

Estimation of Soil Organic Carbon Using Artificial Neural Networks and Regression Models

Ergün KAHVECİ¹ and Salih MALKOÇOĞLU^{2*}

¹Tokat Gaziosmanpaşa University, Niksar Vocational School, Department of Forestry, Tokat, TÜRKİYE

²Muğla Sıtkı Koçman University, Köyceğiz Vocational School, Department of Forestry, Muğla, TÜRKİYE

*Corresponding Author: salihm@mu.edu.tr

Received Date: 26.06.2025

Accepted Date: 08.01.2026

Abstract

Aim of study: This study aimed to estimate the soil organic carbon (SOC) content in red pine (*Pinus brutia* Ten.) plantation areas under semiarid climatic conditions using multiple linear regression (MLR) and artificial neural network (ANN) models.

Area of study: The research was carried out within the boundaries of the Tokat-Niksar Ayvaz Forest Management Chiefdom, Türkiye.

Material and method: The study material consisted of 92 depth-level sampling units (46 plots × two depths: 0-10 cm and 10-30 cm), each represented by paired disturbed and undisturbed topsoil samples, along with vegetation and topographic indices. Various physical and chemical analyses were performed on the soil samples. In addition to soil properties, topographic and remote sensing indices were used as predictor variables. SOC estimations were carried out using MLR and ANN models.

Main results: According to the findings, SOC content increased with higher wilting point (WP), elevation, and normalized difference vegetation index (NDVI), while it decreased with increasing lime content (CaCO₃), pH, and bulk density (BD). In the MLR model, WP, BD, pH, CaCO₃, moisture stress index (MSI), easterly exposure (EExp.), and valley depth (VDep.) were identified as significant variables, resulting in an adjusted R² value of 0.841. The ANN model showed higher predictive performance with the same variables, yielding an overall R² of 0.937 across all samples and a low relative error (0.043) on the test set.

Research highlights: Both models were found suitable for SOC estimation; however, the ANN model represented complex and nonlinear environmental interactions more effectively.

Keywords: Soil Organic Carbon, ANN, MLR, Semiarid Ecosystem, Remote Sensing

Yapay Sinir Ağı ve Regresyon Modelleri Yardımıyla Toprak Organik Karbonunun Tahmini

Öz

Çalışmanın amacı: Bu çalışma, yarı kurak iklim koşullarına sahip kızılçam (*Pinus brutia* Ten.) ağaçlandırma alanlarında toprak organik karbon (TOK) içeriğini, çoklu doğrusal regresyon (ÇDR) ve yapay sinir ağı (YSA) modelleri kullanarak tahmin etmeyi amaçlamıştır.

Çalışma alanı: Araştırma, Tokat-Niksar Ayvaz Orman İşletme Şefliği sınırları içerisinde yürütülmüştür.

Materyal ve yöntem: Çalışma materyalini, 46 örnek alanında iki derinlik (0–10 cm ve 10–30 cm) düzeyinde oluşturulan toplam 92 örnekleme biriminden alınan eşleştirilmiş bozulmuş ve bozulmamış toprak örnekleri ile vejetasyon ve topoğrafik indisler oluşturmuştur. Toprak örnekleri üzerinde çeşitli fiziksel ve kimyasal analizler yapılmıştır. Toprak özelliklerine ek olarak, topoğrafik ve uzaktan algılama indisleri de tahmin değişkenleri olarak kullanılmıştır. TOK tahminleri ÇDR ve YSA modelleriyle gerçekleştirilmiştir.

Temel sonuçlar: Bulgulara göre, TOK içeriği solma noktası (SN), yükselti ve normalize edilmiş fark vejetasyon indeksi (NDVI) arttıkça artmış; buna karşılık kireç içeriği (CaCO₃), pH ve hacim ağırlığı (HA) arttıkça azalmıştır. ÇDR modelinde SN, HA, pH, CaCO₃, nem stresi indeksi (MSI), doğuya bakı (EExp.) ve vadi derinliği (VDep.) anlamlı değişkenler olarak belirlenmiş ve düzeltilmiş R² = 0.841 elde edilmiştir. YSA modeli ise aynı değişkenlerle daha yüksek bir tahmin performansı göstererek tüm örneklerde R² = 0.937 ve test verisinde düşük göreceli hata (0.043) değeriyle öne çıkmıştır.

Araştırma vurguları: Her iki model de TOK tahmini için uygun bulunmuş; ancak YSA modeli, karmaşık ve doğrusal olmayan çevresel ilişkileri daha etkili biçimde temsil etmiştir.

Anahtar kelimeler: Toprak Organik Karbonu, YSA, ÇDR, Yarı Kurak Ekosistem, Uzaktan Algılama



Introduction

Soils are essential to Earth because they control biological, erosional, hydrological, and geochemical cycles. These cycles are crucial for human food production and for maintaining clean air and water supplies (Keesstra et al., 2016). SOC levels are governed by a complex interplay between climatic drivers, such as thermal regimes and moisture availability, and intrinsic edaphic attributes, including parent material, clay fraction, and cation exchange capacity (Dawson and Smith, 2007). As a pivotal ecological compartment, soil functions as both a source and a sink of organic carbon, harboring two to threefold more SOC than the atmospheric reservoir (Lal, 2004; Scharlemann et al., 2014). Globally, soils sequester an estimated 1500 petagrams (Pg) of organic carbon within the uppermost 1 meter of the soil profile (Batjes, 1996). Consequently, mobilizing even a minor portion of the soil organic carbon pool into greenhouse gases can substantially exacerbate anthropogenic global warming (Melillo et al., 2002; Stockmann et al., 2013; Schuur et al., 2015). Thorough evaluation of SOC spatial heterogeneity, stock magnitudes, and environmental drivers can enhance the parametrization of carbon cycle models aimed at quantifying soil carbon feedbacks under global environmental change scenarios (Yang et al., 2008; Liu et al., 2011; Scharlemann et al., 2014; Li et al., 2018; Zhu et al., 2019).

Different approaches have been used to assess the effects of global warming and climate change on SOC stocks. Models stand out as practical tools for examining carbon stocks and dynamics (Falloon et al., 2002; Falloon and Smith, 2003; Jones et al., 2005; Paustian et al., 2018), making them suitable for C reporting and assessment studies (Muñoz-Rojas et al., 2013). They are essential as decision-support systems, especially in addressing climate change issues (Smith et al., 2005). Modeling allows us to estimate short-, medium-, and long-term trends in SOC dynamics and SOC sequestration under projected scenarios of climate change (Smith et al., 2005; Lucht et al., 2006; Wan et al., 2011), which is of critical importance for taking effective management measures in different land uses. Combining simulation

models with spatial datasets (soils, land use) helps estimate current and future regional SOC stocks and SOC sequestration (Falloon et al., 1998; Batjes, 2006; Hashimoto et al., 2012). In addition, patterns in SOC dynamics related to soil and land use can be analyzed (Muñoz-Rojas et al., 2013). The spatial and temporal heterogeneity of SOC is shaped by numerous biophysical and anthropogenic factors, including land-use and cover variability, climatic regimes, edaphic classifications, topographical gradients, and soil management strategies (Mircholi et al., 2020). Numerous investigations have examined the spatial distribution of SOC in relation to diverse controlling variables, including shifts in land use patterns (Dorji et al., 2014; Stumpf et al., 2018), vegetation and terrain characteristics (Oueslati et al., 2013; Tsui et al., 2013), pedogenic classifications (Wang et al., 2003), climatic conditions (Luo et al., 2017), geomorphic disturbances such as landslides (Restrepo and LaQuay, 2018), and the duration of cultivation (Dalal and Mayer, 1986). Therefore, knowledge of the spatial and temporal variations of SOC is essential for policymakers to make informed decisions on agri-environmental climate change and to evaluate the short- and long-term carbon consequences of alternative management strategies (Grinand et al., 2017; Jebari et al., 2018; Mircholi et al., 2020). Because conventional laboratory analyses are costly, approaches that reduce sampling requirements while still supporting SOC mapping, soil-terrain modeling, and large-scale carbon monitoring are valuable for land-management decisions (McKenzie and Ryan, 1999; Post et al., 2001; Wetterlind et al., 2008; Falahatkar et al., 2016). A range of statistical methods has therefore been applied to predict SOC at the watershed scale, including linear and nonlinear regression, logistic regression, ANNs, genetic algorithms, and decision trees (Lamsal, 2009; Kheir et al., 2010; Ayoubi et al., 2011).

Recently, different statistical methods, including linear and nonlinear regression, logistic regression, ANNs, genetic algorithms, and decision trees, have been used by various scientists to estimate SOC at the watershed scale (Lamsal, 2009; Kheir et al., 2010; Ayoubi et al., 2011). Neural networks offer a

method for modeling complex or nonlinear relationships among data (Ingleby and Crowe, 2001). Nowadays, the development of digital soil mapping technologies (McBratney et al., 2003) and various statistical methods used to model SOC stocks (Minasny et al., 2013) have significantly improved the accuracy of estimates (De Anta et al., 2020).

It is crucial to determine the spatial distribution of soil organic carbon content and the key covariates that govern it, particularly in arid and semiarid climates (Garosi et al., 2022). However, the comparative performance of MLR and ANN models for predicting SOC content specifically in semi-arid red pine (*Pinus brutia*) plantation forests remains unexplored. Future studies on SOC pools should be conducted in a comparable manner, and access to datasets should be facilitated (Bahn et al., 2010; Muñoz-Rojas et al., 2012).

Therefore, to our knowledge, this study provides the first dedicated evaluation of these models for red pine plantation ecosystems in the Central Black Sea region, addressing a broader gap in the comparative assessment of MLR and ANN for SOC estimation in widespread yet understudied ecosystems. The objectives of this study are as follows: (1) to estimate SOC in a red pine plantation with the ANN model using environmental variables and remote sensing indices, (2) to determine SOC by the MLR method, (3) to determine the most critical factors for spatially estimating SOC in the semiarid region of the Central Black Sea based on the ANN model results.

Material and Methods

Study Area

The study area is located within the borders of the Niksar district in the Central Black Sea Region of Türkiye. This study was conducted in a 360-hectare red pine plantation forest. Figure 1 shows the location of the study area and the distribution of the soil sampling plots. The map was created by the authors using ArcGIS 10.8, with base data derived from the ALOS PALSAR DEM (ASF DAAC, 2014) and national administrative boundaries. The plantation work was carried out between 1986 and 1990. The altitudes of the experimental areas where the samples were taken range from 343 to 630 m, and the

aspects are both southerly and northerly. The slope values range between 1% and 33%.

The study area is located within the Central Black Sea Region, specifically within the Central Black Sea Section Mountainous Land Plantation Regions Group, as defined in the regional forest site classification of Türkiye (Kantarıcı, 2005a). However, this area lies within the Kelkit basin and is critical due to its unique vegetation and climate. Due to its topography, although it is located in the Central Black Sea Region, it exhibits ecological characteristics typical of the Mediterranean region, particularly in climate and vegetation.

Niksar district is classified as semi-arid in terms of climate characteristics. The average temperature in the study area is 14.9 °C, and the total annual precipitation is 496.6 mm (Table 1). Red pine (*Pinus brutia* Ten) and Oak (*Quercus* spp.) species are also distributed in the study area.

Table 1. Niksar meteorological station climate data (MGM, 2024)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Annual
Climate parameters	Measurement period (2013-2023)												
Temperature, °C	5.2	7.2	9.7	13.9	18.0	21.6	23.7	24.7	21.0	15.9	10.8	6.6	14.9
Max. temperature, °C	26.3	24.8	29.6	33.5	39.3	39.2	43.9	42.0	43.3	35.4	30.0	24.0	43.9
Min. temperature, °C	-13.3	-10.3	-5.2	-0.8	2.1	8.2	11.0	12.6	5.9	0.0	-3.6	-8.0	-13.3
Number of rainy days per month	11.55	9.09	13.45	11.00	16.00	13.27	6.36	6.18	7.55	7.45	8.45	11.55	121.90
Monthly total precipitation (mm)	60.47	34.05	53.96	41.68	64.65	69.57	17.43	20.45	22.29	29.17	38.63	44.31	496.66

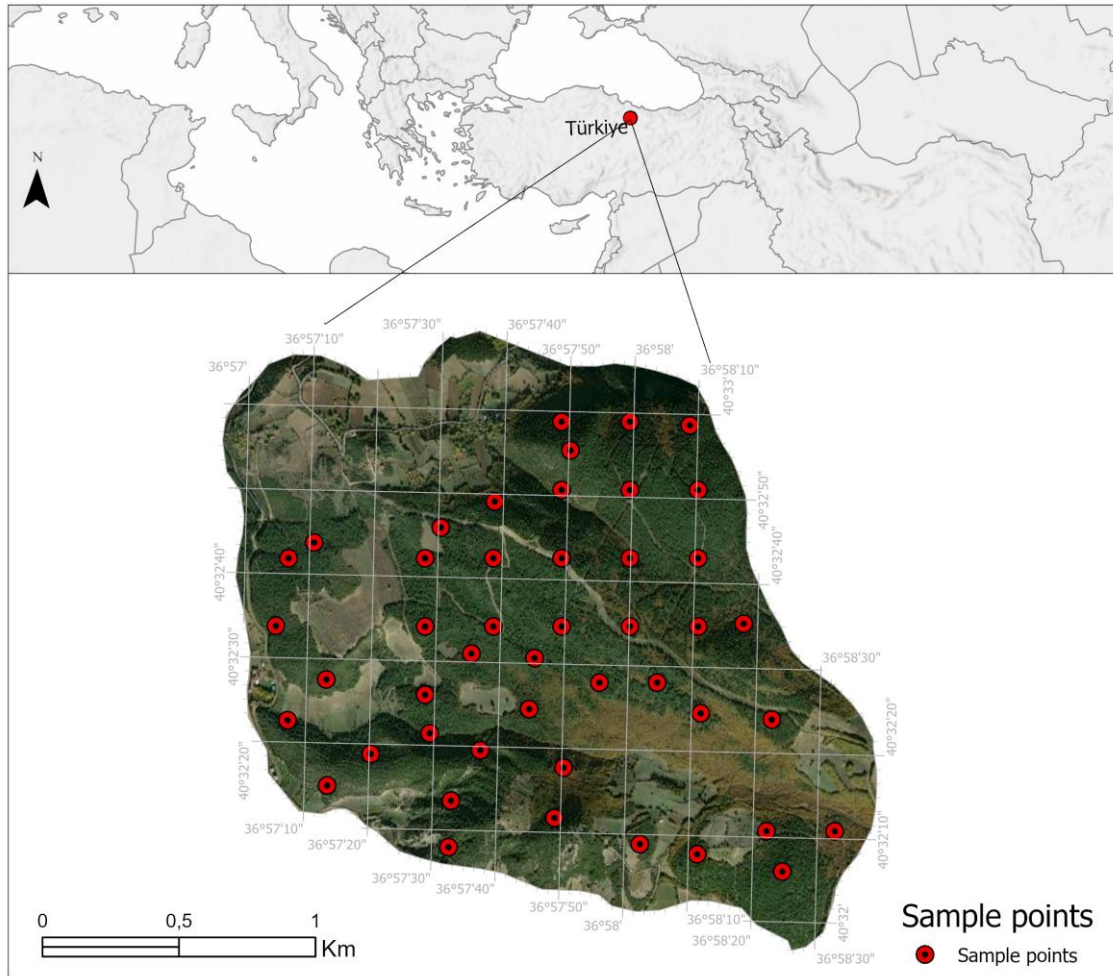


Figure 1. Location of the study area and distribution of the sample plots

Upper Jurassic–Lower Cretaceous grey-white limestones with thin-bedded marl and claystone intercalations (Doğdu Formation) are common between Niksar and Erbaa, where the research area is located (Aktimur et al., 1992). The research area consists of loamy clay and heavy clay- textured soils.

Soil Sampling and Data Analysis

This study was conducted in April 2018, and samples were taken from the topsoil at two depth levels (0-10 cm and 10-30 cm) at each of 46 sampling plots, yielding 92 depth-level sampling units. Soil samples were taken separately as bags (approximately 1 kg) and volume samples (Kantarıcı, 2005b). At each of these 92 units, both a disturbed (bag) and an

undisturbed (cylinder) soil sample were collected.

The soil dataset was considered a baseline for conditions at the time of sampling, given the inherently slow dynamics of the carbon cycle in forest ecosystems. In such systems, the turnover rates of SOC are intrinsically slow and exhibit high temporal stability. In particular, the bulk SOC stock in the 0-30 cm soil layer constitutes a carbon pool with residence times ranging from decades to centuries (Schmidt et al., 2011). Significant changes in SOC stocks generally occur only in response to substantial disturbances, such as fire (Certini, 2005) or land-use change, or under the influence of long-term climatic and management trends (Lal, 2004). In the study area, no large-scale fires or land degradation events have been recorded since 2018, as confirmed by local Forest Management Directorate reports and satellite imagery.

The absence of sensitivity analyses on factors such as changes in litter inputs is a limitation. Because SOC content in the 0–30 cm soil layer was measured only once, the effects of annual changes in litter inputs on SOC content and its spatial patterns could not be assessed. However, because bulk SOC in this layer is dominated by stable fractions, such annual variations are unlikely to alter the primary spatial patterns identified (Lützow et al., 2006).

The soil samples were dried to an air-dry state, ground, and then passed through a 2 mm sieve before being prepared for analysis. The mechanical analysis (texture) of the soils was determined according to Bouyoucos's hydrometer method (Bouyoucos, 1962), the BD was determined according to Blake and Hartge (1986) after soil samples were taken with a steel cylinder and dried in an oven at 105 °C, and the moisture constants (field capacity (FC), WP, available water capacity (AWC)) were determined according to the ceramic plate pressure device method of Soil Moisture Equipment Corp. (Gülçur, 1974; Karaöz, 1989). pH and EC were determined by the glass electrode method (Gülçur, 1974) in a 1:2.5 soil-water mixture, while the organic carbon amounts were determined by the Walkley-Black wet combustion method (Walkley, 1947; Gülçur, 1974). CaCO₃ contents of the soils were determined by the

Scheibler calcimeter method (Çağlar, 1949; Hızalan and Ünal, 1966).

Predictor Variables and Dataset Preparation

Topographic variables were derived from the ALOS PALSAR DEM (12.5 m resolution; ASF DAAC, 2014). Altitude, slope, aspect-based metrics (TRASP, Northness, Eastness), and VDep were calculated. The first group of indices was generated in ArcGIS 10.8 using the Geomorphometric and Gradient Metrics Toolbox (Evans et al., 2014), while VDep was calculated in SAGA GIS using the VDep module (Conrad et al., 2015; Böhner and Antonić, 2009). These variables are listed in Table 2 and mapped in Figure 2.

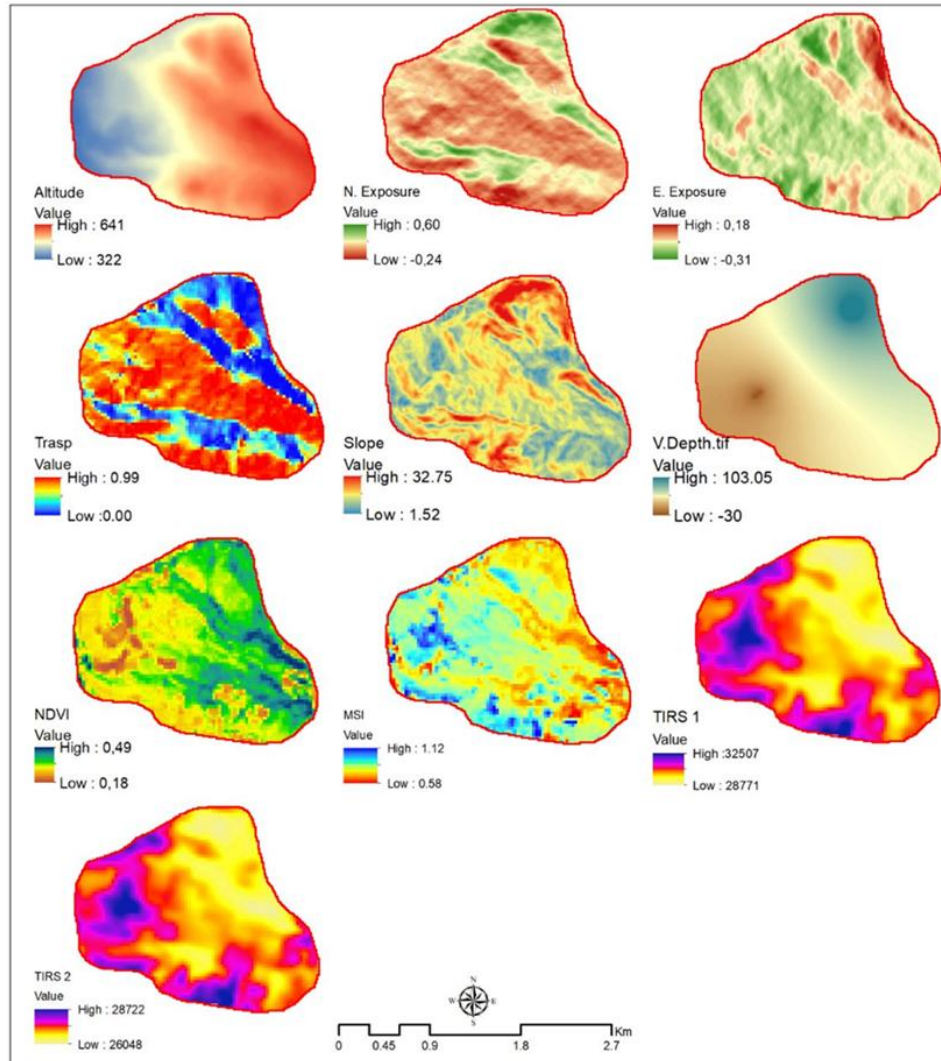


Figure 2. Spatial distribution of topographic and spectral variables used in the study
Map legends show values across the entire study area; the ranges in Table 2 refer to the 46 sampling plots.

Table 2. Topographic variables used in the study

Index	Description	Range	Reference
Altitude	Elevation above sea level (m) derived from the DEM.	Sampling plots: 342.61-629.86 m	Moore et al., 1993
TRASP	Transformed Aspect: linear transformation of aspect ranging from 0 (north-northeast, associated with cooler and moister conditions) to 1 (south-southwest, associated with hotter and drier conditions).	North-northeast (the coolest and wettest), south-southwest (the hottest and driest). From 0 to 1	Roberts and Cooper, 1989; Evans et al., 2014
NExp.	Northness (cosine of aspect) indicates exposure along a north-south axis, which affects temperature and moisture.	Relation to north (1 to -1)	Stage, 1976
EExp.	Eastness refers to the balance of solar radiation from morning to afternoon, which influences wind and moisture.	Relation to east (1 to -1)	Stage, 1976
Slope	Slope steepness in percent (%), computed from the DEM (planar gradient).	Sampling plots: 1.52-32.75%	Moore et al., 1993
VDep.	Vertical valley depth: modeled ridge level minus surface elevation (proxy for valley confinement/relative position).	Sampling plots: 2.98-103.05 m	Conrad et al., 2015; Böhner and AntoniĆ, 2009

Vegetation indices were derived from Landsat 8 OLI/TIRS imagery acquired in April 2018 under low cloud conditions. The dataset included NDVI, MSI, and two thermal raster variables (TIRS1 and TIRS2; Table 3). The thermal raster pixel values were used as

numerical predictors without conversion to physical temperature. Images were downloaded from USGS Earth Explorer and processed in ArcGIS using standard band math.

Table 3. Vegetation indices and thermal raster predictors, with their formulas and data sources

Index	Formula	Description	Reference
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	Normalized difference vegetation index; an indicator of green vegetation and photosynthetic activity; the most widely used vegetation index.	Rouse et al., 1974
MSI	SWIR1/NIR	Moisture Stress Index: higher MSI indicates greater vegetation water stress.	Rock et al., 1986
TIRS-1	-	Unconverted pixel values of the first thermal raster are used as a numerical predictor.	USGS, 2024
TIRS-2	-	Unconverted pixel values of the second thermal raster are used as a numerical predictor.	USGS, 2024

All raster layers (DEM 12.5 m; OLI/TIRS 30 m) were aligned to a standard projection and grid using the Snap Raster option. DEM data were resampled to the Landsat resolution. Finally, raster values were extracted to plot centroids using the Extract Multi Values to Points tool (Esri, 2020), generating plot-level predictors for modeling.

Linear Regression Analysis

SPSS 20.0 was used to analyze the data (IBM Corp., 2011). In this research, answers are sought to questions such as “Is there a relationship between dependent and independent variables?”, “If there is a relationship, what is the strength of the relationship?”, “Is it possible to make future predictions about the dependent variable?” and “If certain conditions are controlled, what effects does a specific variable or group of variables have on other variables?” (Alpar, 2013). Regression analysis using a single independent variable is referred to as simple (univariate) regression analysis, while analysis using more than one independent variable to predict a single dependent variable is referred to as multiple regression analysis (Tabachnick and Fidell, 1996; Köksal, 1977; Büyüköztürk, 2018). In multiple regression analysis, the effects of the independent variables on the dependent variable are estimated simultaneously (Ünver and Gamgam, 1999). The multiple regression

analysis model is formulated as shown in Equation 1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (1)$$

Y: Dependent variable,

β_0 : Intercept, $\beta_1, \beta_2, \dots$,

β_n : Coefficients of independent variables,

$X_1, X_2, \dots, X_n$: Independent variables (BD, Sand, Silt, Clay, FC, WP, AWC, pH, EC, CaCO_3 , Alt., Slope, TRASP, NExp., EExp., VDep., NDVI, MSI, TIRS1, and TIRS2)

ϵ : Error term (part that the model cannot predict).

After the linear relationships between SOC and variables were verified by correlation, MLR models were developed to estimate SOC as a function of selected variables (BD, FC, WP, AWC, pH, EC, CaCO_3 , Alt., EExp., TIRS1, VDep., NDVI, and MSI) that were significant at the 0.05 level. In this regression model for SOC values, the F-statistic and the coefficients were significant at the 95% confidence level ($p < 0.05$). Stepwise regression was used to determine the substantial and optimal variability in the model. Stepwise MLR has been used effectively to select informative predictors for estimating vegetation parameters (Kokaly and Clark, 1999). The presence of multicollinearity among variables was evaluated in detail by VIF and tolerance (T) values; $\text{VIF} < 10$ and $T > 0.1$ were considered acceptable (Garson, 2012).

Testing Regression Assumptions

To ensure the robustness of the regression analyses, key assumptions of the multiple linear regression model were carefully examined and verified. The linearity of relationships between the predictors and SOC was evaluated using scatterplots, while residual normality was assessed through Q-Q plots. Homoscedasticity was checked by plotting standardized residuals against predicted values, and error independence was verified using the Durbin-Watson test (Garson, 2012). Potential multicollinearity was assessed using the Variance Inflation Factor (VIF) and tolerance statistics, with thresholds of $VIF < 10$ and $tolerance > 0.1$ indicating acceptable multicollinearity (Hair et al., 2013).

ANN Analysis

ANN is an information processing technology inspired by biological nervous systems (Haykin, 2009). ANNs also produce interconnected nerve cells, similar to those found in biological neural networks. In a prediction problem that can be solved with an ANN based on a cause-and-effect relationship, the inputs are usually independent variables, and the output is the dependent variable. The specified functional relationship can be written as shown in Equation 2.

$$y = f(x_1, x_2, x_3, \dots, x_k) \quad (2)$$

In the functional relationship, $x_1, x_2, \dots, x_k$ are the independent variables and y is the dependent variable. Therefore, ANN can be

considered as a nonlinear regression model. The most commonly used type of ANN in predictions is the Multi-Layer Perceptron (MLP). MLP has a feed-forward structure consisting of an input layer, an output layer, and one or more hidden layers between them (Figure 3) (Hamzacebi, 2011).

For most complex nonlinear problems, a single hidden layer is generally sufficient. Selecting too few or too many neurons in the hidden layer may lead to underfitting or overfitting, respectively. Moreover, an excessive number of hidden neurons can significantly increase the training time. It has been reported that using a smaller network structure with fewer hidden layers often enhances learning efficiency and model generalization (Goodfellow et al., 2016). The optimal number of nodes in the hidden layer can be determined through a trial-and-error approach (Gong, 1999).

Before the analysis, the data were normalized using a standardized transformation. Of the 92 samples, 72 (78.3%) were used for training and 20 (21.7%) for testing. The training-testing split was performed at the sampling-plot level. Both depth-specific records (0–10 and 10–30 cm) from each plot were assigned to the same partition. Thus, 36 plots (72 records) were used for training and 10 plots (20 records) for testing, with no plot represented in both partitions. The ANN model consisted of one hidden layer with five hidden neurons, employing a hyperbolic tangent activation function in the hidden layer and an identity function in the output layer.

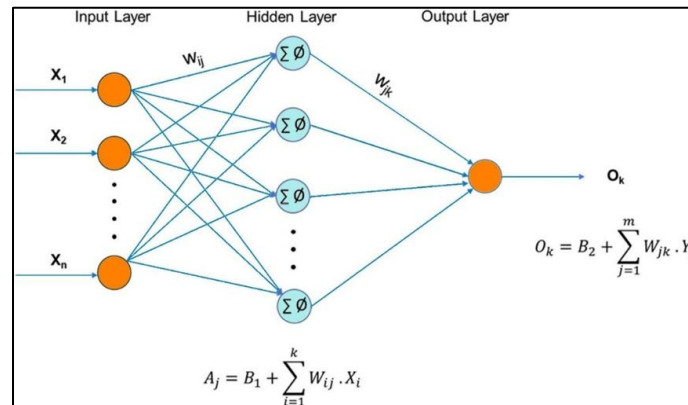


Figure 3. Structure of the MLP ANN used for SOC estimation (adapted from Usta and Yilmaz, 2023)

Variable Selection Strategy

To identify the most relevant predictors of SOC, we first examined pairwise correlations between SOC and all candidate environmental variables. Only those showing significant relationships ($p < 0.05$) were retained for further modeling. A stepwise MLR procedure was then applied to this reduced set. As a result, seven predictors (WP, BD, pH, CaCO_3 , EExp., VDep., and MSI) were identified as significant in the final MLR model. For consistency and comparability, the same subset of variables was also used as input in the ANN modeling.

Although more advanced feature selection algorithms, such as Recursive Feature Elimination (RFE) (Guyon et al., 2002) or LASSO regression, can be practical for large and high-dimensional datasets, they are less suitable for small sample sizes ($n = 92$ samples from 46 plots) and limited predictor spaces. In such cases, variable selection may be unstable, particularly when predictors are collinear (Dormann et al., 2013). Therefore, a combination of correlation screening and stepwise MLR was preferred to ensure robust and interpretable variable selection consistent with the data scale.

Model Validation

Model performance was evaluated using the adjusted coefficient of determination (R^2) for both models, together with the root mean square error (RMSE) for the MLR model, calculated using Equation (3) (Chai and Draxler, 2014). The RMSE provides a direct estimate of modeling error in the original measurement units, with lower values indicating better predictive performance. For the ANN model, the discrepancy between training and testing errors was instead evaluated using the RMSE difference ratio (Equation 4), given that the SPSS Neural Network module reports prediction error on a rescaled (standardized) dependent-variable scale rather than in the original SOC units.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (xi - \bar{xi})^2} \quad (3)$$

Where xi is the predicted value, and $\bar{xi}$ is the observed value. Here, predicted values are obtained from the MLR and MLP ANN

models, while observed values are the laboratory-measured SOC values (%).

The ANN model's tendency toward overfitting was evaluated using the ratio of the difference between training and testing RMSE values, as given in Equation 4.

$$\text{RMSE}_{\text{dr}} = \frac{|\text{RMSE}_{\text{train}} - \text{RMSE}_{\text{test}}|}{\text{RMSE}_{\text{train}}} \times 100 \quad (4)$$

The RMSE difference ratio was used to describe the discrepancy between training and testing errors. Because the ratio in Equation (4) uses an absolute difference, its direction was assessed by comparing the two RMSE values separately. Higher testing error relative to training error was interpreted as indicating a larger generalization gap (Goodfellow et al., 2016; Zhang et al., 2016). Because the RMSE difference ratio was used descriptively, no fixed percentage threshold was imposed.

Result and Discussions

Result of Descriptive Statistics

Descriptive statistics of soil parameters, topographic index, remote sensing parameters, and changes in SOC are given in Table 4. SOC content in the afforestation area varied between 1.85% and 9.84%. Remote sensing data and SOC showed normal distributions, as confirmed by the Kolmogorov-Smirnov (K-S) test and skewness values. In line with the descriptive statistics presented in Table 4, the general trends of soil, topography, and remote sensing variables in the research area were evaluated. In terms of soil properties, organic carbon content is 6.03%, clay ratio is 70.78%, sand is 20.72%, silt is 8.41%, pH is 7.22, electrical conductivity (EC) is $334.08 \mu\text{S cm}^{-1}$, FC is 40.50%, WP is 31.16% and BD is 1.09 g/cm^3 on average. These values reveal that the soils are heavy-textured, slightly alkaline, and rich in organic matter.

The average values of the topographic variables are 509.79 m in elevation (ALT.), 13.47% in slope, and 46.57 m in VDep. This indicates that the area has a medium slope and a topographically diverse structure. From remote sensing data, NDVI is 0.33, MSI is 0.78, TIRS1 is 29,914.48, and TIRS2 is 26,909.15. The TIRS1 and TIRS2 values represent unconverted thermal raster values

and cannot be interpreted as surface temperatures.

Regression Assumptions

The residual analysis of the MLR model confirmed that the fundamental regression assumptions were satisfied. The distribution of residuals against standardized predicted values exhibited a random, pattern-free structure, indicating that the assumption of

constant variance (homoscedasticity) was met. Furthermore, the Q-Q plot showed that the residuals closely followed the theoretical normal distribution, confirming the normality assumption. When evaluated against the criteria described in the Methods section, these results confirmed that the model met the essential assumptions of regression analysis (Osborne and Waters, 2002; Garson, 2012).

Table 4. Descriptive statistics of the input variables

	Mean	Std. Deviation	Min.	Max.	Range	Skewness Statistic	Std. error
Soil Properties							
Organic Carbon%	6.03	1.91	1.85	9.84	7.98	-0.347	0.251
Bulk Density (g/cm ³)	1.09	0.16	0.67	1.44	0.77	-0.115	0.251
Sand%	20.72	11.46	5.00	62.00	57.00	1.248	0.251
Silt%	8.41	3.40	4.00	24.00	20.00	2.139	0.251
Clay%	70.78	10.90	34.00	88.00	54.00	-1.203	0.251
Field Capacity%	40.50	5.95	23.40	52.55	29.15	-0.584	0.251
Wilting Point%	31.16	7.03	15.29	46.94	31.65	-0.374	0.251
Available Water Capacity%	9.89	3.56	3.36	20.76	17.40	0.719	0.251
pH	7.22	0.47	5.28	7.74	2.46	-2.155	0.251
EC ($\mu\text{s cm}^{-1}$)	334.08	155.65	122.50	1121.00	998.50	3.155	0.251
CaCO ₃ %	12.30	15.00	0.00	58.98	58.98	1.313	0.251
Topography							
Altitude (m)	509.79	78.93	342.61	629.86	287.26	-0.485	0.251
North Exposure	0.02	0.21	-0.24	0.60	0.84	1.150	0.251
East Exposure	-0.13	0.12	-0.31	0.18	0.50	0.520	0.251
Transformed aspect (TRASP)	0.63	0.37	0.00	1.00	1.00	-0.526	0.251
Slope%	13.47	6.17	1.52	32.75	31.23	1.099	0.251
Valley Depth (m)	46.57	24.14	2.98	103.05	100.07	0.230	0.251
Remote Sensing							
Normalized difference vegetation index	0.33	0.08	0.18	0.49	0.31	-0.017	0.251
Moisture stress index	0.78	0.13	0.59	1.12	0.52	1.192	0.251
Thermal infrared 1	29914.48	868.53	28771.00	32507.00	3736.00	0.953	0.251
Thermal infrared 2	26909.15	623.43	26048.00	28722.00	2674.00	0.878	0.251

The last column reports the standard error of skewness for 92 depth-level records; plot-level predictors are repeated across the two depths of each of the 46 profiles.

Result of Correlation Analysis

The correlation analysis showed significant relationships between topographic index, soil parameters, remote sensing indices, and SOC. The correlation coefficients between BD, FC, WP, AWC, pH, EC, CaCO₃, ALT, EExp, VDep, NDVI, MSI, and TIRS1

and SOC values were significant at the 0.01 and 0.05 levels. However, the correlation coefficients between Sand, Silt, Clay, Slope, NExp., TRASP, and TIRS2 and SOC values were not significant at the 0.05 level (Figure 4).

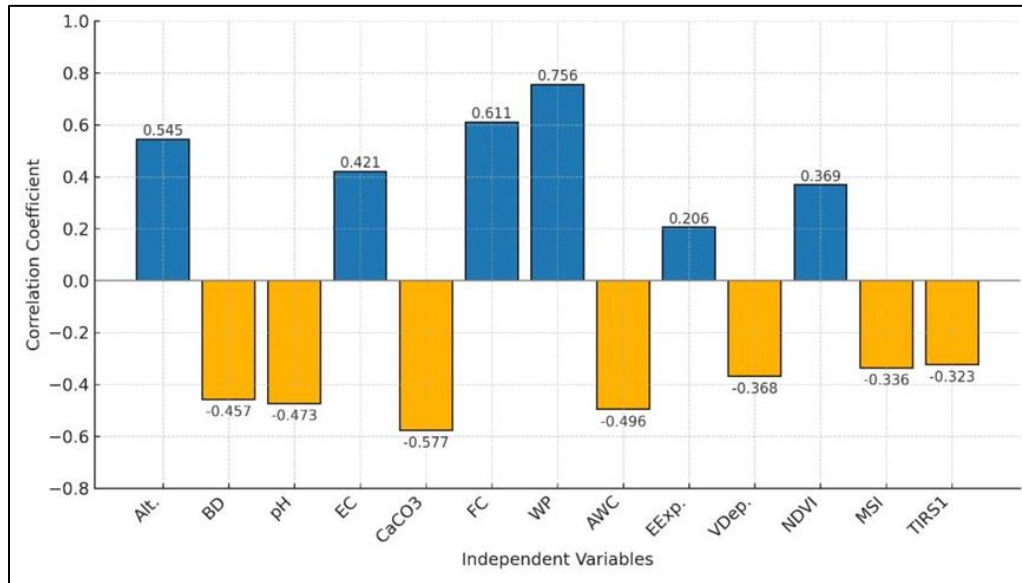


Figure 4. Correlations between independent variables and SOC

According to the correlation analysis, the strongest positive relationship with SOC was observed for WP ($r = 0.756$, $p < 0.01$). FC, ALT., EC, and NDVI also showed significant positive correlations. These findings can be attributed to the effect of organic matter on increasing the soil's water-holding capacity. This is consistent with the well-documented effect of SOC on soil water retention at both field capacity and wilting point (Rawls et al., 2003). On the other hand, variables such as pH, CaCO₃, BD, and MSI show a negative correlation with SOC. Turnover of soil organic matter has been shown to be faster at higher soil pH (Leifeld et al., 2013), which may partly explain the negative association between pH and SOC. The negative bivariate correlation between CaCO₃ and SOC likely

reflects co-varying site conditions in this dataset rather than a direct suppressive mechanism, as detailed further in the model comparison below.

Result of Regression Analysis

After verifying the linear relationships between SOC and the variables using correlation analysis, multiple linear regression models were developed to estimate SOC as a function of selected parameters (BD, WP, pH, CaCO₃, EExp., VDep., and MSI) that were significant at the 0.05 level. In this regression model of SOC values, the F-statistic and coefficients were significant at the 95% confidence level ($p < 0.05$; adj. $R^2 = 0.841$) (Table 5).

Table 5. MLR model to predict SOC

MODEL	Performance		ANOVA		Predictors	Coefficient		t-Test		Collinearity Statistics	
	adj. R^2	RMSE	F	p		B	β	t	p	Tol.	VIF
SOC	0.841	0.789	69.835	<0.001	BD	-1.668	-0.144	-3.000	0.004	0.761	1.315
					WP	0.242	0.891	12.886	<0.001	0.365	2.741
					pH	-0.634	-0.156	-2.955	0.004	0.629	1.590
					CaCO ₃	0.039	0.310	3.969	<0.001	0.286	3.502
					EExp.	2.272	0.137	3.199	0.002	0.945	1.058
					VDep.	-0.011	-0.136	-2.655	0.009	0.661	1.512
					MSI	-2.242	-0.151	-3.308	0.001	0.840	1.191

Result of ANN Analysis

MLP ANN learned the complex relationship between inputs and outputs using the input layer predictors (BD, WP, pH, CaCO₃, EExp., VDep., and MSI) and the output layer (SOC). The hidden layer consists of five neurons, and the output layer contains a single neuron (SOC). The learning rate is 0.4, the maximum training time is 15 minutes, and the model will be trained until a successive step is taken without a decrease in error. The activation functions of the hidden and output layers are the Hyperbolic tangent and Identity functions, respectively. A single hidden layer was preferred for estimating SOC using MLP ANN modeling (Figure 5). The MLP ANN model summary is given in Table 6.

In this study, the ANN model achieved a high coefficient of determination and exhibited a very low relative error on the test

data (RE = 0.043), indicating strong predictive performance. In the literature, similar levels of training and testing errors are considered indicative of a model's generalization capacity, whereas significant discrepancies between these values may suggest potential issues with generalization (Kuhn and Johnson, 2013). Therefore, the model should be further validated under varying conditions and tested on larger, more heterogeneous datasets to ensure its robustness.

Table 6. MLP ANN model summary

Variables		Case processing summary		Model summary	
		N	Percent	Sum of squares error	Relative error
SOC	Training	72	78.3	2.504	0.071
	Testing	20	21.7	0.524	0.043

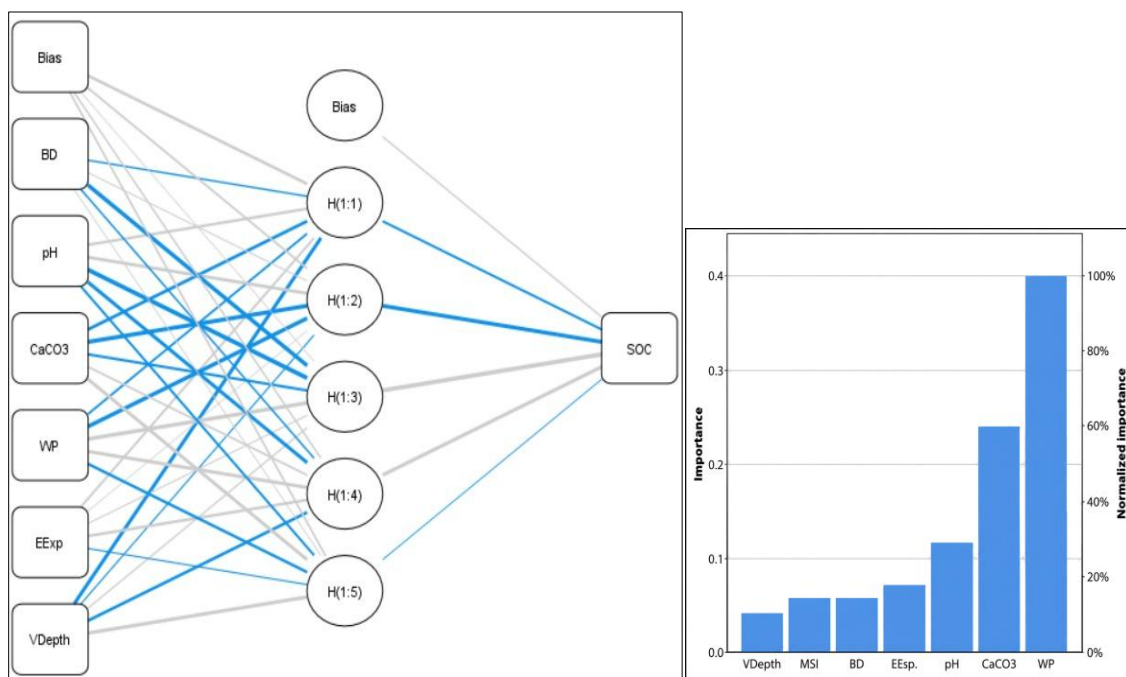


Figure 5. (a) Architecture of the MLP ANN model and (b) normalized importance of the predictor variables

Model Comparison

This study aimed to determine the environmental factors affecting the spatial distribution of SOC, estimate SOC with MLR and MLP ANN models, and compare the results. For Figure 6, measured SOC values and the corresponding MLR and ANN

predictions were summarized as the arithmetic mean of the two depth-specific records (0–10 and 10–30 cm) for each sampling plot. Thus, the figure displays one measured and one prediction value per model for each of the 46 plots; model fitting was based on the 92 depth-specific records.

Stepwise MLR analysis achieved a high explanatory power for the effects of WP, BD, MSI, EExp., CaCO_3 , pH, and VDep on SOC (adj. $R^2 = 0.841$). The significant variables included in the model indicate that these parameters are strong determinants in SOC estimation, consistent with regional evidence that land use/land cover and soil properties significantly affect SOC stocks (Yılmaz and Dengiz, 2021). The adverse effects of BD and pH in the regression analysis support the correlation findings and confirm that these variables contribute to reducing SOC levels.

The same seven predictor variables were used in the ANN model, which demonstrated higher explanatory power ($R^2 = 0.937$). This rate indicates high success in environmental modeling studies (Lamichhane et al., 2019). This situation reveals that ANN can outperform MLR analysis, owing to its ability to model complex, nonlinear relationships. The success of ANN is especially prominent in ecosystems where multiple environmental factors jointly and interactively affect SOC. Similarly, studies such as those by Were et al. (2015) and Taghizadeh-Mehrjardi et al. (2016) have reported that ANN models perform comparably to or better than other machine learning methods in estimating SOC. The ANN yielded a lower relative error in the testing partition (0.043) than in the training partition (0.071; Table 6). The RMSE difference ratio calculated from the partition-specific sums of squares was 13.2% (Equation 4). Testing error did not increase in this split; however, predictive performance for independent sampling plots remains to be evaluated.

WP, BD, soil pH, CaCO_3 , EExp., VDep., and MSI, which are common predictors used in MLR and ANN models, were evaluated using different methodological approaches for estimating SOC. In the MLR model, WP stood out as the strongest predictor with the highest positive regression coefficient and statistical significance level ($t = 12.886$, $p < 0.001$); BD, pH, and MSI contributed to the reduction of the SOC level with negative coefficients, whereas CaCO_3 entered the model with a positive partial coefficient

despite its negative bivariate correlation with SOC ($r = -0.577$), indicating a suppression effect arising from its shared variance with other predictors (tolerance = 0.286); this positive partial effect is consistent with the role of calcium in stabilizing soil organic carbon once the confounding influence of pH and bulk density is statistically removed (Oades, 1988; Rowley et al., 2018). EExp and VDep were included in the model as lower effective but significant variables. In the ANN model, these variables were modeled with weighted connections between layers, rather than linear coefficients. In the ANN model, WP and CaCO_3 had the highest normalized importance, whereas VDep and MSI contributed least (Figure 5b). This difference is related to the capacity of ANN to learn nonlinear relationships and interactions between variables (Bishop, 2006; Haykin, 2009). Although WP is the primary explanatory element in both models, the ANN model can capture nonlinear effects and interactions among predictors and, in this respect, offers a more flexible and robust approach than MLR. The results obtained in this study indicate that both MLR and ANN models used to estimate SOC exhibit high explanatory power.

Future research should integrate sensitivity and scenario analyses to better quantify the response of SOC stocks to key climatic and ecological drivers. In particular, litter type and decomposition kinetics are essential factors influencing SOC dynamics. Fast-decaying litter can increase microbial necromass carbon, thereby reducing the temperature sensitivity of SOC decomposition (Li et al., 2025). Similarly, litter quality effects on soil carbon gain can differ markedly with land-restoration status, as high-quality litter has been shown to boost SOC gain more strongly in unrestored than in restored pine plantations (Li et al., 2024). Furthermore, because more than 70% of global SOC is stored below 20 cm and deep SOC pools respond differently to warming and precipitation change than surface layers (Hicks Pries et al., 2023), extending future work beyond the present study's 0–30 cm sampling depth would

substantially improve the predictability of long-term carbon sequestration in this ecosystem. Such analyses are especially critical for understanding SOC stock responses in semiarid forest ecosystems under climate change scenarios.

The R^2 value (0.937) obtained with the ANN model falls within the upper range of accuracies reported for machine-learning-based SOC models (Lamichhane et al., 2019). On the other hand, the adjusted R^2 value (0.841) obtained with the MLR model used in our study is considered quite successful for linear regression models. Yet it remains lower than the explanatory power of the ANN model. The literature generally reports lower performance of conventional MLR-based

SOC estimation compared with modern machine learning methods. For instance, Forkuor et al. (2017) found that machine learning models performed marginally better than MLR in predicting soil properties, including SOC, from spectral and terrain predictors, mainly because MLR cannot handle non-linear relationships. Collectively, these studies (Forkuor et al., 2017; Taghizadeh-Mehrjardi et al., 2016) indicate that although MLR models can provide moderate explanatory power, ANN and other machine learning approaches can offer better predictive performance for SOC estimation in varied topographic and soil contexts.

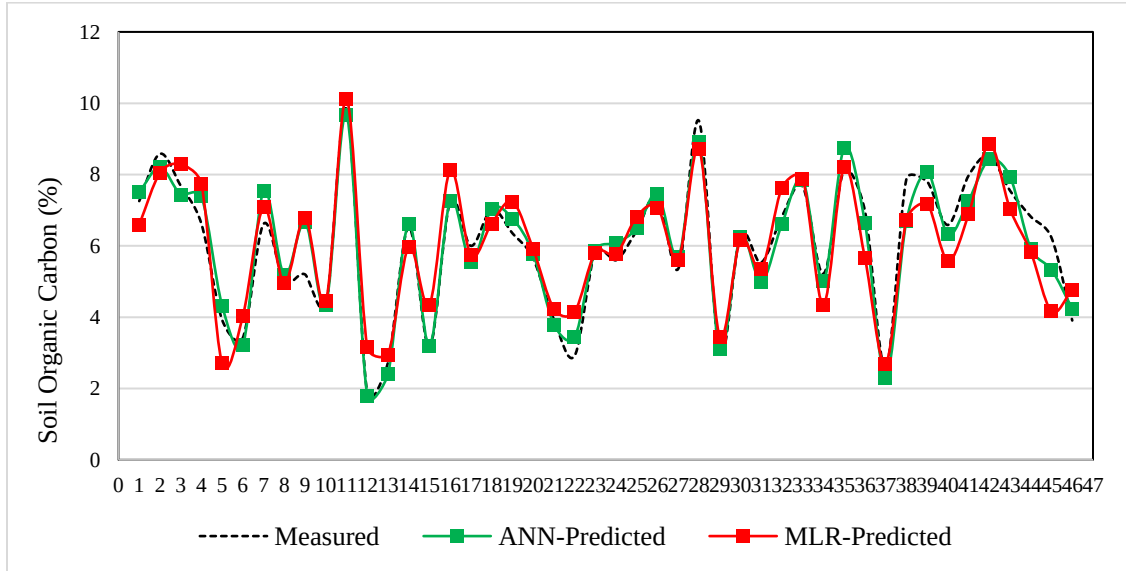


Figure 6. Plot-level arithmetic means of measured SOC and SOC predicted by the MLR and MLP ANN models across the 0–10 and 10–30 cm depth intervals ($n = 46$ plots)

The distribution graphs were obtained by taking topsoil samples from the sample areas. Soil samples (92) from 0-10 and 10-30 cm depths (46 sampling areas) were used in ANN

modeling. The ANN model achieved a high coefficient of determination ($R^2 = 0.937$) between the predicted SOC and the measured SOC (Figure 7).

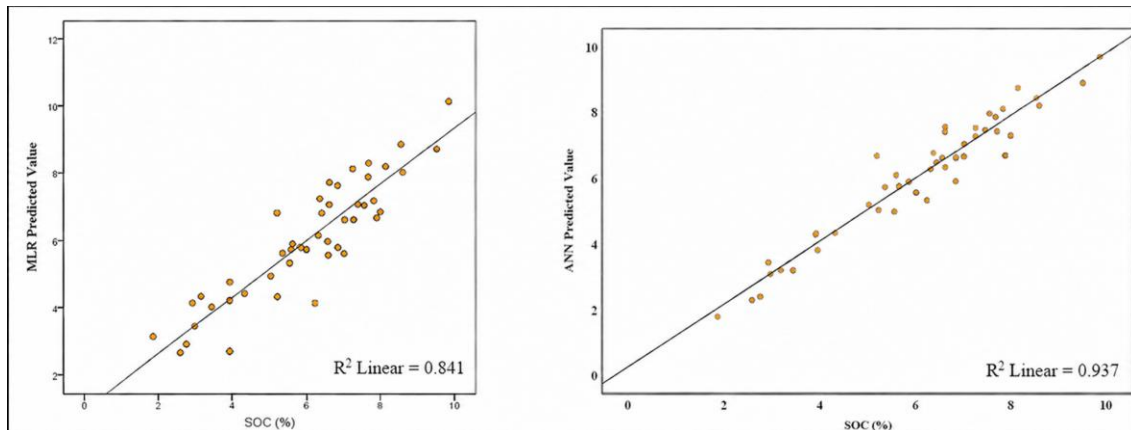


Figure 7. Predicted versus measured SOC values: (a) MLR and (b) MLP ANN

Conclusion

This study demonstrated the comparative effectiveness of MLR and ANN models for predicting SOC in a semi-arid forest ecosystem. The findings revealed that both models achieved high prediction accuracy; however, the ANN model outperformed MLR by better capturing complex, nonlinear interactions among environmental variables ($R^2 = 0.937$).

Based on soil properties, topographic parameters, and remote sensing indices, the study identified WP, BD, pH, CaCO_3 , EExp., VDep., and MSI as the most influential predictors of SOC distribution. WP emerged as the strongest positive contributor in both models, while BD, pH, and MSI showed negative associations with SOC. These findings underscore the significant influence of soil physical and chemical properties, as well as vegetation-related variables, on soil carbon dynamics.

Beyond predictive accuracy, the results have practical implications for sustainable soil carbon management, particularly in light of the ongoing challenges posed by climate change. Using such models as decision-support tools can help optimize land-use strategies and improve estimates of regional SOC content.

Future studies may enhance prediction performance by incorporating spatial resolution, land-use history, and deeper soil layers. In this regard, advanced machine learning approaches such as ANNs offer a

flexible and robust framework for modeling SOC dynamics across diverse ecosystems.

The core assumption of this study is that there have been no significant disturbances (e.g., fires, clear-cutting) in the area since 2018. This was confirmed through available records and remote sensing imagery. Future research should validate these models with new, contemporary soil samples and establish long-term monitoring plots. This would not only confirm the temporal stability of the observed patterns but also enable quantification of SOC stock changes over time, providing a more dynamic understanding of carbon sequestration in this ecosystem.

Ethics Committee Approval

N/A

Peer-review

Externally peer-reviewed.

Author Contributions

Conceptualization: E.K., S.M.; Investigation: E.K., S.M.; Material and Methodology: E.K., S.M.; Visualization: S.M.; Writing-Original Draft: E.K., S.M.; Writing-review & Editing: E.K., S.M.; Other (statistical analysis and mapping): S.M. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

The authors declared that this study has received no financial support.

References

- Aktimur, H. T., Ateş, Ş., Yurdakul, M. E., Tekirli, M. E. & Keçer, M. (1992). Niksar-Erbaa ve Destek dolayının jeolojisi. *Maden Tetkik ve Arama Dergisi*, 114, 25-36.
- Alpar, R. (2013). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Detay Yayıncılık.
- ASF DAAC. (2014). ALOS PALSAR High Resolution Radiometric Terrain Corrected Product. NASA Alaska Satellite Facility Distributed Active Archive Center. <https://doi.org/10.5067/Z97HFCNKR6VA>
- Ayoubi, S., Shahri, A. P., Karchegani, P. M. & Sahrawat, K. L. (2011). Application of artificial neural network (ANN) to predict soil organic matter using remote sensing data in two ecosystems. *Biomass and remote sensing of biomass*, 10, 181-196.
- Bahn, M., Kutsch, W. L. & Heinemeyer, A. (2010). Synthesis: emerging issues and challenges for an integrated understanding of soil carbon fluxes. In *Soil Carbon Dynamics: an Integrated Methodology*, 257-271. <https://doi.org/10.1017/CBO9780511711794>
- Batjes, N. H. (1996). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 47(2), 151-163. <https://doi.org/10.1111/j.1365-2389.1996.tb01386.x>
- Batjes, N. H. (2006). Soil carbon stocks of Jordan and projected changes upon improved management of croplands. *Geoderma*, 132(3-4), 361-371. <https://doi.org/10.1016/j.geoderma.2005.05.013>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Blake, G. R. & Hartge, K. H. (1986). *Bulk density. Methods of soil analysis: Part 1 Physical and mineralogical methods*, 5, 363-375. <https://doi.org/10.2136/sssabookser5.1.2ed.c13>
- Böhner, J. & Antonić, O. (2009). Land-surface parameters specific to topo-climatology. *Developments in Soil Science*, 33, 195-226. [https://doi.org/10.1016/S0166-2481\(08\)00008-1](https://doi.org/10.1016/S0166-2481(08)00008-1)
- Bouyoucos, G. J. (1962). Hydrometer method improved for making particle size analyses of soils 1. *Agronomy journal*, 54(5), 464-465. <https://doi.org/10.2134/agronj1962.00021962005400050028x>
- Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı*. Pegem Atıf İndeksi, 001-214.
- Certini, G. (2005). Effects of fire on properties of forest soils: A review. *Oecologia*, 143(1), 1-10. <https://doi.org/10.1007/s00442-004-1788-8>
- Chai, T. & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E. & et al. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific model development*, 8(7), 1991-2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Çağlar, K. (1949). *Toprak Bilgisi*. Ankara Üniversitesi. Ziraat Fakültesi Yayınları No: 985. Ankara.
- Dalal, R. C. & Mayer, R. J. (1986). Long term trends in fertility of soils under continuous cultivation and cereal cropping in southern Queensland. I. Overall changes in soil properties and trends in winter cereal yields. *Soil Research*, 24(2), 265-279. <https://doi.org/10.1071/SR9860265>
- Dawson, J. J. & Smith, P. (2007). Carbon losses from soil and its consequences for land-use management. *Science of the total environment*, 382 (2-3), 165-190. <https://doi.org/10.1016/j.scitotenv.2007.03.023>
- De Anta, R. C., Luís, E., Febrero-Bande, M., Galiñanes, J., Macías, F., Ortiz, R. & Casás, F. (2020). Soil organic carbon in peninsular Spain: Influence of environmental factors and spatial distribution. *Geoderma*, 370, 114365. <https://doi.org/10.1016/j.geoderma.2020.114365>
- Dorji, T., Odeh, I. O., Field, D. J. & Baillie, I. C. (2014). Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. *Forest Ecology and Management*, 318, 91-102. <https://doi.org/10.1016/j.foreco.2014.01.003>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G. & et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Esri. (2020). ArcGIS Desktop: Release 10.8. Environmental Systems Research Institute, Redlands, CA, USA.

- Evans, J. S., Oakleaf, J., Cushman, S. A., & Theobald, D. M. (2014). *An ArcGIS toolbox for surface gradient and geomorphometric modeling* (Version 2.0-0). <http://evansmurphy.wix.com/evansspatial>
- Falahatkar, S., Hosseini, S. M., Ayoubi, S. & Salmanmahiny, A. (2016). Predicting soil organic carbon density using auxiliary environmental variables in northern Iran. *Archives of Agronomy and Soil Science*, 62(3), 375-393. <https://doi.org/10.1080/03650340.2015.1051472>
- Falloon, P. & Smith, P. (2003). Accounting for changes in soil carbon under the Kyoto Protocol: need for improved long-term data sets to reduce uncertainty in model projections. *Soil Use and Management*, 19(3), 265-269. <https://doi.org/10.1111/j.1475-2743.2003.tb00313.x>
- Falloon, P. D., Smith, P., Smith, J. U., Szabo, J., Coleman, K. & Marshall, S. (1998). Regional estimates of carbon sequestration potential: linking the Rothamsted Carbon Model to GIS databases. *Biology and Fertility of soils*, 27(3), 236-241. <https://doi.org/10.1007/s003740050426>
- Falloon, P., Smith, P., Szabo, J. & Pasztor, L. (2002). Comparison of approaches for estimating carbon sequestration at the regional scale. *Soil Use and management*, 18(3), 164-174. <https://doi.org/10.1111/j.1475-2743.2002.tb00236.x>
- Forkuor, G., Hounkpatin, O. K., Welp, G. & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1), e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- 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 Regional*, 29, e00513. <https://doi.org/10.1016/j.geodrs.2022.e00513>
- Garson, G. D. (2012). *Testing statistical assumptions*.
- Gong, P. (1999). Inverting a canopy reflectance model using a neural network. *International Journal of Remote Sensing*, 20(1), 111-122. <https://doi.org/10.1080/014311699213631>
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.
- Grinand, C., Le Maire, G., Vieilledent, G., Razakamanarivo, H., Razafimbelo, T. & Bernoux, M. (2017). Estimating temporal changes in soil carbon stocks at ecoregional scale in Madagascar using remote-sensing. *International Journal of Applied Earth Observation and Geoinformation*, 54, 1-14. <https://doi.org/10.1016/j.jag.2016.09.002>
- Gülçür, F. (1974). *Toprağın Fiziksel ve Kimyasal Analiz Metodları*, İ.Ü. Orman Fakültesi Yayınları, Yayın No: 201, İstanbul.
- Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1), 389-422. <https://doi.org/10.1023/A:1012487302797>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). *Multivariate data analysis* (7th ed., Pearson New International Edition). Pearson Education Limited.
- Hamzacebi, C. (2011). *Artificial neural networks*. Ekin Publication, Bursa.
- Hashimoto, S., Ugawa, S., Morisada, K., Wattenbach, M., Smith, P. & Matsuura, Y. (2012). Potential carbon stock in Japanese forest soils-simulated impact of forest management and climate change using the CENTURY model. *Soil Use and Management*, 28(1), 45-53. <https://doi.org/10.1111/j.1475-2743.2011.00372.x>
- Haykin, S. (2009). *Neural networks and learning machines*, 3/E. Pearson Education India.
- Hicks Pries, C. E., Ryals, R., Zhu, B., Min, K., Cooper, A., Goldsmith & et al. (2023). The Response of Deep Soil Organic Carbon to Global Change. *Annual Review of Ecology, Evolution, and Systematics*, 54(1), 375-401. <https://doi.org/10.1146/annurev-ecolsys-102320-085332>
- Hızalan, E. & Ünal, H. (1966). *Topraklarda önemli kimyasal analizler*. AÜ Ziraat Fakültesi Yayınları, 278, 5-7.
- IBM Corp., (2011). IBM SPSS Statistics for Windows, Version 20.0. IBM Corp, Armonk, NY.
- Ingleby, H. R. & Crowe, T. G. (2001). Neural network models for predicting organic matter content in Saskatchewan soils. *Outlook*, 1(2), 3.
- Jebari, A., Del Prado, A., Pardo, G., Rodriguez Martin, J. A. & Álvaro-Fuentes, J. (2018). Modeling regional effects of climate change on soil organic carbon in Spain. *Journal of Environmental Quality*, 47(4), 644-653. <https://doi.org/10.2134/jeq2017.07.0294>
- Jones, C., McConnell, C., Coleman, K., Cox, P., Falloon, P., Jenkinson, D. & Powlson, D. (2005). Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in

- soil. *Global Change Biology*, 11(1), 154-166. <https://doi.org/10.1111/j.1365-2486.2004.00885.x>
- Kantarıcı, M. D. (2005a). Türkiye'nin yetiştirme ortamı bölgesel sınıflandırması ve bu birimlerdeki orman varlığı ile devamlılığının önemi. İstanbul Üniversitesi Yayın No: 4558, Orman Fakültesi Yayın No: 484, İstanbul.
- Kantarıcı, M. D. (2005b). *Orman Ekosistemleri Bilgisi*. İÜ Yayın, (4594).
- Karaöz, M. Ö. (1989). Toprakların su ekonomisine ilişkin bazı fiziksel özelliklerinin laboratuvarında belirlenmesi yöntemleri. *Journal of the Faculty of Forestry*, Istanbul University, 39(2), 133-144.
- Keesstra, S. D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P. & et al. (2016). Forum paper: The significance of soils and soil science towards realization of the UN sustainable development goals (SDGS). *SOIL*, 2(2), 111-128. <https://doi.org/10.5194/soil-2-111-2016>
- Kheir, R. B., Greve, M. H., Bøcher, P. K., Greve, M. B., Larsen, R. & McCloy, K. (2010). Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. *Journal of Environmental Management*, 91(5), 1150-1160. <https://doi.org/10.1016/j.jenvman.2010.01.001>
- Kokaly, R. F. & Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, 67(3), 267-287. [https://doi.org/10.1016/S0034-4257\(98\)00084-4](https://doi.org/10.1016/S0034-4257(98)00084-4)
- Köksal, B. A. (1977). *İstatistik analiz metodları*. Boğaziçi Üniversitesi Yayınları.
- Kuhn, M. & Johnson, K. (2013). Applied predictive modeling. New York: Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science*, 304(5677), 1623-1627. <https://doi.org/10.1126/science.1097396>
- Lamichhane, S., Kumar, L. & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, 352, 395-413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- Lamsal, S. (2009). Visible near-infrared reflectance spectroscopy for geospatial mapping of soil organic matter. *Soil Science*, 174(1), 35-44. <https://doi.org/10.1097/SS.0b013e3181906a09>
- Leifeld, J., Bassin, S., Conen, F., Hajdas, I., Egli, M. & Fuhrer, J. (2013). Control of soil pH on turnover of belowground organic matter in subalpine grassland. *Biogeochemistry*, 112(1-3), 59-69. <https://doi.org/10.1007/s10533-011-9689-5>
- Li, R., Wang, C., Lv, C., Zhou, T., Yin, S. & Zhou, Z. (2025). Fast-decaying tree litter reduces the temperature sensitivity of soil carbon decomposition by increasing microbial necromass carbon. *Geoderma*, 454, 117185. <https://doi.org/10.1016/j.geoderma.2025.117185>
- Li, S., Jiang, Y., Lyu, M., Deng, C., Deng, W. & et al. (2024). High-quality litter exerts a greater effect on soil carbon gain in unrestored than restored pine plantations. *Science of The Total Environment*, 952, 175975. <https://doi.org/10.1016/j.scitotenv.2024.175975>
- Li, Y., Wang, X., Niu, Y., Lian, J., Luo, Y. & et al. (2018). Spatial distribution of soil organic carbon in the ecologically fragile Horqin Grassland of northeastern China. *Geoderma*, 325, 102-109. <https://doi.org/10.1016/j.geoderma.2018.03.032>
- Liu, Z., Shao, M. A. & Wang, Y. (2011). Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture, Ecosystems & Environment*, 142(3-4), 184-194. <https://doi.org/10.1016/j.agee.2011.05.002>
- Lucht, W., Schaphoff, S., Erbrect, T., Heyder, U. & Cramer, W. (2006). Terrestrial vegetation redistribution and carbon balance under climate change. *Carbon Balance and Management*, 1(1), 6. <https://doi.org/10.1186/1750-0680-1-6>
- Luo, Z., Feng, W., Luo, Y., Baldock, J. & Wang, E. (2017). Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. *Global Change Biology*, 23(10), 4430-4439. <https://doi.org/10.1111/gcb.13767>
- Lützw, M. V., Kögel-Knabner, I., Ekschmitt, K., Matzner, E., Guggenberger, G., Marschner, B. & Flessa, H. (2006). Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions—a review. *European Journal of Soil Science*, 57(4), 426-445. <https://doi.org/10.1111/j.1365-2389.2006.00809.x>
- McBratney, A. B., Santos, M. M. & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- McKenzie, N. J. & Ryan, P. J. (1999). Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89(1-2),

- 67-94. [https://doi.org/10.1016/S0016-7061\(98\)00137-2](https://doi.org/10.1016/S0016-7061(98)00137-2)
- Melillo, J. M., Steudler, P. A., Aber, J. D., Newkirk, K., Lux, H. & et al. (2002). Soil warming and carbon-cycle feedbacks to the climate system. *Science*, 298(5601), 2173-2176. <https://doi.org/10.1126/science.1074153>
- MGM. (2024). *Niksar iklim verileri* (2013-2023). Meteoroloji 10. Bölge Müdürlüğü, Samsun.
- Minasny, B., McBratney, A. B., Malone, B. P. & Wheeler, I. (2013). Digital mapping of soil carbon. *Advances in Agronomy*, 118, 1-47. <https://doi.org/10.1016/B978-0-12-405942-9.00001-3>
- Mircholi, F., Kiani-Harchegani, M., Khaledi Darvishan, A., Falahatkar, S. & Sadeghi, S. H. (2020). Spatial distribution dependency of soil organic carbon content to important environmental variables. *Ecological Indicators*, 116, 106473. <https://doi.org/10.1016/j.ecolind.2020.106473>
- Moore, I. D., Lewis, A. & Gallant, J. C. (1993). Terrain attributes: estimation methods and scale effects. *Modelling change in environmental systems*, 189-214.
- Muñoz-Rojas, M., Jordán, A., Zavala, L. M., de la Rosa, D., Abd-Elmabod, S. K. & Anaya-Romero, M. (2012). Organic carbon stocks in Mediterranean soil types under different land uses (Southern Spain). *Solid Earth*, 3(2), 375-386. <https://doi.org/10.5194/se-3-375-2012>
- Muñoz-Rojas, M., Jordán, A., Zavala, L. M., González-Peñaloza, F. A., de la Rosa, D., Pino-Mejias, R. & Anaya-Romero, M. (2013). Modelling soil organic carbon stocks in global change scenarios: a CarboSOIL application. *Biogeosciences*, 10(12), 8253-8268. <https://doi.org/10.5194/bg-10-8253-2013>
- Oades, J. M. (1988). The retention of organic matter in soils. *Biogeochemistry*, 5(1), 35-70. <https://doi.org/10.1007/BF02180317>
- Osborne, J. W. & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, 8(1). <https://doi.org/10.7275/r222-hv23>
- Oueslati, I., Allamano, P., Bonifacio, E. & Claps, P. (2013). Vegetation and topographic control on spatial variability of soil organic carbon. *Pedosphere*, 23(1), 48-58. [https://doi.org/10.1016/S1002-0160\(12\)60079-4](https://doi.org/10.1016/S1002-0160(12)60079-4)
- Paustian, K., Elliott, E. T. & Killian, K. (1998). Modeling soil carbon in relation to management and climate change in some agroecosystems in central North America. In R. Lal, J. M. Kimble, R. F. Follett & B. A. Stewart (Eds.), *Soil processes and the carbon cycle* (pp. 459-471). CRC Press.
- Post, W. M., Izaurralde, R. C., Mann, L. K. & Bliss, N. (2001). Monitoring and verifying changes of organic carbon in soil. *Climatic Change*, 51(1), 73-99. <https://doi.org/10.1023/A:1017514802028>
- Rawls, W. J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M. & Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. *Geoderma*, 116(1-2), 61-76. [https://doi.org/10.1016/S0016-7061\(03\)00094-6](https://doi.org/10.1016/S0016-7061(03)00094-6)
- Restrepo, C. & LaQuay, G. (2018). Variation in carbon dynamics mediated by landslide activity in a Central American mountainscape. In EGU General Assembly Conference Abstracts, 17184.
- Roberts, D. W. & Cooper, S. V. (1989). Concepts and techniques of vegetation mapping. *Land Classifications Based on Vegetation: Applications for Resource Management*, 90-96.
- Rock, B. N., Vogelmann, J. E., Williams, D. L., Vogelmann, A. F. & Hoshizaki, T. (1986). Remote Detection of Forest Damage: Plant responses to stress may have spectral "signatures" that could be used to map, monitor, and measure forest damage. *Bioscience*, 36(7), 439-445. <https://doi.org/10.2307/1310339>
- Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. In Third NASA Earth Resources Technology Satellite Symposium, 1973, 1, 309-317.
- Rowley, M. C., Grand, S. & Verrecchia, É. P. (2018). Calcium-mediated stabilisation of soil organic carbon. *Biogeochemistry*, 137(1), 27-49. <https://doi.org/10.1007/s10533-017-0410-1>
- Scharlemann, J. P., Tanner, E. V., Hiederer, R. & Kapos, V. (2014). Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management*, 5(1), 81-91. <https://doi.org/10.4155/cmt.13.77>
- Schmidt, M. W., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G. & et al. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, 478(7367), 49-56. <https://doi.org/10.1038/nature10386>
- Schuur, E. A., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W. & et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, 520(7546), 171-179. <https://doi.org/10.1038/nature14338>
- Smith, J. O., Smith, P., Wattenbach, M., Zaehle, S., Hiederer, R. & et al. (2005). Projected

- changes in mineral soil carbon of European croplands and grasslands, 1990–2080. *Global Change Biology*, 11(12), 2141-2152. <https://doi.org/10.1111/j.1365-2486.2005.001075.x>
- Stage, A. R. (1976). An expression for the effect of aspect, slope, and habitat type on tree growth. *Forest Science*, 22(4), 457-460. <https://doi.org/10.1093/forestscience/22.4.457>
- Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N. & et al. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment*, 164, 80-99. <https://doi.org/10.1016/j.agee.2012.10.001>
- Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A. & Schaepman, M. (2018). Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. *Agriculture, Ecosystems & Environment*, 258, 129-142. <https://doi.org/10.1016/j.agee.2018.02.012>
- Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K. & Kerry, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*, 266, 98-110. <https://doi.org/10.1016/j.geoderma.2015.12.003>
- Tsui, C. C., Tsai, C. C. & Chen, Z. S. (2013). Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. *Geoderma*, 209, 119-127. <https://doi.org/10.1016/j.geoderma.2013.06.013>
- Ünver, Ö. & Gamgam, H. (1999). *Uygulamalı istatistik yöntemler*. Siyasal Kitabevi.
- USGS, (2024). Earth Explorer. Available online: <https://earthexplorer.usgs.gov>. (Retrieved date: 27 April 2024).
- Usta, A. & Yilmaz, M. (2023). Prediction of pine mistletoe infection using remote sensing imaging: A comparison of the artificial neural network model and logistic regression model. *Forest Pathology*, 53(1), e12783. <https://doi.org/10.1111/efp.12783>
- Walkley, A. (1947). A critical examination of a rapid method for determining organic carbon in soils-Effect of variations in digestion conditions and of inorganic soil constituents. *Soil Science*, 63, 251-264.
- Wan, Y., Lin, E., Xiong, W., Li, Y. E. & Guo, L. (2011). Modeling the impact of climate change on soil organic carbon stock in upland soils in the 21st century in China. *Agriculture, Ecosystems & Environment*, 141(1-2), 23-31. <https://doi.org/10.1016/j.agee.2011.02.004>
- Wang, S., Tian, H., Liu, J. & Pan, S. (2003). Pattern and change of soil organic carbon storage in China: 1960s-1980s. *Tellus B: Chemical and Physical Meteorology*, 55(2), 416-427. <https://doi.org/10.3402/tellusb.v55i2.16715>
- 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>
- Wetterlind, J., Stenberg, B. & Söderström, M. (2008). The use of near infrared (NIR) spectroscopy to improve soil mapping at the farm scale. *Precision Agriculture*, 9(1), 57-69. <https://doi.org/10.1007/s11119-007-9051-z>
- Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J. & Zhu, B. (2008). Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Global Change Biology*, 14(7), 1592-1599. <https://doi.org/10.1111/j.1365-2486.2008.01591.x>
- Yılmaz, M. & Dengiz, O. (2021). Bazı toprak özellikleri ile ilişkili olarak arazi kullanımı ve arazi örtüsünün toprak organik karbon stokuna etkisi. *Türkiye Tarımsal Araştırmalar Dergisi*, 8(2), 154-167. <https://doi.org/10.19159/tutad.865188>
- Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530. <https://doi.org/10.48550/arXiv.1611.03530>
- Zhu, M., Feng, Q., Zhang, M., Liu, W., Deo, R. C., Zhang, C. & Yang, L. (2019). Soil organic carbon in semiarid alpine regions: the spatial distribution, stock estimation, and environmental controls. *Journal of Soils and Sediments*, 19(10), 3427-3441. <https://doi.org/10.1007/s11368-019-02295-6>