Digital soil mapping of metals and metalloids in croplands using multiple geospatial data and machine learning, implemented in GEE, for the Peruvian Mantaro Valley

Samuel Pizarro Carcausto[1], Jesús Vera Vilchez[1], Joseph Huamani[1], Juancarlos Cruz[2], Sphyros Lastra[2], Richard Solórzano-Acosta[2], Patricia Verástegui Martínez[1]

[1] 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; sam20048130@gmail.com (S.P.); patymarve@gmail.com (P.V.); jvera@lamolina.edu.pe (J.V.); josephrolando.2805@gmail.com (J.H.);

[2] Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; jcruz@inia.gob.pe (J.C); slastrapaucar@gmail.com (S.L.); investigacion_labsaf@inia.gob.pe (R.S.);

Abstract

Quality and safety of the soil are essential to ensure social and economic development and provides the supply of contaminant free food. With agriculture intensification, expansion of urban zones, construction of roads, and mining, some agricultural soils sites become polluted increasing environmental risks to ecosystems functions and human health. Hence the need know the spatial distribution of elements in soils, we mapped 25 elements, namely 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, using various geospatial datasets, such as remote sensing, climate, topography, soil data, and distance, to establish the spatial estimation models of spatial distribution trained trough machine learning model with a supervised dataset of 109 topsoil samples, into Google earth engine platform. Using R^2, RMSE and MAE to assess the prediction accuracy. First Random Forest gave satisfactory results in predicting the distribution of analyzed elements in soil, being improved for some elements when adds more trees. Additionally, each element analyzed has a different combination of environmental covariates as predictor, mainly soil, climate, topographic and distance variables especially croplands close to rivers, with less importance for spectral variables. Our results suggest that is possible to identify polluted soils and improved regulations to minimize harm to environmental health and human health, for short-to-medium-term environmental risk control.

Key words: Random Forest, Soil mapping, Google Earth Engine, machine learning; cloud computing

1 Introduction

Soil in conjunction with air and water, as a fundamental natural resources for agriculture, and becomes the base for most food production, providing 98.8% of the daily calorie consumption [1]. Additionally provides a variety of ecosystem services as the major source of nutrients, and the capacity to remove contaminants from water [2]. Is known that one of the main nutritional deficiencies in humans originates with soil

nutrient limitation, and the crops yield increase has led to a decrease nutritional quality like Ca, Fe and Zn [3]. Additionally, soil can affect human health, when is contaminated by nature or human activities, that involves the addition of chemical elements in toxic levels and translocated to foods, even elements that are essential for life [4]. Then quality and safety of the soil are essential to ensure a social and economic development [5].

With agriculture intensification, expansion of urban zones, construction of roads, and mining, some elements are incorporated into soils contaminating through a wide range of mechanisms, by translocation in pollutants, water, air or incorporated by fertilization, and distributed by erosion and infiltration process, which are affected by the topography, soil factors and the climate, generating polluted soils, increasing environmental risks to ecosystems functions and human health [3,6]. Some elements like Manganese (Mn), Zinc (Zn), Copper (Cu), Iron (Fe), Molybdenum (Mo), nickel (Ni), magnesium (Mg), calcium (Ca) and Boron (B), at relative low concentrations can enhance the growth and development of plants and the human body, but concentrations above optimum levels can affect plant negatively plant development [7]. Other elements like Cadmium (Cd), Lead (Pb), Chromium (Cr) and Arsenic (As) are considered toxic and detrimental for plant and human health [2,4].

With increasingly of spatially datasets available from satellite and models, is possible the use of complex analysis using machine learning models to build spatial prediction frameworks, for soil classes, fertility and heavy metals, using multiple source and scales of spatial data, with reasonably well performed results [5,8,9]. The random forest (RF) [10], as a supervised machine-learning algorithm for regression problems has been widely used in multiple applications for soil mapping considering multiple classes of data [11–15] [16], and involves a combination of multiple decision trees, to predict a value by vote system, and makes no assumptions about the data distribution; and can handle scores and continuous variables simultaneously and has good nonlinear data mining capabilities and generalization capabilities [17]. For this purpose the use of complex models, large volume of multiple sources of spatial data, make it challenging, and demands to use cloud computing, and specifically Google Earth Engine (GEE) [18], to reduce time processing and resources with accurate results.

In the Peruvian Mantaro Valley, some sites are recognized as polluted soils with high content of Cd, Pb, Zn, As and Cu, with relation to irrigated lands by Mantaro river [19,20]. Hence becomes necessarily know the spatial distribution of elements in soils, to identify polluted soils and improved regulations, particularly in developing countries, in order to remediate existing contaminated and minimize harm to environmental health and human health, for short-to-medium-term environmental risk control. To sum up, this study explored a framework for digital mapping for elements content in soil of common that includes 25 elements, namely 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. Various geospatial datasets, such as remote sensing, climate, topography, soil data, and distance, were utilized to establish the spatial estimation models of analyzed elements using a dataset consisting of 109 soil samples.

In these context, the aims of this work are (i) generate digital maps of elements in soil, implemented by the Google Earth Engine (GEE) platform using multiple covariable geospatial data, compared with 25

elements content in soils, (ii) assess and compare the accuracy of multiple random forest models to predict elements values based in spatialized covariables and (iii) identify the most important covariables for predicting elements content in soil. First, we present the material and methods section, where we describe the study area, how soil samples were collected, the covariables acquisition and software processing, statistical and spatial analysis, and validation. Then, we show the results, the elements determined by the laboratory, the correlation analysis between elements and covariates, and the evaluation of RF models. Finally, we discuss the results and present the conclusions.

2 Methodology

2.1 Study Site

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

The MV has the largest agricultural land in the Andes, with almost 43,000 ha of croplands, for potato, wheat, corn, onion, garlic, leafy vegetables and livestock. The sowing is recognized as "big growing season" which starts in September and October and harvesting in February and March. The "small growing season" take place from May to August under irrigation, but it represents only the 10% of the croplands. [21]. The climate is characterized by periods of rain between October and March, and a dry season between April to September, with an average 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 [22].

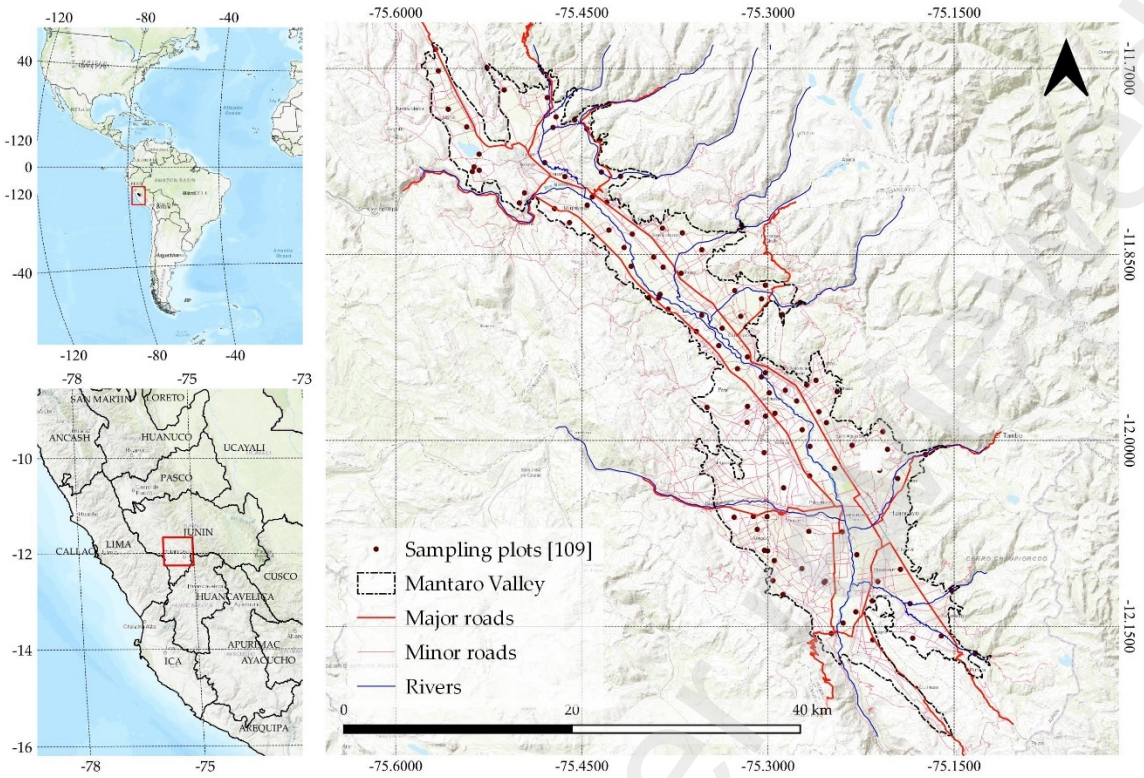

Figure 1. Location of the study area, Mantaro Valley, Junin (Peru).

2.2 Methodological Framework

The methodological framework employed in this study is presented in Figure 2, and described in more detail in the following four methods subsections:

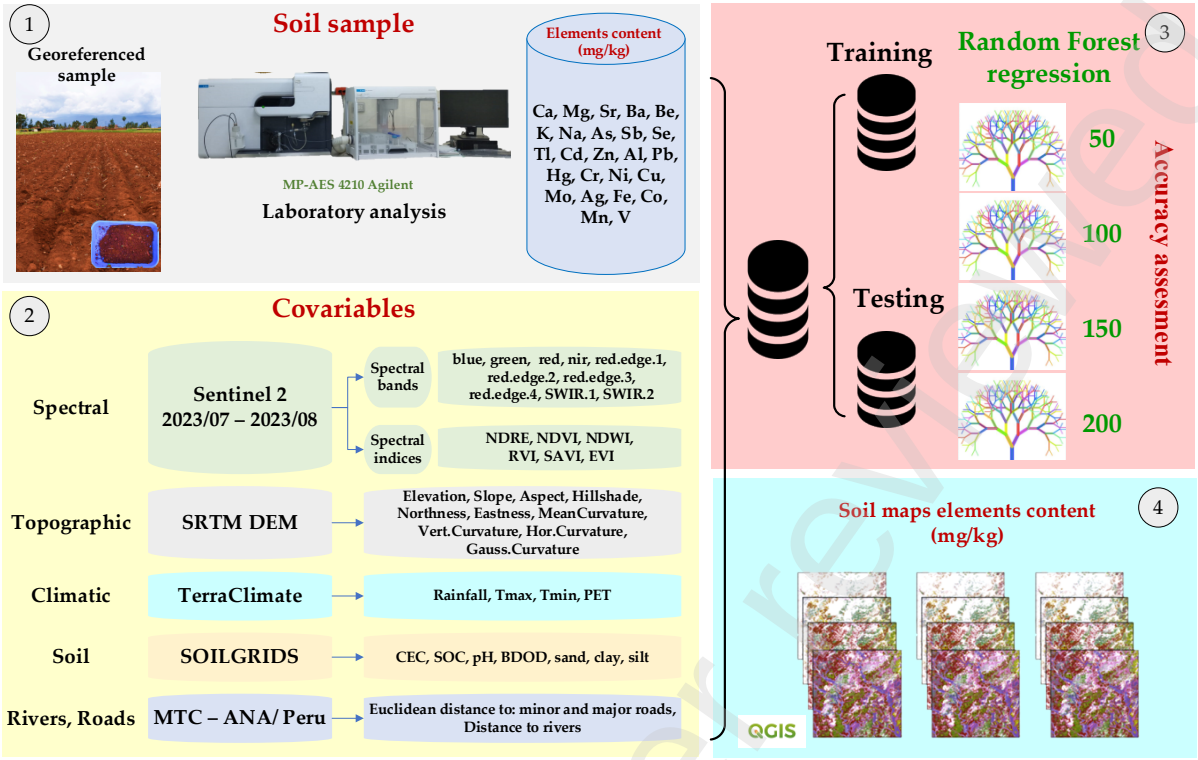

Figure 2. Representation of the methodological framework used in this study.

2.2.1 Land cover Analysis.

To determine de extension of land covers we generate a supervised land cover map based on a Sentinel 2 imagery collection of 2023 April to November, and Random Forest (RF) classification method proposed by Pizarro et al. [23] who combine spectral bands and topographic covariates derived from Digital elevation model (DEM). The land cover training set was composed by nine land cover classes and completed with other reference sources such as the 2015 national vegetation map [24] (croplands, urban zones, barelands, shrublands, wetlands, forest, water bodies and roads) and we added the Sand Bar class beside rivers, using a Google Earth imagery (pan-sharpened QuickBird, GeoEye, and WorldView-2 imagery), and visual interpretation (Figure 2).

We used 9000 pixels samples (1000 for each class) available (80% from field data, and 20% from Google Earth imagery and the 2015 national vegetation map). The samples were split into a validation data set in 30/70% proportion by a stratified random sampling method for each cover class to ensure independence 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, informing how well the map represents what is on the ground and it includes the error of commission which refers to pixels erroneously included in a class [25].

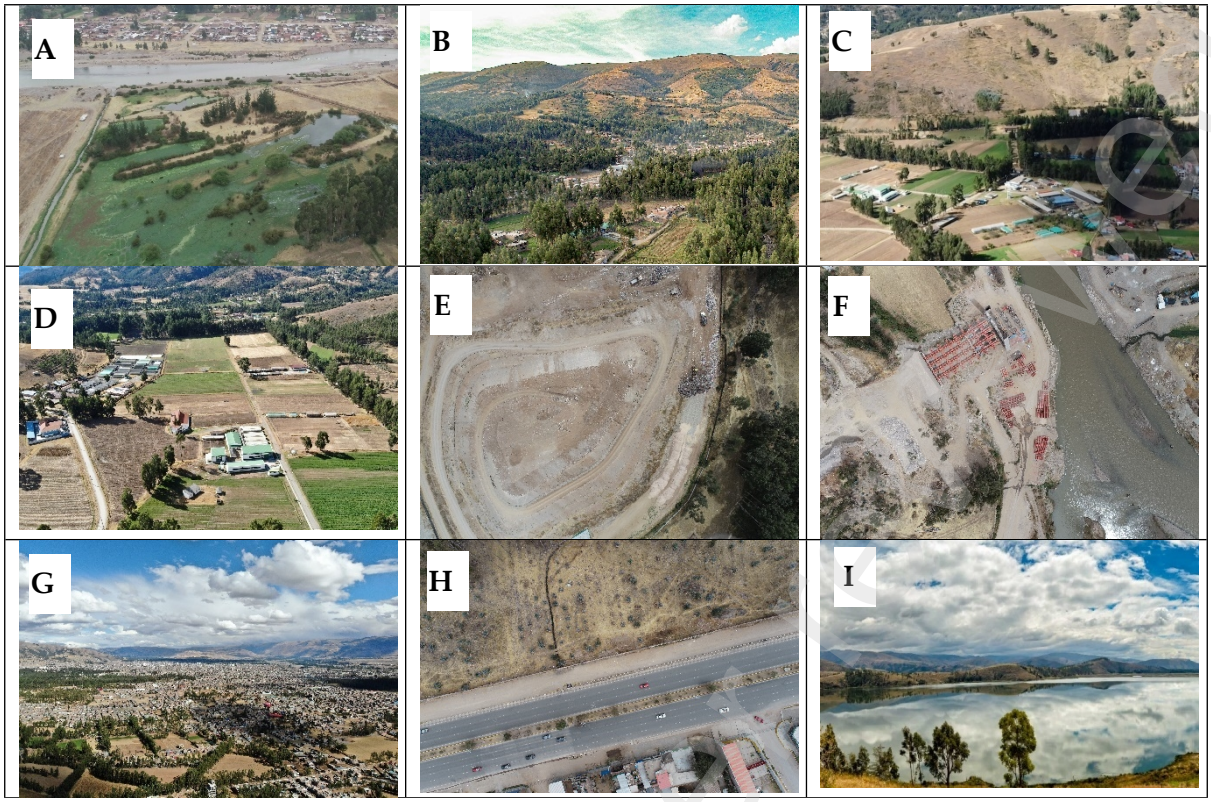

Figure 3. Depiction of land cover classes adopted in the LULC product and found in Mantaro Valley. (A) Wetlands. (B) Forest. (C) Shrubland. (D) Cropland. (E) Bareland. (F) Sand bar. (G) Urban (built up). (H) Roads. (I) Water body.

To generate training data sets for land-cover classes, we used primary data for vegetation land-cover types and completed with other reference sources such as the 2015 national vegetation map (shrublands, water, and bare soil classes), Google Earth imagery (pan-sharpened QuickBird, GeoEye, and WorldView-2 imagery), and visual interpretation.

2.2.2 Field Sampling of Chemical and Physical Soil Parameters

A total of 109 soil samples, composite by four subsamples were collected at 30 cm depth, in agriculture lands along Mantaro valley; each subsample was georeferenced using a Leica Zeno 5 sub metric GPS. In order to ensure randomness, the distance between samples was controlled around 2.5 km.

The metal and metalloids soil contents were assessed by analyzing twenty-five elements, namely 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 at Soil, Water and Foliar Laboratory (LABSAF) of Santa Ana Agricultural Station.

The soil samples after drying at 40 °C by 48 hours, were ground, sieved (2 mm). Then 1 gr of 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 prereduction of arsenic with

potassium iodide (KI) and then by hydride generation. The elements concentrations are expressed in mg / kg.

Finally, two methodologies were used for the analysis of elements: EPA Method 3050A: Acid digestion of sediments, sludge and soils and EPA Method 6020B (SW-846): Inductively Coupled Plasma - Mass Spectrometry, the first was to perform the digestion of each of the samples and the second for the process of analysis of elements content , considering that both methodologies and support equipment, detects the necessary limits for the purposes of analysis.

2.2.3 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 variable

We used Google Earth Engine platform, for build a Sentinel 2 multispectral composed mosaic by median, from the image collection "COPERNICUS/S2_SR", applying a cloud filter for all the scenes, and pan sharpening process for bands B5, B6, B7, B8A, B11 and B12, to the 10 m resolution of bands B2, B3, B4 and B8, for each scene from July to August 2023, for the whole study area.

Additionally, based on previous research like we computed spectral indices, including normalized difference vegetation index (NDVI) [26], normalized difference water index (NDWI) [27], soil-adjusted vegetation index (SAVI) [28] , Enhanced Vegetation Index (EVI) [29], Ratio Vegetation Index (RVI) [30] and Red-edge Normalized Difference Vegetation Index (NDRE) [31], hence the spectral indices have advantages in enhancing soil heavy metals prediction [11,32]. See Table 1 for the calculation formulas.

B. Topographic variable

Topography affects the redistribution of elements in soil [33], therefore the effect of topography factors has to be considered. The topographic variables were derived from Shuttle Radar Topographic Mission (SRTM) digital Elevation model, resampled at 10 m resolution. Trough TAGEE package implemented in GEE [34], we extracted covariates related to terrain that includes slope, aspect, hill shade, eastness, northness, Gaussian curvature, Horizontal curvature, Vertical curvature and Mean Curvature. These variables has been used in other studies in digital soil mapping [9].

C. Climatic variable

Climate has influence on geochemical process in soils and promote soil-to-plant transfer of some elements [35], hence we added climatic variables, including Rainfall, maximum temperature (Temp.min), minimum temperature (Temp.max), and Reference evapotranspiration (PET), were obtained from the TerraClimate repository in GEE [36]. TerraClimate provides monthly gridded climate and climatic data at a spatial resolution of 0.5° from 1958 to 2023. Therefore, the climate data was processed to have average PET, Temp.max and Temp.min, average accumulate Rainfall, for the las 30 years.

D. Soil variable

The soil data were obtained from SoilGrids, 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 in machine learning trained with multiple covariates and soil profile database from the WoSIS [37]. To construct soil variable set, we selected clay, sand, silt, pH, CEC, BDOD, SOC and nitrogen. These soil variables were resampled into 30 m grids resolution.

E. Distance variable.

The major and minor roads inventory was obtained from the Peruvian Transport and communication minister (https://portal.mtc.gob.pe/estadisticas/descarga.html), for the rivers inventory was downloaded from the Peruvian National Water Adminsitration (https://www.geoidep.gob.pe/autoridad-nacional-del-agua-ana). To construct distance variable set, we used the Euclidean distance mapping approach [38]. These variables has been used in other studies in heavy metal in soil mapping [39].

Table 1. All environmental covariates and their abbreviations and sources.

Data type	Covariate	Abbreviation	Source
Spectral	Reflectance of Bands Sentinel 2	B1–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
	Hillshade	Hillshade	Calculated from Elevation
	Eastness	Eastness	Calculated from Elevation
	Northness	Northness	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
	Rainfall	Rainfall	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

2.2.4 Spatial Analysis

The soil sample data were randomly split into training (70%) and validation data (30%). We used a compiled set of the covariates as predictors, and insert it in multiple Random Forest (RF) models from 50 to 200 decision trees, into GEE platform [18] and applied iteratively to each element. We used logic-based machine learning Random Forest model [40] for constructs multiple decision trees that are sampled independently during training, typically improving classification by voting results compared to a single decision tree model. The algorithm makes no assumptions about the data distribution; and can handle scores and continuous variables simultaneously and has good nonlinear data mining capabilities and generalization capabilities [17]. We defined the number of generated decision trees of 100 leaving the other parameters by default. A total of 100 models were built between the combination of predictors and input element data.

2.2.5 Model Validation and Accuracy Assessment

In order to evaluate the performance of the models developed, an accuracy assessment was conducted to evaluate the performance of regression. The coefficient of determination (R^2), the Root Means Square Error (RMSE), and Mean Absolute Error (MAE) were used to compare the accuracy of different models. 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 parameter. MAE and RMSE express the average prediction error in units of the variable of interest. Regarding validation metrics, the closer R^2 is to 1, and the closer RMSE and MAE are to 0, the better the model fit is considered. To select the best model, we used the higher estimation accuracy, and the smaller error by element modeled.

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

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

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (3)$$

where n is the number of samples (individual plot) in the data set, y_i is the measured element, i, y_i i is the predicted element based on the set of covariates, and $\hat{y}_i$ indicates the average of the measured element.

Finally, we used variable importance metrics that consider that every time a split of a node is made on a variable, the impurity criterion for the two descendent nodes is less than the parent node and adding up the decreases for each individual variable over all trees in the forest gives a fast variable importance that is often very consistent with the permutation importance measure. That provided us with an additional method for assessing how each predictor variable enabled accuracy improvements in the optimized soil parameter prediction model, in terms of a normalized percentage contribution.

3 Results

3.1.1 Supervised Classification of Land Cover

In order of importance, croplands have the most extensive area, and includes annual croplands (66.4%), and permanent pastures, urban zones are most concentrated in the south, and includes urban and rural areas (12.3 %), roads include urban zones and rural zones (3.1%). Forests include eucalyptus, pine and native wood species (7.6 %), shrublands area distributed around the valley in piedmont areas (3.9%). Water bodies includes lakes and rivers (1.5%), and the sandbars are distributed along rivers (1.8%), wetlands are located around lakes, and rivers (1.7%). This analysis provided a base reference to consider only the croplands areas sampled for the elements mapped over the period examined (**Figure 4**).

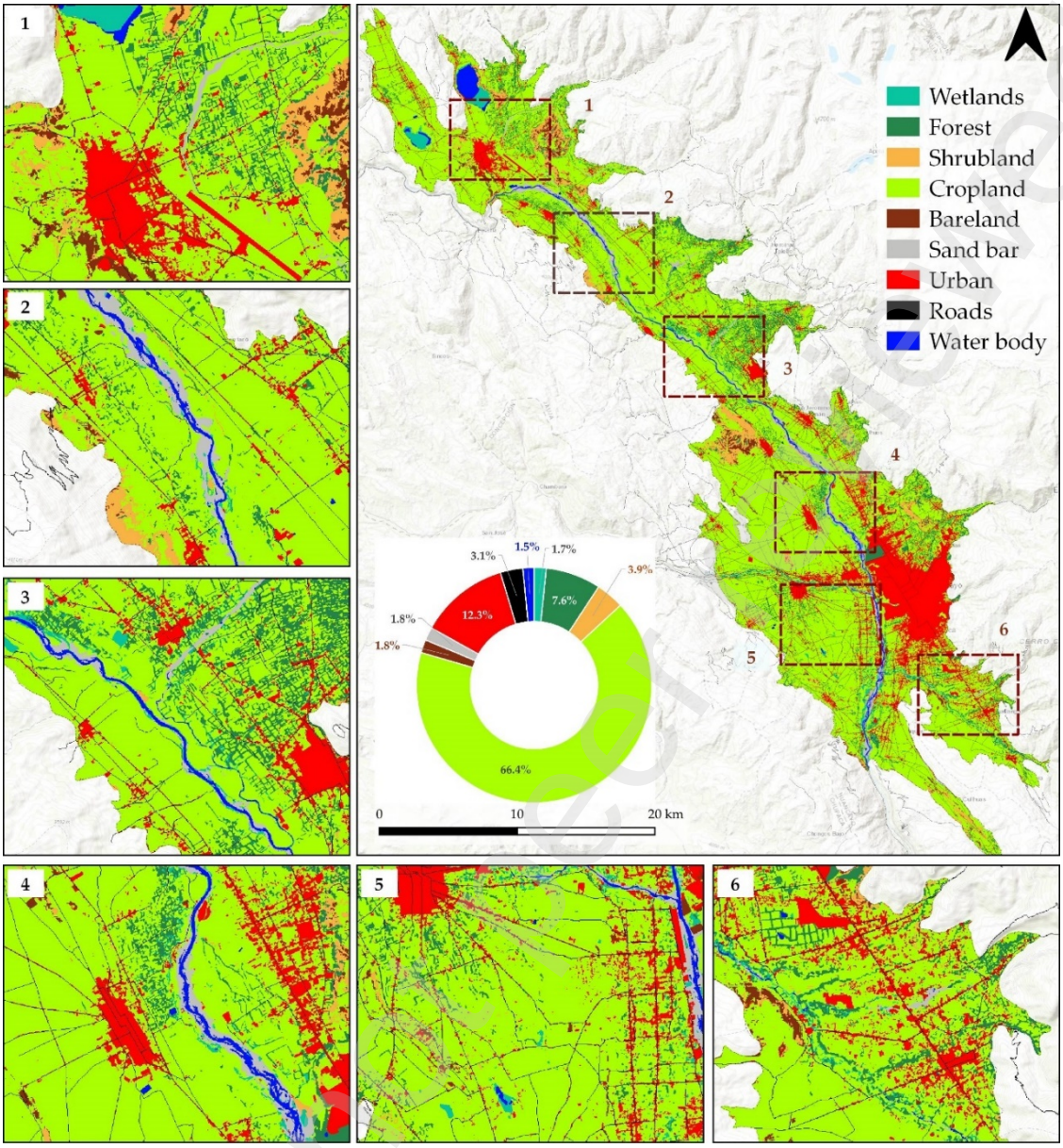

Figure 4. Land cover map of Mantaro Valley, Peru, based on supervised classification for the 2022 – 2023 period.

3.1.2 Summary statistics descriptive of measured elements in soil

The statistical information of elements in soil analyzed are shown in Table 3. The most abundant elements were Fe, Ca, Al with concentrations above 19 000 mg/kg, in the other side the less abundant elements are Be, Hg, Ag, Tl and 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 metals are positively correlated from moderate to high, especially Ag, Pb, Zn and Cu. Metalloids are more correlated with Transition Metals. Co, Fe, Cr, Be, Ni, Al, K, Na, Ba, and V have a CV less than 50%; Mg, Mn and Tl have CV between 5 and 100 %; the other elements have CV higher than 100%. Both CV and Std can

indicate the degree of variability, but the first is not affected by the number of dimensions. The most part of soil elements analyzed showed a high variability reflecting the high variation of soil forming factors in the study region.

Table 3. Descriptive statistics of all elements analyzed (mg/kg).

Type	Element	Mean	Median	Min	Max	Std	C.V (%)
Alkali metals	**Na**	190.12	159.95	65.96	422.42	87.41	45.98
Alkali metals	**K**	2,067.77	1,813.58	956.79	5,279.98	910.82	44.05
Alkali-earth metals	**Ba**	159.42	138.37	56.75	485.57	77.36	48.53
Alkali-earth metals	**Be**	0.94	0.92	0.22	2.11	0.36	38.05
Alkali-earth metals	**Ca**	22,987.90	8,041.85	863.52	159,196.41	28,872.45	125.6
Alkali-earth metals	**Mg**	4,743.03	4,208.62	815.54	14,072.90	2,882.40	60.77
Alkali-earth metals	**Sr**	64.81	35.76	5.4	479.72	81.44	125.66
Transition Metals	**Ag**	0.63	0.1	0	9.2	1.53	243.26
Transition Metals	**Co**	10.24	10.21	3.63	17.52	2.36	23.04
Transition Metals	**Cr**	21.59	19.63	10.4	52.16	7.7	35.67
Transition Metals	**Cu**	83.74	30.34	13.01	980.76	165.68	197.85
Transition Metals	**Fe**	27,444.92	25,916.79	11,643.70	71,509.24	9,036.09	32.92
Transition Metals	**Mn**	917.07	738.81	67.27	5,258.92	725.17	79.07
Transition Metals	**Mo**	1.69	1.1	0.36	12.21	1.77	104.64
Transition Metals	**Ni**	20.39	18.49	9.99	68.02	8.1	39.71
Transition Metals	**V**	45.03	40.75	17.72	168.56	22.03	48.93
Post-transition metals	**Al**	19,064.72	16,594.57	8,375.64	46,054.44	7,632.83	40.04
Post-transition metals	**Cd**	1.9	0.62	0.04	21.8	3.54	186.7
Post-transition metals	**Hg**	0.75	0.24	0	9.57	1.52	203.4
Post-transition metals	**Pb**	146.32	39.36	10.67	1,674.76	321.8	219.93
Post-transition metals	**Tl**	0.49	0.36	0.04	3.3	0.47	97.06
Post-transition metals	**Zn**	691.05	132.35	26.19	7,638.58	1,546.54	223.79
Metalloid	**As**	61.94	31.8	10.38	477.44	93.16	150.4
Metalloid	**Sb**	15.42	1.3	0.46	217.96	45.3	293.86
Metalloid	**Se**	0.48	0	0	4.68	0.83	174.08

3.1.3 Correlation Analysis between Predictors and elements mapped

Out of the 41 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 with the corrplot library [41], in R environment [42]. (**Figure 5**). Most variables exhibited low, but significant correlation with the analyzed elements. Nevertheless, there were noticeable negative

correlations between mineral contents and elevation, as well as climatic covariates. Overall, multiple covariates correlated with the elements of interest, suggesting that incorporating them as predictors could still improve the model's performance to different extents.

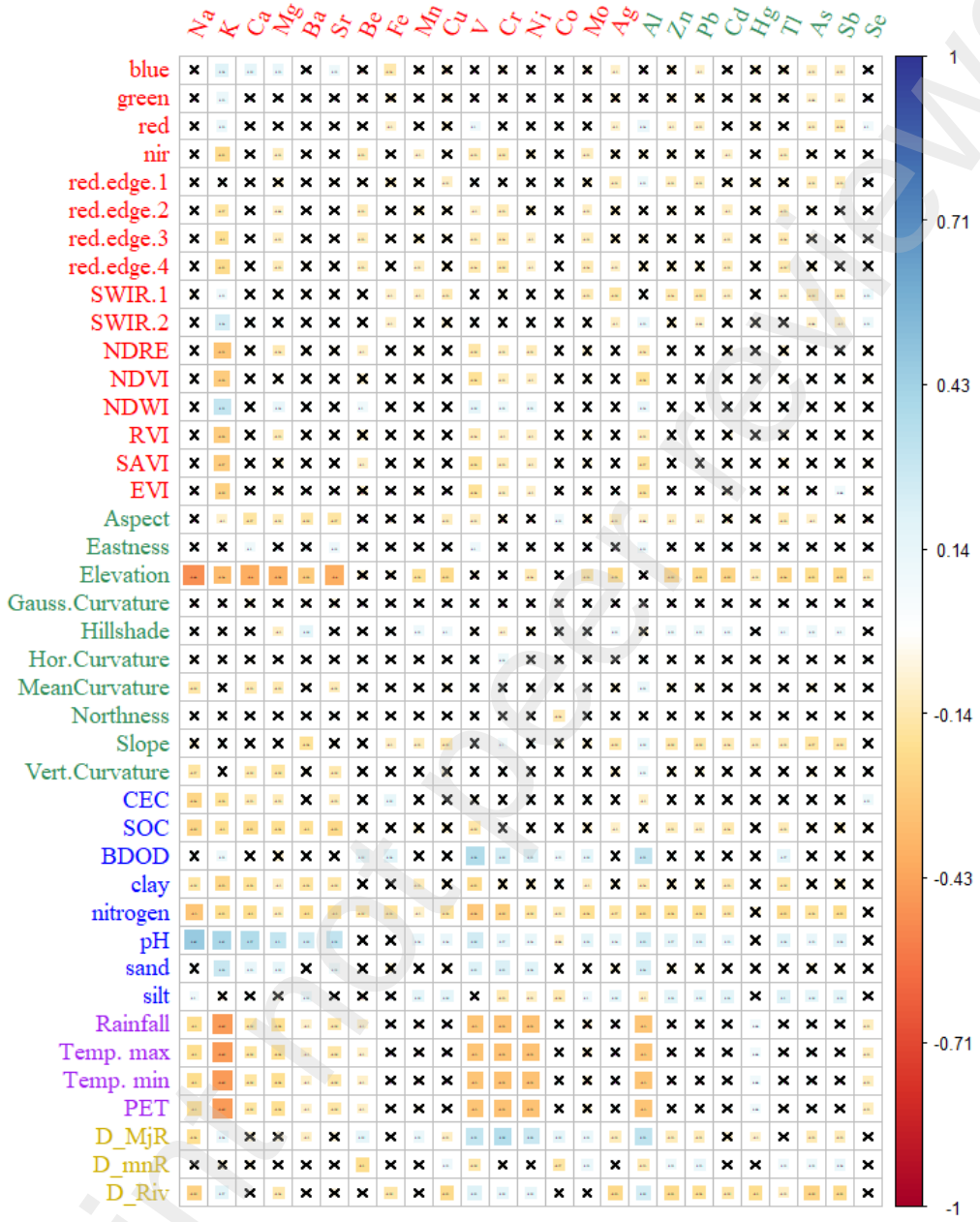

Figure 5. Correlation coefficients between measured soil elements and predictors. r—Pearson's correlation coefficient; Significant at 5% probability; X—non-significant.

3.1.4 Analysis of Modeling Results

The assessment of RF regression models, employing 50, 100, 150, and 200 decision trees, for both training and validation datasets, based on environmental covariates, is presented in Table 4. Overall, most models performed well except for Hg, and Se, where the difference between sets is asymmetric in R-square, RMSE and MAE.

The elements Mo, Sb, Ag, Tl, Ca, Cr, K, Mg, Ni, Sr, As, and Cd had the better performance and smaller error when is trained with 50 decision trees in RF. Fe, Ba, and Co, has better performance when is trained with 100 decision trees. Mn has better performance trained with 150 decision trees. Finally, Pb, Cu, Al, Zn, V, Be, Hg, Na, Se has better performance trained with 200 decision trees. In general, the chosen models exhibited satisfactory predictive capabilities for all the analyzed elements, with a slight advantage observed when incorporating more decision trees for certain elements.. According to the results obtained, we selected the best models for element to generate spatial distribution maps in soil.

Table 4. Evaluation of the prediction effects of the different models in predicting soil properties.

Element	Training				Testing			
	50	100	150	200	50	100	150	200
				R-square				
K	**0.94**	0.93	0.93	0.93	**0.82**	0.82	0.82	0.81
Na	0.93	0.93	0.93	**0.93**	0.85	0.85	0.86	**0.86**
Ba	0.88	**0.89**	0.89	0.89	0.74	**0.73**	0.73	0.73
Be	0.86	0.86	0.86	**0.87**	0.71	0.7	0.71	**0.71**
Mg	**0.90**	0.89	0.89	0.9	**0.82**	0.81	0.81	0.81
Ca	**0.88**	0.88	0.87	0.87	**0.78**	0.77	0.77	0.77
Sr	**0.86**	0.85	0.85	0.85	**0.75**	0.74	0.76	0.76
Hg	0.78	0.78	0.78	**0.79**	0.32	0.36	0.38	**0.39**
Cr	**0.91**	0.90	0.90	0.90	**0.82**	0.8	0.8	0.80
Ni	**0.89**	0.88	0.88	0.88	**0.77**	0.77	0.78	0.77
Cu	**0.85**	0.85	0.85	0.86	**0.71**	0.69	0.69	0.69
Mo	**0.85**	0.83	0.83	0.83	**0.74**	0.73	0.73	0.73
Ag	**0.88**	0.87	0.87	0.87	**0.73**	0.72	0.72	0.71
Fe	0.87	**0.88**	0.87	0.87	0.73	**0.74**	0.73	0.72
Co	0.83	**0.85**	0.85	0.85	0.66	**0.70**	0.7	0.71
Mn	0.81	0.82	**0.83**	0.82	0.65	0.62	**0.61**	0.61
V	0.9	0.9	0.9	**0.90**	0.79	0.78	0.79	**0.79**
Tl	**0.88**	0.87	0.87	0.87	**0.77**	0.74	0.74	0.74
Cd	**0.86**	0.84	0.85	0.85	**0.64**	0.62	0.65	0.65
Zn	**0.86**	0.86	0.86	0.86	**0.72**	0.69	0.70	0.70
Al	0.93	0.93	0.93	**0.93**	0.88	0.86	0.87	**0.87**
Pb	0.85	0.85	0.85	**0.86**	0.69	0.67	0.69	**0.70**
As	**0.87**	0.86	0.86	0.86	**0.69**	0.69	0.7	0.70
Sb	**0.86**	0.86	0.86	0.86	**0.67**	0.66	0.66	0.64
Se	0.76	0.79	0.79	**0.80**	0.48	0.5	0.51	**0.52**
				RMSE				
K	**231.52**	240.63	238.21	241.3	**380.8**	388.41	388.66	393.2
Na	23.52	23.59	22.6	**22.31**	33.39	34.25	33.14	**32.35**
Ba	26.91	**26**	26.13	26.13	39.35	**39.79**	39.91	40.27
Be	0.13	0.13	0.13	**0.13**	0.19	0.2	0.19	**0.19**
Mg	**926.71**	942.03	935.88	927.5	**1236.87**	1250.86	1263.91	1270.95
Ca	**9816.53**	10166.49	10407.64	10490.63	**13659.88**	13786.97	13976.17	13938.81
Sr	**30.58**	31.46	31.25	31.5	**40.7**	41.33	39.52	39.64
Hg	0.71	0.71	0.71	**0.7**	1.25	1.21	1.2	**1.19**
Cr	**2.37**	2.42	2.42	2.42	**3.23**	3.43	3.42	3.42
Ni	**2.73**	2.77	2.79	2.78	**3.85**	3.87	3.76	3.85
Cu	**63.28**	63.56	63.3	62.97	**88.64**	92.5	92.04	92.23
Mo	**0.68**	0.73	0.73	0.73	**0.89**	0.92	0.92	0.92
Ag	**0.53**	0.55	0.54	0.55	**0.8**	0.81	0.8	0.82

Fe	3212.27	**3144.91**	3224.45	3239.06	4682.55	**4615.7**	4716.24	4741.73
Co	0.96	**0.91**	0.91	0.91	1.37	**1.3**	1.3	1.28
Mn	313.49	307.63	**302.07**	305.45	428.34	448.1	**451.48**	452.95
V	7.01	6.98	6.92	**6.83**	10.18	10.26	10	**10.16**
Tl	**0.16**	0.17	0.17	0.17	**0.23**	0.24	0.24	0.24
Cd	**1.31**	1.41	1.38	1.36	**2.11**	2.18	2.1	2.09
Zn	**584.69**	588.02	575.19	574.43	**817.07**	859.48	845.02	848.15
Al	1990.84	2045.32	1983.05	**1963.99**	2619.41	2825.3	2724.79	**2716.83**
Pb	122.85	125.46	123.25	**121.39**	178.81	183.89	178.66	**177.51**
As	**33.54**	34.28	34.2	34.2	**51.59**	52.2	51.26	51.23
Sb	**16.68**	16.75	16.74	16.92	**26.01**	26.54	26.48	26.98
Se	0.4	0.38	0.38	**0.37**	0.6	0.59	0.58	**0.57**
				MAE				
K	**174.29**	175.24	172.73	172.48	**278.47**	276.93	275.08	275.99
Na	18.04	18.19	17.51	**17.31**	25.04	26.32	25.21	**24.84**
Ba	19.49	**18.6**	18.46	18.35	28	**27.93**	27.85	27.92
Be	0.1	0.1	0.1	**0.1**	0.15	0.15	0.15	**0.15**
Mg	**647.87**	632.75	623.8	622.68	**928.92**	906.62	902.88	912.26
Ca	**6720.34**	6790.31	6866.56	6952.69	**10282.84**	10211.56	10302.89	10266.61
Sr	**16.8**	16.85	16.6	16.66	**24.63**	24.58	23.99	24.09
Hg	0.36	0.37	0.36	**0.35**	0.57	0.56	0.55	**0.55**
Cr	**1.7**	1.71	1.72	1.7	**2.41**	2.53	2.53	2.52
Ni	**1.79**	1.79	1.79	1.79	**2.53**	2.6	2.57	2.59
Cu	**32.73**	32.73	32.91	33.01	**48.6**	50.48	50.52	50.47
Mo	**0.37**	0.38	0.38	0.38	**0.54**	0.55	0.54	0.54
Ag	**0.28**	0.29	0.29	0.29	**0.42**	0.45	0.44	0.45
Fe	2067.95	**2055.87**	2093.55	2096.87	3107.56	**3151.03**	3203.67	3209.55
Co	0.66	**0.64**	0.63	0.64	0.96	**0.93**	0.92	0.92
Mn	155.77	154.89	**153.46**	154.39	225.99	239.7	**241.11**	240.87
V	4.33	4.2	4.19	**4.16**	6.31	6.38	6.3	**6.43**
Tl	**0.1**	0.1	0.1	0.1	**0.14**	0.14	0.15	0.14
Cd	**0.63**	0.7	0.68	0.68	**0.96**	1.04	1.02	1.01
Zn	**319.71**	326.04	320.06	317.58	**465.33**	492.94	481.05	476.89
Al	1511.37	1522.51	1487.43	**1467.3**	2063.33	2202.89	2123.66	**2121.09**
Pb	68.76	68.38	67.39	**65.91**	101.28	102.54	99.54	**97.74**
As	**18.37**	18.68	19.01	18.98	**28.44**	28.87	28.67	28.43
Sb	**8.44**	8.55	8.49	8.58	**12.75**	13.21	13.19	13.37
Se	0.27	0.26	0.26	**0.26**	0.4	0.39	0.39	**0.39**

3.1.5 Prediction Results and Relative Importance of the Predictors

Based on the previous results, we selected the best regression models and created maps of the spatial quantitative distribution of each soil property in Figures 6–7. The maps generated by the ten selected models show a gradient concentration of elements differentiated between north and south, and for other elements such as Na, Ca, Mg, Sr, Ba, Tl, Cd, Zn, Pb, Hg, ag, Cu, As and Sb, are concentrated along rivers.

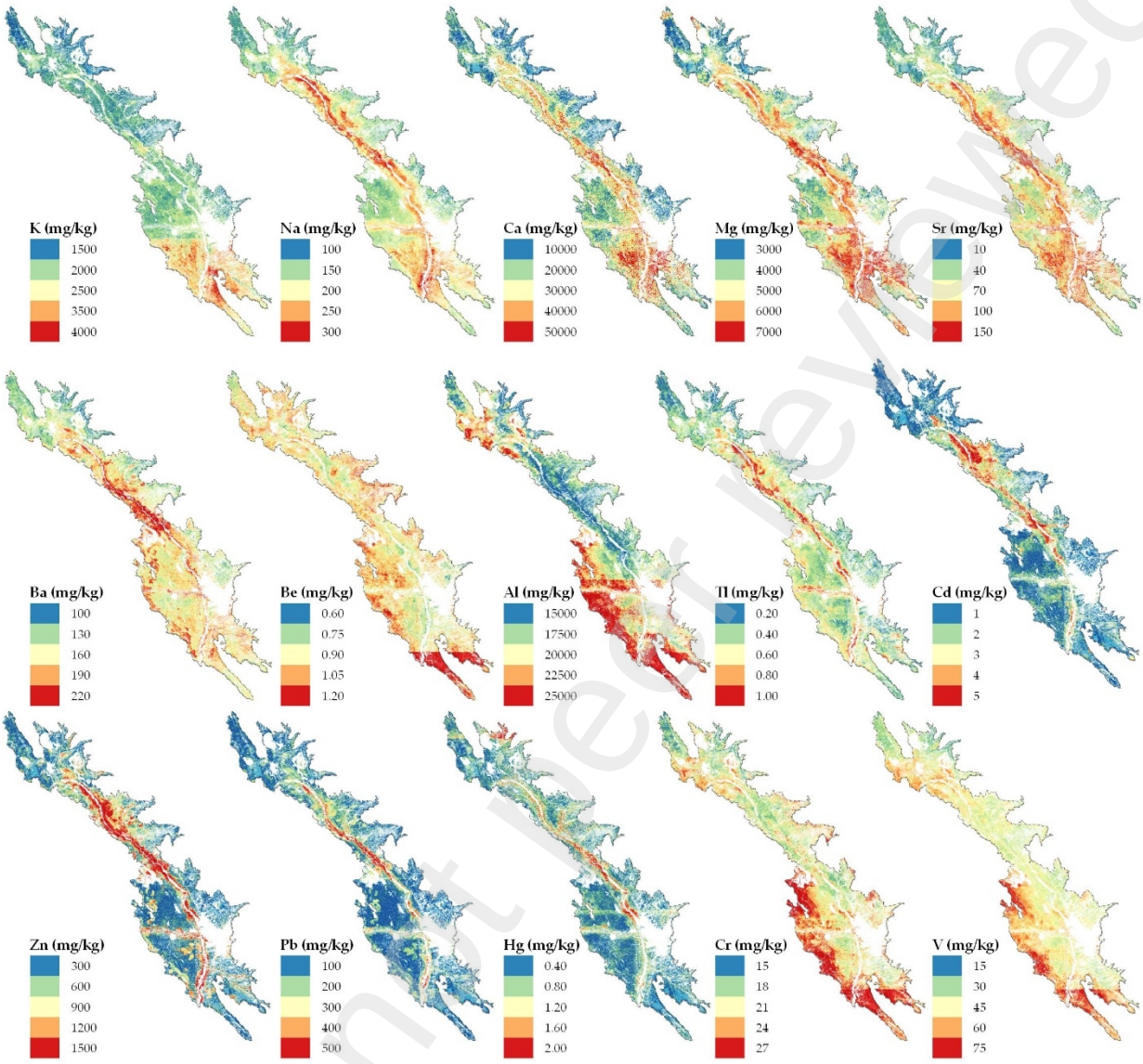

Figure 6. Spatial distribution maps of K and Na (Alkali metals), Ca, Mg, Sr, Ba and Be (Alkali-earth metals), Al, Tl, Cd, Zn, Pb and Hg (Post-transition metals), Cr, V (Transition metals).

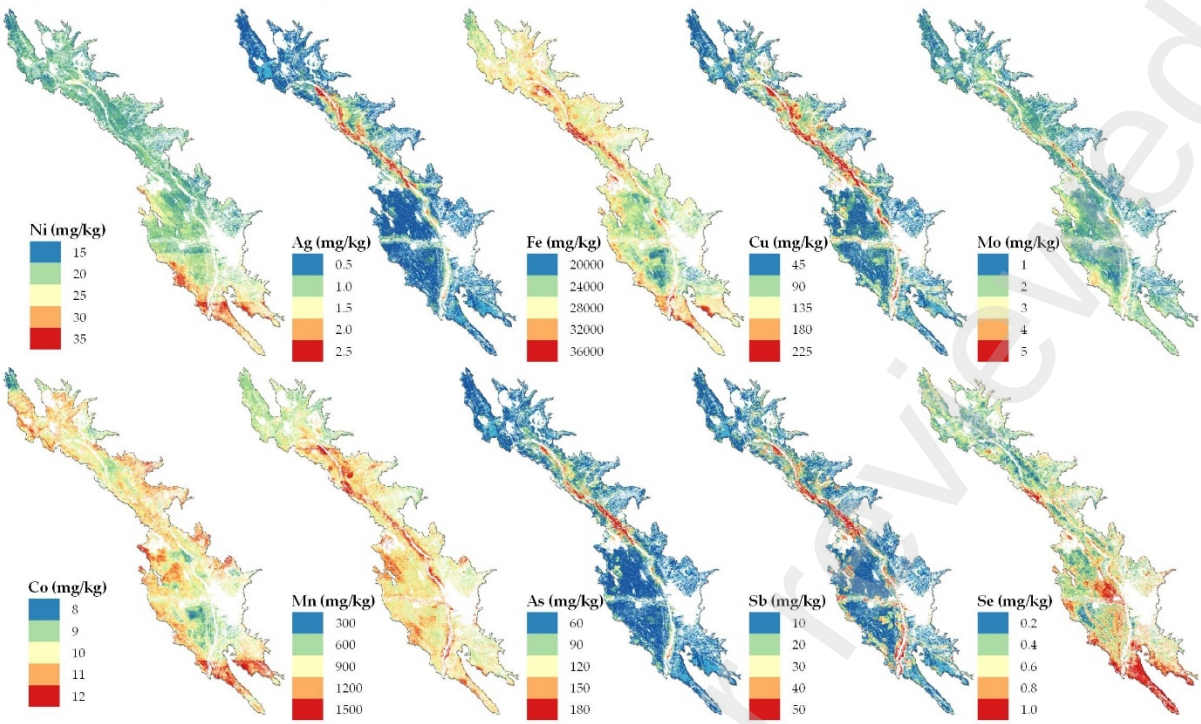

Figure 7. Spatial distribution maps of Ni, Ag, Fe, Cu, Mo, Co and Mn (Transition metals), As, Sb and Se (Metalloids).

3.1.6 Importance of covariables

We calculated the relative importance of 41 environmental covariates (note that the importance value has been converted to percentage) for the selected models by element and grouped by type of elements, with the highest accuracy and small errors (**Figure 8)**. For alkali and alkali - earth metals (Na, K, Mg, Ba, Ca, Be and Sr), the most important covariates were elevation, pH, distance to rivers and temperature variables, associated to nitrogen, CEC and clay content. The most important covariates in transition metals (Cr, V, Ni, Ag, Fe, Cu, Mo, Co and Mn) were distance to rivers and roads associated to soil bulk density and temperatures, with SOC, pH, silt and nitrogen content. In post-transition metals (Al, Tl, Cd, Zn, Pb and Hg), the elevation, distance to rivers, bulk density, sand content and pH were the most important covariables and for Cd, the rainfall, and temperatures specifically were important. For metalloids, nitrogen content, distance to rivers and roads, elevation, pH, rainfall and temperatures were the most important covariates.

It is also worth noting that topographic derived covariates and spectral covariates, have medium importance in all cases, but elevation reveals as good predictor.

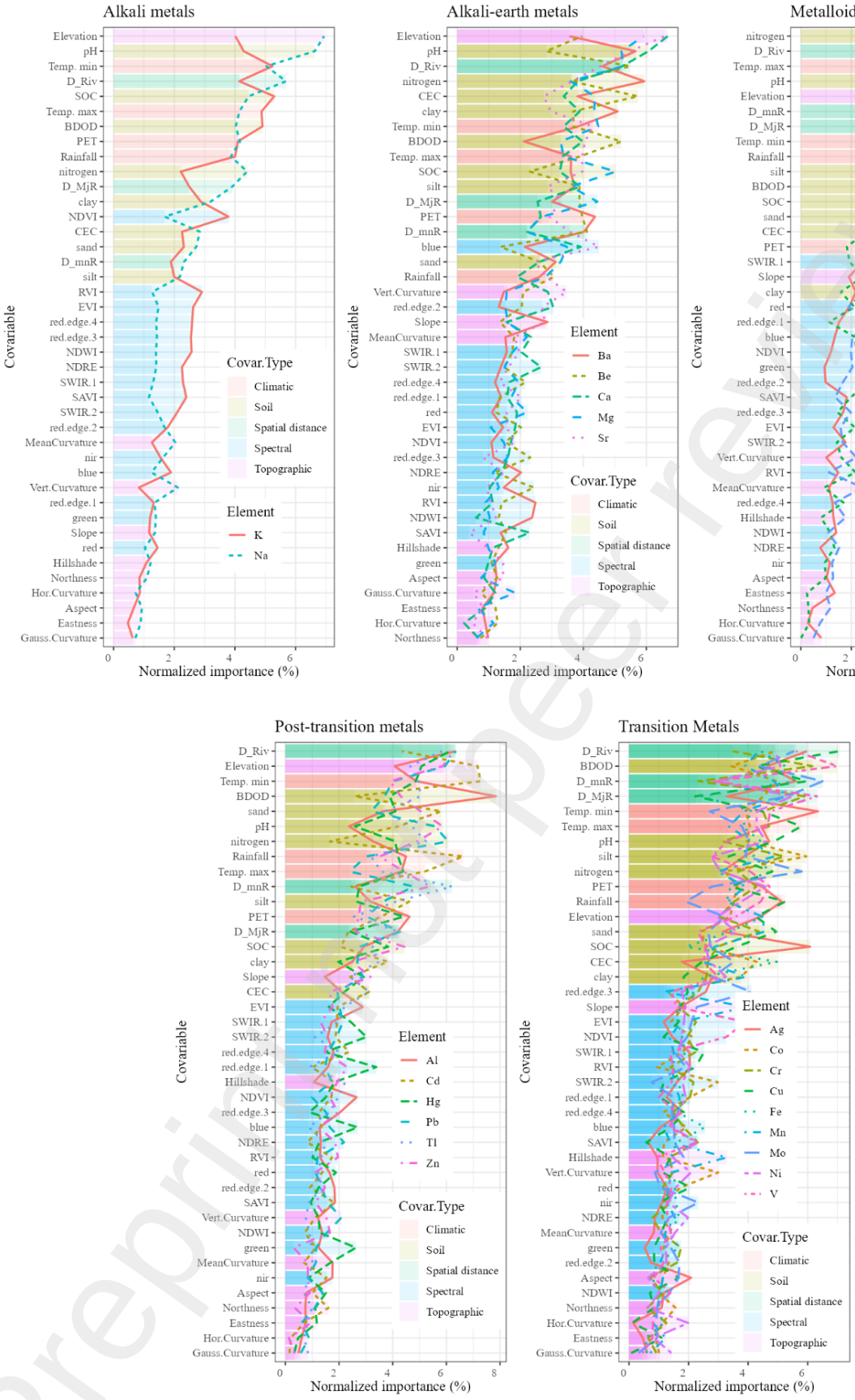

Figure 8. Relative importance of 41 environmental covariables on the alkali metals (A), alkali-earth metals (B), metalloid (C), post-transition metals (D) and transition metals (E) content.

4 Discussion

Here we present the first maps of alkali, alkali-earth, transition metals, post transition metals, and metalloids for the croplands of Mantaro Valley. First , we found a positive correlation between some elements, especially heavy metals, which indicate an aggregation degree of these elements, like the results obtained by Liu et al. (2023) [9] and Zhou et al. (2021) [43].

High variability of elements along different areas, from north-to-south and close to river, were determined for the elements analyzed. This can be explained by the complex interaction of variables in soil formation, even if Mantaro Valley is composed mostly by debris and transported material [44]. Topography affects the soil formation, , trough erosion, runoff and infiltration processes, and are related to chemical and physical properties, in consequence their spatial variability [6]. Zgłobicki [45], found that in areas with high slope exposed to erosion the concentration of elements in soils was low, and more concentrated at foot of slopes or bottoms of depressions. Climate plays an important role un pedogenetic process, and events like heavy rainfall, can transport elements from abandoned mining dams, geological deposits, biological degradation of organic matter, atmospheric deposition and accelerate industrial and domestic discharge [46].

Distance to river and had high relative importance in the mapping elements, even Pearson correlation coefficients showed moderate and low relationship with some elements. This indicated that water used for irrigating croplands can transport elements like heavy metals and another pollutants [47], product of human activities, that might result in metals accumulation in the area close rivers.

Distance to roads, has a relative medium importance in mapping models generated, hence the most croplands in Mantaro valley has low traffic. In the same way is recognized that some heavy metals like Pb, was often accumulated in surface soils near roads released from oil combustion, with implication of accumulation in plants growing in that soils [48]. Additionally other anthropogenic activities, such as agricultural and change land use management, could be other factors of element's accumulation, especially for some heavy ketals like Cu, contained in fungicides an turn an important source [49]

Another soil elements like SOC, sand, lime and clay content are linked with metals and metalloids [16], and the addition of manure immobilize some elements by formation of complexes with metal ions, in fact change the pH and affect the accumulation of metals in soils [50]

Spectral covariates had light importance, cause these variables are more sensitive with vegetation growth status, being other variables more important in the prediction of elements content. This result indicates that metals and metalloids concentration are highly affected by topography and human activities [51].

In the Peruvian environment standard of agriculture soils [52], for As the maximum permissible limit is 50 mg/kg, and in this study we report areas with concentration until 477.44 mg/kg. For Pb the maximum permissible limit is 50 mg/kg and we found areas with 1,674.76 mg/kg. Cd reach max concentrations of 21.8 mg/kg, above the 1.4 mg/kg permissible. Hence the predicted soils maps can become in an efficient tool for in order to identify polluted areas, that mostly of them are located around rivers.

5 Conclusions

This study contributes to the current state of art in digital soil mapping of elements in soils, using multiple geospatial data, including remote sensing data, climate data, topographic data, soil data and distance data for the Peruvian Mantaro Valley, trough random forest algorithm and computed in the cloud using Google Earth Engine with satisfactory results. For the study area, there are different combination of environmental covariates in estimate groups elements content, mostly soil, climate and topographic and distance variables, with less importance for spectral variables. The Random Forest modelling approach, gave satisfactory results in predicting the distribution of analyzed elements in soil, being improved for some elements when adds more trees. Then the application of machine learning algorithms with ground-truth data augmentation is effective in the mapping of soil elements. These findings suggest the implementation of spatial monitoring scheme for elements in soil, especially for toxics is feasible using environmental covariates. Finally, more research needs to be done along Mantaro Valley on the translocation of toxic metals to foods and its implication on public health. Also, the influence of contaminated water sources used in crops irrigation.

Author Contributions: S.P. and P.V. designed the methodology; J.V., and J.H. provided and validated field data, S.P., and S.L. performed the data processing; S.P. and P.V analyzed the data; R.S. and J.C. funding acquisition, S.P., and P.V. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

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" of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government.

Acknowledgements

Santa Ana's LABSAF teams for providing infrastructure and equipment for the soil data collection and laboratory analysis.

Conflicts of 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.

6 References

1. Kopittke, P.M.; Menzies, N.W.; Wang, P.; McKenna, B.A.; Lombi, E. Soil and the Intensification of Agriculture for Global Food Security. *Environ. Int.* **2019**, *132*, 105078, doi:10.1016/j.envint.2019.105078.

2. Brevik, E.C.; Slaughter, L.; Singh, B.R.; Steffan, J.J.; Collier, D.; Barnhart, P.; Pereira, P. Soil and Human Health: Current Status and Future Needs. *Air, Soil Water Res.* **2020**, *13*, doi:10.1177/1178622120934441.

3. Silver, W.L.; Perez, T.; Mayer, A.; Jones, A.R. The Role of Soil in the Contribution of Food and Feed. *Philos. Trans. R. Soc. B Biol. Sci.* **2021**, *376*, doi:10.1098/rstb.2020.0181.

4. Steffan, J.J.; Brevika, E.C.; Burgessa, L.C.; Cerdà, A. The Effect of Soil on Human Health: An Overview. *Eur J Soil Sci.* **2018**, *69*, 159–171, doi:doi:10.1111/ejss.12451.

5. Chen, T.; Chang, Q.; Liu, J.; Clevers, J.G.P.W.; Kooistra, L. 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.* **2016**, *565*, 155–164, doi:10.1016/j.scitotenv.2016.04.163.

6. Wu, Y.; Zhou, L.; Meng, Y.; Lin, Q.; Fei, Y. Influential Topographic Factor Identification of Soil Heavy Metals Using GeoDetector: The Effects of DEM Resolution and Pollution Sources. *Remote Sens.* **2023**, *15*, 1–21, doi:10.3390/rs15164067.

7. Rashid, A.; Schutte, B.J.; Ulery, A.; Deyholos, M.K.; Sanogo, S.; Lehnhoff, E.A.; Beck, L. Heavy Metal Contamination in Agricultural Soil: Environmental Pollutants Affecting Crop Health. *Agronomy* **2023**, *13*, 1–30, doi:10.3390/agronomy13061521.

8. Taghizadeh-Mehrjardi, R.; Minasny, B.; Toomanian, N.; Zeraatpisheh, M.; Amirian-Chakan, A.; Triantafilis, J. Digital Mapping of Soil Classes Using Ensemble of Models in Isfahan Region, Iran. *Soil Syst.* **2019**, *3*, 1–21, doi:10.3390/soilsystems3020037.

9. Liu, Q.; Du, B.; He, L.; Zeng, Y.; Tian, Y.; Zhang, Z.; Wang, R.; Shi, T. Digital Soil Mapping of Heavy Metals Using Multiple Geospatial Data: Feature Identification and Deep Neural Network. *Ecol. Indic.* **2023**, *154*, doi:10.1016/j.ecolind.2023.110863.

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

11. Omondi, E.; Boitt, M. 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.* **2020**, *12*, 597–619, doi:10.4236/jgis.2020.126035.

12. Moradpour, S.; Entezari, M.; Ayoubi, S.; Karimi, A.; Naimi, S. Digital Exploration of Selected Heavy Metals Using Random Forest and a Set of Environmental Covariates at the Watershed Scale. *J. Hazard. Mater.* **2023**, *455*.

13. Lachaud, A.; Adam, M.; Mišković, I. Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. *Minerals* **2023**, *13*, doi:10.3390/min13081073.

14. Pizarro, S.; Pricope, N.G.; Figueroa, D.; Carbajal, C.; Quispe, M.; Vera, J.; Alejandro, L.; Achallma, L.; Gonzalez, I.; Salazar, W.; et al. Implementing Cloud Computing for

the Digital Mapping of Agricultural Soil Properties from High Resolution UAV Multispectral Imagery. *Remote Sens.* **2023**, *15*, 3203, doi:10.3390/rs15123203.

15. Wu, W.; Li, Y.; Yan, M.; Yang, L.; Lei, J.; Liu, H. Bin Surface Soil Metal Elements Variability Affected by Environmental and Soil Properties. *PLoS One* **2021**, *16*, 1–14, doi:10.1371/journal.pone.0254928.

16. Thomas, E.; Atkinson, R.; Zavaleta, D.; Rodriguez, C.; Lastra, S.; Yovera, F.; Arango, K.; Pezo, A.; Aguilar, J.; Tames, M.; et al. The Distribution of Cadmium in Soil and Cacao Beans in Peru. *Sci. Total Environ.* **2023**, *881*, 163372, doi:10.1016/j.scitotenv.2023.163372.

17. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil Organic Carbon and Texture Retrieving and Mapping Using Proximal, Airborne and Sentinel-2 Spectral Imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103, doi:10.1016/j.rse.2018.09.015.

18. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27, doi:10.1016/j.rse.2017.06.031.

19. Custodio, M.; Peñaloza, R.; Ochoa, S.; Cuadrado, W. 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.* **2021**, *28*, 5952–5962, doi:10.1016/j.sjbs.2021.06.054.

20. Munive, R.; Gamarra, G.; Munive, Y.; Puertas, F.; Valdiviezo, L.; Cabello, R. Lead and Cadmium Uptake by Sunflower from Contaminated Soil and Remediated with Organic Amendments in the Form of Compost and Vermicompost. *Sci. Agropecu.* **2020**, *11*, 177–186, doi:10.17268/SCI.AGROPECU.2020.02.04.

21. Fujimoto, A.; Miyaura, R.; Ugas, R. 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.* **2004**, *49*, 1–16.

22. Instituto Geofísico del Perú *Atlas Climático de Precipitación y Temperatura Del Aire En La Cuenca Del Río Mantaro*; Fondo Editorial del Consejo Nacional del Ambiente - CONAM, Ed.; Lima - Perú, 2005;

23. Pizarro, S.E.; Pricope, N.G.; Vargas-Machuca, D.; Huanca, O.; Ñaupari, J. Mapping Land Cover Types for Highland Andean Ecosystems in Peru Using Google Earth Engine. *Remote Sens.* **2022**, *14*, doi:10.3390/rs14071562.

24. Ministerio del Ambiente (MINAM) Mapa Nacional de Cobertura Vegetal - Memoria Descriptiva Available online: https://www.gob.pe/institucion/minam/informes-publicaciones/2674-mapa-nacional-de-cobertura-vegetal-memoria-descriptiva (accessed on 15 January 2022).

25. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46, doi:10.1016/0034-4257(91)90048-B.

26. Rouse, J..; Haas, R..; Schell, J..; Deering, D.. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Proceedings of Third Earth Resources

Technology Satellite Symposium; Remote Sensingcenter, Texas A&M hivemity, Colfegp Station, Texas: Washington, DC, 1974; Vol. 351, p. 309.

27. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432, doi:10.1080/01431169608948714.

28. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126, doi:10.1016/0034-4257(94)90134-1.

29. Huete, A.R.; Didan, K.; Miura, T.; Rodriguez, E..; Gao, X.; Ferreira, L.. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213.

30. Pearson R. L.; Miller L. D. Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie, Pawnee National Grasslands, Colorado,. In Proceedings of the Proceedings of the 8th International Symposium on Remote Sensing of the Environment II; Department of Watershed Sciences-College of Forestry and Natural Resources-Colorado State University: Colorado, 1972; pp. 1355–1379.

31. Gitelson, A.; Merzlyak, M.N. 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.* **1994**, *143*, 286–292, doi:10.1016/S0176-1617(11)81633-0.

32. Peng, Y.; Wang, L.; Zhao, L.; Liu, Z.; Lin, C.; Hu, Y.; Liu, L. Estimation of Soil Nutrient Content Using Hyperspectral Data. *Agric.* **2021**, *11*, doi:10.3390/agriculture11111129.

33. Wu, Y.; Zhou, L.; Meng, Y.; Lin, Q.; Fei, Y. Influential Topographic Factor Identification of Soil Heavy Metals Using GeoDetector : The Effects of DEM Resolution and Pollution Sources. *Remote Sens.* **2023**, *15*, 1–21, doi:https://doi.org/10.3390/rs15164067.

34. Safanelli, J.L.; Poppiel, R.R.; Chimelo Ruiz, L.F.; Bonfatti, B.R.; de Oliveira Mello, F.A.; Rizzo, R.; Demattê, J.A.M. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. *ISPRS Int. J. Geo-Information* **2020**, *9*, doi:10.3390/ijgi9060400.

35. Cornu, J.Y.; Denaix, L.; Lacoste, J.; Sappin-Didier, V.; Nguyen, C.; Schneider, A. 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.* **2016**, *23*, 2997–3007, doi:10.1007/s11356-015-5432-4.

36. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958-2015. *Sci. Data* **2018**, *5*, 1–12, doi:10.1038/sdata.2017.191.

37. Poggio, L.; De Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil* **2021**, *7*, 217–240, doi:10.5194/soil-7-217-2021.

38. Danielsson, P.E. Euclidean Distance Mapping. *Comput. Graph. Image Process.* **1980**, *14*, 227–248, doi:10.1016/0146-664X(80)90054-4.

39. Zhang, W.; Liu, M.; Li, C. Soil Heavy Metal Contamination Assessment in the Hun-Taizi River Watershed, China. *Sci. Rep.* **2020**, *10*, 1–10, doi:10.1038/s41598-020-65809-0.

40. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324.

41. Wei, T.; Simko, V. Corrplot: Visualization of a Correlation Matrix (Version 0.84) 2017, 18.

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

43. Zhou, W.; Yang, H.; Xie, L.; Li, H.; Huang, L.; Zhao, Y.; Yue, T. Hyperspectral Inversion of Soil Heavy Metals in Three-River Source Region Based on Random Forest Model. *Catena* **2021**, *202*, 105222, doi:10.1016/j.catena.2021.105222.

44. Martínez Vargas, A. *ESTUDIO DE LA GEOLOGÍA REGIONAL DE LOS VALLES DEL MANTARO Y TARMA*; 1978;

45. Zgłobicki, W. Impact of Microtopography on the Geochemistry of Soils within Archaeological Sites in SE Poland. *Environ. Earth Sci.* **2013**, *70*, 3085–3092, doi:10.1007/s12665-013-2368-1.

46. Custodio, M.; Álvarez, D.; Cuadrado, W.; Montalvo, R.; Ochoa, S. Potentially Toxic Metals and Metalloids in Surface Water Intended for Human Consumption and Other Uses in the Mantaro River Watershed, Peru. *Soil Water Res.* **2020**, *15*, 237–245, doi:10.17221/152/2019-SWR.

47. Chira, J.; Vargas, L.; Calderón, C.; Arcos, F.; Mogrovejo, M.; La Cruz, C. De Heavy Metals and Their Impact on Surface Waters of the Mantaro River Basin, Junin, Peru. *Int. J. Hydrol.* **2022**, *6*, 88–93, doi:10.15406/ijh.2022.06.00306.

48. Wang, F.; Guan, Q.; Tian, J.; Lin, J.; Yang, Y.; Yang, L.; Pan, N. Contamination Characteristics, Source Apportionment, and Health Risk Assessment of Heavy Metals in Agricultural Soil in the Hexi Corridor. *Catena* **2020**, *191*, 104573, doi:10.1016/j.catena.2020.104573.

49. Mouazen, A.M.; Nyarko, F.; Qaswar, M.; Tóth, G.; Gobin, A.; Moshou, D. Spatiotemporal Prediction and Mapping of Heavy Metals at Regional Scale Using Regression Methods and Landsat 7. *Remote Sens.* **2021**, *13*, 1–23, doi:10.3390/rs13224615.

50. De Temmerman, L.; Vanongeval, L.; Boon, W.; Hoenig, M.; Geypens, M. Heavy Metal Content of Arable Soils in Northern Belgium. *Water. Air. Soil Pollut.* **2003**, *148*, 61–76, doi:10.1023/A:1025498629671.

51. Peng, Y.; Kheir, R.B.; Adhikari, K.; Malinowski, R.; Greve, M.B.; Knadel, M.; Greve, M.H. Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data. *Remote Sens.* **2016**, *8*, 1–19, doi:10.3390/rs8121003.

52. MINAM *Estándares de Calidad Ambiental (ECA) Para Suelo*; El Peruano, 2017; pp. 1–4;.

Fig. S1. Correlation matrix between the analyzed elements in soil.