Check for updates

Different approaches to estimating soil properties for digital soil map integrated with machine learning and remote sensing techniques in a sub-humid ecosystem

Fikret Saygın · Hasan Aksoy · Pelin Alaboz · Orhan Dengiz

Abstract Today, data mining has become a relevant topic in digital soil mapping. In this current study, prediction of some soil properties and their spatial distribution were examined by machine learning algorithms (Support Vector Machine, Artificial Neural Network) using reflectance values of Triplesat satellite image bands in Vezirköprü district of Samsun province. The band data obtained from different wavelengths revealed positive correlations between the electrical conductivity and calcium carbonate equivalent contents of the soils. The support vector machine algorithm was the most successful to estimate the textural fractions, organic matter, electrical conductivity, and calcium carbonate equivalent contents of the soils using the bands obtained from satellite images. The mean absolute error for estimating sand, silt, and clay contents by support vector machine was 4.05%, 3.05%, and 3.66%, respectively. Texture classes were determined with an accuracy of 82% with support vector machine and 60% with artificial neural network. In all estimations, the highest percentage of error was for calcium carbonate equivalent content with very low estimation reliability. The mean absolute percentage of error values for this property are 101.13% and 51.61% for artificial neural network and support vector machine, respectively. Also, in both algorithms, the most successfully estimated soil property was clay fraction of soils. It was also investigated the spatial distribution of actual and estimated values using various interpolation methods (Kriging, inverse distance weighting-, radial basis function). Considering the spatial distributions, it was determined that the most successful method was kriging for sand, silt, and clay contents and inverse distance weighting for electrical conductivity, calcium carbonate equivalent, and organic matter contents. According to our findings, it is concluded that successful estimations and spatial distributions can be made by the support vector machine algorithm using band data from different wavelengths.

Keywords Soil structure · Artificial neural network · Satellite image · Interpolation methods

F. Saygın
Faculty of Agriculture Sciences and Technology, Plant Production and Technology Department, Sivas University of Science and Technology, Sivas, Turkey

H. Aksoy
Department of Forestry and Forest Products, Ayvacık Vocational School, Sinop University, Sinop, Turkey

P. Alaboz (✉)
Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Isparta University of Applied Sciences, Isparta, Turkey
e-mail: pelinalaboz@isparta.edu.tr

O. Dengiz
Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Ondokuz Mayıs University, Samsun, Turkey

🖄 Springer

Introduction

Soil is a crucial natural environment for plant growth with the presence of three-dimensional layers located on the earth. Many physical, chemical, and biological properties of soils impact their sustainability and productivity. In addition, soil properties play a critical role in not only terrestrial ecosystem services but also in erosion, irrigation, drainage management, land planning, soil improvement, and soil management (Wu et al., 2018). Among the physical properties of soils, one of the most significant genetic properties is textural fractions. The textural properties of soils are the most frequently used parameter for developing pedotransfer functions (Alaboz & Işıldar, 2019), with a crucial role in all properties of soil, including water holding capacity, soil fertility, soil air, and drainage. Determining the spatial distributions for the textural fractions of soils is also important for plant production (Shahriari et al., 2019; Wu et al., 2018). Moreover, organic matter content improves the physical and chemical properties of the soil, regulates the nutrient supply, affects soil fertility and sustainability, and is a critical property for the functioning of the ecosystem due to its impact on the global carbon cycle (Dong et al., 2021). The EC and calcium carbonate equivalent contents of soils are also some of the parameters that should be followed for the rehabilitation and management of problematic soils. Determining the unique properties of soils provides convenient management and allows for efficient use of resources. Better decision-making about the use and management of natural resources depends on gathering accurate information by identifying changes (Akinyemi et al., 2017; Aksoy & Kaptan, 2021; Feng et al., 2020; Gharbia et al., 2016; Moradi et al., 2020). The soil maps that are produced to ensure the correct and sustainable use of land resources and to show the distribution areas of soils with different properties, and the reports that accompany these maps constitute a database for users (Dengiz et al., 2012). High resolution maps provide convenience in agricultural production, forest management, hydrological analysis, and environmental protection plans (Zhao et al., 2009).

In recent years, advances in technology have increased the interest in digital mapping. Geographical Information Systems and Remote Sensing applications have become increasingly widespread, which is of great significance in monitoring land use and land cover (LULC) on a spatial and temporal scale (Kesgin & Nurlu, 2009; Lu et al., 2011; Sünbül & Tonyaloğlu, 2021). Satellite systems have different spatial and temporal coverage areas, high resolution imaging, and a better capacity for data collection, all of which have increased their use in monitoring LULC changes (Aksoy & Kaptan, 2021). Data from visible, near-infrared, and short-wave infrared (SWIR) bands can be used to validate field-derived datasets that describe the correlations between quantitative soil properties and spatial and temporal information for large areas (Mulder et al., 2011). Studies have revealed a very strong correlation between soil properties and multispectral remote sensing images (Gozdowski et al., 2015; Maselli et al., 2008; Mitran et al., 2019). Digitally soil maps consist of field observations and laboratory analyses combining numerical models with environmental variables to create a geo-based soil information system, then mapping the area and estimating many values based on certain variables. Generating soil maps using satellite images provides a synoptic view of the area (de Castro Padilha et al., 2020; Grunwald et al., 2011; Mulder et al., 2011) and allows the examination of temporal changes in soil properties (Fathololoumi et al., 2020).

Today, machine learning algorithms are one of the most widely used estimation methods in digital soil mapping (Alaboz et al., 2021; Wang et al., 2021). Machine learning is the process of learning from data using the necessary algorithms and formulations and reaching a level to make a decision about the relevant issue (Öztemel, 2012). Researchs have shown that technological approaches like machine learning methods help obtain reliable results (Alaboz et al., 2021; Yamaç et al., 2020). Silva et al. (2020) reported successful estimation of clay and sand contents of soils using support vector machine (SVM) and silt content using the random forest algorithm. Dong et al. (2021) stated that the Cubist method outperformed SVM and Artificial Neural Network (ANN) in estimating organic carbon content of soils.

In the current study, we aimed to: i-) determine the predictability of textural fractions (sand, silt, clay), organic matter, calcium carbonate equivalent, and electrical conductivity contents of soils by machine learning algorithms (SVM-ANN) using the reflectance values of 5-band Triplesat satellite image and ii-) determine spatial distribution of observed and estimated values using different interpolation techniques

to find the most suitable model. This study focuses on "I) determining the best interpolation model in spatial distribution maps for basic soil properties, II) evaluating the predictive accuracy of soil properties with SVM and ANN algorithms, III) examining the distribution pattern of maps created with predictive values".

Material and method

Description of the study area

The study area is located in Vezirköprü district of Samsun province located in the Central Black Sea Region and coordinated 41° 2.518'N—35° 32.986'E and 41° 10.234'N—35° 30.087'E (WGS84, Zone-36, UTM-m) (Fig. 1).

The study area size is about 111 km², with an altitude of 240–750 m above the mean sea level (Fig. 2). The slope within the study area often ranges from flat-mild steep to moderately steep; the areas in the northeast and southeast have steep slopes. When the geological characteristics of the study area are examined, it is seen that a large part consists of mudstone and limestone rocks. In addition, agglomerate tuff and andesite rocks are distributed in the southern parts of the area, while sandstone and occasionally alluvial deposits are found in the northern parts.

Regarding climatic conditions, Vezirköprü district has a distinct characteristic between the humid and temperate coastal climate and the continental climate in the interior regions, carrying some thermal and humid characteristics peculiar to the transition zone. The winters are colder than the coast, and the summers are warmer. Long term of meteorological data (1977–2019) show an annual average temperature of 12.5 °C and a precipitation of 724.5 mm. Bölük (2016) reports that the study area is classified as "sub humid," with a precipitation activity index of 57.94 points based on the macroclimate regions of Erinc in Turkey. In
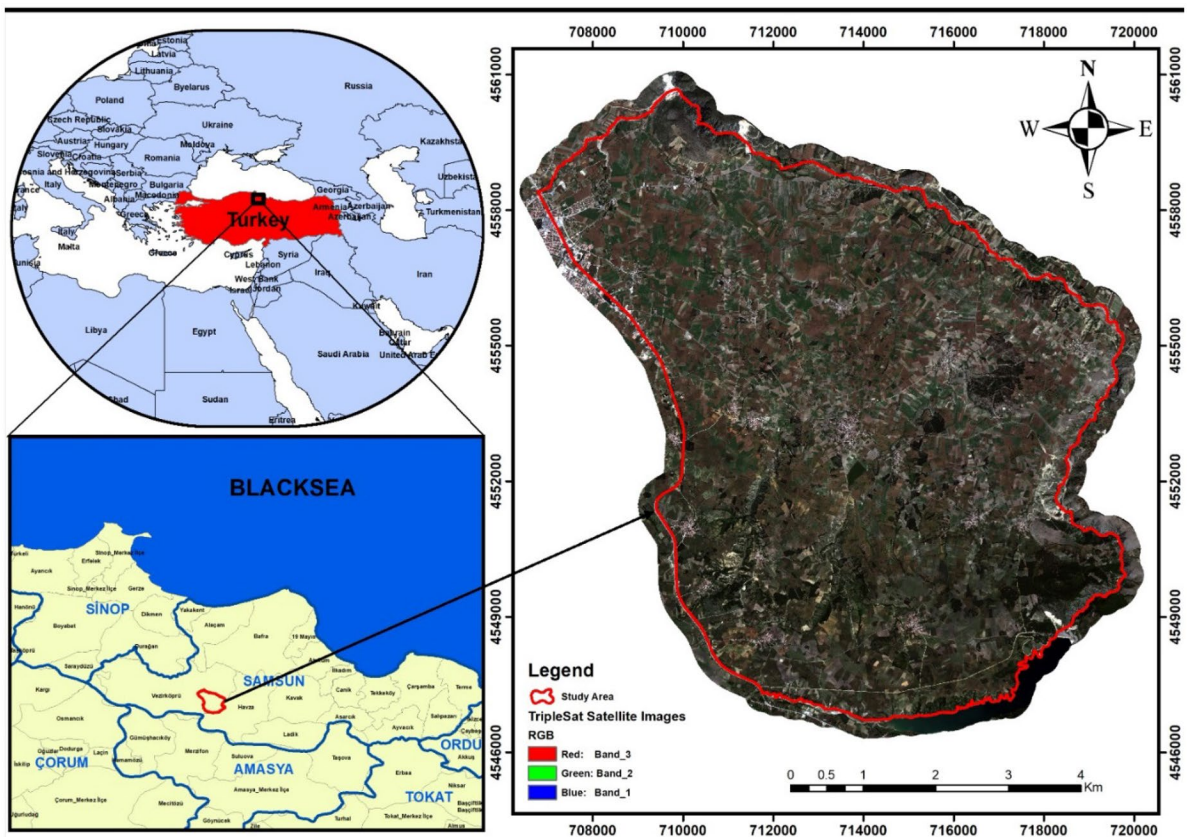

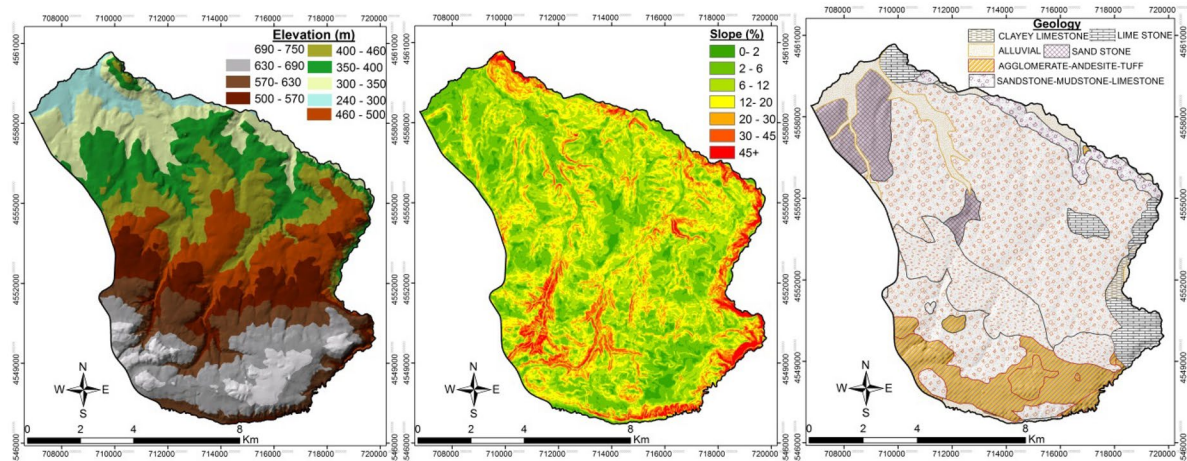**Fig. 1** Location map of the study area

Fig. 2 Elevation, slope and geology maps of the study area

addition, Turan et al. (2018) indicated that Vezirköprü district has a climate that was expressed by *C1B'2db'4* that means sub humid, second-order mesothermal, no excess water, close to marine influence. Besides, according to the Newhall simulation model of the soils within Vezirköprü, the soil temperature is determined as Mesic, and the moisture regime as Xeric (in subgrup: Typic Xeric) (Turan et al., 2018; Van Wambeke, 2000). Intensive irrigated and dry agriculture have been carried out over a large part of the study area, and a very small part is covered by pasture and forest areas.

Soil sampling and laboratory analyses

Soil samples of 0–30 cm depth from 657 points at 400 m x 400 m intervals according to the grid system were collected (Fig. 3). Before the analysis, the soil samples were separated from coarse particles, passed through a 2 mm sieve in the laboratory. They were prepared for physical and chemical analysis. Table 1 shows the protocols of soil physic-chemical analyses on the soil samples.

Satellite imaging and properties

In this study, it was used Triplesat satellite image and performed band separations. Table 2 shows the technical specifications of Triplesat satellite images.

The Triplesat satellite image is dated 04.06.2018. It has 5 bands, including panchromatic and multispectral bands. These bands have spatial resolutions of 0.8 and 3.2 m, respectively. The satellite images have a section width of 23.4 km and a temporal resolution of 97.7 min. The reflectance data for the sample areas were obtained in two steps. In the first step, the satellite image (Triplesat) was calibrated. In the second step, the reflectance values of the pixels were calculated according to the sample points. First, the Triplesat satellite image was calibrated using the QGIS Desktop 3.8.1 software, making atmospheric corrections. In this way, the reflectance values of all bands were obtained for the satellite image. The second step involves obtaining these data in accordance with the areas of the samples. For this, we transformed each band into a vector layer using the ArcGIS Desktop 10.7.1 software. Then, we overlapped this vector layer and obtained the vector data that is suitable for the sample area. The reason for converting from raster to vector is that when the pixels are sectioned without converting according to the sample area, there may be gaps or excess areas remaining at the borders based on the spatial resolution. By converting to vector, we could make identical inferences according to the sample area and obtain data from all pixels corresponding to the boundaries. Then, we converted the layer to raster data again, applied the zonal statistics command, and obtained the average reflectance values for each sample point.

Fig. 3 Soil sampling pattern in the study area
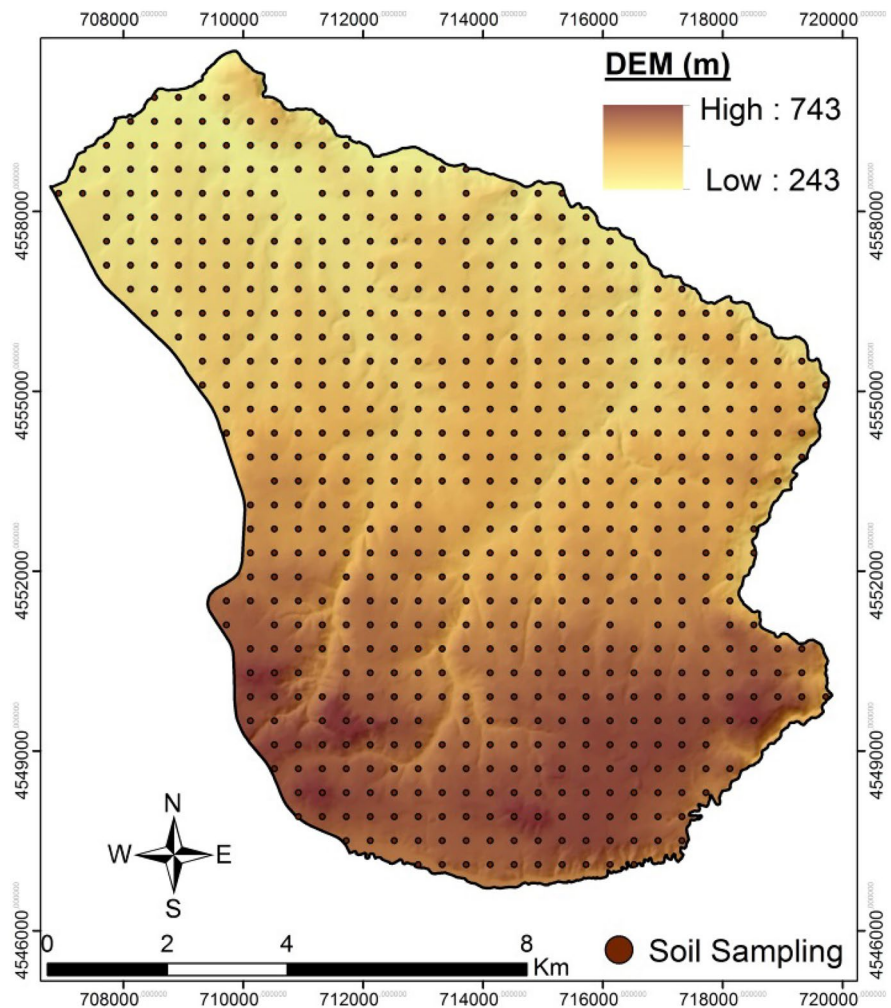
Estimation models (support vector machine-regression and artificial neural network)

Support Vector Machine is a supervised machine learning algorithm that is used both in classification and regression analyses. Depending on the type of data, Kernel functions are used during the operation of the algorithm; this way, both linear and nonlinear classifications are estimated. The purpose of using SVM modeling is to find the hyperplane that gives the maximum margin that separates the vector sets with one-category states on one side of the plane and other categorical states on the other

Table 1 Soil analysis methods and their references

Analysis	References
Soil texture (clay, silt, sand) (%)	Bouyoucos (1951)
Organic matter (%)	Jackson (1958)
calcium carbonate equivalent (%)	Soil Survey Staff (1993)
Electrical conductivity (μmhos cm^{-1})	Richards (1954)

Table 2 Technical properties of Triplesat satellite image

Bands	Electromagnetic area	Wavelength (nm)	Spatial resolution (m)
B1	Blue (B)	440 – 510	3.2
B2	Green (G)	510 – 590	3.2
B3	Red (R)	600 – 670	3.2
B4	Near Infrared (NIR)	760 – 910	3.2
B5	Panchromatic	-	0.8

side (Cortes & Vapnik, 1995). SVM is a machine algorithm that finds the best separating hyperplane in an N-dimensional dataset. It draws a hyperplane according to the reference points and these points are the support vectors. SVM does not always present a linear plane; it is much easier to find hyperplanes in linear planes, and using such lines, classifications can be made (Smola & Schölkopf, 2004). When the plane is not linear, the data is given one more dimension to determine the hyperplane. The Kernel function can transform a nonlinear separation problem in the input space into a linear problem in the property space. Then, the nonlinear problem can be solved linearly in the property space. We used the support vector machine to model the complex correlations between soil properties and band data. Support Vector Regression is a popular machine learning tool with relatively high accuracy and low computational costs (Were et al., 2015). Various soil modelling studies have shown that SVR is promising because of its capacity to handle non-linear correlations and generalization (Pasolli et al., 2011; Sihag et al., 2018; Taghizadeh-Mehrjardi et al., 2020). Though there are many Kernel functions in the literature, the most frequently used ones are linear function, polynomial function, sigmoid function, and radial basis function (Kuzu & Yakut, 2020). The system aims to find a hyperplane that separates the data points with the maximum margin. Equation (1) gives the basis of the SVM vector.

$$f(x) = \omega.\varphi(x) + b \tag{1}$$

where $\varphi(x)$ are non-linear, transformed training data, ω: weight vectors, b: bias stands for the distance of the hyperplane from the original. The coefficients ω and b are estimated by minimizing the regular risk function; to maximize this margin, ω must be minimized.

To arrange unsuitable groups, a C function is needed (Eqs. 2 and 3).

$$R(C) = Cn \, \Sigma n \, i = 1 \, L\varepsilon(yi, f(xi)) + 12 \, \|\omega\|^2 \tag{2}$$

$$L\varepsilon(yi, f(xi)) = \{|f(xi) - yi| - \varepsilon, \\ for |f(xi) - yi| \geq \varepsilon \, (0, \text{ otherwise}) \tag{3}$$

The equation is then written as Eqs. 4 and 5.

$$C\Sigma n \, i - 1(\zeta i, \zeta *i) + 12\|\omega\|^2 \tag{4}$$

$$yi - (\omega\varphi(xi) + bi) \leq \varepsilon + \zeta i \, (\omega\varphi(xi) + bi) \\ - yi \leq \varepsilon + \zeta * i \, \zeta i, \, \zeta * i \geq 0, \, i = 1, \cdots, n \tag{5}$$

where, ξi refers to the error made by that data, called a loose variable.

$12\|\omega\|^2$: the regulatory term that refers to the complexity of the learning machine, and $Cn\Sigma ni = 1L\varepsilon(yi, f(xi))$ is the empirical error. ζi and $\zeta *i$: Slack variables *are* positive and negative errors at the *i*th point, respectively. C is the penalty factor and ε is the loss function.

Optimization problems are solved using Lagrange and Karush–Kuhn–Tucker conditional methods (Eq. 6).

$$L = C\Sigma n \, i = 1(\zeta i, \zeta * i) + 12 \, \| \omega \| \, 2 \, - \Sigma n \, i \\ = \, ai(\omega\varphi(xi) + b - yi + \varepsilon + \zeta i) - \Sigma n \, i \\ = \, 1 \, a * i(yi - \omega\varphi(xi) - b + \varepsilon + \zeta i *) - \Sigma n \, i \\ = 1(\beta i\zeta i + \beta * i\zeta * i) \tag{6}$$

L stands for Lagrangian, and β, $\beta*$, a and a*: Lagrange factors.

Nonlinear Support Vector Regression functions are obtained using Eq. 7.

$$f(x) = \omega.\varphi(x) + b = \Sigma n \, i = \, 1 \, (ai, \, a * i)k(xi, \, x) + b \tag{7}$$

$k(xi, x)$ is the Kernel function.

Here, we used a radial basis function as the Kernel function (Ballabio, 2009) (Eq. 8). Radial basis moves the sample to a higher dimensional space because it is nonlinear. It differs from other Kernel functions in that it is a more compressed Kernel, a valuable feature in model design, restricting the computational training process, and making it more convenient to improve the efficiency of generalization.

$$k(xi, \, x) = exp\left(\sigma \, \| \, xi, x \, \|^2\right) \tag{8}$$

σ is the "spread" of the Kernel.

The data set was evaluated as 70% training and 30% testing. We used the" e1071" and "Caret" packages in the R software (Cost:1, gamma:0.25, epsilon:0.1).

Secondly, we used multilayer feedforward back-propagation networks (MLP). The multilayer sensors in artificial neural networks consist of one input layer, one or more intermediate layers, and one output layer (Çakır, 2019). The first layer is the input layer (the properties that will be used for estimation), the second layer is hidden layers, and the third layer is the output layer (the properties that will be estimated).

In multilayer artificial neural networks, the network architectures with the highest predictive power can be created by making changes in the number of hidden layers and the number of neurons. For creating the estimation with artificial neural networks, the Levenberg–Marquardt algorithm (LM) was used at the 5:10:1 architecture, using feedforward backpropagation. The "nntool" package program was used in MATLAB to obtain the ANN estimations. In artificial neural networks, 70% of the data set was used for model creation, 30% for validation and testing. For evaluating the models, the parameters of root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) coefficient of determination (R^2) and Lin's concordance correlation coefficient (LCC)were used. These statistical indices were also used in similar studies (Mousavi et al., 2022; Rezaei et al., 2023).

Descriptive statistics, correlations, and estimations for soil properties were made on the "R" software. LCC and R^2 values are calculated using the "Desc-Tools and MLmetrics " packages.

In order to evaluate the uncertainty of the model, the variance of the 50 predicted produced from each model was computed for soil properties. The square root of variance is calculated (Standard deviation) (Malone et al., 2011; Sharififar, 2022). The important variables in the models were determined as stated in Kaya et al. (2022) and FAO (2022).

Interpolation analysis and creating spatial distribution maps

Surface (0–30 cm) distribution maps were created based on the texture (sand, silt, clay), organic matter, calcium carbonate equivalent (CCE), and EC content analysis for each soil sample, with coordinates in the grid system at 400 m*400 m, using interpolation models. Intensive agriculture in the region will affect a sustainable management by knowing the basic soil characteristics. Therefore, the focus is on the predictability of basic soil properties and the evaluation of distribution maps. Spatial distribution maps were created for each parameter and the values obtained from the estimation models using different interpolation models in the GIS software ArcGIS. Before the mapping, the data that was not normally distributed were transformed according to the properties. The

Kolmogorov–Smirnov test of normality was applied to soil properties. Kolmogorov–Smirnov test results that did not exhibit a normal distribution were transformed using a logarithmic, root transformation. The skewness value is approximated to the normal distribution by applying square root transformations if it is between 0.5 and 1.0, and logarithm transformations if it is > 1.0. For mapping, the Inverse Distance Weighting (IDW), Radial Basis Function (RBF) called as deterministic approach and Kriging method defined as scholastic approach were used. Kriging methods are Ordinary, Simple, and Universal and their semivariogram models such as Spherical, Exponential, and Gaussian were evaluated in the present study. The pixel size of the distribution maps is 50 m*50 m. Variogram parameters are indicated in Table 5. The most appropriate method was accepted to be the one that gave the lowest root mean square error.

Results and discussion

Soil properties and electromagnetic bands

Table 3 gives the basic descriptive statistics of soil properties. The sand, silt, and clay contents of the soils varied between 13.04–61.93%, 8.63–42.45%, and 4.148–69.198%, respectively. The soils in the study area dominantly had heavy-textured clay (C). Of the sampling points, 14% were clay loam (CL), 5.78% sandy clay loam (SCL), 1.84% loam (L), 1.21% sandy clay (SC), 0.3% silty clay (SiC) and silty loam (SiL), 0.15% was sandy loam (SL), and the remainder was clay (Fig. 4). The EC contents of the soils were non-saline/slightly saline according to Doran and Jones (1996), and the organic matter content was between very low/high according to Hazelton and Murphy (2016). Calcium carbonate equivalent content ranged between 0.624–38.156%.

When the coefficients of variation of soil properties were examined, about 30% differentiations were found compared to the mean in the data set for the textural fractions, while this variation was found to be 40.93%, 64.0% and 68.83% for OM, EC and CCE. OM, EC, and CCE showed approximately 2 to 2.5 times more variation compared to the other textural fractions. According to the coefficient of skewness, the textural fractions showed a distribution close to normal, and the property that was farthest from

Table 3 Descriptive statistics of soil properties

OM Organic matter, *EC* Electrical conductivity, *CCE* calcium carbonate equivalent *St.Dev* Standard deviation, *CV* coefficient of variation, *Min* Minimum, *Max* Maximum

Variable	Mean	St.Dev	CV	Min.	Max.	Skewness	Kurtosis
Sand (%)	31.716	9.000	28.38	13.043	61.933	0.75	0.50
Silt (%)	22.855	5.038	22.04	8.663	42.453	0.49	0.98
Clay (%)	45.428	9.173	20.19	4.148	69.198	-0.51	0.55
OM (%)	2.3807	0.9744	40.93	0.2700	6.2395	1.11	1.69
EC (μmhos cm^{-1})	382.44	244.78	64.00	38.10	2300	3.54	16.81
CCE (%)	10.606	7.300	68.83	0.624	38.156	0.50	-0.16

normal was the EC content. This high value resulted from the low presence of slightly saline soil. Because positive skewness often results from lower than mean data. Most of the data set consists of non-saline soil. While CV alone is not sufficient to determine spatial variability in soil components, it is effective to detect variability in soil properties compared to other parameters like standard deviation (Zhang et al., 2007).

Figure 5 shows the correlations between soil properties and bands. A high correlation between the textural fractions is expected. Regarding the correlations between bands and soil properties, there were high-level correlations due to CCE and EC contents of the soils, rather than due to the textural fractions. Demattê et al. (2018) reported that Bands 5 and 7 (Landsat 7) were the most significant in clay estimation (r > 0.70;

p < 0.01). There were positive correlations between the bands coded as B1, B2, B3, and B5, and EC and CCE contents. The literature reports some differences in reflectance images based on the salt and CCE contents of soils (Aydın et al., 2011). The low-level correlation between band values and textural fractions may have stemmed from the fact that they were mostly from the same texture class. Khalil et al. (2016) examined the correlations between Landsat 8 satellite image reflectance values and soil properties and found significant correlations between silt and bands 2 and 5 ($R^2 = 0.52$) and between clay and bands 4 and 6 ($R^2 = 0.40$) (p < 0.05). Yue et al. (2019) reported that higher hydraulic conductivity and lower soil–water content showed higher reflectance in the band 8 (Landsat-6). The hyperspectral data has been utilized,

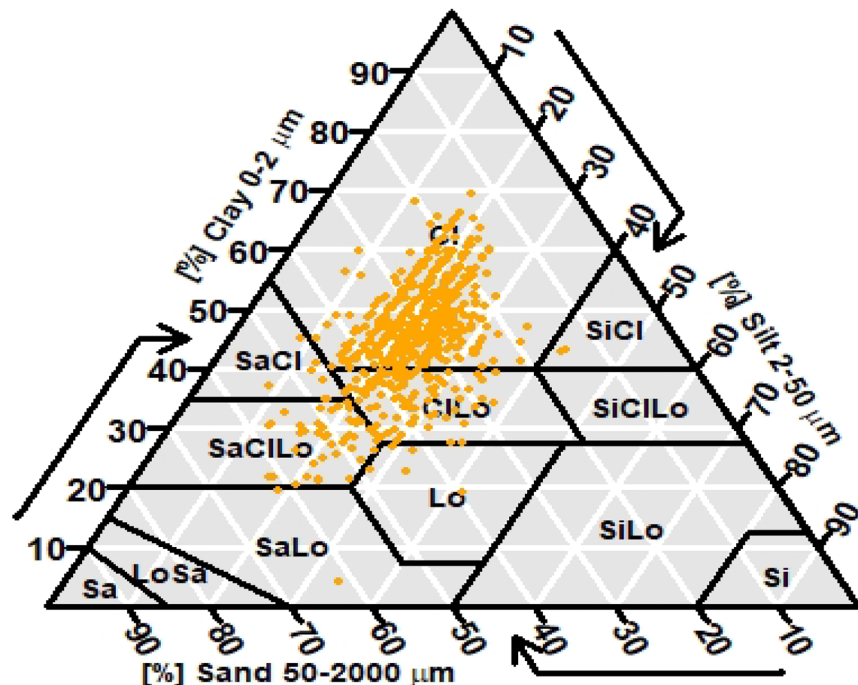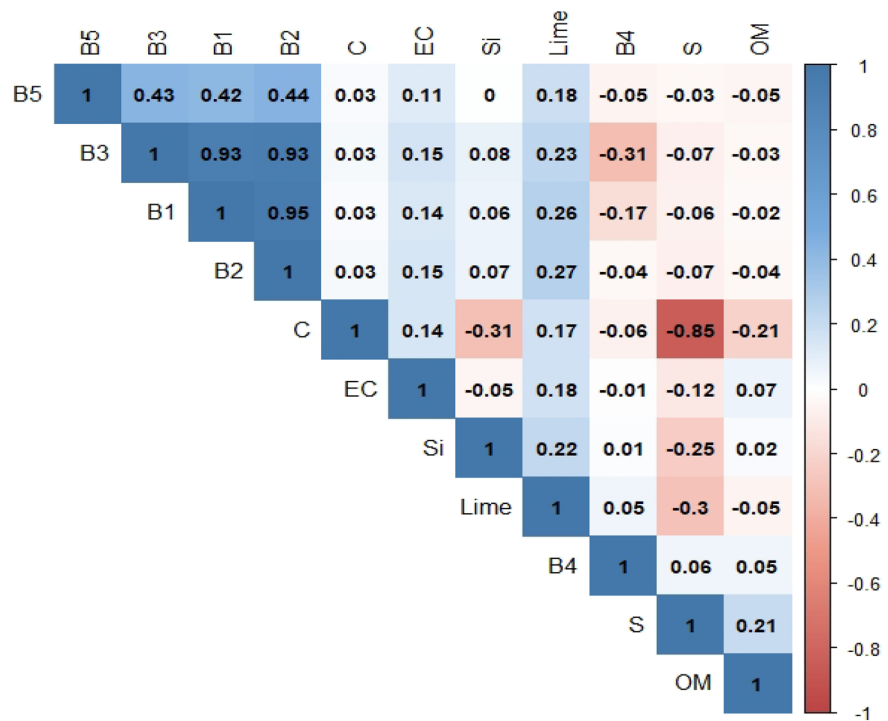**Fig. 4** Soil texture class distributions. Sa sand, Si silt, CI clay, Lo loamy

Fig. 5 Correlation coefficients between soil properties and bands. S sand, Si silt, C clay, Lime calcium carbonate equivalent, OM organic matter, EC electrical conductivity; B1, B2, B3, B4, and B5, blue, green, red, near infrared, and panchromatic electromagnetic area

incorporating the spectral reflectance of soil and soil-covered features that accurately represent the original pixels of satellite images (Bangelesa et al., 2020).

The CCE and EC content in the surface horizons of the soils impact the reflectance. As the carbonate minerals in the soils become quantitatively more concentrated, there will be more concentration of white color, making the radiometric reflectance of electromagnetic energy more intense. Basically, carbonate minerals and salts are light and white in color and they reflect the sun's rays intensely. Besides, as the sand content of the soil increases, the electromagnetic reflectance also increases. Besides the grain diameter size of sands, they reflect most of the electromagnetic wave coming from the sun, because they are often light in color (Özden & Altınbaş, 2005). Baumgardner et al. (1969) reported that organic matter affects the reflectance data of sunlight in locations where the organic matter content is more than 2%.

Evaluation of estimation models

In the study, all Triplesat satellite electromagnetic bands (B1, B2, B3, B4, B5) were used to evaluate the predictive accuracy of the models. Table 4 shows the rates of error for soil properties obtained using different bands during training and testing in SVM and ANN estimation values. The RMSE values for estimating the sand content of the soils using the SVM algorithm were 4.91% and 5.01%; with the ANN algorithm, this was about 1.5 times higher. We found a similar variability for the MAE and MAPE values. The MAE values for estimating the silt content were very close at 3.07% and 3.05% in training and testing, respectively, and they were 3.814% and 3.631% with ANN. Despite the lower rate of error in SVM, they both had estimations with close error levels. As with the sand content, more successful estimations were obtained with SVM for silt content, and the estimation rate with ANN was nearly 1.5 times higher than SVM. The MAPE values for estimating the textural fractions of the soils were mostly similar among the algorithms. However, there were significant differences in MAPE values for OM, CCE, and EC, and the lowest rate of error for these properties was determined for the SVM algorithm. According to Lewis (1982), a model is classified as highly accurate with a MAPE value below 10%, good with a value of 10–20%, acceptable- reasonable for 20–50%, and inaccurate or faulty for above 50%.

Table 4 Predictive accuracy of models

Model	Properties	Training					Testing				
		RMSE	MAE	MAPE %	R^2	LCC	RMSE	MAE	MAPE %	R^2	LCC
SVM	Sand	5.01	4.18	14.3	0.65	0.77	4.91	4.05	12.86	0.68	0.79
	Silt	3.84	3.07	14.27	0.44	0.60	3.84	3.05	14.36	0.55	0.67
	Clay	5.53	4.65	9.66	0.77	0.83	4.549	3.66	8.64	0.75	0.85
	OM	0.46	0.36	18.52	0.76	0.87	0.47	0.37	20.24	0.76	0.85
	CCE	2.7	2.06	65.36	0.73	0.71	2.9	2.24	51.61	0.64	0.72
	EC	44.18	34.87	13.01	0.72	0.78	44.74	35.77	11.63	0.86	0.88
ANN	Sand	8.391	6.388	19.72	0.12	0.35	8.505	6.668	23.05	0.11	0.37
	Silt	4.518	3.631	18.03	0.04	0.09	4.928	3.814	18.152	0.08	0.10
	Clay	8.53	6.8	17.61	0.86	0.88	9.14	7.165	18.92	0.85	0.88
	OM	0.90	0.674	36.16	0.10	0.20	0.91	0.683	36.51	0.10	0.22
	CCE	6.78	5.77	99.98	0.11	0.26	6.85	5.54	101.13	0.11	0.22
	EC	163.34	136.09	70.53	0.03	0.06	240.26	145.07	44.19	0.03	0.06

SVM Support vector machine, *ANN* Artificial neural network, *RMSE* Root mean squared error, *MAE* Mean absolute error, *MAPE* Mean absolute percentage of error, R^2 coefficient of determination, *LCC* Lin's concordance correlation coefficient, *EC* Electrical conductivity *OM* Organic matter, *CCE* calcium carbonate equivalent

According to Lewis (1982), the SVM algorithm model is classified as good for estimating sand, silt, OM, and EC, very good for clay, and incorrect or faulty for CCE. The ANN model is classified as good for estimating sand, silt, and clay, acceptable for OM, and incorrect or faulty for EC and CCE. In our estimations, the most effective algorithm was found to be SVM. Considering the distribution of CCE contents, a positive skewness draws attention. A positive skewness in distribution is explained as mean > median > mode; therefore, the values are often lower than the mean. Overall, an estimation error of ± 3 units in values lower than the mean resulted in a high MAPE value. Also, for estimating the texture classes from the proportional distributions of the textural fractions, SVM had an 82% successful estimation, and ANN had a 60% successful estimation. The higher rate of error in ANN may be due to the small number of properties in the input data, but the predictive accuracy may change based on the algorithm and the number of neurons (Alaboz et al., 2021).

Rossel et al. (2016) state that a LCC value of 1 signifies perfect agreement. An LCC value greater than 0.9 indicates excellent agreement, and a value ranging from 0.80 to 0.90 suggests good agreement. Moderate agreement is achieved when LCC values fall between 0.65 and 0.80, whereas values below 0.65 represent poor

agreement. According to the LCC values, Clay, OM, EC are classified as "good", others as " "moderate" agreement by SVM algorithm. Only clay is classified as "good agreement" with ANN. R^2 values are determined at lower levels than LCC. Similarly, differences in R^2 and LCC were found in Mousavi et al., 2022. Similar trends were observed in LCC and R^2, values.

Wu et al. (2018) found that SVM was more successful than decision tree and ANN in estimating texture class by machine learning algorithms. Heung et al. (2016) compared the effect of various machine learning algorithms for classification in digital soil mapping studies and reported that model selection greatly affects the outputs. Xing et al. (2018) reported that SVM was successful in estimating soil temperature, and Jiang et al. (2019) found that SVM was more effective than ANN in estimating soil salinity. Numerous studies support that SVM, a nonlinear estimation method, can successfully provide better estimation accuracy during the training and validation stages in terms of various criteria (Deiss et al., 2020). Again, Lamorski et al. (2008) reported that a 3-parameter SVM estimation was equally accurate with or more accurate than an eleven-parameter ANN. Panneerselvam et al. (2023) reported that, the artificial neural network model successfully predicted both low and high-frequency soil chemical properties within satellite images.

Fig. 6 Relative importance of bands reflectance
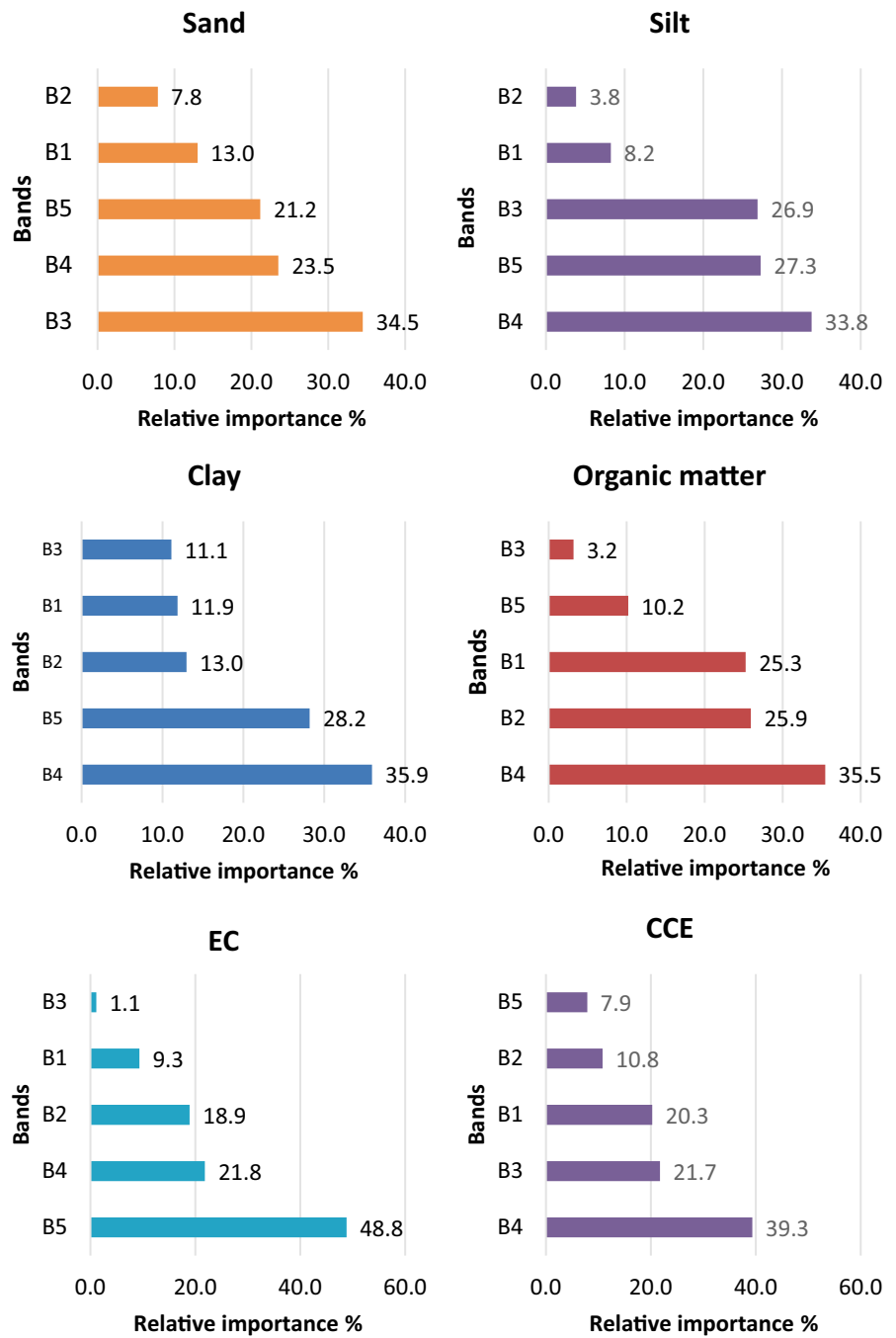
Relative importance of bands

The relative importance % of the bands on the models are given in Fig. 6. The most effective band in the estimation of the clay, silt organic matter contents and CCE contents was determined as near infrared (B4), and the relative importance % was determined as 33.8, 35.9, 35.5 and, 48.8. respectively. The most effective band in estimating EC contents was determined as panchromatic. B3 and B1 contributed models with similar weights. For Sand, the relative importance is determined in the highest red region (B3), while it is similar in B4 and B5.

The stretching and bending of NH, OH, and CH groups are connected to the near-infrared (NIR) regions of the electromagnetic spectrum (Viscarra Rossel & Behrens, 2010). NIR wavelengths have identified as important for estimation of soil organic carbon (Miloš & Bensa, 2017). Keshavarzi et al. et al. (2022) reported that Landsat 8 OLI near infrared band is the most important environmental variable in the digital mapping of the soil particle fraction. Red and blue bands are used in the indices used for the prediction of soil properties. It is known that these bands are effective on the reflection characteristics of soils and the differences in reflection may be caused by physical properties (Kaya et al., 2022).

The lowest contribution rate for the CCE property was obtained with the B5 band. In the estimation of organic matter and clay properties, it was determined as the B3 band, which provided the lowest contribution rate. B2 band was determined for sand and silt. As in this study, Stamatiadis et al. (2005) also determined important relationships between soil surface reflection and soil properties. The blue and green bands exhibited higher associations with soil properties than

Fig. 7 Sand, silt and clay maps (**A**: spatial distribution of the observed values, **B**: spatial distribution of the predicted values with SVM, **C**: spatial distribution of the predicted values with ANN)

the red and NIR bands. Linear correlation was found with soil water content, OM, K and P, blue and green bands. NIR band was found to be more effective in estimating clay from texture fractions, and Sørensen and Dalsgaard (2005) reported that NIR band could be used in clay and silt estimation. The NIR band predicts soil colour and water content. Studies related to the soil water content of the NIR band have been presented (Yin et al., 2013). Since the clay content is the most effective fraction on the soil water content, it is expected that the contribution rate of the B4 band is high. In a study by Shrestha (2006); The mid-infrared band and near-infrared were correlated with the EC values of the soil. Karavanova et al. (2001) reported that water-soluble salts are negatively correlated with spectral reflectance, and carbonates are positively correlated. He stated that the dominant bands, which are highly correlated with salt content, will vary depending on the effects of environmental conditions.

Table 5 Preferred models and semivariogram models' values for interpolation analysis

Properties	Determined model	RMSE	Lag size	C_0	A(m)	C_1
Sand	Kriging-Ordinary-Exponential	8.05	201.30	47.48	1811.92	23.89
Sand-SVM	Kriging-Simple-Exponential	5.58	815.70	0.78	4611.09	0.30
Sand-ANN	Kriging-Simple-Spherical	3.49	887.89	0.91	6767.11	0.94
Silt	Kriging-Simple-Spherical	4.72	317.54	0.76	2236.76	0.23
Silt-SVM	Kriging-Simple-Exponential	2.62	1508.49	0.93	2748.80	0.18
Silt-ANN	Kriging-Simple-Exponential	1.15	214.52	0.84	1901.40	0.09
Clay	Kriging-Ordinary-Exponential	8.05	233.50	45.10	1865.21	30.18
Clay-SVM	Kriging-Ordinary-Spherical	6.09	167.22	26.67	1147.24	12.10
Clay-ANN	Kriging-Simple-Spherical	3.91	781.84	0.94	6867.02	0.11
OM	IDW- 2. Degree[a]	0.87				
OM-SVM	Kriging-Ordinary-Exponential	0.69	166.33	0.31	1803.86	0.21
OM-ANN	Kriging-Simple-Exponential	0.33	542.36	0.93	5112.57	0.13
EC	Kriging-Simple-Gaussian	242.44	798.59	1.13	5259.68	0.12
EC-SVM	Kriging-Simple-Gaussian	232.09	773.59	0.77	5712.54	0.29
EC-ANN	IDW- 1. Degree[a]	43.45				
CCE	Kriging-Simple-Exponential	5.64	837.60	0.24	5337.96	0.65
CCE -SVM	Kriging-Simple-Spherical	5.16	887.89	0.51	6696.86	0.48
CCE -ANN	IDW- 2. Degree[a]	2.40				
Texture	Kriging-Universal-Exponential	0.45	100.531	0.06	736.85	0.14
Texture SVM	Kriging-Universal- Gaussian	0.38	1508.49	0.12	18,101.93	0.11
Texture ANN	Kriging-Simple- Exponential	0.55	574.92	0.96	4864.14	0.01

C0 nugget variance, *C1* partial sill, *A* range of influence (m), *RMSE* Root mean squared error for the theoretical semivariogram and crossvariogram models

[a]Variogram parameters are not obtained in the IDW method. *CCE* calcium carbonate equivalent, *EC* Electrical conductivity, *OM* Organic matter
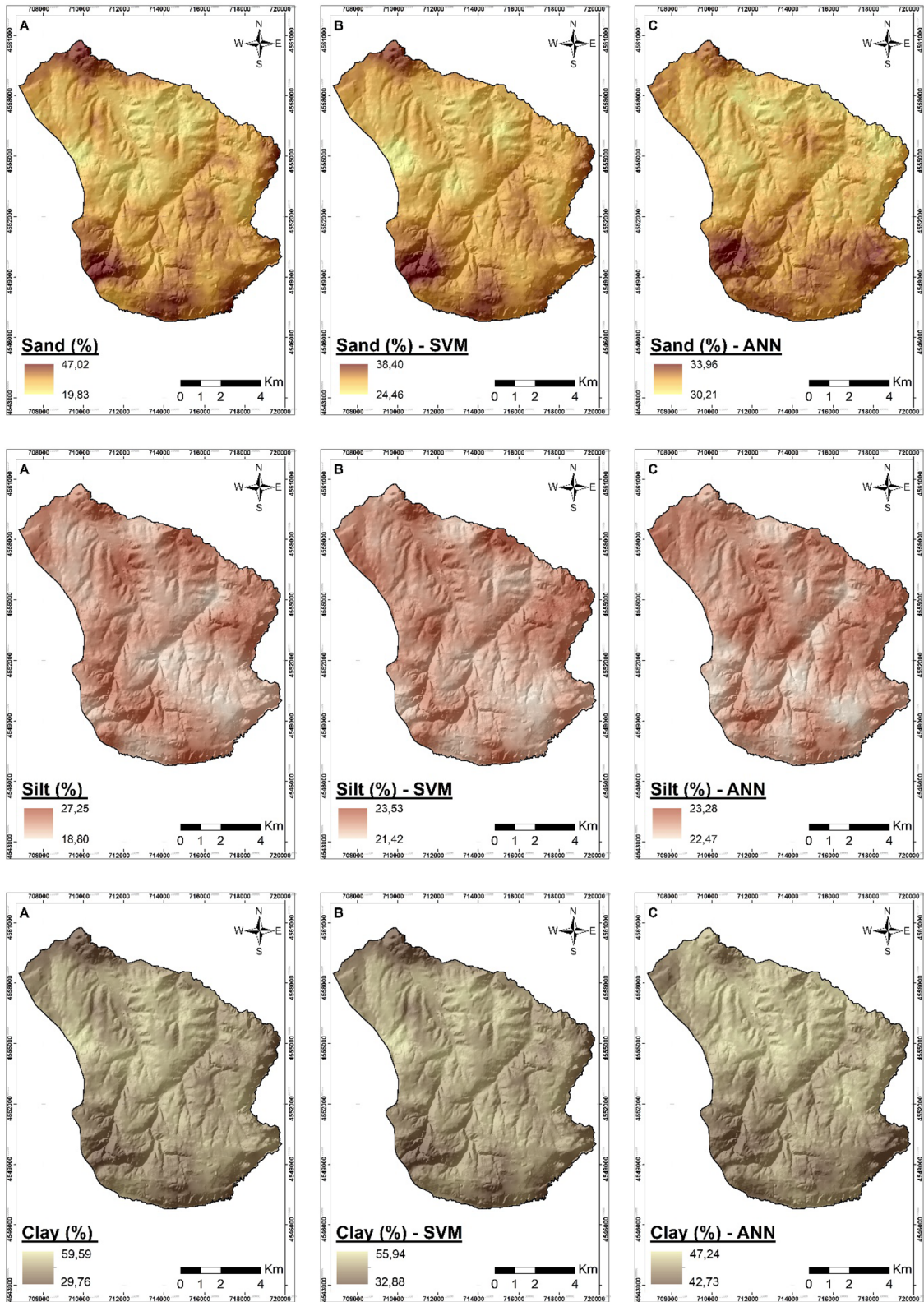
Spatial distribution of soil properties and estimated models

Table 5 shows the most suitable interpolation methods and their RMSE values for the actual soil properties and for estimating these properties by both algorithms. Among the interpolation methods, simple and ordinary kriging were more effective than IDW in the spatial distribution of soil properties. In the present study, for ANN estimates and distribution maps of organic matter, EC, and CCE contents, IDW was the most successful method. Gotway et al. (1996) reported that interpolation based on organic matter content gave more accurate results with the IDW method. In the IDW approach, the estimated value in the grid system decreases with increasing distance from the current point. Besides, in cases of a high coefficient of variation (CV), choosing a distance power as 1 gives the most accurate result, but this value is too weak for datasets with a low CV, so a distance power of 2 or 3 can be chosen for these datasets (Gotway et al., 1996). Here, we determined the properties with higher CVs than textural fractions using OM, CCE, and EC. In the kriging method, the distance is determined using semivariograms. The weights are independently chosen to minimize the estimated mean squared error (Gotway et al., 1996). Despite being difficult to estimate and model, the kriging semivariogram is a preferred

Fig. 9 Organic matter, EC and CaCO₃ CCE maps (**A**: spatial distribution of the observed values, **B**: spatial distribution of the predicted values with SVM, **C**: spatial distribution of the predicted values with ANN)

geostatistical method that is found to be very useful by scientific research (Gotway et al., 1996). Mitran et al. (2019) reported that the RMSE values for the spatial variability of sand, silt, and clay contents were lower in Regression kriging compared to multilinear regression. Seyedmohammadi et al. (2019) reported that, for a methodology that provides the spatial variation of soil texture using fuzzy logic and geostatistics, the accuracy of the distribution map produced by the kriging method was acceptable.

For the estimation data of the soil properties and models among different interpolation approaches, it was selected the method with the lowest RMSE and created the distribution maps. Considering the actual distribution of sand of the soils and the spatial distribution maps of SMV and ANN in Fig. 7, we see that the south-western and northern parts of the study area have higher sand content than the other areas, and the clay content is higher in the alluvial land part, with the decreasing slope from south to north.

Regarding the distribution maps of the texture classes, the distribution map of the actual data shows a loamy texture, with coarser characteristics,
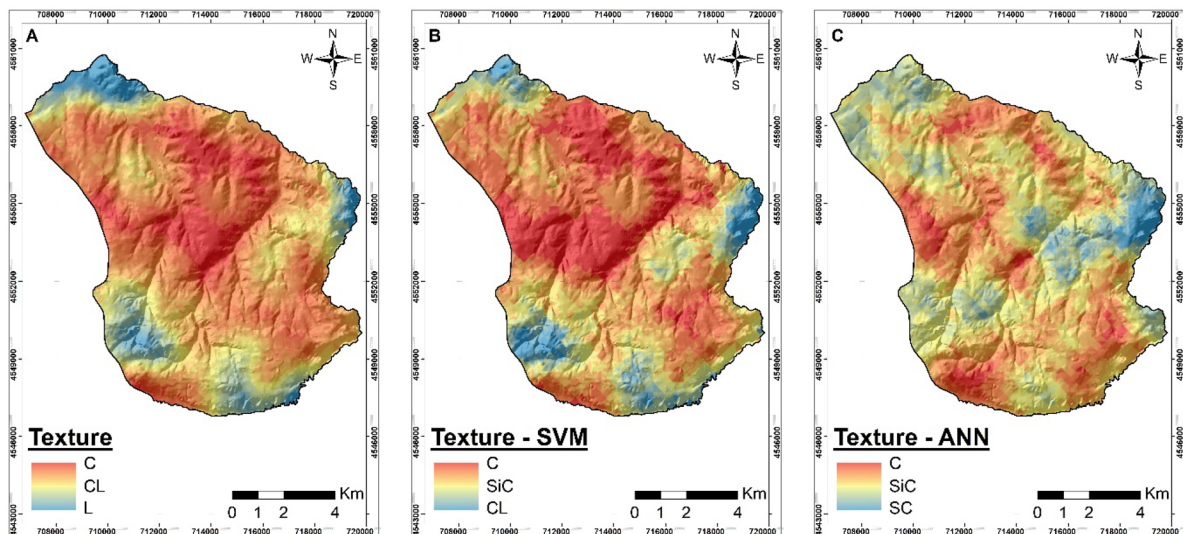

Fig. 8 Texture classes maps (C: Clay, CL: Clay loam, L: Loam, SiC: Silty clay, SC: Sandy clay) maps (**A**: spatial distribution of the observed values, **B**: spatial distribution of the predicted values with SVM, **C**: spatial distribution of the predicted values with ANN)
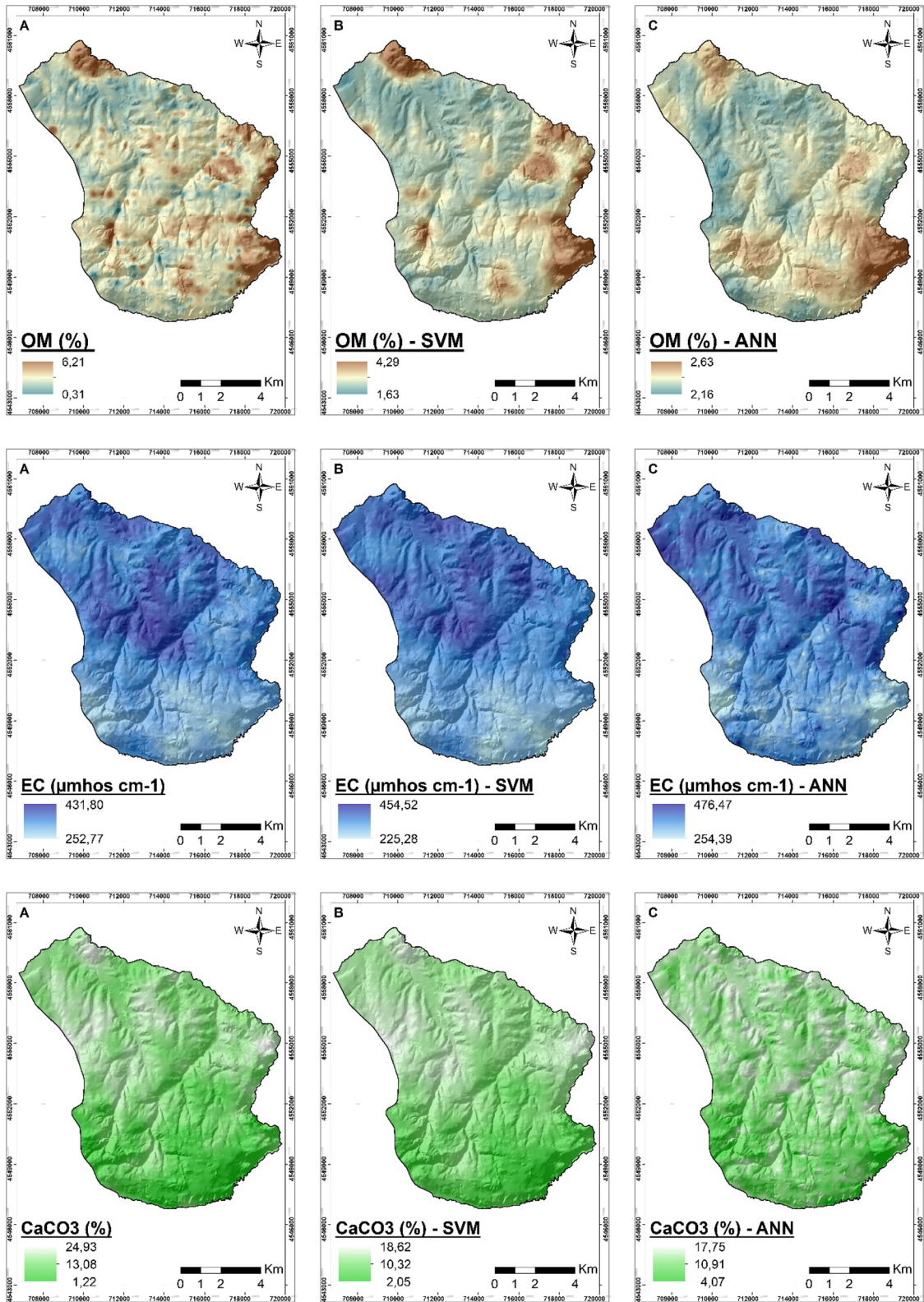
Fig. 10 Sand, silt and clay uncertainty maps (**D**: spatial distribution of the SVM model, **E**: spatial distribution of the ANN model)
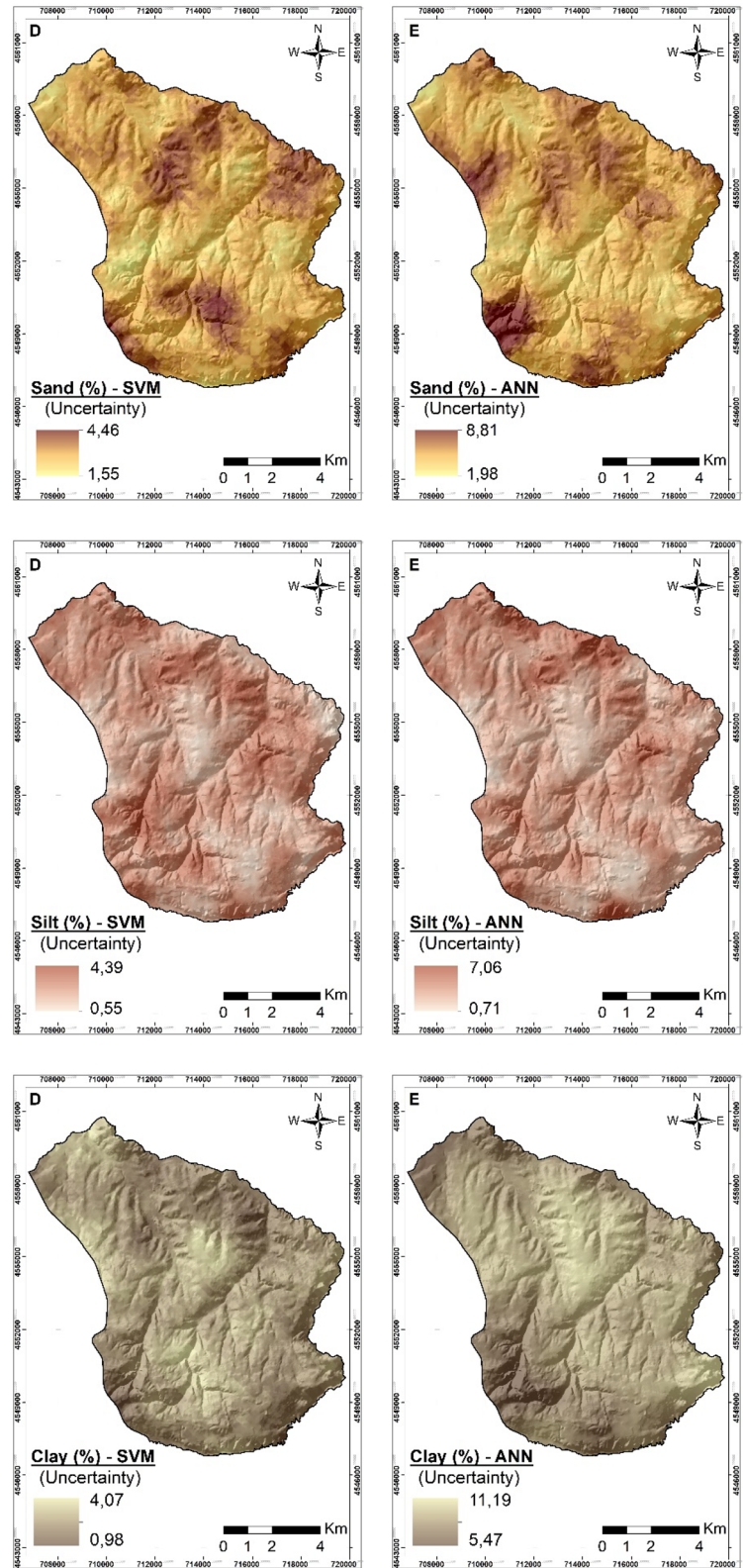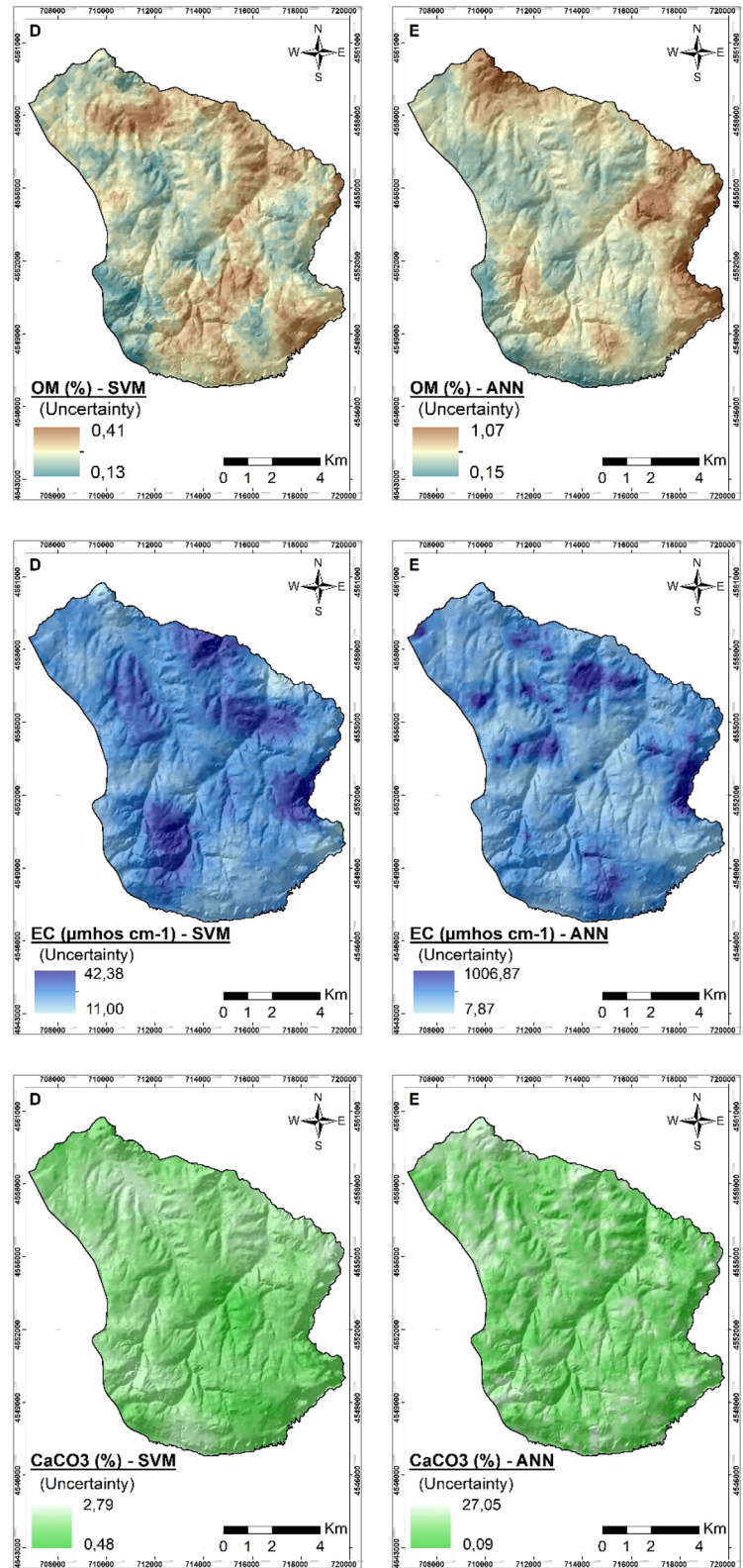
Fig. 11 Organic matter, EC and CaCO₃ CCE uncertainty maps (**D**: spatial distribution of the SVM model, **E**: spatial distribution of the ANN model)

particularly in the northern part. Moreover, most of the area has a clay and clay loam structure. The map produced by SVM is largely similar to the distribution map produced with actual data. The distribution map obtained by SVM shows that clay texture dominates the central and north-western parts, but towards the south, the sand fractions increase, turning into loamy and clay loam (Fig. 8). Comparing the map produced by the estimations of ANN and the distribution map of the actual data, a low level of accuracy was obtained.

(Fig. 9) provides the actual values and spatial distribution maps by SMV and ANN for the organic matter, EC, and calcium carbonate equivalent (CCE) contents of the soils. For the OM distribution maps, the actual data and the distribution map obtained by SVM show a very similar pattern, although the estimation obtained by ANN is quite low in accuracy. Looking at the OM distribution maps, we see that OM is high due to forest cover in the east, southeast, and rather smaller the north areas. The case is similar for EC and CCE, and the EC and CCE contents tend to increase towards the north.

In addition, the SVM accuracy was found to be much higher than the models used in the estimation of the texture class distribution of the soils, and the distribution maps created with SVM prediction values for the other examined properties in general showed closer results compared to the distribution maps created with ANN prediction values.

The uncertainty maps obtained from the predicted values are given in Figs. 10 and 11. The lowest standard deviations in uncertainty maps were obtained with the SVM model.

When the uncertainty maps of the sand, silt, clay, calcium carbonate equivalent and EC prediction values of the soils are examined, it is seen that the standard deviation is generally high for the features with a high variation range. However, the standard deviations were generally high for low-level estimates of organic matter. Organic matter contents are at lower levels compared to the other properties examined. This difference can be considered as an indication that the model makes predictions with higher error rates in estimating low values.

Conclusion

In this research, we correlated the reflectance values of Triplesat Satellite image bands with soil texture (sand, silt, and clay), organic matter, electrical conductivity, and calcium carbonate equivalent contents in soils and determined their estimation accuracies using machine learning algorithms like support vector machine and artificial neural network. Positive correlations were found between the band reflectance values and the electrical conductivity, and calcium carbonate equivalent contents of the soils, suggesting that band data can be used to evaluate areas with high electrical conductivity, and calcium carbonate equivalent content. For estimating basic soil properties with band data, we obtained more successful estimations with the support vector machine algorithm. Also, we observed the closest pattern to the spatial distribution of the actual values in the estimations obtained by support vector machine. It has been determined that near infrared and red bands are effective in the estimation of basic soil properties. The present study findings show that support vector machine was successful in estimating the correlations between band values and soil properties, indicating that it can be used safely for other studies. The maps obtained in this way can be used on both national and global scales and make significant contributions in revealing the overall properties of soil. The soil maps obtained by support vector machine and band values can later be used as baseline information for preventing or mitigating significant negative impacts on agricultural productivity.

Declarations

Conflict of interest   The author declares no conflict of interest.

References

Akinyemi, F. O., Pontius, R. G., Jr., & Braimoh, A. K. (2017). Land change dynamics: Insights from Intensity Analysis

applied to an African emerging city. *Journal of Spatial Science, 62*(1), 69–83. https://doi.org/10.1080/14498596.2016.1196624

Aksoy, H., & Kaptan, S. (2021). Monitoring of land use/land cover changes using GIS and CA-Markov modeling techniques: A study in Northern Turkey. *Environmental Monitoring and Assessment, 193*(8), 507. https://doi.org/10.1007/s10661-021-09281-x

Alaboz, P., Başkan, O., & Dengiz, O. (2021). Computational intelligence applied to the least limiting water range to estimate soil water content using GIS and geostatistical approaches in alluvial lands. *Irrigation and Drainage, 70*(5), 1129–1144. https://doi.org/10.1002/ird.2628

Alaboz, P., & Işıldar, A. A. (2019). Evaluation of pedotransfer functions (PTFs) for some soil physical properties. *Turkish Journal of Science and Engineering, 1*(1), 28–34.

Aydın, G., L. Atatanır, A., Yorulmaz, Y., Kurucu, H. S. Öztürk, K., Kızılkaya, K., & Kaptan, M. A. (2011). Estimation of Soil Parameters by Nirs (Near Infrared) Spectroscopy. *Soil and Water Symposium* 25–27.

Ballabio, C. (2009). Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma, 151*, 338–350. https://doi.org/10.1016/j.geoderma.2009.04.022

Bangelesa, F., Adam, E., Knight, J., Dhau, I., Ramudzuli, M., & Mokotjomela, T. M. (2020). Predicting soil organic carbon content using hyperspectral remote sensing in a degraded mountain landscape in lesotho. *Applied and Environmental Soil Science, 2020*, 1–11. https://doi.org/10.1155/2020/2158573

Baumgardner, M. F., Kristof, S., Johannsen, C. J., & Zachary, A. (1969). Effects of organic matter on the multispectral properties of soils. In *Proceedings of the Indiana Academy of science* (Vol. 79, pp. 413–422).

Bölük, E. (2016). According to Erinç Climate Classification Turkish Climate, Ministry of Forestry and Water Management General Directorate of Meteorology, Ankara.

Bouyoucos, G. J. (1951). A Recalibration of the Hydrometer Method for Making Mechanical Analysis of Soils. *Agronomy Journal, 43*(9), 434–438. Portico. https://doi.org/10.2134/agronj1951.00021962004300090005x

Çakır, F. S. (2019). Artificial neural networks. Nobel publications 2nd Edition. Ankara.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

de Castro Padilha, M. C., Vicente, L. E., Demattê, J. A., Loebmann, D. G. D. S. W., Vicente, A. K., Salazar, D. F., & Guimarães, C. C. B. (2020). Using Landsat and soil clay content to map soil organic carbon of oxisols and Ultisols near São Paulo, Brazil. *Geoderma Regional, 21*, e00253. https://doi.org/10.1016/j.geodrs.2020.e00253

Deiss, L., Margenot, A. J., Culman, S. W., & Demyan, M. S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma, 365*, 114227. https://doi.org/10.1016/j.geoderma.2020.114227

Demattê, J. A. M., Guimarães, C. C. B., Fongaro, C. T., Vidoy, E. L. F., Sayão, V. M., Dotto, A. C., & Santos, N. V. D. (2018). Satellite Spectral Data on the Quantification of Soil Particle Size from Different Geographic

Regions. *Revista Brasileira de Ciência do Solo, 42*. https://doi.org/10.1590/18069657rbcs20170392

Dengiz, O., Erel, A., Erkoçak, A., & Durmuş, M. (2012). Basic soil properties, classification and mapping of Kuskonagi Basin. *Journal of Agriculture Faculty of Ege University, 49*(1), 71–82.

Dong, Z., Wang, N., Liu, J., Xie, J., & Han, J. (2021). Combination of machine learning and VIRS for predicting soil organic matter. *Journal of Soils and Sediments, 21*(7), 2578–2588. https://doi.org/10.1007/s11368-021-02977-0

Doran, J. W., & Jones, A. J. (1996). Methods for assessing soil quality, vol. 49. SSSA special publication. Madison, WI: ASA.

FAO. (2022). Soil Organic Carbon Mapping Cookbook. Available online: https://www.fao.org/documents/card/en/c/I8895EN/ (accessed on 7 june 2023).

Fathololoumi, S., Vaezi, A. R., Alavipanah, S. K., Ghorbani, A., Saurette, D., & Biswas, A. (2020). Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. *Science of the Total Environment, 721*, 137703. https://doi.org/10.1016/j.scitotenv.2020.137703

Feng, Y., Lei, Z., Tong, X., Gao, C., Chen, S., Wang, J., & Wang, S. (2020). Spatially-explicit modeling and intensity analysis of China's land use change 2000–2050. Journal of Environmental Management, 263, 110407. https://doi.org/10.1016/j.jenvman.2020.110407

Gharbia, S. S., Alfatah, S. A., Gill, L., Johnston, P., & Pilla, F. (2016). Land use scenarios and projections simulation using an integrated GIS cellular automata algorithms. *Modeling Earth Systems and Environment, 2*, 1–20. https://doi.org/10.1007/s40808-016-0210-y

Gotway, C. A., Ferguson, R. B., Hergert, G. W., & Peterson, T. A. (1996). Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal, 60*(4), 1237–1247. https://doi.org/10.2136/sssaj1996.03615995006000040040x

Gozdowski, D., Stępień, M., Samborski, S., Dobers, E. S., Szatyłowicz, J., & Chormański, J. (2015). Prediction accuracy of selected spatial interpolation methods for soil texture at farm field scale. *Journal of Soil Science and Plant Nutrition, 15*(3), 639–650. https://doi.org/10.4067/s0718-95162015005000033

Grunwald, S., Thompson, J. A., & Boettinger, J. L. (2011). Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Science Society of America Journal, 75*(4), 1201–1213. https://doi.org/10.2136/sssaj2011.0025

Hazelton, P., & Murphy, B. (2016). *Interpreting soil test results: What do all the numbers mean*. CSIRO Publishing.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma, 265*, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014

Jackson, M. L. (1958). *Soil Chemical Analysis*. Prentice Hall Inc.

Jiang, J., Wen, Z., Zhao, M., Bie, Y., Li, C., Tan, M., & Zhang, C. (2019). Series arc detection and complex load recognition based on principal component analysis and support vector machine. *IEEE Access, 7*, 47221–47229.

Karavanova, E.I, Shrestha, D.P., & Orlov, D.S. (2001). Application of remote sensing techniques for the study of soil

salinity in semi-arid Uzbekistan. In Response to Land Degradation, Bridges EM, Hannam ID, Oldeman LR, de Vries FWTP, Scherr SJ, Sombatpanit S (eds). Oxford and IBHPublishing Co. Pvt. Ltd.: New Delhi; 261–273.

Kaya, F., Schillaci, C., Keshavarzi, A., & Başayiğit, L. (2022). Predictive Mapping of Electrical Conductivity and Assessment of Soil Salinity in a Western Türkiye Alluvial Plain. *Land, 11*(12), 2148. https://doi.org/10.3390/land11122148

Kesgin, B., & Nurlu, E. (2009). Land cover changes on the coastal zone of Candarli Bay, Turkey using remotely sensed data. *Environmental Monitoring and Assessment, 157*, 89–96. https://doi.org/10.1007/s10661-008-0517-x

Keshavarzi, A., del Árbol, M. Á. S., Kaya, F., Gyasi-Agyei, Y., & Rodrigo-Comino, J. (2022). Digital mapping of soil texture classes for efficient land management in the Piedmont plain of Iran. *Soil Use and Management, 38*(4), 1705–1735. https://doi.org/10.1111/sum.12833

Khalil, R. Z., W. Khalid, & M. Akram. (2016). Estimating of soil texture using landsat imagery: A case study of Thatta Tehsil, Sindh. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3110–3113). IEEE.

Kuzu, B. S., & Yakut, S. G. (2020). Examination of Financial Failure Estimates According to Technological Density with the Help of Support Vector Machines. *Osmaniye Korkut Ata University Journal of Economics and Administrative Sciences, 4*(2), 36–54.

Lamorski, K., Pachepsky, Y., Sławiński, C., & Walczak, R. T. (2008). Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. *Soil Science Society of America Journal, 72*(5), 1243–1247. https://doi.org/10.2136/sssaj2007.0280n

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods* (p. 40). Butterworths Publishing.

Lu, D., Moran, E., Hetrick, S., & Li, G. (2011). Land-Use and Land-Cover Change Detection. (2011). *Advances in Environmental Remote Sensing,* 273–288. https://doi.org/10.1201/b10599-18

Malone, B. P., McBratney, A. B., & Minasny, B. (2011). Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma, 160*(3–4), 614–626. https://doi.org/10.1016/j.geoderma.2010.11.013

Maselli, F., Gardin, L., & Bottai, L. (2008). Automatic mapping of soil texture through the integration of ground, satellite and ancillary data. *International Journal of Remote Sensing, 29*(19), 5555–5569. https://doi.org/10.1080/01431160802029651

Miloš, B., & Bensa, A. (2017). Prediction of soil organic carbon using VIS-NIR spectroscopy: Application to Red Mediterranean soils from Croatia. *Eurasian Journal of Soil Science, 6*(4), 365–373. https://doi.org/10.18393/ejss.319208

Mitran, T., Solanky, V., Janakirama Suresh, G., Sujatha, G., Sreenivas, K., & Ravisankar, T. (2019). Predictive mapping of surface soil texture in a semiarid region of India through geostatistical modeling. *Modeling Earth Systems and Environment, 5*, 645–657. https://doi.org/10.1007/s40808-018-0556-4

Moradi, F., Kaboli, H. S., & Lashkarara, B. (2020). Projection of future land use/cover change in the Izeh-Pyon Plain of Iran using CA-Markov model. *Arabian Journal of Geosciences, 13*, 1–17. https://doi.org/10.1007/s12517-020-05984-6

Mousavi, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement, 201*, 111706. https://doi.org/10.1016/j.measurement.2022.111706

Mulder, V. L., De Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma, 162*(1–2), 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Özden, N., & Altınbaş, Ü. (2005). Research on Determining of Reflection Characters of Different Soil Taxonomic Units by Utilizing Remote Sensing Technique. *Journal of Agriculture Faculty of Ege University, 42*(2), 143–153.

Öztemel, E. (2012). *Artificial Neural Networks* (3rd ed.). Papatya publishing.

Panneerselvam, B., Muniraj, K., Pande, C., & Ravichandran, N. (2023). Prediction and evaluation of groundwater characteristics using the radial basic model in Semi-arid region, India. *International Journal of Environmental Analytical Chemistry, 103*(6), 1377–1393. https://doi.org/10.1080/03067319.2021.1873316

Pasolli, L., Notarnicola, C., & Bruzzone, L. (2011). Estimating soil moisture with the support vector regression technique. *IEEE Geoscience and Remote Sensing Letters, 8*(6), 1080–1084. https://doi.org/10.1109/LGRS.2011.2156759

Rezaei, M., Mousavi, S. R., Rahmani, A., Zeraatpisheh, M., Rahmati, M., Pakparvar, M., ... & Cornelis, W. (2023). Incorporating machine learning models and remote sensing to assess the spatial distribution of saturated hydraulic conductivity in a light-textured soil. *Computers and Electronics in Agriculture,* 209, 107821. https://doi.org/10.1016/j.compag.2023.107821

Richards, L. A. (1954). Diagnosis and improvement of saline and alkali soils. *Agriculture Handbook No, 60*, 105–106.

Rossel, R. V., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., ... & Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Rossel, R. V., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma, 158*(1–2), 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Seyedmohammadi, J., Navidi, M. N., & Esmaeelnejad, L. (2019). Geospatial modeling of surface soil texture of agricultural land using fuzzy logic, geostatistics and GIS techniques. *Communications in Soil Science and Plant Analysis, 50*(12), 1452–1464. https://doi.org/10.1080/00103624.2019.1626870

Shahriari, M., Delbari, M., Afrasiab, P., & Pahlavan-Rad, M. R. (2019). Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: A case of southeastern Iran. *Catena*, 182, 104149. https://doi.org/10.1016/j.catena.2019.104149

Sharififar, A. (2022). Accuracy and uncertainty of geostatistical models versus machine learning for digital mapping of soil calcium and potassium. *Environmental Monitoring and Assessment,* 194(10). https://doi.org/10.1007/s10661-022-10434-9

Shrestha, R. P. (2006). Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand. *Land Degradation & Development, 17*(6), 677–689. https://doi.org/10.1002/ldr.752

Sihag, P., Tiwari, N. K., & Ranjan, S. (2018). Support vector regression-based modeling of cumulative infiltration of sandy soil. *ISH Journal of Hydraulic Engineering*, 1–7. https://doi.org/10.1080/09715010.2018.1439776

Silva, S. H. G., Weindorf, D. C., Pinto, L. C., Faria, W. M., Acerbi Junior, F. W., Gomide, L. R., de Mello, J. M., de Pádua Junior, A. L., de Souza, I. A., Teixeira, A. F. dos S., Guilherme, L. R. G., & Curi, N. (2020). Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. *Geoderma, 362*, 114136. https://doi.org/10.1016/j.geoderma.2019.114136

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222. https://doi.org/10.1023/b:stco.0000035301.49549.88

Soil Survey Staff. (1993). Soil Survey Manual, USDA. Handbook No: 18 Washington D.C.

Sørensen, L. K., & Dalsgaard, S. (2005). Determination of Clay and Other Soil Properties by Near Infrared Spectroscopy. *Soil Science Society of America Journal, 69*(1), 159. https://doi.org/10.2136/sssaj2005.0159

Stamatiadis, S., Christofides, C., Tsadilas, C., Samaras, V., Schepers, J. S., & Francis, D. (2005). Ground-Sensor Soil Reflectance as Related to Soil Properties and Crop Response in a Cotton Field. *Precision Agriculture, 6*(4), 399–411. https://doi.org/10.1007/s11119-005-2326-3

Sünbül, V., & Tonyaloğlu, E. E. (2021). Determination of Land Use / Land Cover Change in the Case of Kaş District of Antalya. *International Journal of Eastern Anatolia Science Engineering and Design, 3*(2), 376–387.

Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., & Scholten, T. (2020). Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. *Remote Sensing, 12*(7), 1095. https://doi.org/10.3390/rs12071095

Turan, M., Dengiz, O., & Turan, İD. (2018). Determination of soil moisture and temperature regimes for Samsun province according to Newhall model. *Turkish Journal of Agricultural Research, 5*(2), 131–142.

Van Wambeke, A. R. (2000). *The Newhall Simulation Model for estimating soil moisture and temperature regimes*. Cornell University Ithaca, NY.

Wang, Z., Du, Z., Li, X., Bao, Z., Zhao, N., & Yue, T. (2021). Incorporation of high accuracy surface modeling into machine learning to improve soil organic matter mapping. *Ecological Indicators*, 129, 107975. https://doi.org/10.1016/j.ecolind.2021.107975

Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators, 52*, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028

Wu, W., Li, A.-D., He, X.-H., Ma, R., Liu, H.-B., & Lv, J.-K. (2018). A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture, 144*, 86–93. https://doi.org/10.1016/j.compag.2017.11.037

Xing, L., Li, L., Gong, J., Ren, C., Liu, J., & Chen, H. (2018). Daily soil temperatures predictions for various climates in United States using data-driven model. *Energy, 160*, 430–440. https://doi.org/10.1016/j.energy.2018.07.004

Yamaç, S. S., Şeker, C., & Negiş, H. (2020). Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. *Agricultural Water Management, 234*, 106121. https://doi.org/10.1016/j.agwat.2020.106121

Yin, Z., Lei, T., Yan, Q., Chen, Z., & Dong, Y. (2013). A near-infrared reflectance sensor for soil surface moisture measurement. *Computers and Electronics in Agriculture, 99*, 101–107. https://doi.org/10.1016/j.compag.2013.08.029

Yue, J., Tian, J., Tian, Q., Xu, K., & Xu, N. (2019). Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS Journal of Photogrammetry and Remote Sensing, 154*, 216–230. https://doi.org/10.1016/j.isprsjprs.2019.06.012

Zhang, X.-Y., & SUI, Y.-Y., Zhang, X.-D., Meng, K., & Herbert, S. J. (2007). Spatial Variability of Nutrient Properties in Black Soil of Northeast China. *Pedosphere, 17*(1), 19–29. https://doi.org/10.1016/s1002-0160(07)60003-4

Zhao, Z., Chow, T. L., Rees, H. W., Yang, Q., Xing, Z., & Meng, F.-R. (2009). Predict soil texture distributions using an artificial neural network model. *Computers and Electronics in Agriculture, 65*(1), 36–48. https://doi.org/10.1016/j.compag.2008.07.008