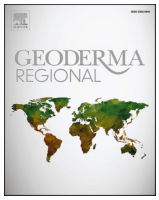
Geoderma Regional 38 (2024) e00820

Contents lists available at ScienceDirect

Geoderma Regional

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



Spatial variability of clay minerals in a semi-arid region of Turkiye

Hikmet Günal [a,*], Nurullah Acir [b]

[a] Harran University, Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Sanliurfa, Turkiye
[b] Kırsehir Ahi Evran University, Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Kırşehir, Turkiye

ARTICLE INFO

Keywords:
Clay mineral
Covariate
Modeling
Spatial distribution
Prediction
Artificial intelligence

ABSTRACT

Clay minerals are the source of many chemical and physical properties that influence the provision of soil-based ecosystem services. This study aimed to identify the most significant soil characteristics contributing to the spatial variability of clay minerals in a semi-arid region of Turkiye. Additionally, the study assessed the predictive capabilities of Classification and Regression Tree (CART), Random Forest Regression (RF) and eXtreme Gradient Boosting Regression (XGBoost) in estimating soil clay mineral content. Smectite+vermiculite (SMVR) was the most abundant clay mineral in the study area, followed by illite and kaolinite. Hyperparameter tuning significantly improved model accuracy, with root mean square error (RMSE) reductions ranging from 2.53% to 97.3%. The machine learning algorithms demonstrated varying performances in spatial prediction accuracy. The RF model achieved the lowest RMSE (8.587%) and the highest R^2 values (0.796) for predicting SMVR. The XGBoost outperformed other models for kaolinite (RMSE: 4.814%, R^2:0.713) and illite (RMSE:7.368%, R^2:0.613). Exchangeable cations, particularly magnesium (Mg) and calcium (Ca), were identified as crucial factors influencing the spatial distribution of clay minerals. Among these, M concentration had the strongest influence on predicting both SMVR (38.1%) and illite (26.3%). Conversely, for kaolinite prediction, Ca concentration played the most significant role (38.7%), followed by Mg (19.93%). In conclusion, this study demonstrates the effectiveness of machine learning models, particularly XGBoost which achieved the lowest RMSE for all clay minerals investigated. These models offer a valuable tool for predicting clay mineral content in the Kazova Plain. The findings highlight the importance of parent material, weathering processes, and specific soil properties, such as exchangeable cations, in shaping clay mineral distribution. This knowledge not only contributes to a deeper understanding of soil formation in semi-arid environments but also practical applications. For instance, by predicting the abundance of SVMR, known for its high cation exchange capacity land managers can develop targeted strategies for optimizing fertilizer application in the Kazova Plain.

1. Introduction

Soils consist of minerals (products of rock weathering), organic matter and water. The parent material, essentially the starting point for soil formation, exerts a strong influence on a mineralogical composition of soil. This, in turn, affects various soil properties and processes. The type of parent material dictates the initial mineral content, and through weathering over time, contributes to the development of specific mineral compositions in the soil (Daher et al., 2019). The clay mineral composition, an intrinsic attribute of a soil, is primarily influenced by parent material (Spinola et al., 2022; Silva et al., 2024).

Due to their high surface area, clay minerals are the most active mineral component in soils. The inherent structural characteristic allows them to play a crucial role in shaping the physical, chemical, and biological characteristics of the soil. Their high surface area allows clay minerals interact with other soil components, facilitating functions such as long-term storage of organic matter, retention of water, nutrients, and even heavy metals. Understanding these contributions is essential for soil dynamics, biogeochemical processes, and ecosystem functioning (Viscarra Rossel, 2011; Sarkar et al., 2018; Wakeel and Ishfaq, 2022). Knowledge gained about clay minerals in soils is invaluable for improving existing soil management practices and promoting sustainable agricultural production.

A comprehensive understanding of soil properties, encompassing physical, chemical, mineralogical, and biological soil aspects, is essential for ensuring the sustainable use and effective management of agricultural soils. Clay minerals play a particularly significant role in shaping these characteristics. High surface area clay minerals influence

* Corresponding author.
E-mail address: hikmetgunal@harran.edu.tr (H. Günal).

https://doi.org/10.1016/j.geodrs.2024.e00820
Received 20 March 2024; Received in revised form 31 May 2024; Accepted 1 June 2024
Available online 3 June 2024

soil properties like porosity, plasticity, erodibility, and fertility by binding soil particles, promoting aggregate stability, and improving overall porosity (Kasanin-Grubin, 2013). Furthermore, the high surface area and negatively charged nature of clay minerals contribute significantly to soil functionality. They facilitate the retention of essential plant nutrients and water (Araújo et al., 2014). Knowledge of clay mineral composition of soils has proven valuable in various aspects of soils science, like predicting the fate of heavy metals (Behroozi et al., 2021), understanding soil behavior particularly its swelling and shrinkage properties (Zolfaghari et al., 2015; Puppala et al., 2016), and its hardness (Crevelin and Bicalho, 2019), structural stability (Yilmaz et al., 2005; Pessoa and Libardi, 2022), workability (Barnes, 2013), hydraulic conductivity (Zhu et al., 2019) and ultimately soil erosion (Elliot and Flanagan, 2023). The importance of clay mineral composition in shaping various soil properties that significantly contribute to ecosystem services, such as nutrient cycling and water retention is well established (Pereira et al., 2018; Al-Tawaha et al., 2021). For instance, the high cation exchange capacity of smectite clays allows them to hold onto essential plant nutrients, promoting efficient nutrient cycling in the soil.

X-ray diffraction (XRD) analysis has been traditionally used to determine clay mineral composition in soil samples. This technique relies on the analysis of peak intensities to estimate clay mineral concentrations within a sample (Whittig and Allardice, 1986). The XRD analysis of clay minerals requires specialized equipment and skilled personnel, leading to high operational costs and time-consuming laboratory procedures. This necessitates exploring alternative methods for rapid and cost-effective assessment of clay mineral composition of soils. Therefore, this study explores the potential of machine learning models to address this gap by offering a faster and more cost-effective approach for the prediction of clay mineral composition of soils using readily available soil data.

Soil-environmental conditions, such as parent material, climate, drainage, and topography, as along with inherent soil properties like cation exchange capacity (CEC) specific surface area (SSA) and available or potassium content, influence the formation and distribution of clay minerals within a field or region (Galán and Ferrell, 2013). Despite this, few studies have explored estimating clay mineral composition using readily available soil data, including CEC, SSA, potassium content, and visible and near-infrared reflectance (VNIR) spectroscopy (Zhao et al., 2018, 2020). Various techniques, including partial least squares regression and continuum-removal-based multiple linear regression, have been employed in these studies (Zhao et al., 2020). However, it's important to note that regression equations used for predicting soil properties, especially clay mineral composition, tend to be more accurate in regions with homogenous soil formation factors, such as parent material, climate, and topography. This aligns with the findings of Pachepsky and Rawls (1999) who suggest that predictive models perform better for soils formed under similar conditions. In addition, machine learning algorithms, a branch of artificial intelligence, have emerged as a powerful tool for researching and estimating soil characteristics, including clay mineral composition. These models offer a significant advantage over traditional regression methods, particularly for soils with diverse formation factors. Machine learning models, including artificial neural network (ANN) models (Chittoori and Puppala, 2011; Shahrokh et al., 2023), feature selection algorithms and adaptive neuro-fuzzy inference systems (ANFIS) Najafi-Ghiri et al. (2019), have been successfully employed to predict the clay mineral composition of soils. These models rely on various soil-related data, such as physical, chemical, and spectral properties. Shahrokh et al. (2023) used a combination of soil physico-chemical properties, environmental data from a Landsat-8 operational land imager (OLI) and a digital elevation model (DEM) to prepare digital clay type maps in Darab district, southern Iran.

Machine learning offers a compelling alternative for clay mineral prediction. Compared to traditional methods, it enables faster analysis, handles complex soil datasets, and incorporates various data sources to improve the accuracy (Najafi-Ghiri et al., 2019). Despite their advantages, its application in clay mineral prediction remains underdocumented. This study addresses this gap by utilizing machine learning to predict clay types in soil from readily available variables. Our aim is to achieve faster and more cost-effective analysis with high-resolution digital maps. Additionally, we explored optimizing future clay prediction models by investigating hyperparameter tuning for machine learning algorithms.

2. Materials and methods

The study area, known as Kazova Plain, is situated in Tokat province of Turkiye, (40° 28′ 41“N, 36° 59’ 25” E, Fig. 1), and spans approximately 20,600 ha, comprising croplands, pastures, and orchards. Major crops such as wheat, corn, sunflower, sugar beet, tomato and alfalfa are grown in rotation. The region experiences an average total annual precipitation of 446 mm with an average temperature of 12.4 °C. The Köppen-Geiger classification categorizes the study area's climate as Csa (Warm-summer Mediterranean Climate). This classification falls within the C group, indicating temperate climates with average annual temperature exceeding 10 °C. The ‘sa’ subgroup represents a Warm-summer Mediterranean climates, characterized by an annual temperature above 10 °C and the driest months following winter. This perfectly aligns with the climate of study area, as evidenced by the average annual temperature of 12.4 °C and the driest months occurring after winter. Additionally, Soil Taxonomy classifies the study area as having an Ustic moisture regime, indicating periods of dryness during the growing season, and a Mesic temperature regime, signifying moderate annual temperature (Soil Survey Staff, 2014). The soils in the study area are classified in Entisols, Mollisols, Inceptisols and Alfisols orders (Durak et al., 2006; Günal, 2006).

To capture the variability of soil properties across the study, a grid-based sampling approach was employed. The area was divided into 1 × 1 km squares, and soil samples were collected from 0 to 30 cm depth at each grid corner using a GPS for precise location recording. Additional samples were taken at 100, 200, 300 and 500 m intervals between grid corners to account variations in soil properties over short distances. These additional sampling points targeted variability resulting from differences in parent material, land use and slope. A total of 400 soil samples were collected, including the interval samplings, with an average sampling distance of 460 m and a sampling density of 50 ha per soil sample. Intersects between grid corners were chosen to further ensure coverage of diverse soil conditions.

2.1. Soil sampling to determine clay mineral composition

Due to resource limitations (time, labor, and cost), a subset of 96 soil samples was chosen for detailed clay mineral analysis. To ensure the analysis captured the influence of key factors like geomorphological surfaces, land-use types, and parent materials, these samples were strategically selected to represent the diversity of the study area, particularly considering floodplain deposition patterns. In floodplains, like the Kazova Plain, coarser sediments are typically deposited near the active river channel (riverbank physiographic unit) due to higher flow velocities. Conversely, finer sediments settle farther away as water flow slows down (Anonymous, 2023). Following this principle, samples were collected from distinct floodplain zones within Kazova Plain, targeting the riverbank physiographic unit near the Yeşilırmak River, the river terraces further downstream, and lastly the clay deposits farthest from the river. To examine the influence of deposition patterns on clay mineral composition, a total of 72 soil samples were collected from distinct floodplain zones within Kazova Plain (Fig. 1). The parent material also plays a significant role in clay mineral composition. Therefore, an additional 24 samples were collected from colluvial lands bordering the floodplain with slopes ranging from 2 to 6%. These colluvial samples were further divided into two groups based on parent material: 11 from

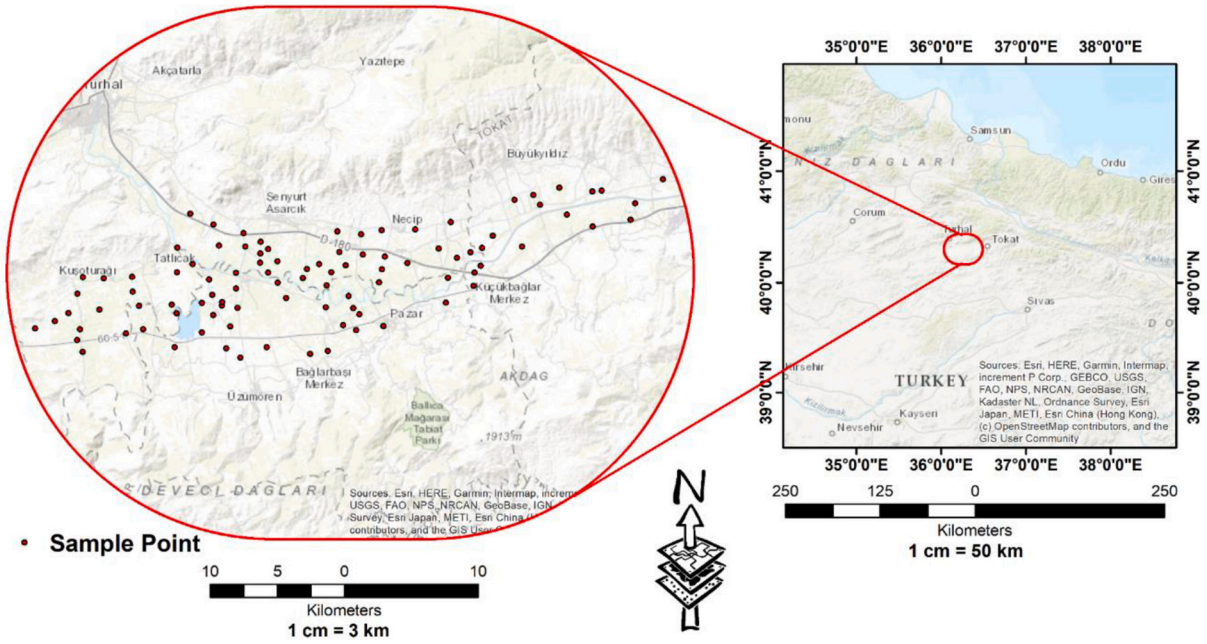


Fig. 1. Location of study area and distribution of soil samples within the study area.

serpentine schist and 13 samples from limestone.

2.2. Soil analysis

All soil samples were air-dried, ground, and sieved through a 2 mm sieve prior to further analysis. Available potassium was extracted using 1 N NH_4OAC, and the concentration was determined using flame atomic absorption spectrometry (Thomas, 1982). Slowly available potassium was extracted using the nitric acid method (Helmke and Sparks, 1996). Organic matter content was determined using the wet combustion method (Nelson and Sommers, 1996). Cation exchange capacity (CEC) was determined using the ammonium acetate method (Sumner and Miller, 1996). Exchangeable calcium and magnesium were extracted by the ammonium acetate method (Thomas, 1982). Particle size distribution was determined by the hydrometer method (Gee and Bauder, 1986). The soil reaction (pH) and electrical conductivity (EC) were measured in saturated paste (Rhoades, 1982). Calcium carbonate content was determined using the calcimeter method (Allison and Moodie, 1965).

2.3. Clay mineral composition and specific surface area of soil samples

Soil samples were separated into sand (2000–50 μm), silt (50–2 μm), and clay (<2 μm) fractions to determine clay mineral compositions, following the method described by Jackson (1975). This fractionation involved removing organic matter, calcium carbonate and iron oxides from the samples. The remaining soil suspension then underwent wet sieving to isolate the sand fraction by continuous washing with distilled water through a 0.053-mm sieve until the exiting liquid run clear. This step ensured all fine particles smaller than the sieve mesh size were removed. The clay fraction (<2 μm) was separated from silt fraction using the sedimentation technique with a 1 N $MgCl_2$ solution as a coagulant. After settling, the clay suspension was carefully transferred to a centrifuge tube for further treatment to differentiate clay minerals. Initially the clay fraction was treated with Mg-25 °C, and K-25 °C. To further differentiate vermiculite and smectite, Mg-saturated samples were subsequently solvated with ethylene glycol (Mg-EG). Finally, X-ray diffraction (XRD) patterns were obtained for K-25 °C, Mg-25 °C and Mg-EG treated samples. The clay samples were oriented on glass slides for XRD analysis. The analysis was performed using a Shimadzu WRD-600 diffractometer (TUBITAK MAM, Gebze, Turkiye) equipped with a copper X-ray tube ($\lambda = 1.5405$ Å). Samples were scanned from 2° 2Θ to 15° 2Θ at a rate of 2° per minute, generating diffraction intensity curves used to identify the clay mineral composition.

The relative abundance of clay minerals was estimated using a semi-quantitative approach. This method involves measuring the area under specific peaks in XRD patterns. These peaks correspond to characteristic reflections of different clay minerals. (XRD) patterns. These peaks correspond to characteristic reflections of different clay minerals. The measured peak areas are then used to calculate the relative percentages of each clay mineral, assuming that smectite, vermiculite, and kaolinite together represent 100% of the clay fraction (<2 μm) in the soil samples (Biscaye, 1965; Hopper, 1981). However, a limitation exists in differentiating between smectite and vermiculite using this method. Their similar X-ray diffraction patterns, especially after ethylene glycol (EG) solvation, make it challenging to quantify them separately. Therefore, for this study, smectite and vermiculite were combined and reported as smectite+vermiculite.

Ethylene glycol monoethyl ether (EGME) is a polar liquid molecule often used to assess the specific surface area (SSA) of soils. EGME acts as a probe molecule, and its ability to form a monolayer on the surface of the soil particles allows for the calculation of SSA. This calculation considers the number of EGME molecules adsorbed and their occupied area on the soil surface. The SSA of soil samples was determined using EGME and the amount of EGME used was converted to SSA (Carter et al., 1986).

2.4. Prediction of clay mineral compositions

2.4.1. Machine learning algorithms

This study employed three tree-based machine learning models (Classification and Regression Tree (CART), Random Forest Regression (RF) and eXtreme Gradient Boosting Regression (XGBoost)) to estimate the clay mineral composition of soils using the dataset, and the performance of each model was subsequently evaluated. However, identifying a single, universally optimal model for this task remains challenging. Therefore, the objective was to optimize the performances of all models across the board. Hyperparameters, which are crucial determinants of model performance, were adjusted to achieve this goal. The models' performance could be significantly improved by setting optimal

hyperparameter values.

Exhaustive evaluation of all possible hyperparameter values is an impractical due to their inherent indeterminacy. This necessitates employing optimization techniques to identify the most effective configuration. Grid Search Cross Validation, implemented within the Scikit Learn library in Python, was used in this study for this purpose. As highlighted by Brownlee (2021), machine learning models that are highly sensitive to hyperparameter settings, based on the lowest mean standard error, often benefit the most from tuning. These models are sometimes referred to as "post-tuning (AT)" models in the literature.

2.4.1.1. Classification and Regression Tree (CART). The CART algorithm is a popular tool for developing predictive machine learning models, specially designed at handling complex structures present in a dataset. Its strength lies in transforming them into clear, tree-like decision structures. CARTs works iteratively splitting the heterogeneous datasets into increasingly homogeneous subgroups based on a predetermined target variable (Abakay et al., 2024). This hierarchical model reveals relationships between various factors (covariates) and the target variable (Bittencourt and Clarke, 2003). The process of constructing a decision tree typically involves two main stages. The first stage entails the development and formation of the tree structure, including decision points (nodes) and possible outcomes (branches). The second stage involves the application of pruning techniques to optimize the tree by trimming irrelevant branches to reduce prediction errors (Sun et al., 2023).

2.4.1.2. Random Forest Regression (RF). The RF is an ensemble learning model based on the decision tree algorithm, originally proposed by Breiman (2001). This model has gained significant popularity in the field of spatial prediction of soil properties (Garosi et al., 2022; Saidi et al., 2022)). The inclusion of covariates plays a significant role in the RF algorithm. The RF is preferred when there is a need to reduce the number of variables until only the most effective ones remain in the prediction variance of the dataset. Due to its ability to effectively identify optimal variables, the RF algorithm is extensively employed. The RF algorithms are derived from the bootstrapping method, a statistical technique that utilizes classification and regression tree analysis to estimate the dependent variable (Iqbal et al., 2018). The RF algorithm optimizes the collection of decision tree models using the dataset (Cutler et al., 2012). Each tree within the ensemble is trained using distinct bootstrap samples from the training dataset and functions independently as a regression model. The final output of each tree corresponds to the regression function's output, aligning with the average of the outputs of each individual tree (Prasad et al., 2006). Samples that do not conform to the Booststrap framework are employed to evaluate the precision of the decision trees. The architecture comprises a sequence of decision trees constructed using the residuals obtained from preceding trees or the prediction errors. This approach focuses on progressively identifying and addressing instances with the highest prediction error, ultimately leading to more robust final output (Chen and Guestrin, 2016; Xue and Wu, 2020).

2.4.1.3. eXtreme Gradient Boosting Regression (XGBoost). The XGBoost model introduces complexity through its numerous hyperparameters, necessitating a tuning process (Putatunda and Rama, 2018). These hyperparameters share similarities with other tree-based models, but also include XGBoost-specific ones designed to prediction variability, improve accuracy, prevent overfitting (Park and Ho, 2021; Ramosaj and Pauly, 2019). In this study, we focus on tuning colsample_bytree, learning_rate, max_depth, and n_estimator, with discussion section exploring their impact on the accuracy of predictions (Wade, 2020).

2.4.2. Selection of variables with stepwise backward elimination

Backward stepwise selection is a variable selection method that begins with a multiple linear regression incorporating all variables of interest. It then iteratively eliminates variables with *p*-values >0.3 until all remaining variables in the model have p-values <0.3 (Saurabh Pal, 2021). Various approaches regarding the implementation of the cut-off rule based on the p significance value are discussed in the literature (Chowdhury and Turin, 2020). In the present study, a p threshold of 0.3 was set. The training and accuracy assessment of the prediction algorithms used datasets containing diverse clay types with statistically significant *p*-values.

2.4.3. Accuracy assessment

To assess the effectiveness of the machine learning models in predicting clay type content, the dataset was split into two 70% training data and 30% testing data. We employed four key metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2), and Lin's Concordance Correlation Coefficient (CCC). These metrics were calculated by comparing the estimated clay type content in test dataset with the actual values (Eq1, 2, 3 and 4). Lower RMSE and MAE values, and higher R^2 and CCC values all indicate better model performance. The computation of RMSE, MAE and R^2 values was outlined in the equations below (Somaratne et al., 2005);

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - E_i)^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|M(x_i) - E(x_i)| \tag{2}$$

$$R^2 = \frac{\sum_{i=1}^{N}(M_i - M_{mean}) \times (E_i - E_{mean})}{\sqrt{\sum_{i=1}^{N}(M_i - M_{mean})^2 \sum_{i=1}^{N}(E_i - E_{mean})^2}} \tag{3}$$

In Eqs. 1, 2 and 3, *Mi* represents the actual clay type content, *Ei* is the estimated clay type content, and *n* is and the total number of samples, respectively.

Lin (1989) introduced the Concordance Correlation Coefficient (CCC) (Eq. 4) as an improvement over the limitations of the Pearson coefficient for assessing model goodness-of-fit. The CCC goes beyond simple correlation by considering both bias and dispersion, providing a more nuanced evaluation of how well model predictions align with observations (Lin, 1989) (Eq. 4).

$$\rho_c = \frac{2\sigma_{12}}{(\mu_1 - \mu_2)^2 + \sigma_1^2 + \sigma_2^2} \tag{4}$$

In Eq. 4, ρc represents the CCC. Here, as two variables conform to Gaussian statistics, their means were denoted as μ_1 and μ_2, and standard deviations as σ_1 and σ_2, respectively. CCC values within the range of 0.21 to 0.40 indicate a fair correlation, 0.41 to 0.60 signifies a moderate correlation, 0.61 to 0.80 suggests a substantial correlation, and 0.81 to 1.00 represents an almost perfect correlation (Steichen and Cox, 2002).

The concept of Relative Improvement (RI) involves comparing RMSE values obtained before tuning (BT) and after tuning (AT) to assess the impact of tuning on prediction performance. The RI quantifies the extent to which tuning contributes to the improvement of RMSE values. The RI value was calculated using Eq. 5 (Mishra et al., 2010).

$$RI = \frac{RMSE_{AT} - RMSE_{BT}}{RMSE_{AT}} X100\% \tag{5}$$

In the Eq. 5, $RMSE_{AT}$ and $RMSE_{BT}$ are the RMSE of a machine learning algorithm after tuning and the RMSE of the same machine learning algorithm before tuning, respectively.

3. Results and discussion

3.1. Descriptive statistics

Descriptive statistics of the soil properties are presented in Table 1. Calcium (Ca) was dominant exchangeable cation, with an average content of 42.33 me 100 g^{-1}. Clay content exhibited a wide range varying between 12.5% to 51.3%. Similarly, sand content showed significant variation, ranging between 7.5% and 65.0%. Calcium carbonate content also varied considerably across the study area, with values between 2.25% and 21.34% (average:8.16%). The coefficient of variation (CV) is a useful measure of variability in soil properties. Wilding (1985) suggested a classification system based on CV, where values <15% indicate low variability, 15–35% indicate moderate variability, and >35% indicate high variability. Calcium carbonate content exhibited a high variability (CV:43.4%), indicating significant differences among samples. This variability is likely due to the influence of parent material. Soils located in the western sloping lands formed on calcareous colluvium (transported rock by water and gravity). These soils had a redder color, higher calcium carbonate content and were predominantly clayey in texture. In contrast, soils developed from other parent materials likely have lower carbonate content and potentially different textures. The content of slowly available potassium ranged from 226.5 to 3135.4 mg kg^{-1} (average: 925.8 mg kg^{-1}), while readily available potassium exhibited variation between 17.1 and 1688.2 mg kg^{-1} (average:152.8 mg kg^{-1} (Akbas et al., 2017). The high CV values for potassium content highlight significant spatial heterogeneity in its distribution across the soil samples. This variation could be due to several factors, including past application of potassium fertilizers, differences in underlying parent materials, or the natural soil formation processes. Cation exchange capacity (CEC), a crucial factor influencing plant nutrient retention and availability, also varied considerably across the study area. Values ranged from 8.19 to 44.47 me 100 g^{-1}, with an average of 21.92 me 100 g^{-1} (Table 1). The CV values for specific surface area (SSA, CV: 39.8%) and clay mineral composition (illite: 58.1%, smectite+vermiculite (SMVR): 21.0%, kaolinite: 35.2%) indicate significant variability across the soil samples. This variation can influence various soil properties. For instance, the high variability in illite suggests substantial differences in its presence among the samples, potentially affecting soil stability and nutrient retention. Similarly, the moderate to high variability observed in SMVR and kaolinite can influence aggregate stability and CEC of the soils (Table 1).

Descriptive statistics of the clay minerals in relation to parent materials are given in Table 2. As expected, the parent material significantly affects clay mineral composition (Hunt et al., 2021). The SMVR content in the clay size particles ranged from 36.9% to 88.8%, with an average of 64.9%. The SMVR content demonstrated low variability (CV:18.2% and 18.1%), suggesting a relatively stable presence in alluvial and colluvium (serpentinite schist) soils. The narrower range implies a more uniform distribution within alluvial and colluvium parent materials. In contrast, colluvium (limestone) soils displayed moderate variability in SMVR content (CV:27.4%) with a wider range compared to alluvium and colluvium (serpentinite schist) soils. This suggests a less uniform distribution of SMVR in limestone-derived colluvium. The variability of SMVR, illite, and kaolinite contents in soils formed on serpentinite parent material is likely associated with the locations from which the soil samples were taken. Hseu et al. (2007) investigated how landscape position affects clay mineral transformations in serpentinite soils of eastern Taiwan. Their study examined four soil types (Entisol, Vertisol, Alfisol, and Ultisol) formed from the same material but occupying different topographic positions (summit, shoulder, backslope, and on the footslope). The research highlighted the significant influence of both landscape and weathering on the development of distinct soil orders and clay mineral compositions. The findings revealed a clear trend along the topographic sequence: younger soils at higher elevations were dominated by SMVR and serpentine minerals, while relatively older soils on the backslope exhibited an increase in vermiculite content.

Illite and kaolinite content in the study area ranged from 4.4% to 43.6% (illite) and 3.6% to 34.2% (kaolinite), with respective averages of 16.3% and 18.8%. Consistent with Shahrokh et al. (2023), who linked the presence of illite in arid and semi-arid regions to parent rock, our study showed significant variability in illite content across different parent materials. Alluvium exhibited the highest variability, indicating potential compositional differences within this geological substrate. Colluvial soils, originating from limestone and serpentinite schist, showed moderate to high variation in illite content, although not as pronounced as observed in alluvial soils. Kaolinite content also displayed moderate variability within both alluvial and colluvial (derived from serpentinite schist) soils. This suggests some consistency in kaolinite distribution across these soil types, but also moderate degree of variation. Shahrokh et al. (2023) investigated the factors controlling kaolinite content in soils of semi-arid and arid regions. Their findings suggest that kaolinite formation is influenced by both the presence of parent materials (inherited) and in-situ development of fibrous clay minerals within the fine clay fraction. The observed patterns of kaolinite content in the studied soils strongly support this explanation, indicating that kaolinite content in these soils has a dual source consisting of inherited from parent material and newly formed within soil itself.

Across northern and western parts of the study area, soils had similar SMVR content. However, the greatest difference in clay mineral composition was recorded in illite and kaolinite contents. Illite predominates over kaolinite in soils formed on colluvium derived from limestone. This dominance was particularly pronounced in soils with the highest concentration of slowly available potassium (Table 2). This observation implies a potential link between clay particles and

Table 1
Descriptive statistics of soil properties and clay mineral contents.

Soil Properties	Min	Max	Mean	Std.Dev	CV*	Skewness	Kurtosis
Available K (mg kg^{-1})	17.1	1688.2	152.8	140.92	92.2	4.64	39.42
Slowly Available K (mg kg^{-1})	228.5	3135.4	925.8	485.23	52.4	1.22	1.68
CEC (me 100 g^{-1})	8.19	44.47	21.92	8.34	38.0	0.98	3.79
Clay (%)	12.50	75.00	40.88	13.21	32.3	0.45	2.58
Calcium Carbonate (%)	2.25	21.34	8.16	3.54	43.4	1.14	4.98
pH	7.07	9.47	8.16	0.26	3.2	0.67	5.04
Organic Matter (%)	0.83	6.02	2.09	0.98	46.8	1.79	3.72
Exc. Ca (me 100 g^{-1})	18.37	71.65	42.33	11.59	27.4	0.28	2.22
Exc. Mg (me 100 g^{-1})	0.49	2.98	1.76	0.55	31.3	−0.13	2.55
Specific Surface Area (m^2 g^{-1})	22.60	269.53	116.31	46.29	39.8	0.45	0.81
Smectite+Vermiculite (%)	36.92	88.81	64.89	13.64	21.0	−0.48	−1.01
Illite (%)	4.44	43.55	16.28	9.46	58.1	0.86	−0.38
Kaolinite (%)	3.63	34.16	18.83	6.62	35.2	0.52	0.00

* CV. Coefficient of Variation (%); CEC: Cation Exchange Capacity

Table 2
Descriptive statistics for clay content of soil samples collected from different parent materials.

Parent Material	Clay Type	Min	Max.	Mean	Stand. Dev.	CV*	Skewness	Kurtosis
	Smectite+Vermiculite	38.1	88.8	68.2	12.4	18.2	−0.882	−0.259
Alluvium	Illite	4.4	43.5	14.7	9.5	64.6	10.260	0.511
	Kaolinite	5.5	34.0	17.1	5.2	30.1	0.720	10.184
	Smectite+Vermiculite	36.9	84.4	55.6	15.3	27.4	0.611	−0.786
Colluvial (Limestone)	Illite	12.0	36.4	23.8	8.5	35.7	−0.131	−10.419
	Kaolinite	3.6	29.8	20.6	7.6	37.1	−10.090	0.534
	Smectite+Vermiculite	40.4	66.6	54.1	9.8	18.1	−0.219	−10.500
Colluvial (Serpentinite Schist)	Illite	8.9	25.8	17.9	6.2	34.8	0.037	−10.620
	Kaolinite	17.1	34.2	28.1	6.6	23.4	−0.727	−10.233

* CV: Coefficient of Variation (%)

potassium fixation, as indicated by the elevated levels of slowly soluble potassium concentration. Similar to the dominance of illite in limestone colluvial soils, Stahr et al. (2010) investigated soils derived from Paleozoic and Mesozoic limestones in North Laos, Thailand, and Vietnam. They found a sequence of dominant clay minerals in these limestone-derived soils, with illite transitioning to vermiculite and kaolinite.

In floodplain areas with gentle slopes (between 0 and 2%), the average illite content in soils was 14.7%. This finding highlights the subtle distribution of illite based on specific topographical conditions. Consistent with our findings, Shahrokh et al. (2023) proposed that the water transport of illite-rich clay particles could contribute to the increased illite content in these low-lying regions. Floodplain development process significantly influence the mineralogical composition of soils, particularly the content of clay minerals (Sokolova et al., 2013). The detailed examination of illite distribution across diverse parent materials helps us understand the intricate interplay between soil mineral composition and geological origins. This knowledge offers valuable perspectives for developing tailored soil management strategies in arid and semi-arid environments.

Illite had a (001) maximum peak at approximately 10 Å for all treatments (Fig. 2–5). The key difference between smectite and illite lies in their ability to accommodate ethylene glycol (EG) molecules within their crystal structure. Smectite can hold two EG layers, causing a significant peak shift upon solvation, while illite can only accommodate one layer. This expansion behavior, evident in of Mg-smectite peak shifts from 14 to 15 Å to 17–18 Å upon EG treatment (Moore and Reynolds, 1997). Both smectite and vermiculite exhibited a peak in the 14–15 Å basal layer spacing on the Mg-25 °C specimens. Following EG saturation, samples containing vermiculite maintained a peak at 14 Å, while those containing smectite displayed a significant shift to a peak around 17–18 Å (Fig. 2, 4 and 5). This shift in peak position upon EG treatment is indicative of smectite's ability to expand and accommodate additional

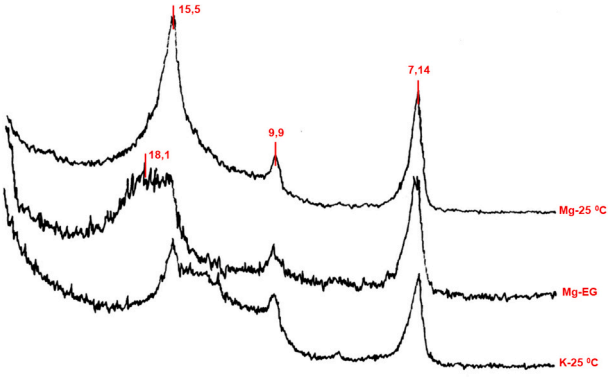


Fig. 2. XRD patterns of the clay fraction of soil sample 168 located on flat to nearly flat sloping lands in floodplain; d-spacings are given in nm. Mg-EG is glycerol solvation, Mg-25 °C is Mg-saturation at 25 °C, K-25 °C is the K-saturation at 25 °C.

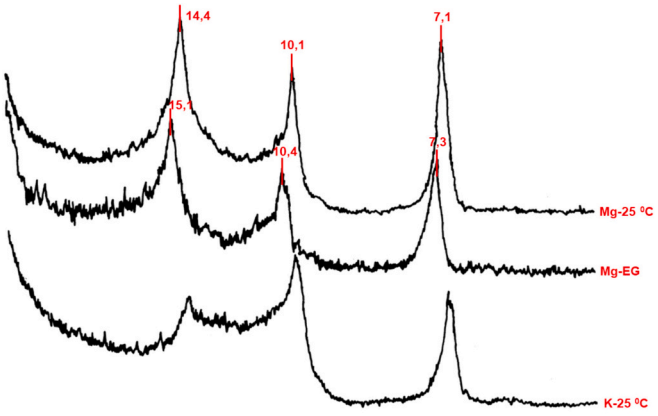


Fig. 3. XRD patterns of the clay fraction of soil sample 132 located on flat to nearly flat sloping lands in floodplain; d-spacings are given in nm. Mg-EG is glycerol solvation, Mg-25 °C is Mg-saturation at 25 °C, K-25 °C is the K-saturation at 25 °C.

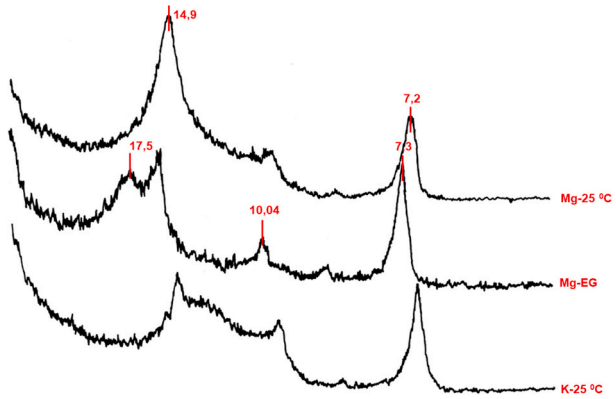


Fig. 4. XRD patterns of the clay fraction of soil sample 202 located on flat to nearly flat sloping lands in floodplain; d-spacings are given in nm. Mg-EG is glycerol solvation, Mg-25 °C is Mg-saturation at 25 °C, K-25 °C is the K-saturation at 25 °C.

EG molecules within its structure. Conversely, the absence of the 17 Å peak in EG-treated Mg-saturated samples strongly suggest a limited presence of smectite mineral in those particular soil samples (Fig. 3). High-temperature treatment (500 °C) is a well-established method for differentiating chlorite from kaolinite in clay samples by exploiting their structural differences. Kaolinite decomposes at such temperatures, leaving behind its basal reflection, while chlorite retains its complete structure. The presence of a peak at 7 Å in samples heated to 500 °C is often indicative of chlorite (Moore and Reynolds, 1997; Dixon and Jackson, 1959). Due to the absence of high-temperature treatments, the

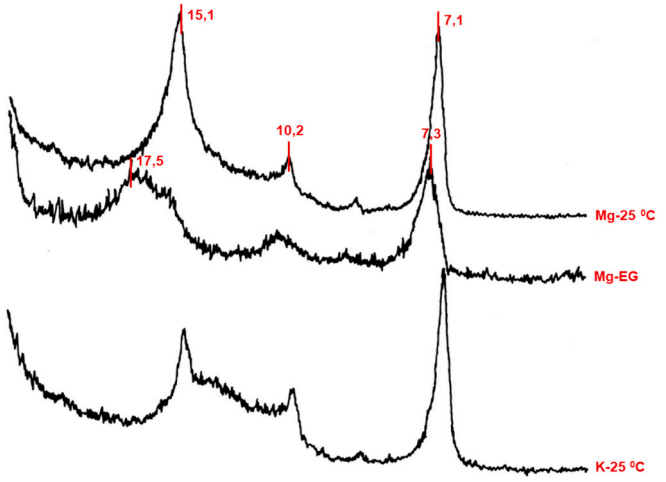


Fig. 5. XRD patterns of the clay fraction of soil sample 316 located on flat to nearly flat sloping lands in floodplain; d-spacings are given in nm. Mg-EG is glycerol solvation, Mg-25 °C is Mg-saturation at 25 °C, K-25 °C is the K-saturation at 25 °C.

semiquantitative analysis in this study was focused exclusively to clearly identifiable, pure phases visible in the XRD patterns. Interstratified minerals, although indicated by the XRD patterns in Fig. 2, 4 and 5, were not included in the quantification process.

3.2. Model parameters of thematic soil maps

Thematic maps of the covariates used for spatial estimation of clay mineral contents in machine learning algorithms were created by Ordinary Kriging method. Ordinary kriging optimal parameters of physical and chemical soil properties maps are presented in Table 3. Prior to modeling, a logarithmic transformation was applied to normalize data that deviated from a normal distribution. These transformed data were then used in the models. The ratio of nugget semivariance to total semivariance was used to classify the spatial dependence of variables. Following Camberdella et al. (1994), a ratio below 25% indicates strong spatial dependence, a value between 25% and 75% suggests moderate dependence, and a ratio exceeding 75% implies weak spatial dependence. As all soil properties had ratios lower than 25%, we can conclude that they all displayed strong spatial dependence within the study area. Soil properties with strong spatial dependence in the study area were SSA, available K and slowly available K, with nugget to sill ratios of 8.4%, 14.5% and 22.2%, respectively. Clay (53.2%), pH (45%) and calcium carbonate content (41.4%) were the soil properties with moderate spatial dependence.

The R^2 and RMSE values used to evaluate the prediction success of Ordinary Kriging spatial interpolation method are given in Table 3. The highest R^2 values were obtained for SSA (0.95), available K (0.96) and slowly available K (0.92), which exhibited a strong spatial dependence among all properties. The clay content had the lowest R^2 value (0.68), and slowly available K had the highest RMSE value (155.90 mg kg^{-1}). The mean value for the slowly available K dataset was 925.82 mg kg^{-1}, which can be considered as the mean error. The lowest RMSE value was obtained for pH with a mean error of 0.224 (Table 3).

3.3. Feature selection

Identifying the most relevant variables for machine learning models predicting spatial distribution of clay mineral types is crucial for accurate predictions. Stepwise Backward Elimination technique was used to identify the variables with the most significant impact on the model. To identify the most important soil properties for predicting SMVR, illite, and kaolinite contents, Ordinary Least Squares (OLS) Regression analysis was performed. This analysis yielded R^2 values of 0.936, 0.863, and 0.929 for predicting SMVR, illite, and kaolinite contents be, respectively. The identified most influential covariates for predicting SMVR content in the study area were slowly available potassium, available potassium, SSA, exchangeable Mg and Ca (Table 4). The SSA emerged as a crucial covariate ($p < 0.05$) for predicting of SMVR, illite, and kaolinite clay minerals. Cation exchange capacity and calcium carbonate content were identified as the only significant soil covariates explaining illite content (Table 4). Calcium carbonate is a fundamental constituent that significantly affects soil formation and its chemical and physical attributes. Its abundance serves as a natural buffering agent, increasing soil pH levels, affecting the rate of soil weathering (Drewnik et al., 2014), and contributes to the soil's CEC, facilitating the retention of positively charged nutrients and often resulting in elevated base saturation (Wilford et al., 2015). Therefore, it is reasonable to anticipate that the existence of calcium carbonate, originating from both lithogenic and pedogenic sources, primarily sourced from carbonate-rich slope deposits, will exert an influence on the formation and alteration processes of clay minerals (Emadi et al., 2008). The findings align with Razzaghi et al. (2021) who demonstrated the influence of calcium carbonate

Table 4
The result of Stepwise Backward Elimination Feature Selection.

Feature	Smectite + Vermiculite	Illite	Kaolinite
Specific Surface Area (m^2 g^{-1})	* (0.016)	* (0.000)	* (0.001)
Slowly Available K (mg kg^{-1})	* (0.015)	* (0.000)	–
Available K (mg kg^{-1})	* (0.020)	* (0.004)	–
Exchangeable Mg (me 100 g^{-1})	* (0.000)	* (0.000)	* (0.000)
Exchangeable Ca (me 100 g^{-1})	* (0.000)	–	* (0.000)
Clay (%)	* (0.000)		* (0.000)
Calcium Carbonate (%)	–	* (0.000)	–
Cation Exchange Capacity (me 100 g^{-1})	–	* (0.011)	–
pH	–	–	* (0.000)

* **P > |t|**

Table 3
Optimal parameters of spatial distribution models for thematic soil maps.

Soil Properties	Model	Nugget (Co)	Sill (Co + C)	Spatial Dep. (%)	Range (m)	R^2	RMSE	Trans
Available K	Stable	0.07	0.49	14.5	3771.7	0.96	37.20	Log
Slowly Available K	Stable	0.06	0.26	22.2	2301.6	0.92	155.90	Log
Clay	K-Bessel	123.68	235.39	53.3	23,333.4	0.68	6.10	None
Calcium carbonate	Gaussian	0.07	0.18	41.4	3290.8	0.74	2.77	Log
Exc. Calcium	Pentaspherical	26.87	188.09	**14.7**	248,158.7	0.91	3.73	None
Exc. Magnesium	J-Bessel	0.18	0.55	**32.3**	7420.1	0.7	1.58	Log
pH	Gaussian	0.03	0.07	45.0	538.5	0.85	0.22	None
Specific Surface Area	Stable	145.97	1740.26	8.4	2009.8	0.95	36.93	None

content on predicting CEC which is closely associated with soil clay mineral types. Their research identified a gap in existing CEC models – the omission of $CaCO_3$ as an input variable, potentially limiting their accuracy in calcareous soils. This limitation was addressed by evaluating several CEC models. Models lacking $CaCO_3$ overestimated CEC for samples with high CEC values. A model incorporating both $CaCO_3$ and hygroscopic water content provided the most accurate CEC estimates. These findings highlight the importance of including CaCO3, alongside potentially other clay mineral composition-related variables, as suggested by Razzaghi et al. (2021), to enhance CEC model accuracy, particularly for calcareous soils.

Soil pH, a measure of soil acidity or alkalinity, significantly influences clay mineral composition. Significant positive correlation ($p < 0.05$) was recorded between soil pH and kaolinite abundance, while slowly available K was played a significant role in predicting SMVR and illite contents. This aligns with Watanabe et al. (2006), who identified soil pH as a key factor governing clay mineral distribution in humid Asia. They proposed that acidic environments (low pH) promote kaolinite formation, while Al-OH species activity and parent material influence the presence of other clay minerals like smectite. Additionally, alongside pH, the most significant soil properties for explaining kaolinite in the study area included SSA, exchangeable Mg and Ca and clay content (Table 4).

3.4. Spatial distribution models of clay mineral composition

3.4.1. Tunning results for machine learning algorithms

Table 5 presents the results of hyperparameter tuning optimized using grid search method to achieve the lowest RMSE value in CART, RF and XGBoost estimation algorithms. The max_depth, among the CART hyperparameters, refers to the maximum depth of the tree. This tuning aimed to improve both model performance in avoiding overfitting and achieving accurate estimation. Among the CART hyperparameters, max_depth, which refers to maximum depth of the tress, was optimized o 8, 7 and 3 for predicting SMVR, kaolinite and illite, respectively. Similarly, min_samples_split, defining the minimum number of samples required to split a node in the regression tree, was optimized to 7, 14 and 23 for SMVR, kaolinite and illite, respectively. Finally, the max_leaf_nodes hyperparameter, representing the maximum number of leaves in the regression tree, was optimized to 3 for SMVR, 4 for kaolinite and 6 for illite (Table 5).

The max_feature parameter, signifying the maximum number of features considered for a single tree in RF supervised machine learning, was set at 3 for all three clay minerals. However, n_estimator hyperparameter, indicating the number of decision trees and model size, was differed between RF models: established at 100 for RF_{SMVR} and 200 for both $RF_{Kaolinite}$ and RF_{Illite} (Table 5). Similarly, XGBoost, known as multi-tree approach, utilized the colsample_bytree parameter to determine the proportion of covariates used in constructing each tree. This parameter was configured as 0.4 for SMMR and illite, while kaolinite used a value of 1 (Table 5). Additionally, the learning_rate parameter, defining the reduction in the eta value during each iteration through Grid Search Cross Validation, was set to 0.01 for XGBoost models for predicting kaolinite and illite contents (Table 5). The careful adjustment of hyperparameters ensures a refined and optimized strategy tailored to each clay mineral type during the modeling phase.

Table 5
Hyperparameter tunning results.

Model	Hyperparameters	Smectite + Vermiculite	Kaolinite	Illite
CART	max_leaf_nodes	3	4	6
	min_samples_split	7	14	23
	max_depth	8	7	3
RF	max_features	3	3	3
	n_estimators	100	200	200
	max_depth	4	9	6
XGBoost	colsample_bytree	0.4	1	0.4
	learning_rate	0.1	0.01	0.01
	max_depth	3	3	3
	n_estimators	1000	300	500

The relative abundance of clay minerals in the study area, estimated by CART, RF and XGBoost models is shown in Fig. 6–8. The SMVR emerged as the predominant clay mineral in the central and western regions, as highlighted in the spatial distribution maps generated through RF and XGBoost predictions (Fig. 6b and c). In contrast, the northern and southern edges, characterized by gentle to moderate slopes, exhibited a dominance of illite and kaolinite clay minerals. These spatial variation of illite and kaolinite reflects distinct parent materials. The southern edges are characterized with calcareous colluvium, while the northern edges are derived from serpentinite. The central and west regions have soils formed from the weathering of alluvial deposits laid down by the Yesilirmak River (Gunal et al., 2008). Our findings align with a study by Baghernejad (2000) on clay mineral composition in Fars Province soils, which emphasized a high degree of similarity in mineral composition across various geomorphic surfaces, with variations observed in their proportions. The study also reported a decrease in illite and chlorite content along a slope gradient accompanied by an increase in montmorillonite and palygorskite content.

Fig. 6 presents spatial distribution maps for SMVR content generated by CART, RF and XGBoost models. The CART model (Fig. 6a) produced a map with histogram values ranging from 41% to 70% (mean:61.5%, standard deviation: 7.1%). The RF model map (Fig. 6b) exhibited higher mean (67.4%) and lower standard deviation (3.5%) compared to the CART model. Finally, XGBoost model map for SMVR (Fig. 6c) exhibited histogram values ranging from 47% to 68%, with a mean of 64.1% and a standard deviation of 3.1%.

Figs. 7a-c depict the spatial distribution of kaolinite clay content across Kazova Plain as predicted by CART, RF and XGBoost models. Despite slight variation in mean values (16.7% for CART, 17.5% for RF and 17.2% for XGBoost) and standard deviations (3.95 for CART, 2.09 for RF and 2.41 for XGBoost), all three algorithms revealed a high degree of similarity in predicted spatial distribution patterns. The maps consistently show lower kaolinite content in the central and western regions, while soils in the northern part derived from serpentinite colluvium exhibited higher kaolinite content. This spatial variation suggests the potential influence of parent material on clay mineral composition across the Kazova Plain.

The spatial distribution of illite content is presented in in Figs. 8a, b and c. The mean predicted illite clay content were 18.3%, 14.5% and 13.7%, for CART, RF and XGBoost models, respectively. Standard deviation values were 13.0% for CART, 3.0% for RF and 4.8% for XGBoost models.

3.4.2. Accuracy assessment of machine learning algorithms

Grid Search Cross Validation with MSE-optimization successfully addressed overfitting in several models ($SMVR_{CART}$, $SMVR_{XGBoost}$, $Kaolinite_{CART}$, $Illite_{CART}$ and $Illite_{XGBoost}$) models. Models initially exhibited overfitting tendencies with RMSE converging to 0 and R^2 converging to 1. However, post-tuning, these values showed convergence between test and training data (Table 6). The $Kaolinite_{XGBoost}$ model, which displayed pre-tuning overfitting, achieved reliable prediction of kaolinite content with a test dataset RMSE of 4.8% and R^2 of 0.713 after tuning.

Tuning the models through Grid Search Cross Validation effectively improved their prediction accuracy. Overfitting models ($SMVR_{CART}$, $SMVR_{XGBoost}$, $Illite_{CART}$ and $Illite_{XGBoost}$) showed significant improvement in RMSE after tuning, ranging from 3.2% to 97.3% (Table 6). The RF algorithm achieved the lowest overall RMSE (8.6%) for predicting SMVR content. Similarly, CART performed well for kaolinite and illite predictions, while XGBoost algorithm excelled in predicting both kaolinite (4.8% RMSE) and illite clay (7.4% RMSE) (Table 6). These results highlight the effectiveness of hyperparameter tuning in enhancing model performance for clay content prediction. Our findings

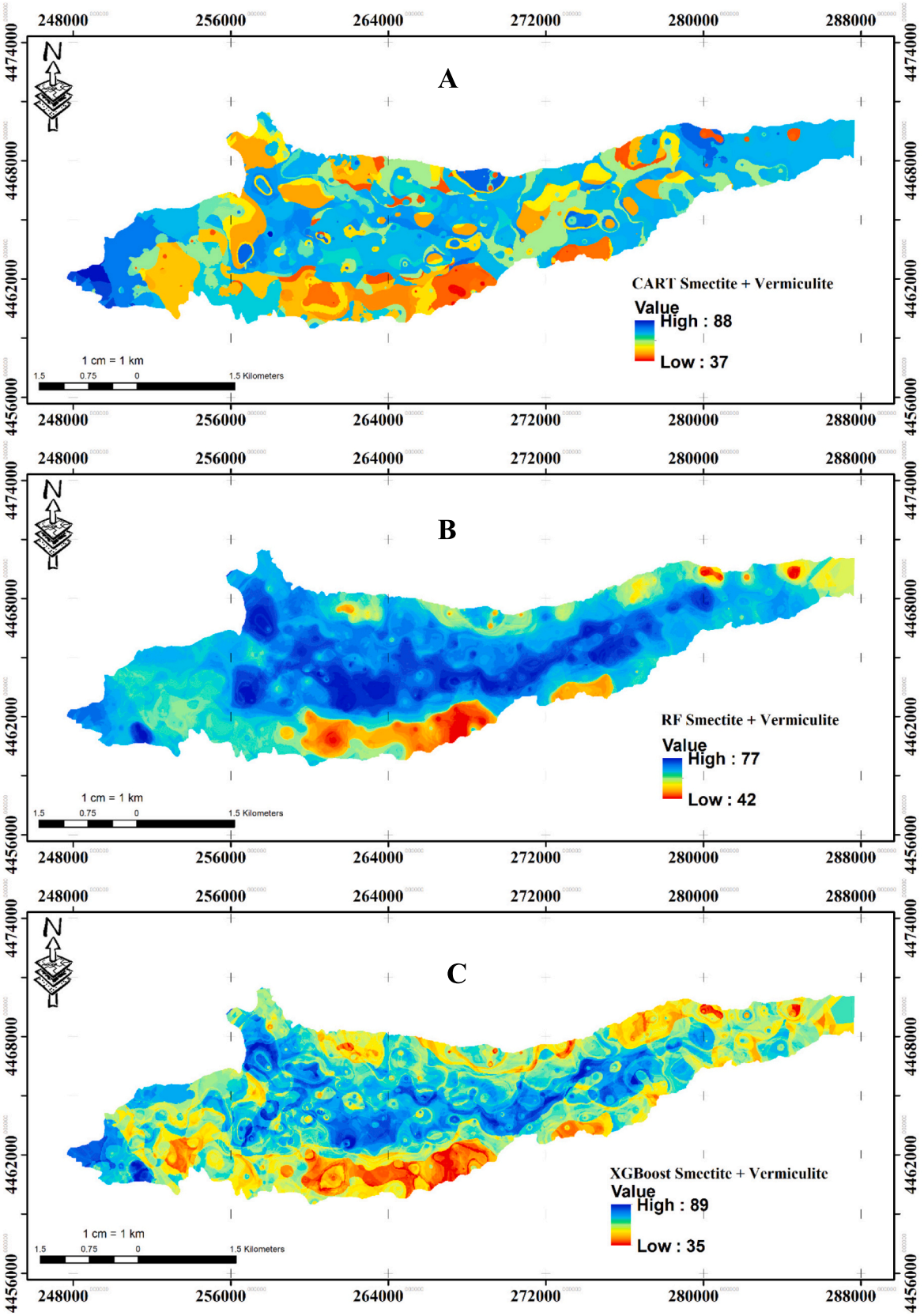


Fig. 6. Spatial distribution maps of smectite+vermiculite contents obtained with a.) CART, b.) RF and c.) XGBoost algorithms.

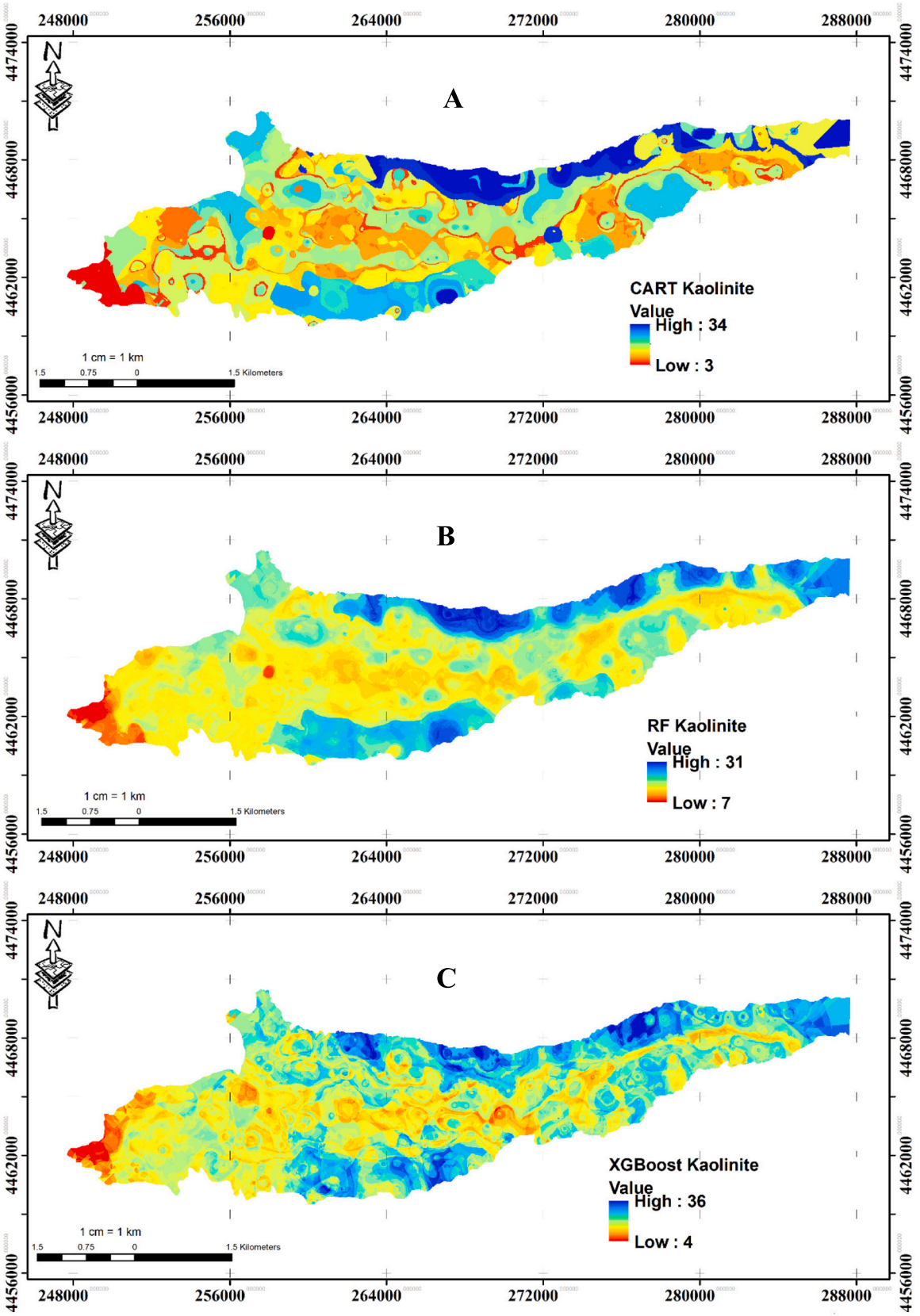


Fig. 7. Spatial distribution maps of kaolinite contents obtained with a.) CART, b.) RF and c.) XGBoost algorithms.

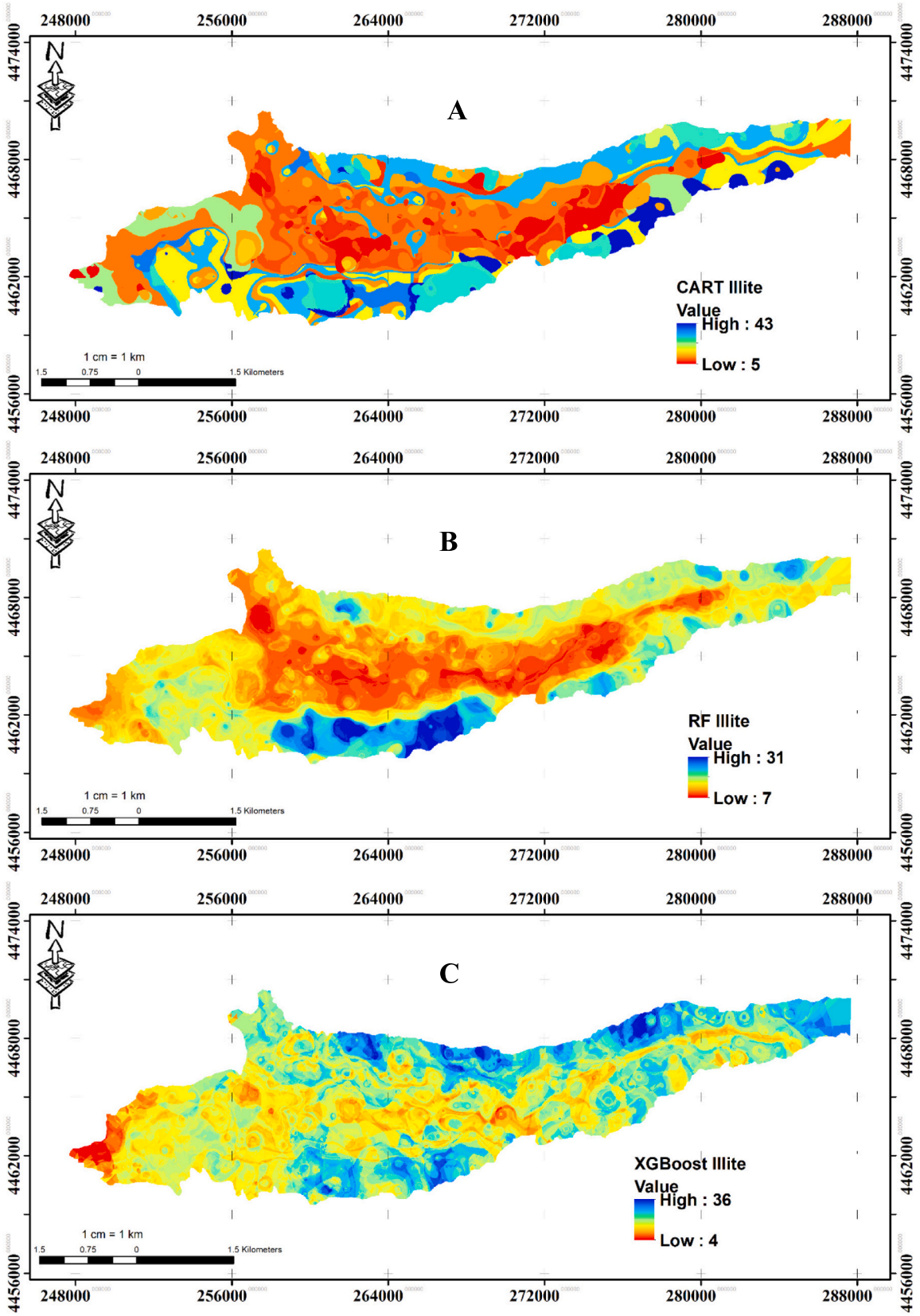


Fig. 8. Spatial distribution maps of illite contents obtained with a.) CART, b.) RF and c.) XGBoost algorithms.

Table 6
Accuracy Assessment of Supervised Machine Learning Algorithms.

Model		Tuning*	Training			Test			RI (%)
			MAE	RMSE (%)	R^2	MAE	RMSE (%)	R^2	
Smectite+ Vermiculite	CART	BT	0.000	0.000	1.000	9.317	12.078	0.523	
		AT	6.870	8.931	0.742	8.363	10.332	0.685	**30.8**
	RF	BT	3.156	4.022	0.953	7.451	9.396	0.749	
		AT	2.939	3.533	0.964	7.077	**8.587**	0.796	6.3
	XGBoost	BT	0.000	0.001	1.000	7.627	10.133	0.699	
		AT	0.844	1.269	0.995	7.927	9.810	0.722	**3.2**
Kaolinite	CART	BT	0.000	0.000	1.000	4.957	6.070	0.468	
		AT	3.294	4.070	0.775	4.631	5.938	0.503	7.4
	RF	BT	1.609	2.009	0.950	4.233	5.316	0.632	
		AT	2.444	2.943	0.889	4.146	5.236	0.648	2.53
	XGBoost	BT	0.537	0.740	0.993	4.336	5.461	0.607	
		AT	0.000	0.001	1.000	3.669	**4.814**	0.713	17.6
Illite	CART	BT	0.000	0.000	1.000	8.882	11.904	0.261	
		AT	6.952	6.701	0.601	7.163	7.994	0.515	**97.3**
	RF	BT	3.824	1.571	0.927	7.330	7.972	0.512	
		AT	0.471	0.852	0.996	6.861	7.801	0.548	6.9
	XGBoost	BT	0.000	0.001	1.000	6.679	8.159	0.484	
		AT	2.892	3.552	0.927	6.467	**7.368**	0.613	**26.6**

* BT is before tuning; AT is after tuning; RI is the relative importance.

align with Najafi-Ghiri et al. (2019) demonstrated the potential of machine learning models for clay mineral prediction. Similar to their study using ANFIS, which exhibited satisfactory performance for illite, chlorite, palygorskite, and smectite (R^2: 0.67–0.89, MAE: 0.004–0.012, RMSE: 0.001–0.028), XGBoost proved to be powerful tool in this study. It is important to note that ANFIS showed lower accuracy for palygorskite and smectite compared to other clay types, suggesting potential limitations in predicting these specific clay types.

Examining agreement between predicted and measured clay content, Lin's Concordance Correlation Coefficient (CCC) revealed significant correlation (0.61–0.80) for RF and XGBoost models predicting SWMR content, while CART model for kaolinite showed a fair correlation (0.25). This indicates that RF and XGBoost may be superior for predicting SMVR in these soils, highlighting the effectiveness of machine learning for clay content prediction with XGBoost achieving the lowest RMSE for both kaolinite (4.8%) and illite (7.4%). Hyperparameter tuning significantly improved the agreement between predicted and measured kaolinite content in CART model with CCC increasing by 47% (Table 7). In contrast, the already high CCC for RF model predicting SMVR content showed minimal improvement (2%) after tuning. The CCC values were generally lower than the R^2 values shown in Table 6. This difference suggests that a lower proportion of the variability in the predicted data is not explained by the observed values according to Lin (1989). The significant improvement in CCC for kaolinite prediction with CART highlights its potential for addressing challenging prediction tasks. On the other hand, the minimal impact of tuning on the RF model for SVMR underscores its overall efficacy for predicting the specific clay mineral content.

Table 7
The Results of Lin's Concordance Correlation Coefficients (LCCC).*

		LCCC		
		BT	AT	RI (%)
CART	Illite	0.30	0.43	0.31
	Kaolinite	0.25	0.47	0.47
	Smectite+Vermiculite	0.43	0.54	0.21
RF	Illite	0.39	0.46	0.16
	Kaolinite+Vermiculite	0.39	0.45	0.13
	Smectite	0.61	0.62	0.02
XGBoost	Illite	0.43	0.50	0.15
	Kaolinite	0.38	0.46	0.17
	Smectite+Vermiculite	0.62	0.64	0.04

* BT is before tuning; AT is after tuning; RI is the relative importance.

3.4.3. *Relative importance of covariates*

Fig. 9 illustrates the significance of different covariates in explaining the variability of predicted values. Exchangeable Mg emerged as the most critical covariate influencing SMVR prediction (38.1% of explained variance), aligning with findings in arid regions by Khormali and Abtahi (2003) and Abbaslou et al. (2013). These studies attributed increased smectite formation in arid and semi-arid regions of Iran to high Mg concentrations (alongside Si and Ca). Smectite clay minerals are typically formed through the weathering of rocks containing specific Mg concentrations or through interaction between weathering residue solutions and Mg present in water (Galán and Ferrell, 2013). The SSA emerged as the third most important covariate for predicting SMVR content (Fig. 9a). This is likely related to the high surface area of smectites, which reach up to 810 m^2 g^{-1} due to their internal and external structures (Carter et al., 1986). Conversely, available potassium had the least impact on SMVR content, with a significance level of 7.0% (Fig. 9a).

Examining key factors influencing clay mineral content prediction across soil samples (Fig. 9), kaolinite content prediction is most influenced by Ca (38.47%), Mg (19.93%), pH (16.33%), clay content (14.5%), and SSA (10.77%) (Fig. 9b). This differs from findings in calcareous soils by Shahrokh et al. (2023), where clay content, slope height, pH, and sand content were most important for kaolinite prediction. Similar to the smectite, illite content prediction is primarily driven by exchangeable Mg content, explaining approximately 26.3% of the observed variance. This aligns with findings of Shahrokh et al. (2023) and supports the of exchangeable Mg content in illite formation. Illite's structure, a 2:1 type phyllosilicate, incorporates Mg (alongside Al or Fe) within its octahedral layer (Brigatti et al., 2013). This finding reinforces the notion that the presence of Mg promotes the formation of various layer silicates (Khormali and Abtahi, 2003). Calcium carbonate content and slowly available potassium content hold the second and third highest significance in explaining the variance of illite clay mineral, while SSA had the least impact (12.03%) (Fig. 9c). Najafi-Ghiri et al. (2019) identified different key covariates for predicting the illite mineral content of calcareous soils in the south of Iran. The most effective covariates determined using Ranker feature selection method were CEC, sand, total potassium, silt, and the agricultural second-degree index.

The CEC serves as a valuable indicator for estimating specific clay mineral contents within a soil sample due to their varying CEC values. Kaolinite and smectite, two distinct clay types, can significantly soil CEC (Moore and Reynolds, 1997). However, despite its importance, CEC

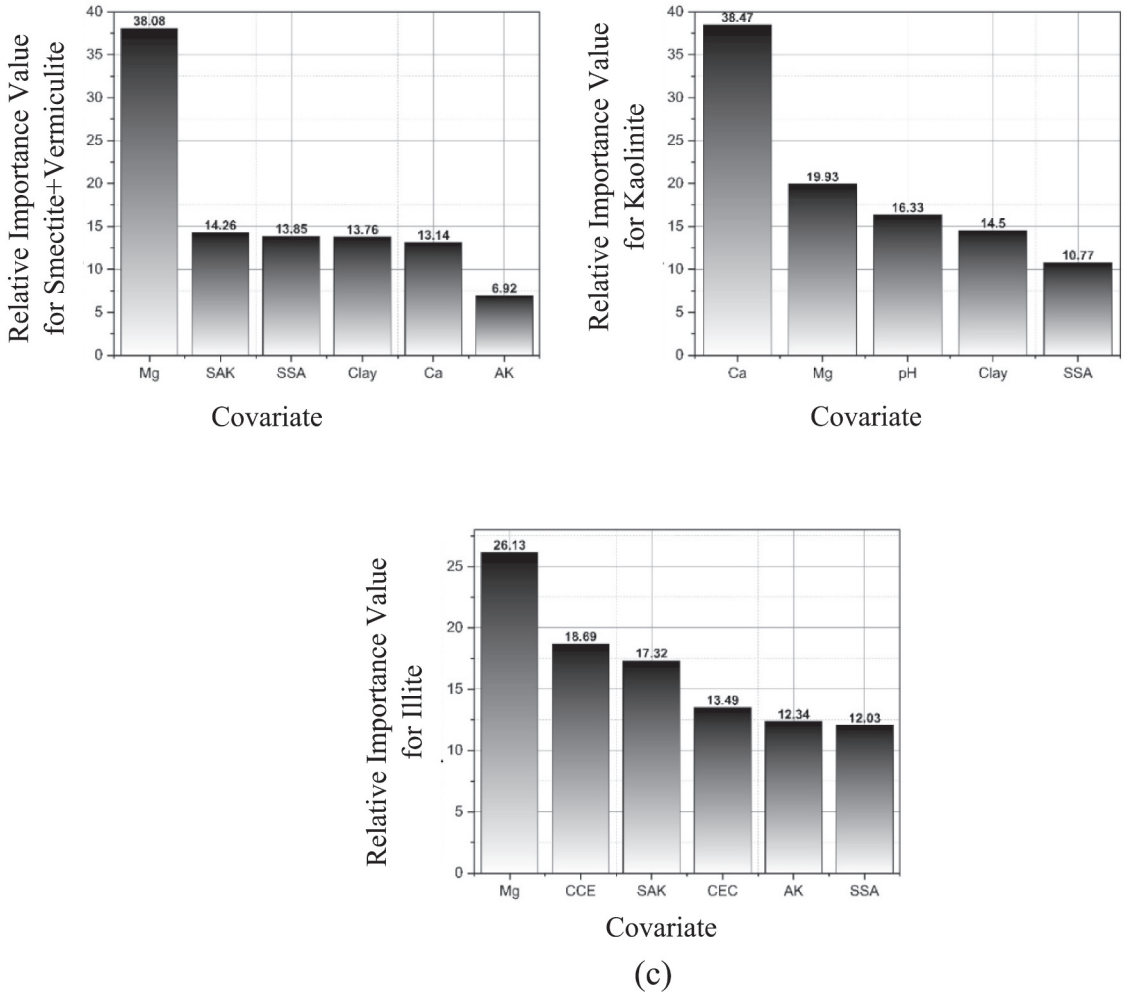


Fig. 9. Importance of covariates in the prediction of (a) Smectite+Vermiculite, (b) Kaolinite and (c) Illite.

inclusion as a covariate did not significantly impact the estimation of SMVR and kaolinite clay minerals. The CEC was identified as the fourth most influential covariate in predicting illite clay variance (Fig. 9c). This finding highlights the intricate relationship between CEC and various clay minerals. It emphasizes that the combined effects and contributions of different clay types to soil properties can vary significantly.

4. Conclusions

The study area revealed a distinct spatial distribution of kaolinite, illite and SMVR across the study area. The SMVR displayed higher content in the central and western regions, while illite and kaolinite were more prevalent in the northern and southern boundaries. This observed distribution pattern is likely influenced by the parent materials and geomorphic positions of the soils, which in turn contribute to the composition and abundance of clay minerals.

Three machine learning algorithms, including CART, RF and XGBoost, were explored for effectiveness in predicting clay mineral composition. The models exhibited varying levels of accuracy, with RF and XGBoost generally outperforming CART. This highlights the strengths of ensemble methods (like RF) and gradient boosting (like XGBoost) in capturing complex relationships within the data. Additionally, the tuning process played a crucial role in optimizing model performance and minimizing overfitting issues in some models. The impressive performance of RF and XGBoost algorithms strongly underscores the significant potential of machine learning in clay mineral mapping. This success paves the way for further investigation of even more sophisticated algorithms and their suitability for application across larger geographical scales. To ensure the generalizability of these models, validation across various a wider range of soil types and climatic conditions is highly recommended. Such validation efforts would provide valuable insights for broader application in diverse landscapes, ultimately leading to a more comprehensive understanding of clay mineral distribution across various environments.

A key finding of this study is the identification of influential factors predicting clay mineral contents, specific surface area, available potassium, and exchangeable calcium and magnesium were recognized as significant covariates, underscoring their impact on the spatial distribution of clay minerals in the studied soils. Investigating the mechanisms underlying the influence of specific surface area, potassium, and exchangeable cations on clay mineral distribution could provide deeper insights into soil weathering processes. For future studies, a more detailed exploration of these covariates, as well as considering a broader spectrum of environmental factors, including climate, vegetation, and drainage conditions, could provide a more holistic understanding of clay mineral distribution. By incorporating these additional variables and investigating underlying mechanisms, future research can refine our understanding of the complex interactions influencing clay mineral composition in various geological settings.

Authors' contributions

Hikmet Gunal designed the study and prepared the manuscript, Nurullah Acir carried out all the laboratory analysis.

Funding

This study was funded by the Scientific and Technological Research Council of Turkiye (Grant No: TOVAG 107 O 879).

CRediT authorship contribution statement

Hikmet Günal: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nurullah Acir:** Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

This study was derived from the MSc. thesis of Nurullah Acir, the second author. We appreciate the help of Dr. Mirac Kilic on artificial intelligence modeling.

References

Abakay, O., Kılıç, M., Günal, H., Kılıç, O.M., 2024. Tree-based algorithms for spatial modeling of soil particle distribution in arid and semi-arid region. Environ. Monit. Assess. 196 (3), 264. https://doi.org/10.1007/s10661-024-12431-6.

Abbaslou, H., Abtahi, A., Peinado, F.J.M., Owliaie, H., Khormali, F., 2013. Mineralogy and characteristics of soils developed on Persian gulf and Oman sea basin, southern Iran: implications for soil evolution in relation to sedimentary parent material. Soil Sci. 178 (10), 568–584. https://doi.org/10.1097/SS.0000000000000022.

Akbas, F., Gunal, H., Acir, N., 2017. Spatial variability of soil potassium and its relationship to land use and parent material. Soil Water Res. 124, 202–211. https://doi.org/10.17221/32/2016-SWR.

Allison, L.E., Moodie, C.D., 1965. Carbonate. In: Black, C.A., et al. (Eds.), Methods of Soil Analysis, Part 2. Agronomy Series No. 9. American Society of Agronomy, Madison, pp. 1379–1400. https://doi.org/10.2134/agronmonogr9.2.2ed.

Al-Tawaha, A.R.M., Günal, H., Křeček, J., Zamfir, R.H.C., Patel, H.K., Vyas, R.V., Celik, I., Al-Ramamneh, E.A.D., 2021. Soil fertility decline under climate change. In: Sustainable Soil and Land Management and Climate Change. CRC Press, pp. 127–145. https://doi.org/10.1201/9781003108894.

Anonymous, 2023. Overbank deposits. In: National Geographic. Accessed on. https://education.nationalgeographic.org/resource/overbank-deposits/. Accessed 25 July 2023.

Araújo, S.R., Wetterlind, J., Demattê, J.A.M., Stenberg, B., 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. Eur. J. Soil Sci. 65 (5), 718–729. https://doi.org/10.1111/ejss.12165.

Baghernejad, M., 2000. Variation in soil clay minerals of semi-arid regions of Fars province, Iran. Iran Agric. Res. 192, 165–180. https://doi.org/10.22099/IAR.2000.4316.

Barnes, G.E., 2013. An apparatus for the determination of the workability and plastic limit of clays. Appl. Clay Sci. 80, 281–290. https://doi.org/10.1016/j.clay.2013.04.014.

Behroozi, A., Arora, M., Fletcher, T.D., Western, A.W., Costelloe, J.F., 2021. Understanding the impact of soil clay mineralogy on the adsorption behavior of zinc. Int. J. Environ. Res. 153, 559–569. https://doi.org/10.1007/s41742-021-00334-0.

Biscaye, P.E., 1965. Mineralogy and sedimentation of recent deep-sea clay in the Atlantic Ocean and adjacent seas and oceans. Geol. Soc. Am. Bull. 76 (7), 803–832. https://doi.org/10.1130/0016-7606(1965)76 [803: MASORD]2.0.CO;2.

Bittencourt, H.R., Clarke, R.T., 2003. Use of classification and regression trees CART to classify remotely-sensed digital images. In: International Geoscience and Remote Sensing Symposium IGARSS, pp. 3751–3753. https://doi.org/10.1109/igarss.2003.1295258.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brigatti, M.F., Galan, E., Theng, B.K.G., 2013. Structure and mineralogy of clay minerals. In: Developments in Clay Science, vol. 5. Elsevier, pp. 21–81. https://doi.org/10.1016/S1572-4352(05)01002-0.

Brownlee, J., 2021. XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn. Machine Learning Mastery. https://machinelearningmastery.com/gradient-boosting-with-scikit-learn-xgboost-lightgbm-and-catboost/.

Camberdella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., Konopka, A.E., 1994. Field scale variability soil properties in central Iowa soils. Soil Sci. Soc. Am. J. 58, 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x.

Carter, D.L., Mortland, M.M., Kemper, W.D., 1986. Specific Surface. Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods. SSSA Book Series: 5. Madison, WI: Soil Science Society of America. https://doi.org/10.2136/sssabookser5.1.2ed.c16.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17-August, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chittoori, B., Puppala, A.J., 2011. Quantitative estimation of clay mineralogy in fine-grained soils. J. Geotech. Geoenviron. Eng. 13711, 997–1008. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000521.

Chowdhury, M.Z.I., Turin, T.C., 2020. Variable selection strategies and its importance in clinical prediction modelling. Family Med. Commun. Health 8, e000262. https://doi.org/10.1136/fmch-2019-000262.

Crevelin, L.G., Bicalho, K.V., 2019. Comparison of the Casagrande and fall cone methods for liquid limit determinations in different clay soils. Rev Bras Ciênc Solo 43, e0180105. https://doi.org/10.1590/18069657rbcs20180105.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In: Ensemble Machine Learning. Springer, New York, New York, NY, pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5.

Daher, M., Schaefer, C.E.G.R., Fernandes Filho, E.I., Francelino, M.R., Senra, E.O., 2019. Semi-arid soils from a topolithosequence at James Ross Island, Weddell Sea region, Antarctica: chemistry, mineralogy, genesis, and classification. Geomorphology 327, 351–364. https://doi.org/10.1016/j.geomorph.2018.11.003.

Dixon, J.B., Jackson, M.L., 1959. Mineralogical analysis of soil clays involving vermiculite chlorite-kaolinite differentiation. Clay Clay Miner. 8, 274–286. https://doi.org/10.1346/CCMN.1959.0080124.

Drewnik, M., Skiba, M., Szymański, W., Zyła, M., 2014. Mineral composition vs. soil forming processes in loess soils—a case study from Kraków (southern Poland). Catena 119, 166–173. https://doi.org/10.1016/j.catena.2014.02.012.

Durak, A., Gunal, H., Akbas, F., Kilic, S., 2006. Long term cultivation effects on soil properties in pastures. Asian J. Chem. 183, 1953–1952.

Elliot, W.J., Flanagan, D.C., 2023. Estimating WEPP cropland soil erodibility from soil properties. In: Soil Erosion Research under a Changing Climate, American Society of Agricultural and Biological Engineers. January 8–13, 2023, Aguadilla, Puerto Rico, USA, p. 1. https://www.fs.usda.gov/rm/pubs_journals/2023/rmrs_2023_elliot_w001.pdf.

Emadi, M., Baghernejad, M., Memarian, H., Saffari, M., Fathi, H., 2008. Genesis and clay mineralogical investigation of highly calcareous soils in semi-arid regions of Southern Iran. J. Appl. Sci. 8, 288–294. https://doi.org/10.3923/jas.2008.288.294.

Galán, E., Ferrell, R.E., 2013. Genesis of clay minerals. In: Developments in Clay Science, 5, pp. 83–126. Elsevier. https://doi.org/10.1016/B978-0-08-098258-8.00003-1.

Garosi, Y., Ayoubi, S., Nussbaum, M., Sheklabadi, M., 2022. Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of Iran. Geoderma Reg 29, e00513. https://doi.org/10.1016/j.geodrs.2022.e00513.

Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods—SSSA Book Series no 5, Madison, Wisconsin, pp. 383–412. https://doi.org/10.2136/sssabookser5.1.2ed.c15.

Günal, H., 2006. Pedogenesis and Classification of Soils Located on Two Consecutive Topographies. J. Agric. Fac. Gaziosmanpaşa Univ. 3 (2), 59–68.

Gunal, H., Akbas, F., Ozgoz, E., Unlukara, A., Yıldız, H., Kurunç, A., Cetin, M., Ersahin, S., 2008. Preparation of an up-to-date database necessary for sustainable agricultural production in Kazova. TUBITAK Project Report, TOVAG 1050266 in Turkish. https://search.trdizin.gov.tr/tr/yayin/detay/607447/.

Helmke, P.A., Sparks, D.L., 1996. Lithium, sodium, potassium, rubidium, and cesium. In: Bigham, J.M. (Ed.), Methods of Soil Analysis, Part 3, Chemical Methods. Soil Science Society of America/American Society of Agronomy, Madison, pp. 551–574. https://doi.org/10.2136/sssabookser5.3.c19.

Hopper, R.W., 1981. A semi-quantitative x-ray diffraction technique for estimation of smectite, illite, and kaolinite. MSc thesis. Montana State University-Bozeman, College of Agriculture.

Hseu, Z.Y., Tsai, H., Hsi, H.C., Chen, Y.C., 2007. Weathering sequences of clay minerals in soils along a serpentinitic toposequence. Clay Clay Miner. 55 (4), 389–401. https://doi.org/10.1346/CCMN.2007.0550407.

Hunt, A.G., Egli, M., Faybishenko, B. (Eds.), 2021. Hydrogeology, Chemical Weathering, and Soil Formation. American Geophysical Union. ISBN: 978-1-119-56396-9.

Iqbal, F., Lucieer, A., Barry, K., 2018. Poppy crop capsule volume estimation using UAS remote sensing and random forest regression. Int. J. Appl. Earth Obs. Geoinf. 73, 362–373. https://doi.org/10.1016/j.jag.2018.06.008.

Jackson, M.L., 1975. Soil Chemical Analysis: Advanced Course, 2nd Edition 10th Printing. Published by M. L. Jackson, Madison, WI.

Kasanin-Grubin, M., 2013. Clay mineralogy as a crucial factor in badland hillslope processes. Catena 106, 54–67. https://doi.org/10.1016/j.catena.2012.08.008.

Khormali, F., Abtahi, A., 2003. Origin and distribution of clay minerals in calcareous arid and semi-arid soils of Fars Province, southern Iran. Clay Miner. 38 (4), 511–527. https://doi.org/10.1180/0009855023740112.

Lin, L.I.K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268. https://doi.org/10.2307/2532051.

Mishra, U., Lal, R., Liu, D., Van Meirvenne, M., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil Sci. Soc. Am. J. 74, 906–914. https://doi.org/10.2136/sssaj2009.0158.

Moore, D.M., Reynolds, R.C., 1997. X-Ray Diffraction and the Identification and Analysis of Clay Minerals. Oxford University Press, New York.

Najafi-Ghiri, M., Mokarram, M., Owliaie, H.R., 2019. Prediction of soil clay minerals from some soil properties with use of feature selection algorithm and ANFIS methods. Soil Res. 577, 788–796. https://doi.org/10.1071/SR18352.

Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. In: Sparks, D.L. (Ed.), Methods of Soil Analysis, Part 3, Chemical Methods—SSSA Book Series no 5, pp. 961–1010. https://doi.org/10.2136/sssabookser5.3.c34. Madison, Wisconsin.

Pachepsky, Y.A., Rawls, W.J., 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. Soc. Am. J. 636, 1748–1757.

Park, Y., Ho, J.C., 2021. Tackling overfitting in boosting for noisy healthcare data. IEEE Trans. Knowl. Data Eng. 33, 2995–3006. https://doi.org/10.1109/TKDE.2019.2959988.

Pereira, P., Bogunovic, I., Muñoz-Rojas, M., Brevik, E.C., 2018. Soil ecosystem services, sustainability, valuation, and management. Curr. Opin. Environ. Sci. Health 5, 7–13. https://doi.org/10.1016/j.coesh.2017.12.003.

Pessoa, T.N., Libardi, P.L., 2022. Physical-hydric properties of Oxisols as influenced by soil structure and clay mineralogy. Catena 211, 106009.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199. https://doi.org/10.1007/s10021-005-0054-1.

Puppala, A.J., Pedarla, A., Hoyos, L.R., Zapata, C., Bheemasetti, T.V., 2016. A semi-empirical swell prediction model formulated from 'clay mineralogy and unsaturated soil' properties. Eng. Geol. 200, 114–121. https://doi.org/10.1016/j.enggeo.2015.12.007.

Putatunda, S., Rama, K., 2018. A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost. ACM International Conference Proceeding Series 6–10. https://doi.org/10.1145/3297067.3297080.

Ramosaj, B., Pauly, M., 2019. Consistent estimation of residual variance with random forest out-of-bag errors. Stat. Probab. Lett. 151, 49–57. https://doi.org/10.1016/j.spl.2019.03.017.

Razzaghi, F., Arthur, E., Moosavi, A.A., 2021. Evaluating models to estimate cation exchange capacity of calcareous soils. Geoderma 400, 115221. https://doi.org/10.1016/j.geoderma.2021.115221.

Rhoades, J.D., 1982. Cation exchange capacity. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis. Part 2. Agronomy Monograph No. 9, Madison, American Society of Agronomy, pp. 149–157.

Saidi, S., Ayoubi, S., Shirvani, M., Azizi, K., Zeraatpisheh, M., 2022. Comparison of different machine learning methods for predicting cation exchange capacity using environmental and remote sensing data. Sensors 22 (18), 6890. https://doi.org/10.3390/s22186890.

Sarkar, B., Singh, M., Mandal, S., Churchman, G.J., Bolan, N.S., 2018. Chapter 3- clay minerals- organic matter interactions in relation to carbon stabilization in soils. In: Garcia, C., Nannipieri, P., Hernandez, T.B. (Eds.), The Future of Soil Carbon. Academic Press, pp. 71–86.

Saurabh Pal, R.A., 2021. Elimination and backward selection of features P-value technique in prediction of heart disease by using machine learning algorithms. Turkish J. Comput. Math. Educ. TURCOMAT 12, 2650–2665. https://doi.org/10.17762/turcomat.v12i6.5765.

Shahrokh, V., Khademi, H., Zeraatpisheh, M., 2023. Mapping clay mineral types using easily accessible data and machine learning techniques in a scarce data region: a case study in a semi-arid area in Iran. Catena 223, 106932.

Silva, F.M., Silva, S.H.G., Andrade, R., Coblinski, J.A., Inda, A.V., Frosi, G., Lima, S.S.F., Menzes, M.D., Tavares, T.R., Guilherme, L.R.G., Weindorf, D.C., Curi, N., 2024. Proximal sensors for modeling clay mineralogy and characterization of soil textural fractions developed from contrasting parent materials. Catena 241, 108053. https://doi.org/10.1016/j.catena.2024.108053.

Soil Survey Staff, 2014. Soil Survey Staff. Keys to Soil Taxonomy, 12th ed. USDA. United States Department of Agriculture Natural Resources Conservation.

Sokolova, T.A., Tolpeshta, I.I., Rusakova, E.S., Maksimova, Y.G., 2013. Clay minerals in the stream floodplain soils in undisturbed landscapes of the Southern Taiga (with the soil of the state central forest nature and biosphere reserve as an example). Moscow Univ. Soil Sci. Bull. 68, 154–163. https://doi.org/10.3103/S0147687413040078.

Somaratne, S., Seneviratne, G., Coomaraswamy, U., 2005. Prediction of soil organic carbon across different land-use patterns. Soil Sci. Soc. Am. J. 69, 1580–1589. https://doi.org/10.2136/sssaj2003.0293.

Spinola, D., Portes, R., Fedenko, J., Lybrand, R., Dere, A., Biles, F., Trainor, T., Bowden, M.E., D'Amore, D., 2022. Lithological controls on soil geochemistry and clay mineralogy across Spodosols in the coastal temperate rainforest of southeast Alaska. Geoderma 428, 116211. https://doi.org/10.1016/j.geoderma.2022.116211.

Stahr, K., Schuler, U., Haring, V., Clemens, G., Herrmann, L., Zarei, M., 2010. Clay and oxide mineralogy of different limestone soils in Southeast Asia. In: Proceedings of World Congress of Soil Science, Soil Solutions for a Changing World, pp. 37–39. Accessed on January 17, 2024. https://www.iuss.org/19th%20WCSS/Symposium/pdf/1241.pdf.

Steichen, T.J., Cox, N.J., 2002. A note on the concordance correlation coefficient. Stata J Promot Commun Stat Stata 2, 183–189. https://doi.org/10.1177/1536867x0200200206.

Sumner, M.E., Miller, W.P., 1996. Cation exchange capacity and exchange coefficients. In: Sparks, D.L. (Ed.), Methods of Soil Analysis, Part 3, Chemical Methods—SSSA, pp. 1085–1121. https://doi.org/10.2136/sssabookser5.3.c40 book series no 5, Madison, Wisconsin.

Sun, X., Xie, M., Zhou, F., Fu, J., Liu, J., 2023. Multi-objective optimization for combustion, thermodynamic and emission characteristics of Atkinson cycle engine using tree-based machine learning and the NSGA II algorithm. Fuel 342, 127839. https://doi.org/10.1016/j.fuel.2023.127839.

Thomas, G.W., 1982. Exchangeable cations. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, pp. 159–165. https://doi.org/10.2134/agronmonogr9.2.2ed. Part 2. Agronomy Monograph No. 9, Madison, American Society of Agronomy.

Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. J Geophys Res Earth 116F4, 1–15. https://doi.org/10.1029/2011JF001977.

Wade, C., 2020. Hands-on Gradient Boosting with XGBoost and Scikit-Learn, 1st ed. Packt Publishing.

Wakeel, A., Ishfaq, M., 2022. Potassium dynamics in soils. In: Potash Use and Dynamics in Agriculture. Chapter 2. Springer, pp. 7–17. https://doi.org/10.1007/978-981-16-6883-8. ISBN 978-981-16-6882-1.

Watanabe, T., Funakawa, S., Kosaki, T., 2006. Clay mineralogy and its relationship to soil solution composition in soils from different weathering environments of humid Asia: Japan, Thailand and Indonesia. Geoderma 136 (1–2), 51–63. https://doi.org/10.1016/j.geoderma.2006.02.001.

Whittig, L.D., Allardice, W.R., 1986. X-Ray Diffraction Techniques. In: Klute, A. (Ed.), Methods of Soil Analysis Part 2, 2nd Ed. American Society of Agronomy, Madison, WI, pp. 331–362. https://doi.org/10.2136/sssabookser5.1.2ed.c12.

Wilding, L.P., 1985. Spatial variability: Its documentation, accommodation, and implication to soil surveys. In: Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability. Pudoc, Wageningen, The Netherlands.

Wilford, J., De Caritat, P., Bui, E., 2015. Modelling the abundance of soil calcium carbonate across Australia using geochemical survey data and environmental predictors. Geoderma 259, 81–92. https://doi.org/10.1016/j.geoderma.2015.05.003.

Xue, W., Wu, T., 2020. Active learning-based XGBoost for cyber physical system against generic AC false data injection attacks. IEEE Access 8, 144575–144584. https://doi.org/10.1109/ACCESS.2020.3014644.

Yilmaz, K., Çelik, I., Kapur, S., Ryan, J., 2005. Clay minerals, Ca/Mg ratio and Fe-Al-oxides in relation to structural stability, hydraulic conductivity, and soil erosion in southeastern Turkey. Turk. J. Agric. For. 291, 29–37. https://journals.tubitak.gov.tr/agriculture/vol29/iss1/4.

Zhao, D., Zhao, X., Khongnawang, T., Arshad, M., Triantafilis, J., 2018. A Vis-NIR spectral library to predict clay in Australian cotton growing soil. Soil Sci. Soc. Am. J. 826, 1347–1357. https://doi.org/10.2136/sssaj2018.03.0100.

Zhao, L., Hong, H., Fang, Q., Algeo, T.J., Wang, C., Li, M., Yin, K., 2020. Potential of VNIR spectroscopy for prediction of clay mineralogy and magnetic properties, and its paleoclimatic application to two contrasting quaternary soil deposits. Catena 184, 104239. https://doi.org/10.1016/j.catena.2019.104239.

Zhu, Y., Bennett, J.M., Marchuk, A., 2019. Reduction of hydraulic conductivity and loss of organic carbon in non-dispersive soils of different clay mineralogy is related to magnesium induced disaggregation. Geoderma 349, 1–10. https://doi.org/10.1016/j.geoderma.2019.04.019.

Zolfaghari, Z., Mosaddeghi, M.R., Ayoubi, S., Kelishadi, H., 2015. Soil atterberg limits and consistency indices as influenced by land use and slope position in Western Iran. J. Mt. Sci. 12, 1471–1483. https://doi.org/10.1007/s11629-014-3339-z.