frontiers in Soil Science

TYPE Original Research
PUBLISHED 05 May 2026
DOI 10.3389/fsoil.2026.1770292

Critical edaphic and altitudinal factors influencing cation exchange capacity in coffee-growing soils of northeastern Peru: implications for sustainable fertility management

Henry Díaz-Chuquizuta[1], Luis Fernando Manrique Gonzales[2], Martín Sánchez Ojanasta[2], Juan Pablo Cuevas-Giménez[3], Carlos Carbajal-Llosa[4], Néstor Cuellar-Condori[5], Boris Martínez[6] and Geomar Vallejos-Torres[2,7*]

[1]Estación Experimental Agraria San Ramón, Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Yurimaguas, Alto Amazonas, Peru, [2]Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Estación Experimental Agraria El Porvenir, San Martin, Peru, [3]Estación Experimental Agraria de Baños del Inca, Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Cajamarca, Peru, [4]Centro Experimental La Molina, Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Lima, Peru, [5]Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Puno, Peru, [6]Dirección de Investigacion y Desarrollo Tecnologico, Estación Experimental Agraria San Ramón, Instituto Nacional de Innovación Agraria (INIA), Yurimaguas, Alto Amazonas, Peru, [7]Facultad de Ciencias Agrarias, Universidad Nacional de San Martín, Tarapoto, Peru

OPEN ACCESS

EDITED BY
Bifeng Hu,
Jiangxi University of Finance and Economics, China

REVIEWED BY
Halus Satriawan,
Universitas Almuslim, Indonesia
Char Twesigye,
Kyambogo University, Uganda

*CORRESPONDENCE
Geomar Vallejos-Torres
gvallejos@unsm.edu.pe

RECEIVED 17 December 2025
REVISED 11 April 2026
ACCEPTED 15 April 2026
PUBLISHED 05 May 2026

CITATION
Díaz-Chuquizuta H, Manrique Gonzales LF, Sánchez Ojanasta M, Cuevas-Giménez JP, Carbajal-Llosa C, Cuellar-Condori N, Martínez B and Vallejos-Torres G (2026) Critical edaphic and altitudinal factors influencing cation exchange capacity in coffee-growing soils of northeastern Peru: implications for sustainable fertility management. *Front. Soil Sci.* 6:1770292. doi: 10.3389/fsoil.2026.1770292

Introduction: Effective cation exchange capacity (ECEC) is a key indicator of soil fertility and sustainable soil management assessment in coffee-growing systems.
Methods: This study aimed to identify the principal edaphic and altitudinal factors explaining ECEC variability in 69 soil samples collected from coffee farms in northeastern Peru.
Results: ECEC results exhibited substantial variation, ranging from 0.14 to 55.49 cmol(+)·kg^{-1} (mean = 15.21; SD = 12.47), and were significantly correlated with organic matter (r = 0.71), clay content (r = 0.62), exchangeable acidity (r = −0.63), and altitude (r = 0.33). Principal component analysis accounted for 64.3% of the edaphic variability, identifying Ca^{2+}, pH, Mg^{2+}, and exchangeable acidity as the most influential variables. The Random Forest model demonstrated high predictive accuracy (R^2 = 0.93; root mean square error (RMSE) = 2.1 cmol (+)·kg^{-1}), outperforming the generalized additive model (GAM) and identifying Ca^{2+} as the most important predictor (IncMSE% = 3177.37). A functional altitudinal gradient was also evident: areas above 1150 m.a.s.l. showed higher acidity and aluminium content, whereas areas below 900 m.a.s.l. exhibited greater base saturation and higher ECEC.
Discussion: These findings support the development of sitespecific fertilization strategies and soil−climate zoning, emphasizing the value of integrating multivariate analyses with machine-learning models as key tools for optimizing fertility management and coffee crop productivity in tropical mountain

ecosystems; where soil texture represents a key factor influencing coffee sustainability, as greater nutrient retention capacity and improved nutritional balance are associated with enhanced potential for sustainable production and reduced environmental impact.

KEYWORDS
altitude, cation exchange capacity, multivariate analysis, random forest, soil fertility, soil zoning, sustainable coffee production, tropical soils

1 Introduction

Coffee farming is one of the most important agricultural activities for Peru's economy and rural development. Coffee (*Coffea arabica* L.) propagate sexually and asexually using microtunnels for rooting cuttings (1) and is the country's fourth-largest agricultural export, with annual production exceeding 400,000 tonnes and significantly contributing 7.8% to the gross value of agricultural output (2). This activity directly supports more than 223,000 farming households. It is practiced across 16 regions, with San Martín standing out for its significant contribution to the national production, representing 21.5% of the overall cultivated area (3). Despite its economic relevance, many coffee-growing plantations still rely on inadequate fertilization practices that overlook soil spatial variability, resulting in inefficient nutrient management and compromising long-term sustainability (4).

In this context, detailed knowledge of soil properties and their interactions with environmental factors such as altitude is essential for designing site-specific management strategies (5). Cation exchange capacity (CEC) is a key indicator of soil fertility, as it determines the soil's ability to retain and supply essential nutrients, including Ca^{2+}, Mg^{2+}, K^{+}, and NH_4^{+}, to plants. Its value is influenced by intrinsic factors, such as organic matter content, soil texture, and pH, as well as extrinsic factors, such as altitude, which can affect organic matter accumulation and mineralization (5). In tropical mountainous regions like northeastern Peru, the interaction between soil and climate generates substantial soil heterogeneity, directly influencing coffee-tree fertility and productivity (6, 7).

This complexity requires multivariate analytical approaches, such as principal component analysis (PCA), generalized additive models (GAM), and machine-learning algorithms like Random Forest, to identify latent patterns, homogeneous zones, and key variables relevant to agronomic management (8, 9). This study aims to determine the most influential edaphic and altitudinal factors affecting the CEC of coffee-growing soils in northeastern Peru, using an integrated statistical and geostatistical approach to support the identification of soil typologies and the development of more efficient and sustainable fertilization strategies.

The physico-chemical properties of acidic soils in the districts of Alto Saposoa, Soritor, and Vista Alegre, located in the San Martín region of Peru, significantly influence coffee productivity, confirming the hypothesis that edaphic variability constrains crop performance. In particular, soil organic matter, cation exchange capacity, and the availability of P and K were identified as the main positive regulators of productivity, whereas exchangeable sodium and silt content acted as limiting factors (10). In another study, Ahmed et al. (11) reported both increases and decreases in secondary metabolites and sensory attributes determining coffee quality in response to changes in environmental and management conditions. The most consistent evidence identified through this systematic review highlights two main trends: (1) increasing altitude is associated with improved sensory attributes of coffee; and (2) increased light exposure is associated with a decline in coffee sensory quality.

2 Materials and methods

The present study followed a quantitative, observational, and cross-sectional design, integrating descriptive, inferential, multivariate, predictive, and geostatistical approaches. This design enabled the characterization of soil fertility variability and the modelling of the determinants of effective cation exchange capacity (ECEC) along an altitudinal gradient in coffee production systems.

2.1 Study area

The study was conducted in the Peruvian high jungle, specifically in the districts of San José de Sisa and Santa Rosa, in the province of El Dorado, in the San Martín region, an area well known for its coffee production (Figure 1). The province's climate corresponds to a warm, humid highland jungle environment. The average annual rainfall is approximately 1,157 mm, and annual relative humidity averages 78.5%, typically fluctuating between 77% and 80%. The mean annual temperature is around 25.0 °C, with recorded minimums of 12.5 °C and maximums of 38.4 °C. Regarding seasonal dynamics, the most intense rainfall occurs between January and April, marking the onset of the rainy season, followed by alternating dry and wet periods throughout the remainder of the year. These historical climatic averages were derived from meteorological data recorded at stations operated by the National Service of Meteorology and Hydrology of Peru (12).

2.2 Soil sampling and analysis

RStudio software was used to determine sampling points using the *sample* function from the sp package (13). The R script generated sampling coordinates by integrating a raster layer with high-resolution satellite imagery (e.g., Sentinel-2) as the basemap and applying a stratified sampling design. A total of 69 soil samples were collected at a depth of 30 cm from fields managed by coffee producers.

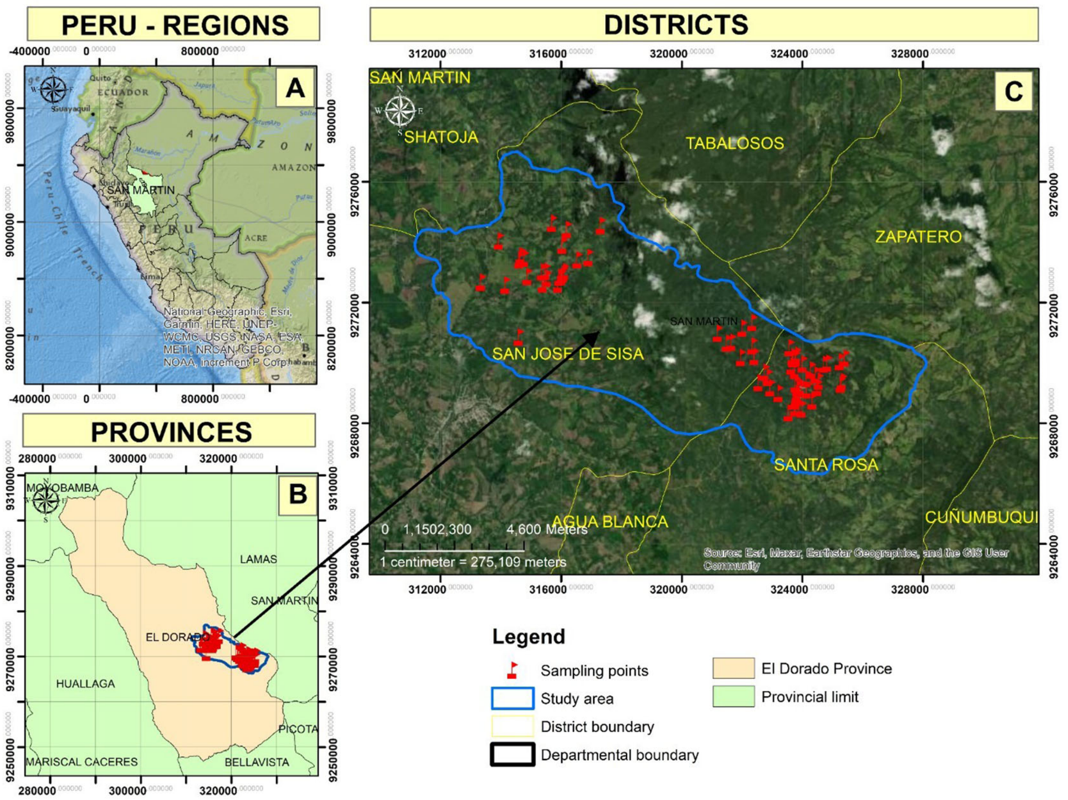

FIGURE 1
Location of the study area in the Peruvian Amazon: **(A)** country and region; **(B)** provincial boundaries; **(C)** district boundaries and sampling points.

The collected samples were analysed at INIA (Instituto Nacional de Innovación Agraria)'s Soil, Water, and Foliar Laboratory network. The variables assessed formed part of a comprehensive soil characterization following established reference methodologies. Soil texture (sand, silt, and clay percentages) was determined using the Bouyoucos hydrometer method (14). Soil pH was measured according to the U.S. Environmental Protection Agency (15) guidelines. Organic matter (OM) content was analysed according to the protocol of the Secretariat of the Environment and Natural Resources (16), while total nitrogen (N) was determined in accordance with the International Organization for Standardization (17). Available phosphorus (P) in both neutral and acidic soils was measured using the method of Bray et al. (18), and available potassium (K) was measured according to Helmke et al. (19). Exchangeable cations (H^+, Al^{3+}, Ca^{2+}, Mg^{2+}, K^+, and Na^+) were quantified following the procedures of Sumner et al. (20). In addition, soil property data with a spatial resolution of 250 m for the 15–30 cm depth interval were retrieved from the SoilGrids database in TIFF format and processed to extract the corresponding values.

2.3 Statistical analysis

All statistical analyses were performed using R software version 4.5.2 (13) within the RStudio environment. Spatial analyses were conducted using the packages sf, terra, gstat, and automap under the same software version, ensuring full computational reproducibility.

2.3.1 Descriptive statistics

For this analysis, exploratory data analysis and visualisation techniques were applied to the soil dataset comprising 69 soil samples from coffee plots. The summary function and the describe function from the psych package (2023) were used to estimate the following:

$$\bar{x} = \frac{1}{n}\sum_{i=1} xi \quad ; \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(xi-\bar{x})^2} \quad ;$$

$$SE = \frac{s}{\sqrt{n}} \quad ; \quad CV(\%) = \left(\frac{s}{\bar{x}}\right)x100$$

Mean ($\bar{x}$), standard deviation (s), standard error (SE), and coefficient of variation (CV) were computed. These metrics enabled the assessment of soil heterogeneity, particularly the detection of extreme soil acidity and variability in organic matter and available cations. The results were visualized using boxplots, histograms, and customised graphics generated with the ggplot2 package (21), which facilitated the identification of outliers and non-normal distributions. This combined methodological approach enabled the detection of preliminary patterns that justified the subsequent application of multivariate techniques.

2.3.2 Pearson correlation

To evaluate the bivariate relationships among soil variables, Pearson's correlation coefficient (r) was computed using the *cor* function in R, following the approach described by Wei et al. (4). This coefficient is calculated as:

$$r = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum(xi - \bar{x})^2} \cdot \sqrt{\sum(yi - \bar{y})^2}}$$

Visualisation was carried out using heatmaps generated with ggplot2 and corrplot, which facilitated the identification of associations between ECEC and other variables such as OM, pH, and exchangeable acidity. Statistical significance of the correlations was assessed using the *cor.test* function.

2.3.3 Principal component analysis

This multivariate analysis used principal component analysis (PCA) via the *prcomp* function to reduce data dimensionality and identify latent patterns. Visualisations were generated using fviz_pca_biplot from the factoextra package. The PCA revealed that the first two principal components accounted for a large proportion of the total variance and enabled the segmentation of edaphic units into distinct clusters, information crucial for agronomic decision-making (13).

2.3.4 Generalized additive models

A generalized additive model (GAM) was fitted using the *gam* function from the mgcv package, employing an identity link function and a Gaussian distribution (22). The model evaluated ECEC as the dependent variable as a function of smoothed predictors (penalized splines) as follows:

$$yi = \alpha + \sum_{j=1}^{p} s_j\,(x_{ij}) + \in_i$$

2.3.5 Random forest model

The Random Forest (RF) model was implemented to predict effective cation exchange capacity (ECEC) as a function of 15 soil and altitude variables, using the randomForest package in R as described by Liaw et al. (23). The algorithm constructs an ensemble of multiple regression trees and aggregates their predictions, allowing it to capture complex non-linear relationships without requiring assumptions of normality or linearity (24).

$$\tilde{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$$

T = number of trees; $h_t(x)$ = prediction of the t-th tree; $\hat{y}$ = final predicted value.

2.3.6 Cross-validation

K-fold cross-validation (k = 5) was applied to evaluate the performance of the GAM and RF models. The coefficient of determination (R^2) and the root mean square error (RMSE) were calculated following standard procedures (25).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_1(x_1) - Z_2\,(x_1))^2}$$

The selection of the analytical methods employed was driven by the high heterogeneity and complexity of the evaluated tropical soils. Descriptive statistics and Pearson correlation analysis were used to identify preliminary patterns and bivariate associations. Principal component analysis (PCA) was applied to reduce dimensionality and detect multivariate fertility gradients. The generalized additive model (GAM) was used to capture potential nonlinear relationships between effective cation exchange capacity (ECEC) and its predictors, whereas the Random Forest model was applied to model complex interactions and rank variable importance without assuming normality. Finally, universal kriging was implemented to model spatial dependence and interpolate ECEC after confirming spatial autocorrelation.

2.4 Spatial analysis

The spatial distribution of cation exchange capacity (CEC) in the study area was determined using a comprehensive geostatistical kriging workflow implemented in R Study version 4.5.2 (13). The analysis began with data preparation, in which soil sampling points within CEC measurements and their geographic coordinates were integrated with the study area boundary and an environmental covariate (elevation). All spatial layers were harmonised to a unified coordinate reference system (WGS 84/UTM Zone 18S, EPSG:32718) using the sf and terra packages. A spatial autocorrelation analysis using Moran's I statistic confirmed that the CEC values exhibited significant spatial clustering ($p < 0.05$), validating the use of geostatistical interpolation methods. For the universal kriging (UK) scenarios, elevation values were extracted for all sampling locations to incorporate this covariate into trend modelling. This variable enabled the capture of potential topographic effects on CEC spatial variability and improved the model's predictive accuracy.

The core of the analysis focused on variogram modelling and kriging interpolation using the gstat package. An empirical variogram was first computed to quantify how the CEC variance changes with distance between sampling locations. Subsequently, theoretical variogram models (spherical, exponential, Gaussian, and Matérn) were automatically fitted using the automap package. The optimal variogram model was selected using weighted least squares, yielding key geostatistical parameters, including the nugget effect (micro-scale or measurement-error variance), partial sill (structured spatial variance), and range (distance of spatial influence). Given the availability of auxiliary environmental data, universal kriging (UK) was applied to incorporate elevation, extracted from NASA JPL (26), as an explanatory trend variable. The fitted variogram model was then used to interpolate CEC values over a regular prediction grid with 50 m spatial resolution across the entire study area. In parallel, the kriging variance was calculated to quantify the spatial uncertainty associated with the predictions.

The model was validated using k-fold cross-validation (k = 5), in which the dataset was iteratively partitioned into k subsets, with one subset used for testing and the remaining k-1 for training. For each fold, the variogram was re-fitted using only the training data, and predictions were generated for the withheld test locations to emulate an independent validation process. Performance metrics, including root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and bias, were calculated to assess the interpolation's accuracy.

3 Results

3.1 Descriptive statistics

The descriptive results for the 69 coffee-growing soil samples revealed substantial variability across the evaluated soils (Table 1). Soil pH averaged 4.87 (SD = 1.14), ranging from 3.4 to 7.5, indicating that most soils are strongly acidic, although some units fall within moderately acidic to neutral conditions. Electrical conductivity (EC) exhibited a mean of 14.06 $dS \cdot m^{-1}$ (SD = 7.37), suggesting highly saline levels. Organic matter (OM) averaged 4.30% (SD = 2.76), with extreme values ranging from 0.7% to 13.9%, indicating soils that vary from very poor to very rich in organic content, likely influenced by differences in altitude and vegetation cover. Available phosphorus (P) exhibited a highly dispersed distribution (M = 10.27 mg kg^{-1}; SD = 12.40), reaching a maximum of 84.6 mg kg^{-1}, suggesting localized fertilization events or natural accumulation processes. Available potassium (K) showed a mean value of 112.56 mg kg^{-1} (SD = 63.49), while total nitrogen (N) remained low (M = 0.21%; SD = 0.13), indicating a limitation of this macronutrient in most soils. Exchangeable acidity averaged 2.83 cmol(+)·kg^{-1}, with high variability (SD = 3.63), confirming the presence of sites with strong active acidity.

Regarding effective cation exchange capacity (ECEC), the soils averaged 15.21 cmol(+)·kg^{-1} (SD = 12.47), with a wide range from nearly negligible levels (0.14) to highly reactive soils (55.49), reflecting pronounced soil heterogeneity across the study area. Exchangeable Ca^{2+}, Mg^{2+}, and K^+ also exhibited high variability, with Ca^{2+} emerging as the dominant base cation (M = 10.54 cmol(+)·kg^{-1}; SD = 12.82). Textural analysis revealed a predominance of loam- to sandy-loam soils, with average proportions of 54.64% sand, 26.11% clay, and 19.25% silt. Altitude varied markedly among sampling plots, ranging from 462 to 1,362 m.a.s.l. (M = 963.32; SD = 255.33), a variable known to strongly influence soil pH and organic matter.

In general, the standard error coefficients were low across most variables, indicating high precision in estimating their mean values. This statistical consistency reinforces the need to evaluate how interactions among pH, OM, soil texture, and altitude influence ECEC and, consequently, the fertility and long-term sustainability of these coffee-growing soils.

3.2 Statistical analysis

Bivariate associations between the main physico-chemical soil properties were analysed using Pearson's correlation coefficient (Figure 2) to identify multivariate patterns that explain the variability of effective cation exchange capacity (ECEC) in coffee plantations in north-eastern Peru. The ECEC showed substantial and statistically significant positive correlations with organic matter (r = 0.71***) and clay content (r = 0.62***), and a moderate but significant association with altitude (r = 0.33*). These results suggest that these three factors are key determinants of cation retention in acidic soils. Likewise, soil pH showed strong negative associations with exchangeable acidity (r = −0.65***), exchangeable aluminium (r = −0.59***), and exchangeable hydrogen (r = −0.51***), confirming its buffering role against acid toxicity in highly leached soils. A positive relationship was also observed between sand content and pH (r = 0.36**), and an inverse association between clay content and pH (r = −0.42***). These patterns reflect the structural influence of soil texture on acidity dynamics.

The organic matter content also showed a strong positive association with altitude (r = 0.53***) and a negative correlation with sand (r = −0.60***), which may be attributed to greater organic matter accumulation rates in higher, more humid environments, where lower mineralization prevails. Similarly, clay content correlated negatively with sand (r = −0.83***), as expected from the inherent balance of soil textural fractions. Overall, the visualization enabled the identification of clusters of highly interrelated variables, supporting the selection of robust predictors for subsequent multivariate analyses such as PCA and GAM. Correlation coefficients were represented using a colour gradient (blue for positive associations and red for negative associations), while significance levels were denoted using asterisks ($p < 0.05$*, $p < 0.01$**, $p < 0.001$***), following APA standards.

3.3 Principal component analysis by soil characteristics

A principal component analysis (PCA) was performed using 11 edaphic and altitudinal variables, thereby identifying latent multivariate patterns that explain soil fertility variability in coffee-growing areas of the El Dorado district. The first two principal components (Dim1 and Dim2) accounted for 71.2% of the total variance in the dataset, with Dimension 1 explaining 45.0% and Dimension 2 explaining 26.2% (Figure 3).

Dimension 1 (the horizontal axis of the biplot) was strongly associated with key fertility variables, including Ca^{2+}, Mg^{2+}, K^+, and effective cation exchange capacity (ECEC). These indicators showed factor loadings above 0.70, indicating strong explanatory power within the first component. Organic matter (OM) also aligned positively with this dimension (loading > 0.65). In contrast, exchangeable acidity and Al^{3+} projected in the opposite direction, with substantial negative loadings (< −0.60), revealing a chemical

TABLE 1 Descriptive statistics of soil physico-chemical properties and altitude.

Variable	N	Mean	SD	Minimum	Maximum	Median	Range	SE
pH	69	4.87	1.14	3.40	7.50	4.60	4.10	0.14
EC (dS/m)	69	14.06	7.37	3.50	36.10	12.90	32.60	0.89
OM (%)	69	4.30	2.76	0.70	13.90	3.50	13.20	0.33
P bray (mg kg^{-1})	69	10.27	12.40	0.40	84.60	6.60	84.20	1.49
Avail K mg kg^{-1}	69	112.56	63.49	20.00	292.26	103.97	272.26	7.64
Total N (%)	69	0.21	0.13	0.00	0.70	0.18	0.70	0.02
Exc Acidity (cmol(+)/kg)	69	2.83	3.63	0.00	13.60	0.70	13.60	0.44
Exc H (cmol(+)/kg)	69	1.61	1.67	0.00	5.40	0.70	5.40	0.20
Exc Al (cmol(+)/kg)	69	1.22	2.17	0.00	9.10	0.00	9.10	0.26
$CaCO_3$ (%)	69	0.30	1.00	0.00	4.70	0.00	4.70	0.12
Exc Ca (cmol(+)/kg)	69	10.54	12.82	0.31	50.29	4.10	49.98	1.54
Exc K (cmol(+)/kg)	69	0.31	0.21	0.00	1.01	0.27	1.01	0.03
Exc Mg (cmol(+)/kg)	69	1.46	1.38	0.11	7.85	0.97	7.74	0.17
Exc Na (cmol(+)/kg)	69	0.13	0.18	0.00	0.60	0.00	0.60	0.02
ECEC (cmol(+)/kg)	69	15.21	12.47	0.14	55.49	11.77	55.35	1.50
Sand (%)	69	54.64	13.37	33.31	79.53	51.38	46.22	1.61
Clay (%)	69	26.11	9.41	10.74	43.59	27.84	32.85	1.13
Silt (%)	69	19.25	5.85	5.93	32.07	19.34	26.14	0.70
Altitude m a.s.l.	69	963.32	255.33	462.00	1362.00	1069.00	900.00	30.74

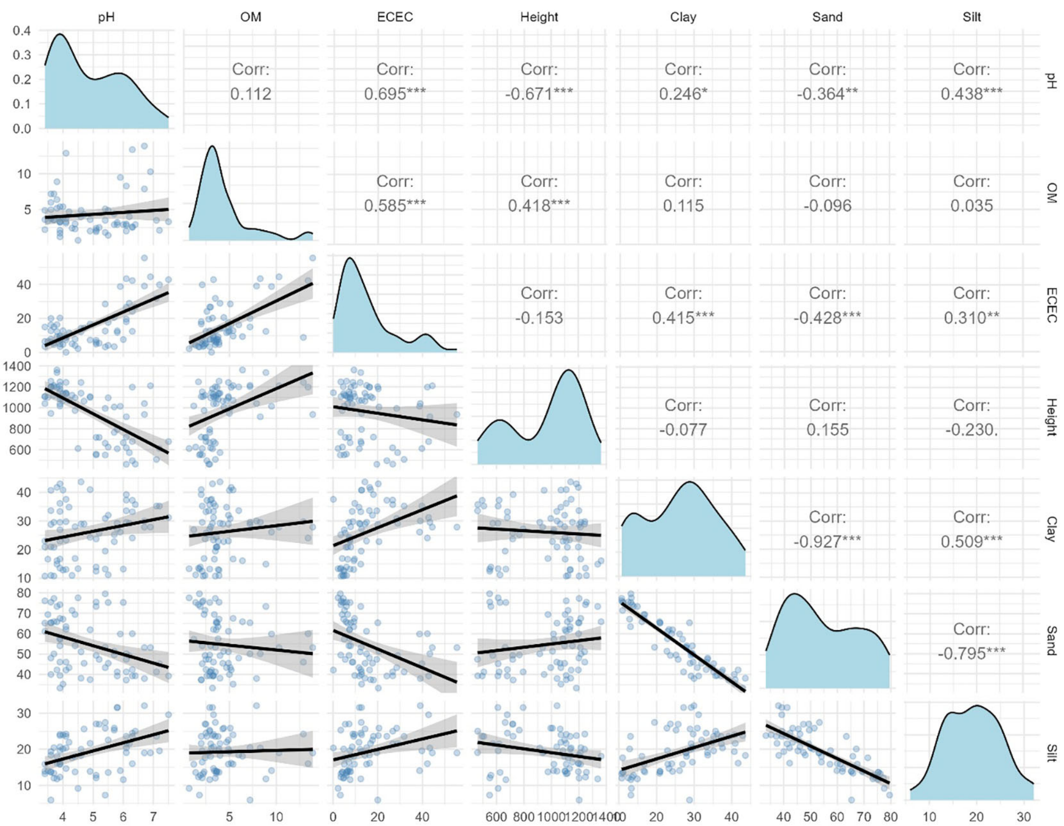

FIGURE 2
Matrix of non-parametric bivariate Spearman correlations between edaphic and altitudinal variables. * → $p < 0.05$ (significant) ** → $p < 0.01$ (very significant) *** → $p < 0.001$ (highly significant)

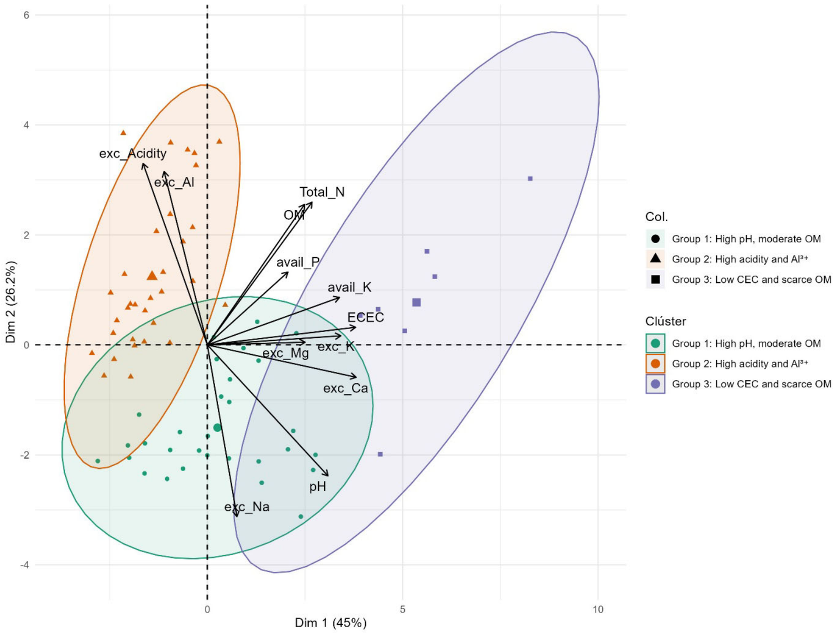

FIGURE 3
Principal component analysis biplot based on edaphic characteristics.

gradient that distinguishes highly acidic soils from those with greater base saturation. Dimension 2, meanwhile, primarily showed variations related to soil acidity and structural properties. Variables such as soil pH and exchangeable Na^+ were projected in the opposite direction to exchangeable acidity and Al^{3+} axes, indicating a possible differentiation of soil units with greater chemical stability. The analysis enabled grouping the observations into three edaphic clusters with quantifiable differences. Group 1 was characterized by higher pH values (M = 5.9) and medium to high ECEC levels (M = 19.6 cmol(+)·kg^{-1}), indicating more balanced chemical conditions. In contrast, Group 3 showed low ECEC values (M = 4.3 cmol(+)·kg^{-1}) and a predominance of active acidity (Al^{3+} and H^+), suggesting lower fertility and greater risk of Al^{3+} toxicity. Group 2, representing an intermediate condition, clustered plots with moderate ECEC but high acidity, highlighting a critical zone for soil management interventions. Overall, these results indicate that ECEC is strongly influenced by the concentrations of Ca^{2+}, Mg^{2+}, and OM, whereas its antagonistic relationship with exchangeable acidity reflects the competition between basic cations and protons for cation exchange sites in the soil. This multivariate classification supports the design of differential fertilization strategies and corrective amendments to optimize the agronomic management of coffee crops.

3.4 Principal component analysis by textural class

Principal component analysis (PCA) stratified by texture class was used to evaluate how different soil fractions (sand, silt, and clay) condition soil composition and, consequently, fertility in coffee-growing soils in the El Dorado district. The resulting biplot (Figure 4) showed a clear differentiation among texture classes, highlighting three dominant categories, clayey loam, sandy clay loam, and loamy sand, with well-defined groupings on the factorial plane.

Clayey loam soils were positioned in the right quadrant of the biplot, closely associated with variables such as ECEC, Ca^{2+}, Mg^{2+}, and OM, indicating favourable chemical behaviour. This group exhibited the highest ECEC levels (M = 28.65 cmol(+)·kg^{-1}, SD = 18.51) and elevated Ca^{2+} concentrations (M = 20.51 cmol (+)·kg^{-1}, SD = 17.99), supporting their greater nutrient retention capacity and active colloidal structure. Moreover, the organic matter content was notably high (M = 6.37%, SD = 3.52), reinforcing the hypothesis that this texture class promotes the accumulation of humic fractions and a higher density of effective surface charges.

In contrast, the loamy sand soils were positioned in the lower-left region of the biplot, far from the edaphic centroid and in the opposite direction of the ECEC and exchangeable base axes. These units were characterized by chemically limiting behaviour, with low ECEC values (M = 4.75 cmol(+)·kg^{-1}), reduced exchangeable Ca^{2+} (M = 3.62 cmol(+)·kg^{-1}), and scarce organic matter content (M = 1.92%). This pattern reflects the limited capacity of the exchange complex to retain nutrients, a typical condition of sandy soils, which exhibit weak aggregation and a minimal proportion of active colloidal fractions.

Sandy clay loam soils and intermediate textures exhibited greater dispersion across the factorial space, reflecting wide heterogeneity in their edaphic properties. Although their ECEC and exchangeable base contents were intermediate, they showed a marked tendency toward acidity, as indicated by their proximity vectors associated with Al^{3+} and exchangeable acidity. This textural pattern analysis reinforces the idea that soils with a higher clay content have superior nutrient retention capacity due to their greater specific surface area. These findings validate the use of PCA as an effective tool for delineating soil units with distinct management requirements, thus supporting agronomic zoning and

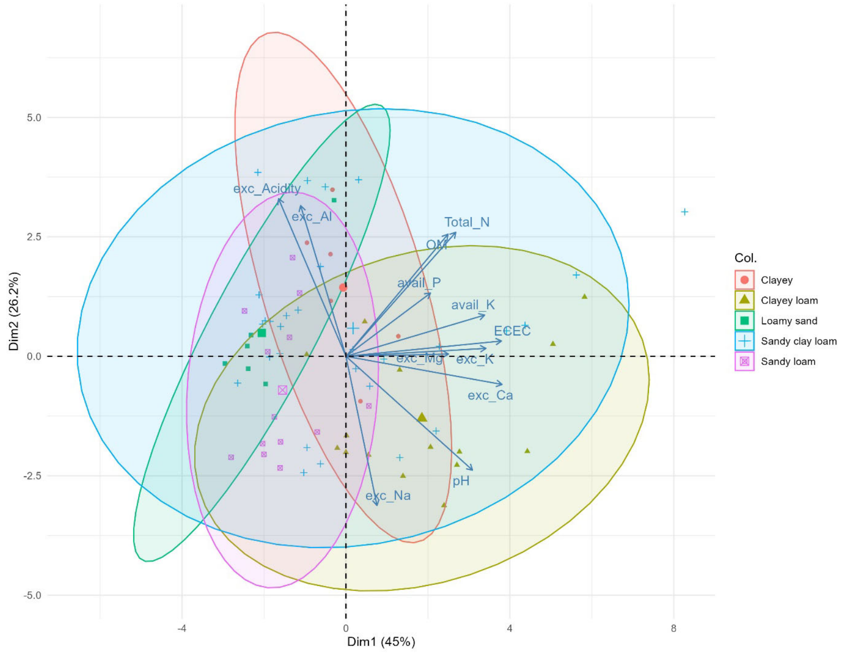

FIGURE 4
Principal component analysis biplot by soil texture class.

the design of site-specific fertilization strategies in coffee-growing systems.

3.5 Principal component analysis by altitude

Regarding altitudinal effects, plots located at high altitudes (>1150 m.a.s.l.) were predominantly clustered in the upper-left quadrant of the biplot, closely associated with vectors representing exchangeable acidity (exc_Acidity) and exchangeable aluminium (Al^{3+}). These soils exhibited more acidic chemical conditions, likely driven by intensified leaching processes under higher precipitation and lower organic matter mineralization rates. Statistically, this group showed lower mean pH values (M = 4.2) and higher Al^{3+} concentrations (M = 2.74 cmol(+)·kg^{-1}), suggesting a potentially restrictive edaphic environment for coffee cultivation in the absence of corrective management practices such as liming.

The medium-altitude group (900–1150 m.a.s.l.) showed greater dispersion in the PCA ordination space, indicating higher edaphic heterogeneity. Nevertheless, these soils were positioned in an intermediate zone between those dominated by active acidity and those enriched in exchangeable bases. This pattern suggests a transitional edaphic condition along the altitudinal gradient, where processes of moderate organic matter accumulation and partial saturation of the exchange complex coexist. Accordingly, this group displayed intermediate ECEC values (M = 12.6 cmol(+)·kg^{-1}) and an average organic matter content of 3.9%, supporting its characterization as a transition zone.

Finally, soils at lower elevations (<900 m.a.s.l.) were concentrated in the lower-right quadrant of the biplot, closely aligned with variables such as Ca^{2+}, Mg^{2+}, K^{+}, ECEC, and pH. This pattern indicates greater availability of exchangeable bases and lower levels of active acidity. These soils exhibited the highest mean ECEC (M = 22.3 cmol(+)·kg^{-1}) and Ca^{2+} content (M = 17.2 cmol(+)·kg^{-1}), suggesting higher fertility and greater cation exchange capacity at lower altitudes. This altitudinal stratification reinforces the hypothesis that altitude acts as an indirect modulator of soil fertility by influencing organic matter accumulation, pH regulation, and cation balance. Consequently, this environmental gradient should be explicitly considered in agroecological zoning and in the design of differential liming practices and fertilization strategies (Figure 5).

3.6 Generalized additive model

A generalized additive model (GAM) with an identity link function and a Gaussian error distribution was fitted to model effective cation exchange capacity (ECEC) as the dependent variable, using 15 soil and altitudinal predictors represented by penalized smoothing splines. The statistical contribution of each predictor was evaluated using the F statistic and the effective degrees of freedom (EDF). The results indicated that exchangeable calcium [s(Ca^{2+})] was the most influential predictor of ECEC (F = 1265.66; EDF = 1.00), with a highly significant non-linear effect. Exchangeable magnesium (s(Mg^{2+})) also exhibited a strong influence (F = 78.26; EDF = 1.00), followed by sand content (s(Sand %); F = 22.24; EDF = 0.41). Moderate effect predictors included s(Silt %) (F = 5.22; EDF = 0.89), s(Clay %) (F = 3.59; EDF = 0.71), s(Al^{3+}) (F = 1.96; EDF = 5.21), s(exchangeable acidity %) (F = 1.83; EDF = 1.82), and s(available P) (F = 1.75; EDF = 3.41). In contrast, several predictors showed limited statistical contribution to the model, including s(Na^{+}) (F = 0.36; EDF = 1.00), s(total N) (F = 0.23; EDF = 1.00), s(pH) (F = 0.23; EDF = 1.00), s(altitude) (F = 0.22; EDF = 1.00), s(K^{+}) (F = 0.14; EDF = 1.00), s(OM) (F = 0.013; EDF = 1.00), and s(available K) (F = 0.006; EDF = 1.00).

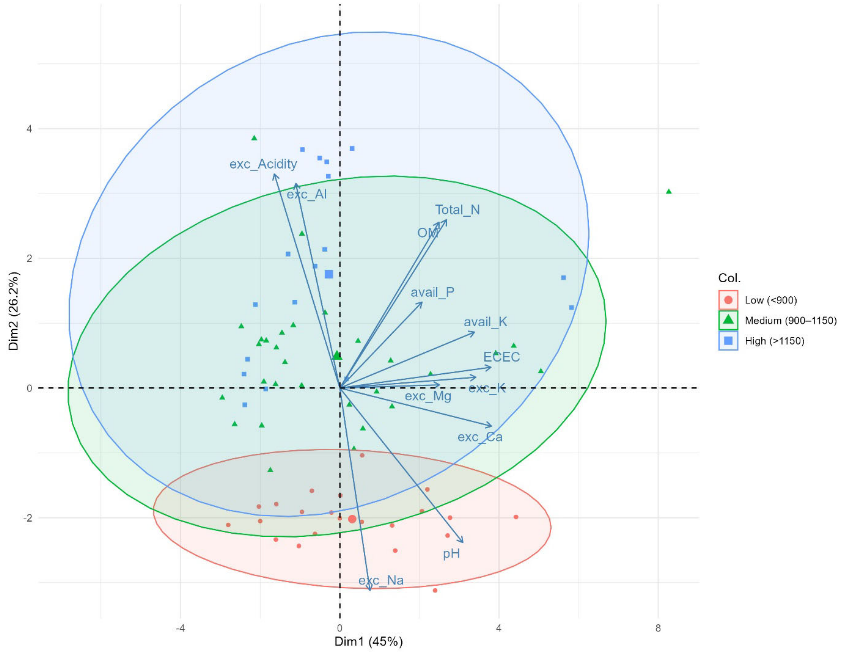

FIGURE 5
Principal component analysis biplot by altitude.

Overall, the model explained a high proportion of the total variability in ECEC (adjusted $R^2 = 0.89$), indicating an adequate fit to the observed data. Parameter estimation was conducted using the restricted maximum likelihood (REML) method, which is well-suited for penalized regression models incorporating multiple smooth terms (Figure 6).

3.7 Random forest model

A Random Forest (RF) model was implemented to predict effective cation exchange capacity (ECEC) in coffee-growing soils of northeastern Peru, using 15 soil and altitudinal variables as predictors. Variable importance was assessed using the percentage increase in mean squared error (IncMSE%), which estimates the loss of model accuracy when the values of a given predictor variable are randomly permuted.

The results indicated that exchangeable calcium (Ca^{2+}) was the most influential variable in predicting ECEC, with an importance value of 3177.37 IncMSE%, highlighting its dominant and strongly non-linear relationship with the response variable. This was followed by soil pH (1638.36 IncMSE%) and exchangeable acidity (1595.95 IncMSE%), both of which also exerted a substantial influence on model performance. Variables with intermediate importance included exchangeable potassium (K^+; 1021.95 IncMSE%), exchangeable magnesium (Mg^{2+}; 667.14 IncMSE%), total nitrogen (N; 526.70 IncMSE%), available potassium (K; 428.54 IncMSE%), and organic matter (OM; 365.97 IncMSE%). In contrast, variables with relatively lower contributions to model accuracy were clay content (%; 212.31 IncMSE%), sand content (%; 169.46 IncMSE%), available phosphorus (P; 128.65 IncMSE%), altitude (99.74 IncMSE%), exchangeable sodium (Na^+; 91.94 IncMSE%), silt content (%; 81.92 IncMSE%), and exchangeable aluminium (Al^{3+}; 70.60 IncMSE%).

From a statistical perspective, the Random Forest model revealed a clear hierarchy in variable importance, dominated by basic cations, particularly Ca^{2+}, and by parameters related to soil reaction (pH and exchangeable acidity). This pattern is characteristic of non-parametric models, which are well-suited to capturing complex non-linear interactions and combined effects among soil properties.

The Random Forest model was fitted with 500 decision trees (ntree = 500), achieving high predictive performance with a validated coefficient of determination (R^2) of 0.93 and a root mean square error (RMSE) of approximately 2.1 units of ECEC. These results indicate satisfactory model fit and strong predictive capability. Finally, prediction variables with comparatively low influence on model performance included the silt fraction [Silt (%), IncMSE = 114.7%] and exchangeable aluminium (Al^{3+}, IncMSE = 101.3%), suggesting a limited direct contribution to ECEC variability within the modelled system.

Overall, the Random Forest model exhibited high predictive performance, with a coefficient of determination (R^2) of 0.93 and a low root mean square error (RMSE) following cross-validation. These results demonstrate the model's robustness and its effectiveness in identifying and ranking critical edaphic factors controlling effective cation exchange capacity (ECEC) in tropical coffee-growing ecosystems (Figure 7).

3.8 Performance cross-validation between the GAM and random forest models

K-fold cross-validation (k = 5) was applied to compare the predictive performance between the generalized additive model (GAM) and the Random Forest (RF) model, both designed to estimate effective cation exchange capacity (ECEC) based on soil physico-chemical properties and altitudinal variables.

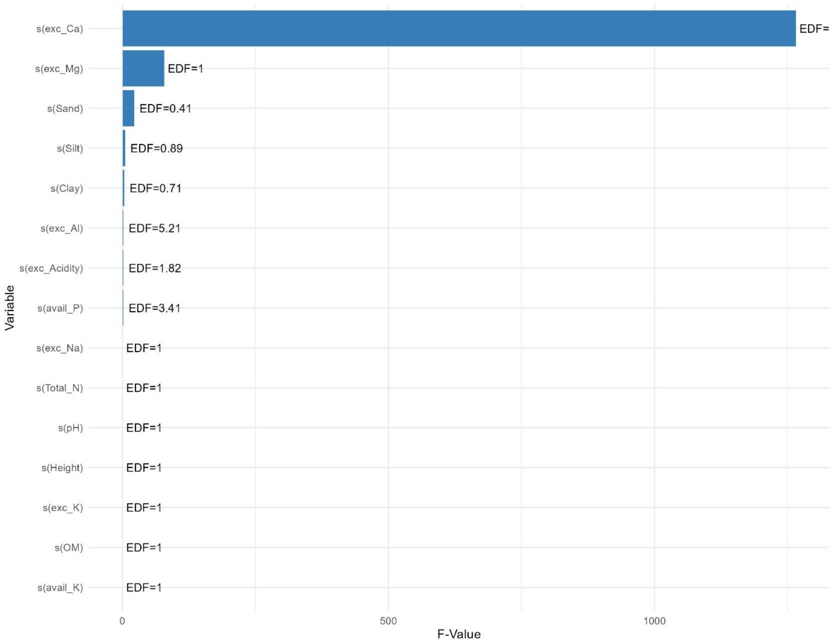

FIGURE 6
Relative importance of predictors in the generalised additive model.

The results demonstrated clear superiority of the Random Forest model, which achieved a mean validated coefficient of determination (R^2) of 0.93 and a root mean square error (RMSE) of approximately 2.1 ECEC units. In contrast, the GAM exhibited slightly lower predictive performance, with a validated R^2 of 0.89 and an RMSE of approximately 3.4 CICe units.

These results indicate that, although both models exhibit high predictive performance, the Random Forest model explains a greater proportion of the variance in ECEC and yields a lower prediction error. This difference suggests that Random Forest captures complex nonlinear relationships and interactions among soil variables more effectively, whereas the GAM is particularly useful for interpreting smoothed individual effects.

3.9 Spatial distribution of cation exchange capacity

Geostatistical analysis of the 69 soil samples revealed a significant spatial structure in the distribution of CEC across the study area. Spatial autocorrelation analysis, assessed using Moran's I

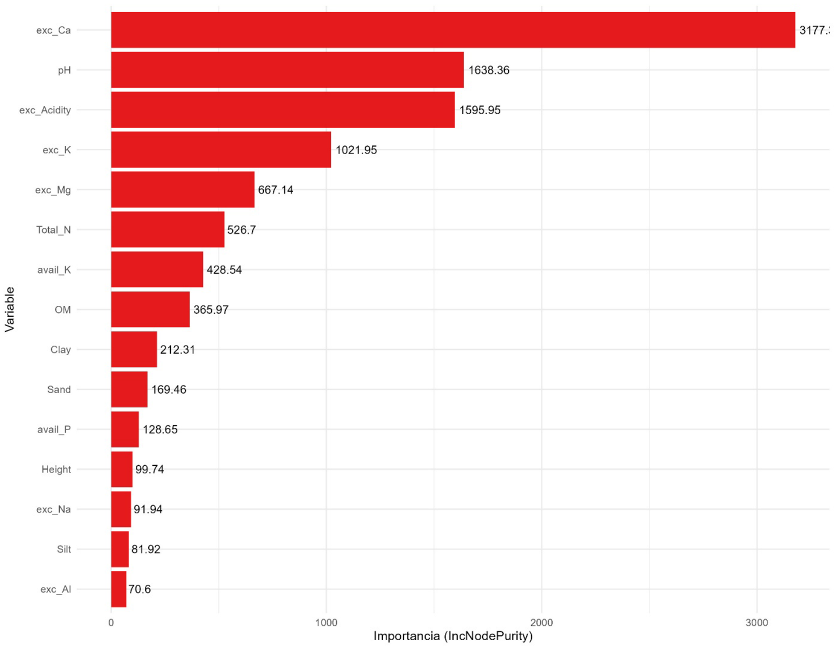

FIGURE 7
Relative importance of predictors in the random forest model.

statistic, yielded a value of 0.1314 ($p = 0.015$), indicating statistically significant positive spatial clustering. This pattern suggests that areas with high CEC values tend to be located near other high-value areas, while low CEC values are similarly spatially associated. These findings confirm the presence of spatial dependence in CEC and validate the use of kriging interpolation methods to predict CEC at unsampled locations.

Variogram modelling identified the exponential model as the best-fitting theoretical structure, characterized by a nugget effect of 49.37 (representing 23.5% of the total variance), a partial sill of 160.44, and a range of 1,199 m (Figure 8). The spatial dependency ratio of 76.5% was classified as strong, indicating that most of the variability in CEC is spatially structured rather than random, thus providing favourable conditions for reliable kriging interpolation. This relatively large range suggests that CEC remains spatially correlated over distances exceeding one kilometre, reflecting the influence of large-scale environmental gradients, likely driven by topographic and edaphic factors across the landscape.

The application of universal kriging (UK), incorporating elevation as an external drift to capture spatial trends, generated continuous prediction surface interpolations of CEC across the entire study area at a 50 m spatial resolution (Figure 9), along with kriging variance maps, which were produced to quantify the spatial uncertainty associated with the predictions (Figure 10). The uncertainty analysis indicated a mean kriging variance of 180.86, with approximately 11.7% of the study area classified as high uncertainty (variance > 228.72). This relatively limited extent of high-uncertainty zones suggests that the sampling design provided generally adequate spatial coverage, resulting in moderate to high confidence in most predictions. Areas of low prediction variance were primarily located near sampling points and in regions where the elevation covariate effectively explained CEC variability. In contrast, zones of higher uncertainty were concentrated in sparsely sampled areas, where spatial extrapolation is inherently less reliable (Figures 9, 10).

Validation of the model using 5-fold cross-validation yielded mixed performance metrics, warranting careful interpretation. The RMSE of 11.57 and the MAE of 8.36 represent moderate prediction errors relative to the observed range of CEC values in the dataset. However, the low coefficient of determination ($R^2 = 0.127$) suggests that the model explains only 12.7% of the total CEC variance, reflecting limited predictive capacity. Despite this, the model demonstrated excellent calibration, with a minimal bias of 0.211 (equivalent to 1.39% of the mean CEC), indicating the absence of systematic over- or underestimation across the prediction range. This unbiased behaviour, together with the strong spatial dependence identified in the variogram analysis, supports the model's predictive utility for identifying broad spatial patterns and delineating management zones. However, individual predictions should be interpreted with caution.

4 Discussion

4.1 Principal component analysis by soil characteristics

The high edaphic variability observed in coffee-growing soils of north-eastern Peru, particularly in pH, organic matter (OM), exchangeable acidity, and ECEC, reflects a marked agroecological complexity that requires multivariate approaches for proper interpretation (27, 28). The mean soil pH (4.87) confirms the predominance of acidic soils, which are negatively associated with Al^{3+} and active acidity, as evidenced by significant correlations ($r = -0.59$ and -0.65, respectively) and negative loadings on the first PCA dimension (Dim1) (29). ECEC showed strong associations with Ca^{2+}, Mg^{2+}, and OM across PCA, GAM, and Random Forest models, reaffirming its role as a key indicator of soil fertility in tropical environments (30, 31). The GAM identified exchangeable Ca^{2+} as the most influential predictor ($F = 1{,}265.66$), while the Random Forest model ranked it as the most important variable (IncMSE% = 3,177.37), surpassing even soil pH and exchangeable acidity. These findings are consistent with previous studies highlighting the dominant role of basic cations in nutrient retention and exchange processes in coffee-growing agroecosystems (32, 33).

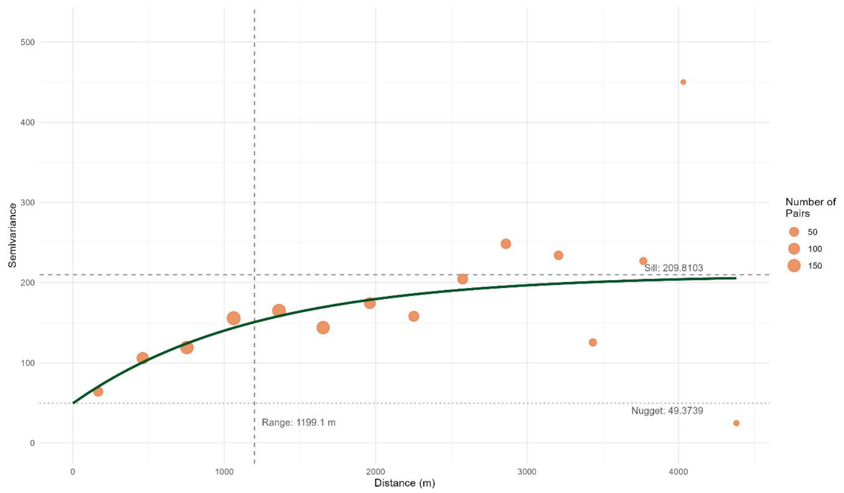

FIGURE 8
Variogram fitted with an exponential model.

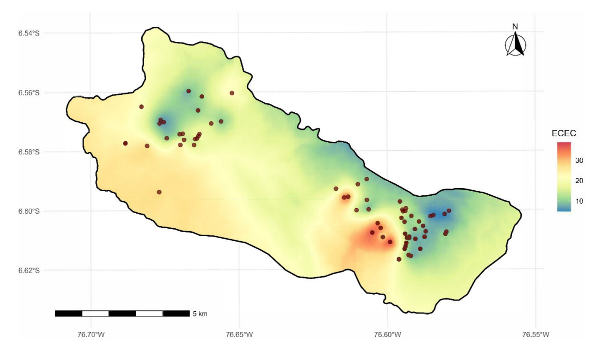

FIGURE 9
Cation exchange capacity interpolation using universal kriging.

Soil texture also strongly influenced ECEC distribution. Clay loam soils exhibited the highest values of ECEC (M = 28.65 cmol(+)·kg^{-1}), OM (M = 6.37%), and exchangeable Ca^{2+} (M = 20.51 cmol(+)·kg^{-1}), supporting their superior colloidal activity and charge retention capacity (34, 35). In contrast, sandy-textured soils were characterized by low ECEC and OM contents, indicating reduced exchange capacity and greater susceptibility to nutrient leaching, a pattern commonly observed in coarse-textured tropical soils (36).

4.2 Principal component analysis by altitude

The altitudinal gradient emerged as another key modulator of soil fertility. High-altitude soils (>1,150 m.a.s.l.) were associated with greater acidity and higher Al^{3+} concentrations, whereas low-altitude soils (<900 m) exhibited higher ECEC and base saturation. This vertical stratification aligns with previous studies that have associated altitude with OM accumulation and soil acidification (37). Furthermore, the positive correlation between altitude and OM (r = 0.53***) supports the hypothesis that higher elevations promote organic matter accumulation due to lower temperatures and reduced mineralization rates (38). Furthermore, temperature appears to be a determining factor influencing the edaphic characteristics of soils under coffee vegetation cover along the altitudinal gradient, likely due to slower decomposition processes (39).

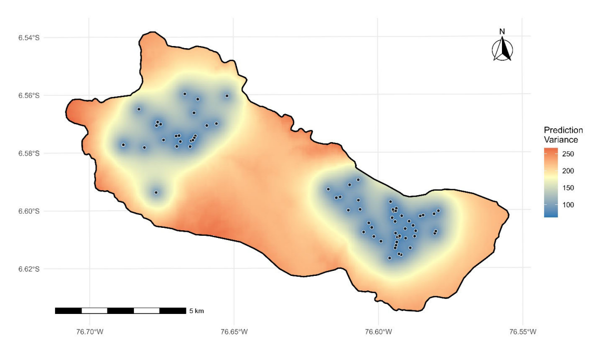

FIGURE 10
Uncertainty map based on kriging variance.

4.3 Performance cross-validation between the GAM and random forest models

Cross-validation between the GAM and RF models demonstrated superior predictive performance of the Random Forest approach (R^2 = 0.93; RMSE = 2.1), suggesting that non-parametric models are better suited to capturing complex non-linear interactions among soil variables (24, 40). Nevertheless, the GAM offered greater interpretability of smoothed individual effects, which is particularly valuable for designing targeted agronomic recommendations and soil management strategies (41).

The variogram fitted to the exponential model showed a moderate nugget effect, indicating acceptable data quality, some short-range variability, and a well-defined spatial structure suitable for kriging. The estimated range, exceeding 1,000 m when defining spatial influence, suggests the presence of large-scale spatial patterns and supports the efficiency of kriging even with relatively dispersed sampling designs. Variograms can also be used to guide sampling strategies; for instance, the sampling interval is often recommended to be less than half of the spatial dependence range. In this context, Kerry et al. (42) proposed defining appropriate sampling intervals for future soil studies using average variogram ranges and standardized mean variograms derived from four parent materials in southern England, adopting half the variogram range as a practical reference.

The uncertainty maps generated in this study provide practical guidance for adaptive sampling strategies by identifying specific locations where additional soil samples would most effectively improve prediction accuracy in future assessments. Digital soil mapping products provide estimates of soil properties that vary by location. These estimates have some spatial uncertainty, which is not evenly distributed across the area. Explicitly quantifying spatial uncertainty enhances mapping quality by enabling the optimization of sampling efforts, particularly in areas with sparse observations or high prediction variance (43).

4.4 Spatial distribution of cation exchange capacity

The low R^2 values obtained during model validation can be attributed to several factors: (1) the high intrinsic variability of CEC, driven by complex soil-forming processes that are not fully captured by elevation alone; (2) potential non-stationarity in the spatial structure across a heterogeneous landscape; and (3) the need of additional environmental covariates required to model spatial trends adequately. Consequently, several improvements are recommended. First, targeted additional sampling in areas of high uncertainty would help reduce prediction variance. Second, incorporating complementary environmental covariates (e.g., terrain derivatives, vegetation indices, or climate variables) could enhance the performance of Universal Kriging (UK). Finally, the application of alternative geostatistical approaches, such as regression kriging, is strongly recommended. Due to its popularity, practicality, and robustness, regression kriging is widely used as a hybrid spatial interpolation technique within the digital soil mapping toolkit to model soil distribution patterns across multiple spatial and temporal scales (44).

One of the limitations of this study was the lack of assessment of soil microfauna, which plays an important role in carbon aggregate formation and soil stabilization (45). Although it is true that the coffee crop is a mycotrophic crop, in this study the fungal potential and its response to the edaphic characteristics at the altitudinal level were not evaluated. Coffee is highly dependent on symbiotic associations with arbuscular mycorrhizal fungi, which enhance nutrient uptake and contribute to shaping soil properties (46). In addition, redundancy analysis revealed that microbial phyla were closely associated with soil pH. These findings suggest that *Coffea arabica* cultivation at mid- to high-elevation zones may favour more sustainable management and improved bean quality (47). Furthermore, edaphic properties are strongly influenced by the continuous input of plant residues across different coffee production systems.

5 Conclusions

This study demonstrated, through a multivariate statistical approach, that effective cation exchange capacity in coffee-growing soils of north-eastern Peru is strongly influenced by key edaphic variables, particularly exchangeable calcium, magnesium, organic matter, soil texture, and altitude. Descriptive statistics revealed high data dispersion, justifying the application of non-linear models and dimensionality reduction techniques. Pearson's correlation analysis showed significant positive associations between ECEC and OM, clay content, and altitude. In contrast, soil pH exhibited strong negative correlations with exchangeable acidity and exchangeable aluminium, indicating a structure of edaphic relationships consistent with leaching and base saturation processes.

The Random Forest (RF) model achieved a validated R^2 of 0.93 and an RMSE of 2.1, clearly outperforming the GAM in predictive capacity. The variable importance ranking identified exchangeable calcium soil pH (1,638.36), and exchangeable acidity as the most influential predictors, which is entirely consistent with the patterns observed in the PCA and Pearson correlation analyses. Five-fold cross-validation (k = 5) confirmed the robustness and stability of the RF model, with low fold-to-fold variability and high accuracy in ECEC estimation. In the spatial modelling framework, the use of elevation as a covariate in Universal Kriging (UK) enabled the generation of continuous ECEC prediction surfaces with good spatial coverage. Only 11.7% of the study area exhibited high prediction uncertainty, indicating that the sampling strategy was generally adequate for reliable spatial interpolation. This assessment is supported by the relatively low mean kriging variance (180.86) and the presence of high-variance areas in unsampled regions. Overall, integrating variogram-based spatial structure with elevation data enabled reliable ECEC predictions, reinforcing the suitability of the adopted sampling and modelling strategy.

Although the model exhibited limited explanatory power, its excellent calibration and moderate prediction errors indicate that the approach is suitable for identifying general spatial patterns and delineating soil management zones. Despite the low R^2 value, the predictions should be interpreted with caution, as the model is more effective at capturing broad trends than local-scale variability. These findings highlight the importance of evaluating geostatistical models not only for predictive accuracy but also for calibration and spatial reliability. The methodology and results of this study can be extrapolated to other tropical regions with similar soil and climatic conditions, providing a robust basis for designing differentiated soil management strategies that optimize fertilizer use efficiency and promote agro-environmental sustainability at the regional scale.

An important aspect of this study is that shaded and high-elevation coffee systems tend to accumulate greater amounts of soil organic matter; consequently, soils exhibit an enhanced capacity to retain nutrients. Therefore, coffee sustainability depends more on ecologically sound soil management than on a single isolated factor, and regenerative practices such as vegetation cover, shade management, and the incorporation of organic residues are essential to maintain the resilience of the production system under climate change and soil degradation scenarios. Furthermore, soil texture represents a key factor influencing coffee sustainability: greater nutrient retention capacity and improved nutritional balance are associated with higher potential for sustainable production and reduced environmental impact.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HD-C: Resources, Funding acquisition, Data curation, Writing – original draft, Project administration, Conceptualization, Investigation, Validation, Methodology. LM: Writing – original draft, Visualization, Validation, Conceptualization, Methodology. MS: Investigation, Resources, Writing – review & editing, Formal analysis. JC-G: Writing – review & editing, Conceptualization, Methodology, Resources. CC-L: Project administration, Methodology, Investigation, Writing – original draft, Resources. NC-C: Visualization, Validation, Methodology, Investigation, Writing – original draft. BM: Conceptualization, Resources, Writing – original draft, Project administration, Data curation. GV-T: Methodology, Investigation, Visualization, Writing – original draft, Validation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The research was funded by the Instituto Nacional de Innovación Agraria, within the framework of the 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.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Solis R, Gomales N, Pezo M, Arévalo L, Vallejos-Torres G. Rooting of sacha inchi (Plukenetia volubilis) juvenile cuttings in microtunnels. *Acta Agronómica*. (2019) 68:35–40. doi: 10.15446/acag.v68n1.72101

2. FAO. *World food and agriculture statistical yearbook 2024*. Food and Agriculture Organization of the United Nations (2024). Available online at: https://www.fao.org/statistics/yearbook (Accessed January 19, 2026).

3. Agraria.pe. *JNC: Perú cuenta con 375–000 hectáreas de café este año* (2022). Available online at: https://agraria.pe/noticias/jnc-peru-cuenta-con-375-000-hectareas-de-cafe-este-ano-37178 (Accessed January 12, 2026).

4. Wei T, Simko V. *R package corrplot: Visualization of a correlation matrix (versión 0.95) [Software]* (2024). Available online at: https://github.com/taiyun/corrplot (Accessed January 09, 2026).

5. Rodríguez Suárez L, Cuarán Pinto SP, Suárez Salazar JC. Soil macrofauna and edaphic properties in coffee production systems in southern Colombia. *Floresta e Ambiente*. (2019) 26:e20180334. doi: 10.1590/2179-8087.033418

6. Posma M, Tampubolon B, Sihombing K, Simanjuntak FN, Irly DR, Sinuraya I, et al. Identification of soil physicochemical, land suitability, and its relationship to Coffee arabica yielding based on plant age groups. In: *Coffee Science*, vol. 18. UFLA, Lavras. (2023). Available online at: https://sbicafe.ufv.br/items/bc9b7cf8-cb64-40ef-b9e5-b06f05219631 (Accessed December 18, 2025).

7. Getachew T, Gebrehiwot T, Getachew Y. Soil temperatures, chemical properties and coffee bean quality along elevation gradients in Ethiopia. *Agron Sustain Dev*. (2022) 42:18. doi: 10.1007/s13593-022-00801-8

8. Onwualu AP, Chukwu O, Ezeh CE. Biochar based fertilizers increase crop yields in acidic tropical soils. *Biofuels Bioprod Biorefin*. (2024) 19(4):1124–42. doi: 10.1002/bbb.2777

9. Revelle W. *psych: Procedures for Personality and Psychological Research*. Northwestern University (2023). Available online at: https://cran.r-project.org/package=psych (Accessed January 16, 2026).

10. Maita SM, Quispe K, Díaz-Chuquizuta H, Rengifo Sanchéz R, Mercado Chinchay R, Cuevas Gimenez JP, et al. Soil spatial variability in high-yield Peruvian Amazon coffee: A geostatistical approach for precision fertilization. *Front Soil Sci*. (2025) 5:1701602. doi: 10.3389/fsoil.2025.1701602

11. Ahmed S, Brinkley S, Smith E, Sela A, Theisen M, Thibodeau C, et al. Climate change and coffee quality: Systematic review on the effects of environmental and management variation on secondary metabolites and sensory attributes of Coffea arabica and Coffea canephora. *Front Plant Sci*. (2021) 12:708013. doi: 10.3389/fpls.2021.708013

12. SENAMHI. *Datos climatológicos históricos de la región San Martín*. Servicio Nacional de Meteorología e Hidrología del Perú (2024). Available online at: https://www.senamhi.gob.pe (Accessed January 26, 2026).

13. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (2023). Available online at: https://www.R-project.org/ (Accessed January 22, 2026).

14. Bouyoucos GJ. Hydrometer method improved for making particle size analyses of soils. *Agron J*. (1962) 54:464–5. doi: 10.2134/agronj1962.00021962005400050028x

15. U.S. Environmental Protection Agency. *Method 9045D: Soil and Waste pH*. Washington, D.C.: U.S. Environmental Protection Agency (2004). pp. 1–5. pp. 1–5.

16. Secretaría de Medio Ambiente y Recursos Naturales – SEMARNAT. NOM-021-RECNAT-2000: Especificaciones de fertilidad, salinidad y clasificación de suelos. Estudios, muestreo y análisis. In: *Diario Oficial de la Federación* Ciudad de México: Diario Oficial de la Federación. (2000). Available online at: https://biblioteca.semarnat.gob.mx/janium/Documentos/Ciga/libros2009/DO2280.pdf (Accessed December 23, 2025).

17. International Organization for Standardization. *Soil quality — Determination of total nitrogen — Modified Kjeldahl method (ISO 11261:1995)* (1995). Available online at: https://www.iso.org/standard/19239.html (Accessed January 22, 2026).

18. Bray RH, Kurtz LT. Determination of total, organic, and available forms of phosphorus in soils. *Soil Sci*. (1945) 59:39–45. doi: 10.1097/00010694-194501000-00006

19. Helmke PA, Sparks DL. Lithium, sodium, potassium, rubidium, and cesium. In: Sparks DL, editor.*Methods of Soil Analysis Part 3: Chemical Methods*. Madison, WI, USA: Soil Science Society of America (1996). p. 551–74. doi: 10.2136/sssabookser5.3.c19

20. Sumner ME, Miller WP. Cation exchange capacity and exchange coefficients. In: Sparks DL, editor.*Methods of Soil Analysis: Part 3, Chemical Methods*. Madison, WI, USA: Soil Science Society of America (1996). p. 1201–29. doi: 10.2136/sssabookser5.3.c40

21. Wickham H. *ggplot2: Elegant graphics for data analysis*. Cham, Switzerland: Springer (2016). doi: 10.1007/978-3-319-24277-4

22. Wood SN. *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation (Versión 1.9-3) [Paquete R]* (2025). Available online at: https://cran.r-project.org/package=mgcv.

23. Liaw A, Wiener M. Classification and regression by randomForest. In: *R News*, vol. 2. Viena: R Foundation for Statistical Computing. (2002). p. 18–22. Available online at: https://CRAN.R-project.org/doc/Rnews/ (Accessed January 01, 2025).

24. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

25. Scikit learn Developers. *Cross validation: evaluating estimator performance. En Reference documentation. Scikit learn* (2024). Available online at: https://scikit-learn.org/stable/modules/cross_validation.html (Accessed January 03, 2026).

26. NASA JPL. NASADEM Merged DEM Global 1 arc second V001 [Data set. *NASA Land Processes Distributed Active Arch Center*. (2020). Available online at: https://www.earthdata.nasa.gov/data/catalog/lpcloud-nasadem-hgt-001 (Accessed November 26, 2025).

27. Padhi SR, Bartwal A, John R, Tripathi K, Gupta K, Wankhede DP, et al. Evaluation and multivariate analysis of cowpea [Vigna unguiculata (L.) Walp] germplasm for selected nutrients—mining for nutri-dense accessions. *Front Sustain Food Syst*. (2022) 6:888041. doi: 10.3389/fsufs.2022.888041

28. Strahlhofer M, Gerzabek MH, Rampazzo N, Couenberg PM, Vera E, Salazar X, et al. Soil fertility changes with climate and island age in Galápagos: New baseline data for sustainable agricultural management. *Front Environ Sci*. (2021) 9:788082. doi: 10.3389/fenvs.2021.788082

29. Dhandapani S, Sng YH, Agisha VN, Suraby EJ, Park BS. Mitigating aluminum toxicity and promoting plant resilience in acidic soil with Penicillium olsonii TLL1. *Front Plant Sci*. (2024) 15:1423617. doi: 10.3389/fpls.2024.1423617

30. Song X, Chen C, Arthur E, Tuller M, Zhou H, Shang J, et al. Cation exchange capacity and soil pore system play key roles in water vapour sorption. *Geoderma*. (2022) 424:116017. doi: 10.1016/j.geoderma.2022.116017

31. Rumpel C, Kögel-Knabner I. Deep soil organic matter—a key but poorly understood component. *Plant Soil*. (2011) 338:143–58. doi: 10.1007/s11104-010-0391-5

32. Antonangelo JA, Culman S, Zhang H. Comparative analysis and prediction of cation exchange capacity via summation: Influence of biochar type and nutrient ratios. *Front Soil Sci*. (2024) 4:1371777. doi: 10.3389/fsoil.2024.1371777

33. Tong L, Li J, Zhu L, Zhang S, Zhou H, Lv Y, et al. Effects of organic cultivation on soil fertility and soil environment quality in greenhouses. *Front Soil Sci*. (2022) 2:1096735. doi: 10.3389/fsoil.2022.1096735

34. Bünemann EK, Bongiorno G, Bai, Creamer RE, De Deyn G, de Goede R, et al. Soil quality—a critical review. *Soil Biol Biochem*. (2018) 120:105–25. doi: 10.1016/j.soilbio.2018.01.030

35. Lal R. Restoring soil quality to mitigate soil degradation. *Sustainability*. (2015) 7:5875–95. doi: 10.3390/su7055875

36. Fageria NK. Role of soil organic matter in maintaining sustainability. *Commun Soil Sci Plant Anal*. (2012) 43:2063–113. doi: 10.1080/00103624.2012.697234

37. de la Cruz-Amo L, Bañares-de-Dios G, Cala V, Granzow-de la Cerda Í, Espinosa CI, Ledo A, et al. Trade-offs among aboveground, belowground, and soil organic carbon stocks along altitudinal gradients in Andean tropical montane forests. *Front Plant Sci*. (2020) 11:106. doi: 10.3389/fpls.2020.00106

38. Zhang B, Xu C, Zhang Z, Hu C, Zhong C, Chen S, et al. Elevational patterns of soil organic carbon and its fractions in tropical seasonal rainforests in karst peak-cluster depression region. *Front Plant Sci*. (2024) 15:1424891. doi: 10.3389/fpls.2024.1424891

39. Vallejos-Torres G, Gaona-Jimenez N, Lozano A, Paredes C, Lozano CM, Alva-Arévalo A, et al. Soil organic carbon balance across contrasting plant cover ecosystems in the Peruvian Amazon. *Chil J Agric Res*. (2023) 83:553–64. doi: 10.4067/s0718-58392023000500553

40. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, et al. Random forests for classification in ecology. *Ecology*. (2007) 88:2783–92. doi: 10.1890/07-0539.1

41. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology. *PeerJ*. (2019) 7:e6876. doi: 10.7717/peerj.6876

42. Kerry R, Oliver MA. Average variograms to guide soil sampling. *Int J Appl Earth Obs Geoinf*. (2004) 5:307–25. doi: 10.1016/j.jag.2004.07.005

43. Stumpf F, Schmidt K, Goebes P, Behrens T, Schönbrodt-Stitt S, Wadoux A, et al. Uncertainty-guided sampling to improve digital soil maps. *Catena*. (2017) 153:30–8. doi: 10.1016/j.catena.2017.01.033

44. Omuto CT, Vargas RR. Re-tooling of regression kriging in R for improved digital mapping of soil properties. *Geosci J*. (2015) 19:157–65. doi: 10.1007/s12303-014-0023-9

45. dos Santos Nascimento M, Barreto-Garcia PAB, Monroe PHM, Pereira MG, Barros WT, Nunes MR. Carbon in soil macroaggregates under coffee agroforestry systems: Modeling the effect of edaphic fauna and residue input. *Appl Soil Ecol*. (2024) 202:105604. doi: 10.1016/j.apsoil.2024.105604

46. Vallejos Torres G, Sánchez T, García MA, Trigoso M, Arévalo LA. Efecto de hongos formadores de micorrizas arbusculares en clones de Coffea arabica, variedad caturra en Perú. *Acta Agronómica*. (2019) 68:278–84. doi: 10.15446/acag.v68n4.72117

47. Ge Y, Zhang F, Xie C, Qu P, Jiang K, Du H, et al. Effects of different altitudes on Coffea arabica rhizospheric soil chemical properties and soil microbiota. *Agronomy*. (2023) 13:471. doi: 10.3390/agronomy13020471