaching 1,674.76 mg/kg. Cadmium (Cd) reached maximum concentrations of 21.8 mg/kg, above the permissible limit of 1.4 mg/kg. These values are consistent with other studies in Mantaro Valley (Castro-Bedriñana et al., 2023; Custodio and Peñaloza, 2021), which reported high concentration of heavy metals in pastures and croplands. Hence the predicted soil maps can serve as efficient tools for spatially identifying polluted areas around all the agricultural spaces. The elevated levels of As, Pb, and Cd pose significant health risks, including potential toxicity to humans and wildlife (Chirinos-Peinado et al., 2022; Tejada-purizaca et al., 2024), and highlight the need for urgent remediation efforts. These findings underscore the importance of using soil maps for environmental monitoring and policy-making to mitigate pollution impacts (Tchounwou et al., 2012). Moreover, such elevated contaminant levels necessitate immediate intervention strategies to protect public health and the environment, including policy measures to control pollution sources and remediation programs to clean affected soils.

5. Conclusion

This study advances the current state of digital soil mapping by using multiple geospatial datasets, including remote sensing, climate, topographic, soil and distance to map elements in the agricultural soils of the Peruvian Mantaro Valley. Utilizing a Random Forest algorithm and cloud computing in Google Earth Engine, we achieved satisfactory results in predicting the spatial distribution of 25 analyzed elements. Our findings demonstrate that a combination of environmental covariates, primarily soil, climate, topographic, and distance variables, significantly influences the distribution of soil elements, with spectral variables playing a lesser role. The Random Forest modeling approach yielded better predictions with hyperparameter tunning, highlighting the effectiveness of machine learning algorithms in conjunction with ground-truth data augmentation for soil element mapping.

This study is the first to map alkali, alkali-earth, transition metals, post-transition metals, and metalloids for the croplands of Mantaro Valley, revealing critical information on soil contamination levels. Notably, concentrations of arsenic (up to 477.44 mg/kg), lead (up to 1,674.76 mg/kg), and cadmium (up to 21.8 mg/kg) exceed Peruvian environmental standards for agricultural soils, posing significant health risks and emphasizing the need for immediate remediation efforts. These maps enable the identification of polluted areas, primarily near rivers, thereby facilitating targeted interventions to mitigate pollution impacts on public health and the environment. Finally, further research is

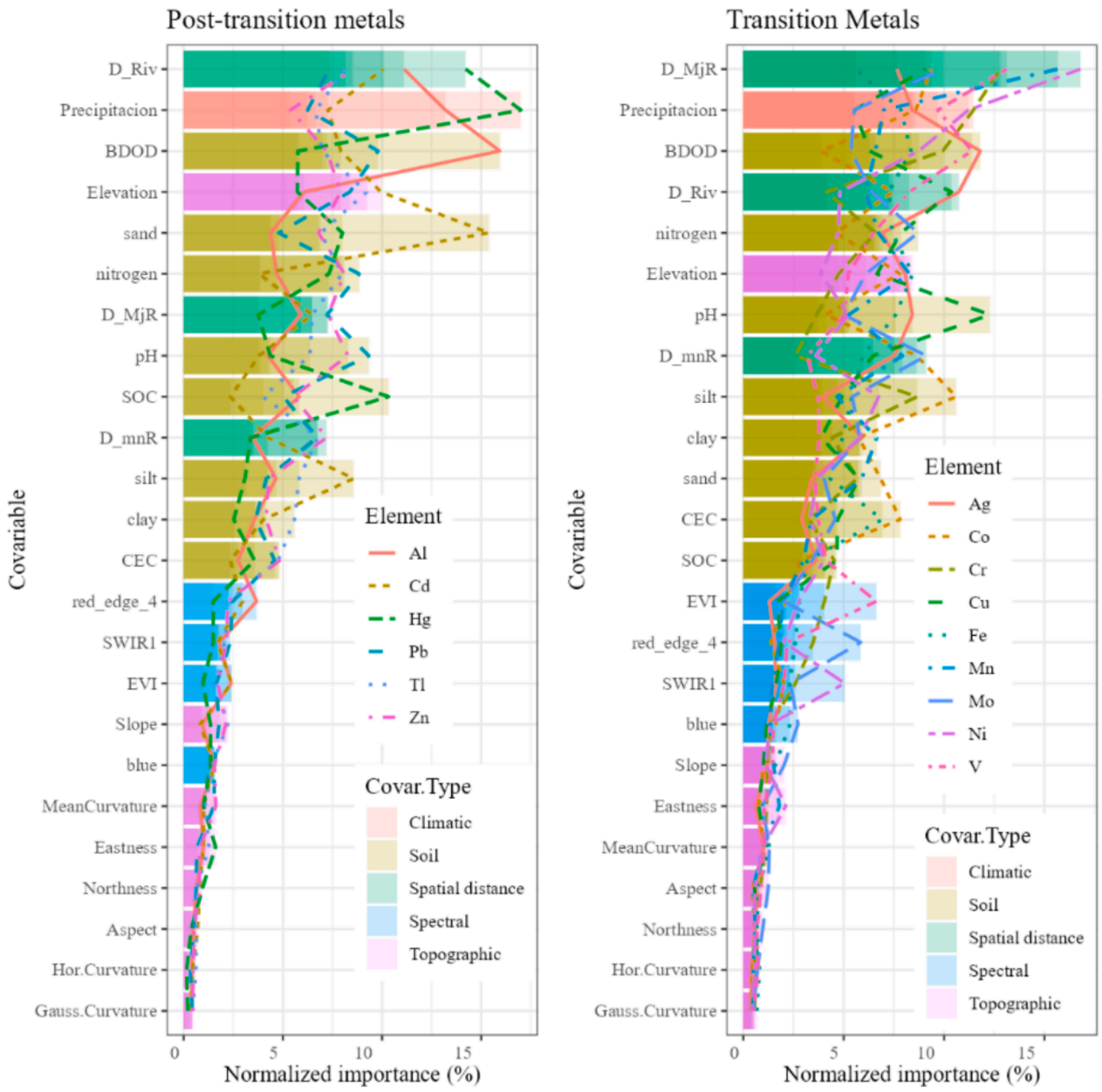


Fig. 8. Relative importance of 24 environmental covariates on post-transition metals (A) and transition metals (B) content.

essential to understand the translocation of toxic metals into food crops, the implications for public health, and the impact of contaminated water sources used for irrigation. Implementing a spatial monitoring scheme based on environmental covariates is feasible and necessary for the continued assessment and management of soil health in the Mantaro Valley and similar regions.

CRediT authorship contribution statement

Samuel Pizarro: . **Narcisa G. Pricope:** Writing – review & editing, Supervision, Formal analysis. **Jesús Vera:** Investigation. **Juancarlos Cruz:** Funding acquisition. **Sphyros Lastra:** Supervision, Formal analysis. **Richard Solórzano-Acosta:** Supervision, Funding acquisition. **Patricia Verástegui Martínez:** Supervision, Methodology, Investigation, Conceptualization.

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.

Acknowledgements

This research was funded by the INIA project “Mejoramiento de los servicios de investigación y transferencia tecnológica en el manejo y recuperación de suelos agrícolas degradados y aguas para riego en la pequeña y mediana agricultura en los departamentos de Lima, Áncash, San Martín, Cajamarca, Lambayeque, Junín, Ayacucho, Arequipa, Puno y Ucayali” CUI 2487112, of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government.

Data availability

Data will be made available on request.

References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Sci. Data 5, 1–12.

Bolan, N., Kunhikrishnan, A., Thangarajan, R., Kumpiene, J., Park, J., Makino, T., Kirkham, M.B., Scheckel, K., 2014. Remediation of heavy metal(loid)s contaminated soils - To mobilize or to immobilize? J. Hazard. Mater. 266, 141–166.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Brevik, E.C., Sauer, T.J., 2015. The past, present, and future of soils and human health studies. Soil 1, 35–46.

Brevik, E.C., Slaughter, L., Singh, B.R., Steffan, J.J., Collier, D., Barnhart, P., Pereira, P., 2020. Soil and Human Health: Current Status and Future Needs. Air. Soil Water Res. 13.

Castro-Bedriñana, J., Chirinos-Peinado, D., Ríos-Ríos, E., Castro-Chirinos, G., Chagua-Rodríguez, P., De La Cruz-Calderón, G., 2023. Lead, Cadmium, and Arsenic in Raw Cow's Milk in a Central Andean Area and Risks for the Peruvian Populations. Toxics 11, 1–19.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.

Chen, T., Chang, Q., Liu, J., Clevers, J.G.P.W., Kooistra, L., 2016. Identification of soil heavy metal sources and improvement in spatial mapping based on soil spectral information: A case study in northwest China. Sci. Total Environ. 565, 155–164.

Chira, J., Vargas, L., Calderón, C., Arcos, F., Mogrovejo, M., De La Cruz, C., 2022. Heavy metals and their impact on surface waters of the Mantaro river basin, Junin. Peru. Int. J. Hydrol. 6, 88–93.

Chirinos-Peinado, D., Castro-Bedriñana, J., Ríos-Ríos, E., Mamani-Gamarra, G., Quijada-Caro, E., Huacho-Jurado, A., Nuñez-Rojas, W., 2022. Lead and Cadmium Bioaccumulation in Fresh Cow's Milk in an Intermediate Area of the Central Andes of Peru and Risk to Human Health. Toxics 10.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Cornu, J.Y., Denaix, L., Lacoste, J., Sappin-Didier, V., Nguyen, C., Schneider, A., 2016. Impact of temperature on the dynamics of organic matter and on the soil-to-plant transfer of Cd, Zn and Pb in a contaminated agricultural soil. Environ. Sci. Pollut. Res. 23, 2997–3007.

Custodio, M., Álvarez, D., Cuadrado, W., Montalvo, R., Ochoa, S., 2020. Potentially toxic metals and metalloids in surface water intended for human consumption and other uses in the Mantaro River watershed. Peru. Soil Water Res. 15, 237–245.

Custodio, M., Peñaloza, R., 2021. Evaluation of the distribution of heavy metals and arsenic in inland wetlands (Peru) using multivariate statistical methods. Ecol. Eng. Environ. Technol. 22, 104–111.

Custodio, M., Peñaloza, R., Ochoa, S., Cuadrado, W., 2021. Human risk associated with the ingestion of artichokes grown in soils irrigated with water contaminated by potentially toxic elements, Junin. Peru. Saudi J. Biol. Sci. 28, 5952–5962.

Danielsson, P.E., 1980. Euclidean distance mapping. Comput. Graph. Image Process. 14, 227–248.

De Temmerman, L., Vanongeval, L., Boon, W., Hoenig, M., Geypens, M., 2003. Heavy metal content of arable soils in Northern Belgium. Water. Air. Soil Pollut. 148, 61–76.

U.S. EPA, 1992. Method 3050A, acid digestion of sediments, sludges, and soils,"EPA Test Methods for Evaluating Solid Waste, Volume IA. Washington, D.C.

U.S. EPA, 2014. Method 6020B (SW-846): Inductively Coupled Plasma-Mass Spectrometry. Washington, DC.

Fujimoto, A., Miyaura, R., Ugas, R., 2004. Cultivation Practices and Economics of the Major Crops in a Central Andean Village, Peru : A Case Study of Pucara in Junin Province in Mantaro Valley. Jour. Agri. Sci. Tokyo Univ. of Agric. 49, 1–16.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. 218, 89–103.

Gitelson, A., Merzlyak, M.N., 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. J. Plant Physiol. 143, 286–292.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Graham, R.D., Welch, R.M., Saunders, D.A., Ortiz-Monasterio, I., Bouis, H.E., Bonierbale, M., de Haan, S., Burgos, G., Thiele, G., Liria, R., Meisner, C.A., Beebe, S. E., Potts, M.J., Kadian, M., Hobbs, P.R., Gupta, R.K., Twomlow, S., 2007. Nutritious Subsistence Food Systems. Adv. Agron. 92, 1–74.

Hengl, T., De Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km - Global soil information based on automated mapping. PLoS One 9.

Huete, A.R., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Instituto Geofísico del Perú, 2005. Atlas climático de precipitación y temperatura del aire en la Cuenca del río Mantaro. Lima - Perú.

Ishwaran, H., Kogalur, U.B., 2010. Consistency of Random Survival Forests. Stat. Probab. Lett. 80, 1056–1064.

Kopittke, P.M., Menzies, N.W., Wang, P., McKenna, B.A., Lombi, E., 2019. Soil and the intensification of agriculture for global food security. Environ. Int. 132, 105078.

Lachaud, A., Adam, M., Mišković, I., 2023. Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. Minerals 13.

Lal, R., 2020. Managing soils for resolving the conflict between agriculture and nature: The hard talk. Eur. J. Soil Sci. 71, 1–9.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.

Liu, Q., Du, B., He, L., Zeng, Y., Tian, Y., Zhang, Z., Wang, R., Shi, T., 2023. Digital soil mapping of heavy metals using multiple geospatial data: Feature identification and deep neural network. Ecol, Indic, p. 154.

Martínez, A., 1978. Estudio de la geología regional de los valles del Mantaro y Tarma. Proyecto especial Programa nacional de pequeñas y medianas irrigaciones plan M.E. R I.S, Lima - Peru.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Minam, 2017. Estándares de Calidad Ambiental (ECA) para Suelo. El Peruano, El Peruano.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311.

Ministerio del Ambiente (MINAM), 2015. Mapa Nacional de Cobertura Vegetal - Memoria descriptiva [WWW Document]. accessed 3.15.23. https://www.gob.pe/institucion/minam/informes-publicaciones/2674-mapa-nacional-de-cobertura-vegetal-memoria-descriptiva.

Moradpour, S., Entezari, M., Ayoubi, S., Karimi, A., Naimi, S., 2023. Digital exploration of selected heavy metals using Random Forest and a set of environmental covariates at the watershed scale. J. Hazard, Mater, p. 455.

Mouazen, A.M., Nyarko, F., Qaswar, M., Tóth, G., Gobin, A., Moshou, D., 2021. Spatiotemporal prediction and mapping of heavy metals at regional scale using regression methods and landsat 7. Remote Sens. 13, 1–23.

Munive, R., Gamarra, G., Munive, Y., Puertas, F., Valdiviezo, L., Cabello, R., 2020. Lead and cadmium uptake by sunflower from contaminated soil and remediated with organic amendments in the form of compost and vermicompost. Sci. Agropecu. 11, 177–186.

Nguyen, H.T.T., Doan, T.M., Tomppo, E., McRoberts, R.E., 2020. Land use/land cover mapping using multitemporal sentinel-2 imagery and four classification methods-A case study from Dak Nong. Vietnam, Remote Sens, p. 12.

Nouri, J., Khorasani, N., Lorestani, B., Karami, M., Hassani, A.H., Yousefi, N., 2009. Accumulation of heavy metals in soil and uptake by plant species with phytoremediation potential. Environ. Earth Sci. 59, 315–323.

Omondi, E., Boitt, M., 2020. Modeling the Spatial Distribution of Soil Heavy Metals Using Random Forest Model—A Case Study of Nairobi and Thirirka Rivers' Confluence. J. Geogr. Inf. Syst. 12, 597–619.

Pearson, R.L., Miller, L.D., 1972. In: Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie, Pawnee National Grasslands, Colorado. and Natural Resources-Colorado State University, Colorado, pp. 1355–1379.

Peng, Y., Kheir, R.B., Adhikari, K., Malinowski, R., Greve, M.B., Knadel, M., Greve, M.H., 2016. Digital mapping of toxic metals in qatari soils using remote sensing and ancillary data. Remote Sens. 8, 1–19.

Peng, Y., Wang, L., Zhao, L., Liu, Z., Lin, C., Hu, Y., Liu, L., 2021. Estimation of soil nutrient content using hyperspectral data. Agric. 11.

Pizarro, S.E., Pricope, N.G., Vargas-Machuca, D., Huanca, O., Ñaupari, J., 2022. Mapping Land Cover Types for Highland Andean Ecosystems in Peru Using Google Earth Engine. Remote Sens. 14.

Pizarro, S., Pricope, N.G., Figueroa, D., Carbajal, C., Quispe, M., Vera, J., Alejandro, L., Achallma, L., Gonzalez, I., Salazar, W., Loayza, H., Cruz, J., Arbizu, C.I., 2023. Implementing Cloud Computing for the Digital Mapping of Agricultural Soil Properties from High Resolution UAV Multispectral Imagery. Remote Sens. 15, 3203.

Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126.

R Core Team, 2021. R: A Language and Environment for Statistical Computing.

Rashid, A., Schutte, B.J., Ulery, A., Deyholos, M.K., Sanogo, S., Lehnhoff, E.A., Beck, L., 2023. Heavy Metal Contamination in Agricultural Soil: Environmental Pollutants Affecting Crop Health. Agronomy 13, 1–30.

Rouse, J.., Haas, R.., Schell, J.., Deering, D.., 1974. Monitoring vegetation systems in the Great Plains with ERTS. In: Proceedings of Third Earth Resources Technology Satellite Symposium. Remote Sensingcenter, Texas A&M hivemity, Colfegp Station, Texas, Washington, DC, p. 309.

Safanelli, J.L., Poppiel, R.R., Chimelo Ruiz, L.F., Bonfatti, B.R., de Oliveira Mello, F.A., Rizzo, R., Demattê, J.A.M., 2020. Terrain analysis in Google Earth Engine: A method adapted for high-performance global-scale analysis. ISPRS Int. J, Geo-Information, p. 9.

Sarkar, B., Mukhopadhyay, R., Ramanayaka, S., Bolan, N., Ok, Y.S., 2021. The role of soils in the disposition, sequestration and decontamination of environmental contaminants. Philos. Trans. R. Soc. B Biol, Sci, p. 376.

Shi, T., Hu, X., Guo, L., Su, F., Tu, W., Hu, Z., Liu, H., Yang, C., Wang, J., Zhang, J., Wu, G., 2021. Digital mapping of zinc in urban topsoil using multisource geospatial data and random forest. Sci. Total Environ. 792, 148455.

Silver, W.L., Perez, T., Mayer, A., Jones, A.R., 2021. The role of soil in the contribution of food and feed. Philos. Trans. R. Soc. B Biol, Sci, p. 376.

Steffan, J.J., Brevika, E.C., Burgessa, L.C., Cerdà, A., 2018. The effect of soil on human health: an overview. Eur J Soil Sci. 69, 159–171.

Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., Zeraatpisheh, M., Amirian-Chakan, A., Triantafilis, J., 2019. Digital mapping of soil classes using ensemble of models in Isfahan Region. Iran. Soil Syst. 3, 1–21.

Tahat, M.M., Alananbeh, K.M., Othman, Y.A., Leskovar, D.I., 2020. Soil Health and Sustainable Agriculture. Sustain. 12, 1–26.

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Molecular, clinical and environmental toxicicology Volume 3: Environmental Toxicology, Molecular, Clinical and Environmental Toxicology.

Tegtmeier, E.M., Duffy, M.D., 2004. External Costs of Agricultural Production in the United States. Int. J. Agric. Sustain. 2, 1–20.

Tejada-purizaca, T.R., Garcia-chevesich, P.A., Ticona-quea, J., Mart, G., Mart, K., Morales-paredes, L., Romero-mariscal, G., Arenazas-rodr, A., Vanzin, G., Sharp, J.O., Mccray, J.E., 2024. Heavy Metal Bioaccumulation in Peruvian Food and Medicinal Products.

Theocharopoulos, S.P., Wagner, G., Sprengart, J., Mohr, M.E., Desaules, A., Muntau, H., Christou, M., Quevauviller, P., 2001. European soil sampling guidelines for soil pollution studies. Sci. Total Environ. 264, 51–62.

Thomas, E., Atkinson, R., Zavaleta, D., Rodriguez, C., Lastra, S., Yovera, F., Arango, K., Pezo, A., Aguilar, J., Tames, M., Ramos, A., Cruz, W., Cosme, R., Espinoza, E., Chavez, C.R., Ladd, B., 2023. The distribution of cadmium in soil and cacao beans in Peru. Sci. Total Environ. 881, 163372.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Rev. 155, 198–230.

Wang, F., Guan, Q., Tian, J., Lin, J., Yang, Y., Yang, L., Pan, N., 2020. Contamination characteristics, source apportionment, and health risk assessment of heavy metals in agricultural soil in the Hexi Corridor. Catena 191, 104573.

Wei, T., Simko, V., 2017. Corrplot: Visualization of a Correlation Matrix (Version 0.84).

Weil, R.R., Brady, N.C., 2017. The Nature and Properties of Soils, 15th ed. Pearson Education.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82.

Wu, Y., Zhou, L., Meng, Y., Lin, Q., Fei, Y., 2023a. Influential Topographic Factor Identification of Soil Heavy Metals Using GeoDetector: The Effects of DEM Resolution and Pollution Sources. Remote Sens. 15, 1–21.

Wu, Y., Zhou, L., Meng, Y., Lin, Q., Fei, Y., 2023b. Influential Topographic Factor Identification of Soil Heavy Metals Using GeoDetector : The Effects of DEM Resolution and Pollution Sources. Remote Sens. 15, 1–21.

Zgłobicki, W., 2013. Impact of microtopography on the geochemistry of soils within archaeological sites in SE Poland. Environ. Earth Sci. 70, 3085–3092.

Zhang, W., Liu, M., Li, C., 2020. Soil heavy metal contamination assessment in the Hun-Taizi River watershed. China. Sci. Rep. 10, 1–10.

Zhou, W., Yang, H., Xie, L., Li, H., Huang, L., Zhao, Y., Yue, T., 2021. Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model. Catena 202, 105222.