International Journal of Agriculture and Wildlife Science

http://dergipark.org.tr/ijaws

Research Article

Using Different Regression Tree Algorithms to Predict Soil Organic Matter with Digital Color Parameters in Soil Profile Wall

Gafur Gözükara[1*] (ID),   **Yasin Altay[2]** (ID)

[1]Eskişehir Osmangazi University, Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Eskişehir
[2]Eskişehir Osmangazi University, Faculty of Agriculture, Department of Animal Science, Eskişehir

*Corresponding author
ggozukara@ogu.edu.tr

Abstract. Soil organic matter has a critical role for the physical, chemical and biological properties of the soil and for sustainable soil and agriculture. Quick and cost-effective prediction of soil organic matter can provide basic data support for precision agriculture. The study area is located in the Muttalip pasture of Tepebaşı, Eskişehir. The soil profile wall (1x1 m) was dug and divided into 10x10 cm raster cell. A total of 100 soil samples were taken from center of each raster cell of the soil profile wall. The field-based and lab-based digital color parameters (CIE Lab) were measured depending on the grid sampling model. The ordinary Kriging interpolation method was used in geostatistical distribution maps of the amount of organic matter (OM) and field-based and lab-based CIE Lab values. CHAID, Ex-CHAID, and CART regression tree algorithms were used to predict the OM with field-based and lab-based CIE Lab values. The OM in the soil profile wall varies between 4.65-10.54% in the topsoils, while it varies between 0.01-0.41% in the subsoils. According to the results, lab-based CIE Lab values obtained high predicting performance and more effective than field-based CIE Lab values. It concluded that the CART algorithm can be used rapidly and economically in prediction OM with high prediction performance (R^2=0.89) with lab-based digital color parameters.

Toprak Profil Duvarında Farklı Regresyon Ağacı Algoritmaları Kullanılarak Sayısal Renk Parametreleri ile Organik Maddenin Tahmin Edilmesi

Özet. Toprak organik maddesi toprağın fiziksel, kimyasal ve biyolojik özellikleri ile sürdürülebilir toprak ve tarım için oldukça kritik bir role sahiptir. Toprak organik maddesinin çabuk ve düşük maliyetle tahmin edilmesi hassas tarım için temel veri desteği sağlayabilir. Çalışma alanı Eskişehir ili Muttalıp merası sınırları içerisinde yeralmaktadır. Toprak profil duvarı (1x1m) kazılmış ve 10x10 cm'lik grid hücrelere bölünmüştür. Toprak profil duvarından herbir grid hücreden grid yöntemi ile toplam 100 adet toprak örneği alınmıştır. Toprak örneklerinde sayısal renk parametrelerinin belirlenmesi grid örnekleme modeline bağlı olarak hem arazi hem de lobaratuvar koşullarında gerçekleştirilmiştir. Profil duvarından arazi ve laboratuvar koşullarında elde edilen CIE Lab değerleri ve organik madde miktarının jeoistatiksel olarak dağılım haritalarında Ordinary Kriging interpolasyon metodu kullanılmıştır. Sayısal renk parametreleri ile organik madde miktarının tahmin edilmesinde CHAID, Ex-CHAID ve CART regresyon ağacı algoritmaları kullanılmıştır. Toprak profil duvarında organik madde miktarı yüzey topraklarda %4.65-10.54 arasında değişirken yüzey altı topraklarda %0.01-0.41 arasında değişmektedir. Araştırma sonuçlarına göre, OM miktarının tahmin edilmesinde laboratuvar koşullarında elde edilen CIE Lab değerlerinin laboratuvar koşullarında elde edilen CIE Lab değerlerinden daha etkilidir. Araştırma, CART algoritmasının laboratuvar koşullarında elde edilen sayısal renk parametreleri ile OM miktarının yüksek başarı performansı (R^2=0.89) ile tahmin edilmesinde hızlı ve ekonomik olarak kullanılabileceğini ortaya çıkarmıştır.

INTRODUCTION

Soil organic matter (OM), one of the most important components of the soil, plays a decisive role in the formation of soil structure and improvement of soil quality (McBratney *et al.,* 2014; Acar *et al.,* 2018; Çelik *et al.,* 2019; Aktaş and Yüksel, 2020). Understanding the dynamic mobility and changes of the OM in the soil surface and profile wall is a fundamental condition for monitoring soil fertility, performing precision agriculture, as well as ensuring the sustainable development of soil and agriculture (Wu *et al.,* 2015; Yılmaz *et al.,* 2019). The determining role of OM on the physical, chemical, and biological properties of soils reveals the necessity of knowing and managing the amount of organic matter. The chemical analysis method performed under laboratory conditions with traditional analysis methods has revealed the necessity of new analysis methods, which are time-consuming, inconvenient, costly, and not sensitive to the environment due to the chemical wastes that result from the analysis (Torrent and Barron, 1993; Nishiyama *et al.,* 2011; Sawada *et al.,* 2013; Gözükara *et al.,* 2021a). Therefore, it seems very difficult to meet the demand for real-time big data in the current agricultural development and targets in a short time, at low cost and environmentally friendly. For this reason, many researchers have focused on the relationship between the physical, chemical, and biological properties of soils and numerical color parameters, which can be alternative to traditional analysis methods (Viscarra Rossel *et al.,* 2006, 2008, Günal *et al.,* 2008; Doi *et al.,* 2010; Budak *et al.,* 2018; Gözükara *et al.,* 2021a).

The Munsell color scale was published in the USA in 1941 (Rice *et al.,*1941; Simonson, 1993) and still maintains its validity and prevalence in determining the color, which is one of the important morphological properties of the soil in field and laboratory conditions (Thwaites, 2002; Gözükara *et al.,* 2019, 2020a, 2020b). Munsell color scale is a method based on subjective observation created by expert knowledge, experience, and interpretation of dominant spectral color (Hue), darkness (Value), and purity (Chroma) values. Although it is based on subjective observation, it still maintains its validity in the definition and classification of horizons and soils in soil profiles in many soil classification systems around the world, including the American soil classification system (Soil Survey Staff, 2010, 2014; Hartemink and Minasny, 2014; Gözükara *et al.,* 2019; Gözükara *et al.,* 2020a, 2020b; Gözükara *et al.,* 2021b). However, many researchers stated that there may be important deficiencies and errors depending on the subjective interpretation of the researcher of the Munsell Color Scale and especially the possibility that some color values may vary from person to person (Post *et al.,* 1993; Moritsuka *et al.,* 2014). In particular, they focused on the CIE Lab colorimetric method, in which the color of the soil is digitized depending on the technique and technology developed in recent years and can be an alternative to the Munsell color scale based on subjective observation (Torrent and Baron, 1993; Günal *et al.,* 2008; Moritsuka *et al.,* 2014; Budak *et al.,* 2018; Gözükara *et al.,* 2021a).

In the CIE Lab colorimetric method, L* value symbolizes the brightness and darkness of the color, a* value symbolizes red and green tones, b* value symbolizes yellow and blue tones (Barrett, 2002; Fan *et al.,* 2017). Digital soil color parameters (CIE Lab) which are obtained easily and low costly in the field and laboratory conditions have been extensively used by soil scientists to predict delineation of horizon boundaries in soil profiles (Soil Survey Staff, 2014), physical soil properties (Günal *et al.,* 2008; Budak *et al.,* 2018; Gözükara *et al.,* 2021a), chemical soil properties (Schulze *et al.,* 1993; Fang *et al.,* 1999; Günal *et al.,* 2008; Budak *et al.,* 2018; Gözükara *et al.,* 2021a), biological and mineralogical soil properties (Torrent *et al.,* 1980; Shen *et al.,* 2006). As can be seen from the literature researches, the researchers focused only on the relationship and prediction of the numerical color parameters (CIE Lab) obtained from the degraded soil samples in the laboratory environment and soil properties. However, studies examining the relationship between numerical color parameters obtained from soil samples in field conditions and soil properties and their effect on prediction performance have been quite limited (Günal *et al.,* 2008). At the same time, the researchers focused on explaining and predicting the relationship between numerical color parameters obtained from soil samples and soil properties and predictive performance using correlation analysis, linear and parabolic regression models (Günal *et al.,* 2008; Moritsuka *et al.,* 2014; Budak *et al.,* 2018; Gözükara *et al.,* 2021a). In cases where some assumptions of linear and nonlinear regression models are not fulfilled, the relationships examined cannot be revealed sufficiently and their reliability also decreases. However, the partial elimination of these disadvantages in the use of data mining algorithms in today's techniques and technologies allows to explain the relationships and to obtain more successful results in obtaining results close to traditional chemical analysis results.

The objectives of this study were to; i-) creating distribution maps of the digital color parameters (CIE Lab) obtained under field and laboratory conditions and organic matter obtained by traditional analysis methods on the soil profile wall, ii-) predict OM in the soil profile wall using digital color parameters (CIE Lab) and determine the most reliable regression tree algorithm

MATERIAL AND METHOD

Study Area and Soil Sampling

The study area is located in Muttalip pasture of Tepebaşı, Eskişehir, TURKEY (latitude 39° 0' 14.08" N, longitude 30° 33' 45.79" E). The research area (pasture land) soils have occurred on almost flat slope, have been formed in the Quaternary period and continue to develop on alluvial materials, and is located at an altitude of 786 m above sea level. The research area's climatic characteristics are harsh with snow in winters, hot, and dry in summers. The study area has mean annual precipitation of 522.2 mm and mean annual temperature of 13.6 °C (*DMİ, 2017*).

The soil profile wall (1 m x 1 m) was dug (1.5m L x 1.5m W x 1.5m D) and divided into a 10 cm x 10 cm raster cell. Every soil sample which collected from the center of each raster cell (covering 80% of the raster cell) (Figure 1). Total 100 soil samples were taken in the soil profile wall based on the grid-sampling model. Air-dried and grounded soil samples sieved using 2 mm diameter for measurement and analysis.
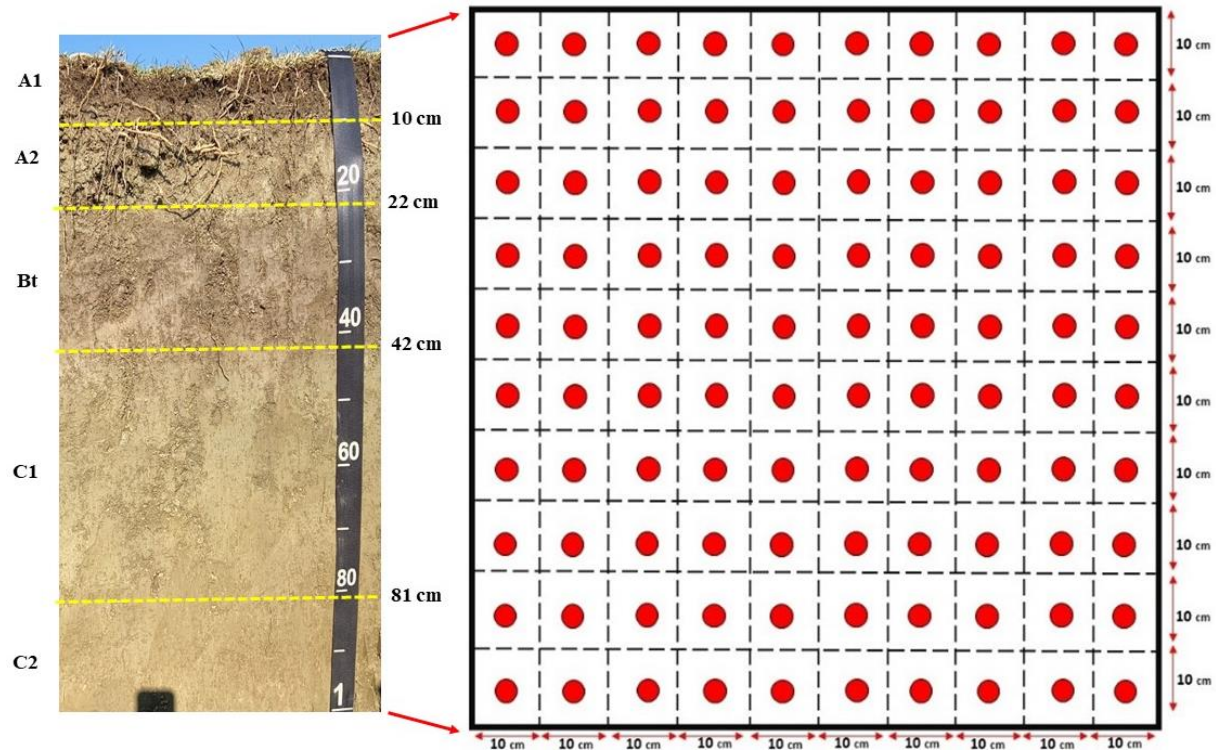

Figure 1. Soil profile wall and soil samples taken with the grid sampling (10 cm * 10 cm) model on the soil profile Wall (red points symbolize soil samples taken from each grid spacing).
*Şekil 1. Toprak profili duvarında grid örnekleme (10 cm * 10 cm) modeli ile alınan toprak örnekleri (kırmızı noktalar her bir grid aralığından alınan toprak örneklerini sembolize etmektedir).*

Digital Color Parameters Measurements and Organic Matter Analysis

Under field conditions, digital color parameters CIE L_1 a_1 b_1 of every raster cell measured into full contact with the soil surface on the soil profile wall using 3NH brand NR10QC model chromometer (3NH, Shenzhen, China). Under lab conditions, air-dried soil samples were measured for CIE L_2 a_2 b_2 digital color parameters by 3NH brand NR10QC model chromometer (3NH, Shenzhen, China) in petri dishes. Digital color parameters based on the field (CIE L_1 a_1 b_1) and lab (CIE L_2 a_2 b_2) were recorded by averaging 5 measurements for each soil sample. The chromameter was calibrated for every 10 soil sample measurements. Soil electrical conductivity (EC) and soil reaction (pH) were measured in a 1:1 soil water suspension (Soil Survey Staff, 2014), soil organic matter (OM) was measured by the Walkley-Black method (Black, 1965). $CaCO_3$ content was measured using a Scheibler calcimeter. Soil texture was determined using the hydrometer method (Bouyoucos, 1953).

Regression Tree Algorithms and Statistical Analysis

Data mining algorithms which are the Chi-squared Automatic Interacion Detector (CHAID), The Exhausive Chi-squared Automatic Interacion Detector (Ex-CHAID), and classification and regression tree (CART) algorithms were used as a regression tree model. In this study, regression tree algorithms were used to obtain model evaluation criteria that best explain the prediction of OM. In order to create the tree structures used, the minimum soil numbers in the parent and child node were taken as 10:5 and cross-validation 10. While OM was use as a

328

dependent variable, numerical color parameters (CIE Lab) determined under field and laboratory conditions were used as independent variables.

Regression tree algorithms are data mining algorithms created by dividing the independent variables into subgroups in terms of dependent variables and presented in the form of a tree structure. The stages of creating regression tree algorithms are combining steps, dividing, and stopping (Altay *et al.,* 2021). Regression trees do not contain any fragmentation at the first stage and only take place in the dependent variable tree structure. This phenomenon where the dependent variable is taken as the basis is the root node. While the root node is divided into two in the CART algorithm, the CHAID and Ex-CHAID algorithms can be divided into two or more parts. The first split of the root node in regression tree algorithms is called the parent node. The basic principle in the creation of the regression tree is to divide it into two child nodes, repetitively, to ensure maximum homogeneity in the response variable. The main purpose of the creation of the tree is that if homogeneity is achieved as much as possible in any child node that is created repetitively in the response variables, the fragmentation process ends in these nodes. This node is called terminal or end node (Oruçoğlu, 2011). In this process; By testing all independent variables included in the tree algorithm, it determines the cut-off value of the independent variable in the new node to ensure the highest homogeneity (Akşahan and Keskin, 2015).

The formation of CHAID, Ex-CHAID and CART algorithms and the linear relationship between color values and OM, Pearson correlation coefficients were made in IBM SPSS 23 package program (IBM Corp. Released, 2015). In the calculation of algorithm evaluation criteria, the "ehaGoF" package was used in the R Studio program (R Core Team, 2020; Eyduran, 2020).

Criteria of Algorithm Performance of Regression Trees

In order to determine the efficiency of regression tree algorithms, Pearson correlation coefficient (PC) (Eq.1) between actual and predicted OM values, Akaike's Information Criterion (AIC) (Eq.2), Corrected Akaike's Information Criterion (CAIC) (Eq.3), Root Mean Square Error (RMSE) (Eq.4), Mean Error (ME) (Eq.5), Standard Deviation Ratio (SDR) (Eq.6), Coefficient of Determination (Rsq) (Eq.7), and Adjusted Coefficient of Determination (ARsq) (Eq.8) were used as algorithm evaluation criteria (Aertsen *et al.,* 2010). These calculations related criteria are given in Equations 1-8.

$$PC = r_{yi,yip} = \frac{Cov(yi,yip)}{S_{yi}\,S_{yip}} \tag{1}$$

$$AIC = n.\ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2\right] + 2k \tag{2}$$

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2} \tag{4}$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2 \tag{5}$$

$$SD_{ratio} = \frac{S_m}{S_d} \tag{6}$$

$$r_{yi,yip}^2 = \frac{Cov(yi,yip)}{S_{yi}\,S_{yip}} \tag{7}$$

$$R_{Adjusted}^2 = 1 - \frac{\left(1-r_{yi,yip}^2\right)(n-1)}{n-p-1} \tag{8}$$

Spatial Distribution of Digital Color Parameters and OM

The distribution of OM and lab-based and field-based dijital color parameters were tested using Kolmogorov-Smirnov test (P>0.01). Simple, ordinary, and universal kriging interpolations methods were used to predict the spatial distributions of CIE L_1 a_1 b_1, CIE L_2 a_2 b_2, and OM through ArcGIS 10.5 (ESRI, USA). Root mean square error (RMSE) was calculated to assess and indicate interpolations model performance for the most accurate kriging interpolation models. Root mean square error (RMSE) (Eq.9) was calculated to assess and indicate interpolation model performance for the most accurate kriging interpolation models. Thus, the lowest RMSE represents the most accurate interpolation model. Zi, Z, and n are indicated predicted value, observed value, and a number of observations, respectively. Since ordinary kriging interpolation methods have the lowest RMSE, all distribution maps were produced by the kriging interpolation method which has commonly used interpolations method (Alaboz *et al.,* 2020, Altunbaş *et al.,* 2020; Dengiz, 2020; Sönmez *et al.,* 2020; Şimşek *et al.,* 2020).

$$RMSE = \sqrt{\frac{\sum (Zi - Z)^2}{n}} \tag{9}$$

RESULTS AND DISCUSSION

Some Pysical and Chemical Properties of Soil Profile

Soil profile were described as A1(0-10 cm), A2 (10-22 cm), Bt (22-42 cm), C1 (42-81 cm), and C2 (81-100 cm) and some soil pysical and chemical properties showed in Table 1. The EC, pH, $CaCO_3$, silt content clearly increased with depth, whereas OM and sand content clearly dicreased with depth. In particular, the pH and EC of soil profile range from 8.29 to 8.64 with moderate alkalinity and range from 2.37 to 5.38 dS m^{-1} with salinity risks. Moreover, the clay accumulation was observed between 22 and 42 cm as Bt horizon.

Table 1. Some pysical and chemical properties of soil profile.
Çizelge 1. Toprak profilinin bazı fiziksel ve kimyasal özellikleri.

Horizons	Depth	EC (dS m^{-1})	pH	OM (%)	CaCO3 (%)	Sand (%)	Silt (%)	Clay (%)	Texture
A1	0-10	3.02	8.29	7.74	7.84	21.62	36.06	42.32	CL
A2	10-22	4.09	8.45	2.66	9.51	23.62	38.06	38.32	CL
Bt	22-42	4.44	8.64	1.69	13.56	6.06	36.44	57.50	C
C1	42-81	5.38	8.61	0.47	16.11	10.42	47.62	41.96	SiC
C2	81-100	2.37	8.40	0.72	18.02	14.42	37.12	37.12	SiCL

Soil Color and Correlation Analysis

Descriptive statistics and distribution of field-based CIE L_1 a_1 b_1, lab-based CIE L_2 a_2 b_2 digital color parameters, and OM showed by histograms in Table 1 and Fig. 2, respectively. Field-based CIE L_1 a_1 b_1 b_2 digital color parameters ranged between 12.25-42-63, 2.52-7.66, and 3.98-14.92, respectively, whereas lab-based CIE L_2 a_2 b_2 digital color parameters ranged between 34.76-49.3, 2.82-5.01, and 11.57-16.87, respectively. OM ranged between 0.01-10.54% with a mean of 1.71%. Therefore, the highest coefficient variation (CV) was observed for OM, indicating a high range (131.41%) (Table 2).

Table 2. Some descriptive statistics of all variables.
Çizelge 2. Tüm değişkenlerin bazı tanıtıcı istatistik değerleri.

Variables	n	Minimum	Maximum	Mean±Std	Std Deviation	Sweekness	Kurtosis	CV (%)
L_1	100	12.25	42.63	34.72±0.55	5.51	-1.77	3.53	15.86
a_1	100	2.52	7.66	3.97±0.11	1.08	1.36	1.50	27.21
b_1	100	3.98	14.92	11.92±0.21	2.08	-1.47	2.27	17.49
L_2	100	34.76	49.31	45.30±0.36	3.56	-1.01	0.05	7.86
a_2	100	2.82	5.01	3.89±0.05	0.45	0.12	-0.45	11.51
b_2	100	11.57	16.84	14.27±0.13	1.33	-0.16	-0.91	9.32
OM(%)	100	0.01	10.54	1.71±0.22	2.24	2.15	4.10	131.41

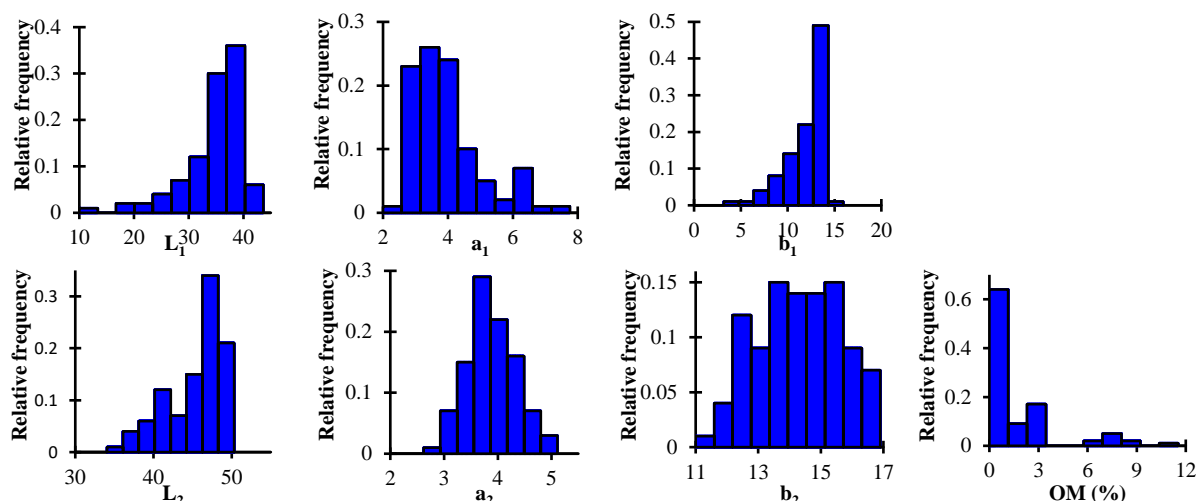CV; coeficiant variations, n; number of soil samples.

Figure 2. Digital color parameters and histogram graphics of OM obtained under field (L_1 a_1 b_1) and laboratory (L_2 a_2 b_2) conditions.

Şekil 2. Arazi (L_1 a_1 b_1) ve laboratuvar (L_2 a_2 b_2) koşullarında elde edilen Sayısal renk parametreleri ve OM'nin histogram grafikleri.

The Pearson correlation coefficients (r) between field-based CIE L_1 a_1 b_1 and lab-based CIE L_2 a_2 b_2 digital color parameters and OM were calculated in Table 3. As expected, lab-based L_2 was highly negative correlated with OM (r = -0.88). Field-based a_1 was highly positively correlated with OM (r = 0.71), while field-based b_1 was highly negatively correlated with OM (r = -0.62). Kirillova *et al.* (2015), Budak *et al.* (2018), and Gözükara *et al.* (2021a) reported that the OM was quite weakly correlation with L digital color parameter. Theoretically, the L digital color parameter value defines the brightness (white =100) or darkness (black = 0) of the color. Therefore, the higher the L value indicates the brighter the color, and the lower L value indicate the darker color. As a result of this theoretical knowledge, a fairly high correlation between L value and OM was expected. Vodyanitski and Kirillova (2016) pointed out that, theoretically, it is possible to determine the organic matter content with the L value which is an indicator of brightness and darkness, but in practice, the low amount of organic matter in the soil sample may cause the correlation to be weakened. In addition, Gözükara *et al.* (2021a) reported that the weak correlation between OM and L was affected by the low decomposition level of OM and the light-colored root fragments of its origin. The study area soils developed under the use of pasture which was close to the surface had dark soil color and a high decomposition level of OM. Therefore, as expected OM was highly correlate with lab-based L digital color parameters.

Table 3. Pearson correlation coefficients between OM and digital color parameters of field-based (CIE L_1, a_1 ve b_1) and lab-based (CIE L_2, a_2 ve b_2).

Tablo 3. Arazi (L_1, a_1 ve b_1) ve laboratuvar (L_2, a_2 ve b_2) koşullarında elde edilen sayısal renk parametreleri ile organik madde arasındaki pearson korelasyon katsayıları.

Soil Property	Field-based			Lab-based		
	L_1	a_1	b_1	L_2	a_2	b_2
OM	-0.64**	0.71**	-0.62**	-0.88**	0.51**	-0.26**

**P<0.01.

Distributions of Digital Color Parameters and OM on Soil Profile Wall

Distribution maps of field-based CIE L_1 a_1 b_1 and lab-based CIE L_2 a_2 b_2 digital color parameters and OM were presented in Fig. 3. In general, field-based and lab-based L_1, L_2, b_1 and b_2 digital color values considerably increase with increasing depth from the surface to the subsoil of the soil profile wall, while a_1 and a_2 digital color values firstly tend to decrease and then increase with the increasing of depth from the soil surface. Measurement of field-based digital color parameters are affected by many factors, mainly the OM, moisture content and structure of the soil depending on the texture. The distribution of field-based CIE L_1 a_1 b_1 digital color parameters in soil profile wall were more heterogeneous than distribution of field-based CIE L_2 a_2 b_2 digital color parameters. Therefore, these results supported researchers. On the contrary, lab-based CIE L_1 a_1 b_1 digital color parameters in the soil profile wall showed a more homogeneous distribution representing different horizon layers horizontally.

In the soil profile wall, OM was very highly and fairly constant in the upper 30 cm but decreased sharply with increasing depth. The fact that the soil profile wall was developed under the use of pasture caused the amount

of OM to be quite high, especially in the surface soils. Koç and İleri, (2016) reported that the amount of OM in Muttalıp pasture was approximately 16%. Our findings are consistent with the findings of the researchers.
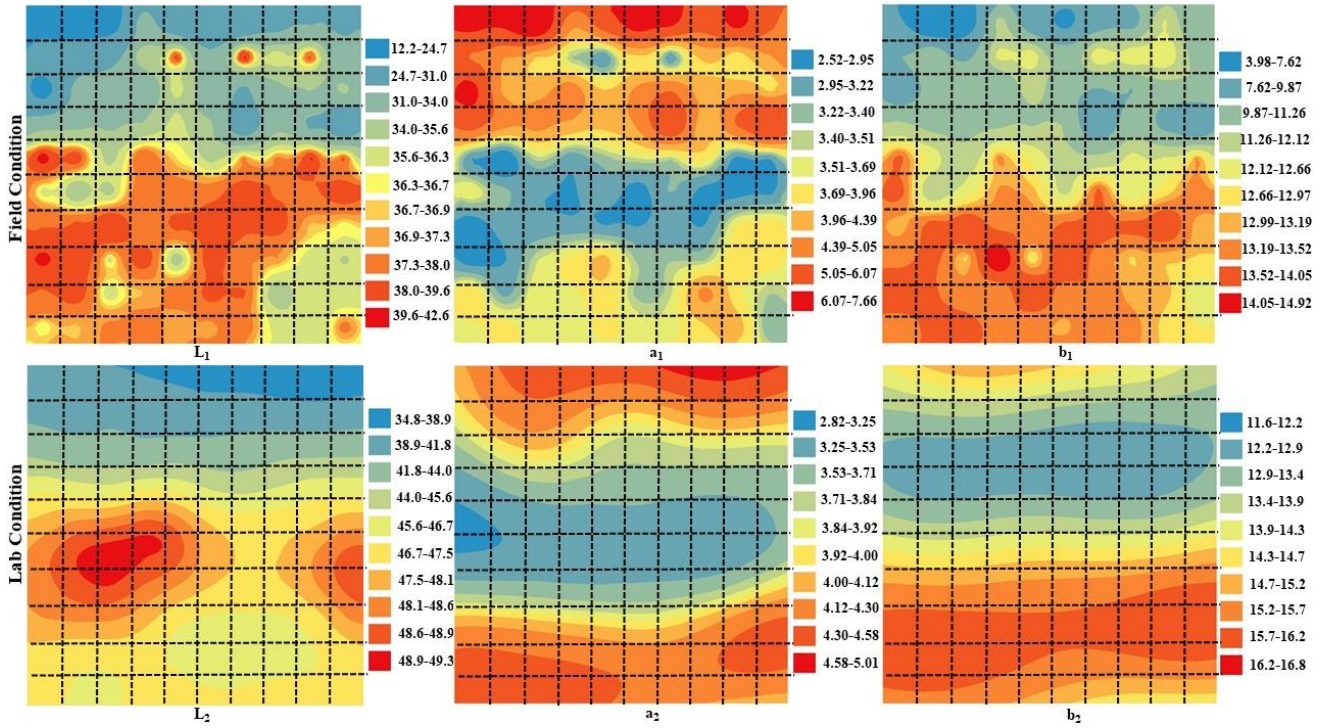

Figure 3. Geostatistical distribution of digital color parameters in soil profile wall obtained under field (L_1 a_1 b_1) and laboratory (L_2 a_2 b_2) conditions.
Şekil 3. Toprak profil duvarında arazi (L_1 a_1 b_1) ve laboratuvar (L_2 a_2 b_2) koşullarında elde edilen sayısal renk parametrelerinin jeoistatistiksel dağılımı.
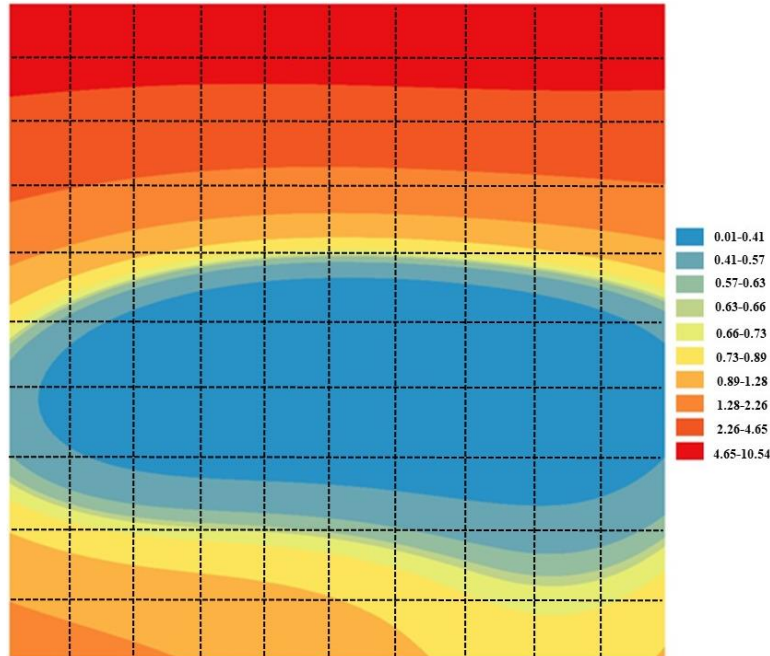

Figure 4. Geostatistical distribution of OM in soil profile wall.
Şekil 4. Toprak profil duvarında OM'nin jeoistatistiksel dağılımı.

Prediction of OM Using Regression Trees Algorithms

According to the maximum depths of the regression trees algorithms to predict of OM created, When used the field-based CIE L_1 a_1 b_1 digital color parameters, the CHAD, Ex-CHAID, and CART algorithms had 1, 1, and 2 parent node, respectively, when used the field-based CIE L_2 a_2 b_2 digital color parameters, the CHAD, Ex-CHAID, and CART algorithms had 2, 1, and 4 parent node, respectively. Field-based a_1 ($5.17 < a_1 \leq 5.17$) and L_1 ($34.44 < L_1 \leq 34.44$) were determined as strong indicators (cut points) using the CART algorithm for predicting OM, while lab-

based L_2 ($39.96 < L_2 \leq 39.96$), a_2 ($3.51 < a_2 \leq 3.51$), and b_2 ($14.35 < b_2 \leq 14.35$) were determined as strong indicators (cut points) using CART algorithm for prediction OM (Figure 5). In the prediction with the CART algorithm of OM with the color characters obtained from the field-based, it predicted the OM amount as 1.10 when $a1 \leq 5.17$ values were taken, and 5.81 when $a1$ values greater than 5.17. On field-based, OM was predicted as 2.14 for $a1 \leq 5.17$ and $L_1 \leq 34.44$, while 0.76 for $a_1 \leq 5.17$ and $34.44 < L_1$. CART algorithm in estimating the amount of OM based on field-based, the highest amount of OM was predicted as 5.81 in case of $5.17 < a_1$. The L_2 was found to be the most important independent variable in the lab-based prediction of OM amount. If the L_2 value was less than or equal to 39.96, the amount of OM was 7.27, and if it was greater than 39.96, the OM was predicted as 1.02. If the L_2 value was less than and equal to 44.34, the amount of OM was predicted as 2.57, and if it was greater than 44.34, the OM was predicted as 0.57. In the lab, OM was predicted to be 0.36 for $39.96 < L_2$ and $a_2 \leq 3.51$ and 0.66 for $39.96 < L_2$ and $3.51 < a_2$. In the laboratory, OM was predicted to be 0.51 for $39.96 < L_2$, $a_2 \leq 3.51$ and $b_2 \leq 14.35$, and 0.13 for $39.96 < L_2$, $a_2 \leq 3.51$ and $14.35 < b_2$. In lab-based CART algorithm prediction, the highest amount of OM was predicted as 7.27.

The highest R^2, PC, and the lowest RMSE, SDR, ME were used to determine for the best prediction algoritm. Using field-based CIE L_1 a_1 b_1 digital color parameters for predication performance was obtained with $R^2 = 0.45$ and RMSE = 1.65 in the CHAID algorithm, $R^2 = 0.51$ and RMSE = 1.54 in the Ex-CHAID algorithm, and $R^2 = 0.56$ and RMSE = 1.48 in the CART algorithm, whereas using lab-based CIE L_2 a_2 b_2 digital color parameters for predication performance was obtained with $R^2 = 0.85$ and RMSE = 0.85 in the CHAID algorithm, $R^2 = 0.85$ and RMSE =0.86 in the Ex-CHAID algorithm, and $R^2 = 0.89$ and RMSE = 0.72 in the CART algorithm (Table 4). According to the results of the research, lab-based digital color parameters had a more predictive performance for predicting OM than field-based digital color parameters. In addition, CART algorithm had more successful prediction performance for OM prediction with lab-based digital color parameters than other regression tree algorithm (Table 4). The regression trees of the CART algorithm created with field-based and lab-based digital color parameters showed in Figure 5. Many researchers have used linear and parabolic regression algorithms to predict the OM using digital color parameters (Günal *et al.,* 2008; Moritsuka *et al.,* 2014; Budak *et al.,* 2018; Gözükara *et al.,* 2021a). In addition, researchers were evaluated between only L digital color parameter and OM in linear regression algorithms. Therefore, with these algorithms, the researchers obtained very low prediction performance for OM ($R^2 = 0.40$-0.65, $R^2 = 0.18$, $R^2 = 0.02$, $R^2 = 0.02$, and $R^2 = 0.03$, respectively). Furthermore, regression tree algorithms have high prediction performance than linear and parabolic regression algorithms, because the regression tree algorithm uses digital color parameters as tree algorithms.

Table 4. Some model performance results of CHAID, Ex-CHAID and CART algorithms.
Çizelge 4. CHAID, Ex-CHAID ve CART algoritmalarının bazı model performans sonuçları.

Criteria of Model Performance	Field-based			Lab-based		
	CHAID	Ex-CHAID	CART	CHAID	Ex-CHAID	CART
RMSE	1.645	1.544	1.475	0.852	0.862	0.723
SDR	0.737	0.691	0.660	0.381	0.386	0.324
PC	0.676	0.722	0.751	0.924	0.923	0.946
ME	0.003	0.002	0.002	0.000	0.002	0.001
Rsq	0.458	0.522	0.564	0.855	0.851	0.895
ARsq	0.447	0.512	0.555	0.852	0.848	0.893
AIC	103.519	90.884	81.673	28.112	25.719	60.787
CAIC	103.642	91.007	81.797	27.988	25.595	60.663

CHAID; The Chi-squared Automatic Interacion Detector, Ex-CHAID; The Exhausive Chi-squared Automatic Interacion Detector, CART; Classification and regression tree
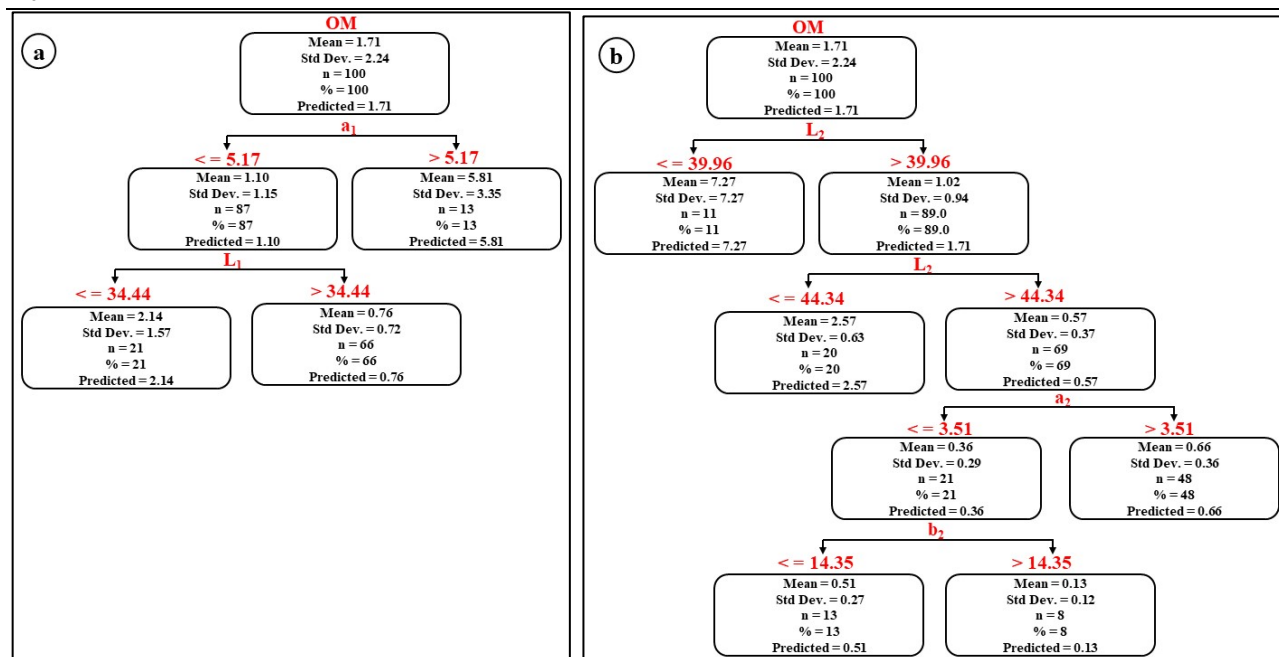
Figure 5. Prediction of the amount of OM using the CART regression tree algorithm of digital color parameters obtained under field (a) and laboratory (b) conditions.

Şekil 5. Arazi (a) ve laboratuvar (b) koşullarında elde edilen sayısal renk parametrelerinin CART regresyon ağacı algoritması kullanılarak OM miktarının tahmin edilmesi.

CONCLUSION

In this current study, OM was predicted by field-based CIE L_1 a_1 b_1 and lab-based CIE L_2 a_2 b_2 digital color parameters using different regression tree algorithms. CART regression tree algorithm can be successfully used to predict organic matter with field-based CIE L_1 a_1 b_1 and lab-based CIE L_2 a_2 b_2 digital color parameters compared to regression tree algorithms. In particular, lab-based CIE L_2 a_2 b_2 digital color parameters have higher prediction performance ($R^2 = 0.89$) using CART regression tree algorithm for predicting OM. Field-based, a_1 ($5.17 < a_1 <= 5.17$) and L_1 ($34.44 < L_1 <= 34.44$) values are absolute indicators for predicting OM, lab-based, combining L_2 ($39.96 < L_2 <= 39.96$).), a_2 ($3.51 < a_2 <= 3.51$) and b_2 ($14.35 < b_2 <= 14.35$) values greatly improved algorithm performance for predicting of OM. As a result of the research, CIE L_2 a_2 b_2 digital color parameters obtained under laboratory conditions can be used to predict OM using the CART regression tree algorithm in a fast, economical, reliable, and highly accurate. These algorithms should be studied and tested at different parent materials and soil types to develop of prediction performance.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this article.

DECLARATION OF AUTHOR CONTRIBUTION

Dr. Gafur GÖZÜKARA designed the field experiments, performed the laboratory analyses, and collected the data. Dr. Yasin ALTAY did all statistics evaluations. All authors read and approved the manuscript.

REFERENCES

Acar, M., Celik, I., & Günal, H. (2018). Effects of long-term tillage systems on aggregate-associated organic carbon in the eastern Mediterranean region of Turkey. *Eurasian Journal of Soil Science*, *7*(1), 51-58.

Aertsen, W., Kint, V., Orshoven, J., Özkan, K., & Muys, B. (2010). Comparison and ranking of different modeling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, *221*, 1119-1130.

Aksahan, R., & Keskin, I. (2015). Determination of the some body measurements effecting fattening final live weight of cattle by the regression tree analysis. *Selçuk Journal of Agriculture and Food Science*, *2*, 53-59.

Aktaş, T., & Yüksel, O. (2020). Effects of vermicompost on aggregate stability, bulk density and some chemical characteristics of soils with different textures. *Journal of Tekirdag Agricultural Faculty*, *17*(1), 1-11.

Alaboz, P., Demir, S., & Dengiz, O. (2020). Determination of spatial distribution of soil moisture constant using different interpolation model case study, Isparta Atabey Plain. *Journal of Tekirdag Agricultural Faculty*, *17*(3), 432-443.

Allison, L. E., Moodie, C. D. (1965). Carbonate, agronomy monograph, methods of soil analysis. Part 2. In: Chemical and Microbiological properties, Agronomy. 9.2. American Society of Agronomy, Wisconsin, pp. 1379-1396.

Altay, Y., Boztepe, S. Eyduran, E., Keskin, İ., Tariq, M. M., Bukhari, F. A. & Ali, I. (2021). Description of factors affecting wool fineness in Karacabey Merino Sheep using Chaid and Mars Algorithms. *Pakistan Journal of Zoology*, *53*(2), 691-697.

Altunbaş, S., Demirel, B. Ç., Gözükara, G., & Erol. S. (2020). Determination of land capability classes of some soils developing on alluvial lands. *International Journal of Agriculture and Wildlife Science*, *6*(3), 638-646.

Barret, L. R. (2002). Spectrophotometric color measurement in situ in well drained sandy soils. *Geoderma*, *108*, 49-77.

Black, C. A. (1965). *Methods of Soil Analysis Part 2*, Amer. Society of Agronomy Inc., Publisher Madison, Wisconsin, U.S.A.

Bouyoucos, G. J. (1953). An improved type of soil hdyrometer. *Soil Science, 76*, 377-378.

Budak, M., Günal, H., Süer, M., & Akbaş, F. (2018). Determination of some physical and chemical characteristics of soil properties from digital color parameters. *Harran Journal of Agriculture and Food Science*, *22*(3), 376-389.

Çelik, İ., Günal, H., Acar, M., Acir, N., Barut, Z.B., & Budak, M. (2020). Evaluating the long-term effects of tillage systems om soil structural quality using visual asessment and classical methods. *Soil Use and Management*, *36*, 223-239.

Dengiz, O. (2020). Soil quality index for paddy fields based on standard scoring functions and weight allocation method. *Archives of Agronomy and Soil Science*, *66*(3), 301-315.

DMİ (Devlet Meteoroloji İşleri), 2017. Eskişehir ili iklim verileri. Devlet Meteoroloji İşleri Genel Müdürlüğü, Ankara.

Doi, R., Wachrinrat, C., Teejuntuk, S., Sakurai, K., & Sahunalu, P. (2010). Semiquantitative color profiling of soils over a land degradation gradient in Sakaerat, Thailand. *Environmenal Monitoring Assessment*, *170,* 301-309.

Eyduran, E. (2019). ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0. https://CRAN.R-project.org/package=ehaGoF.

Fan, Z., Herrick, J. E., Saltzman, R., Matteis, C., Yudina, A., Nocella, N., Crawford, E., Parker, R., & Van Zee, J. (2017). Measurement of soil color: a comparison between smartphone camera and the munsell color charts. *Soil Science Society of America Journal*, *81*(5), 1139-1146.

Fang, X. M., Ono, Y., Fukusawa, H., Pan, B. T., Li, J. J., Guan, D. H., Oi, K., Tsukamoto, S., Torii, M., & Mishima, T. (1999). Asian summer monsoon instability during the past 60.000 years: magnetic susceptibility and pedogenic evidence from the western Chinese Loess Plateau. *Earth and Planetary Science Letters*, *168*, 219-232.

Gozukara, G., Zhang, Y., & Hartemink, A. E. (2021b). Using vis-NIR and pXRF data to distinguish soil parent materials – an example using 136 pedons from Wisconsin, USA. *Geoderma*, *396,* 115091.

Gözükara, G., Altunbaş, S., & Sarı, M. (2019). The effect of spatial change on the properties of soil formed on alluvial fans. *Mediterranean Agricultural Sciences*, *32*(3), 425-435.

Gözükara, G., Altunbaş, S., & Sarı, M. (2020a). Effects of temporal and spatial changes on formation, development and morphology of soil in different physiographs. *Journal of Agriculture Faculty of Ege University*, *57*(2), 277-278.

Gözükara, G., Altunbaş, S., & Sarı, M. (2020b). Temporal and spatial changes in old lake bottom effect on soil formation, development and morphology. *Harran Journal of Agriculture and Food Science*, *24*(1), 96- 110.

Gözükara, G., Demirel, B. Ç., & Altunbaş, S. (2021a). Effect of soil horizons on the relationship between digital color paremeters and soil properties. *Mediterranean Agricultural Sciences*, *34*(1).

Grzesiak, W., & Zaborski, D. (2012). *Examples of the Use of Data Mining Methods in Animal Breeding*, Data Mining Applications in Engineering and Medicine, Adem Karahoca, IntechOpen.

Günal, H., Erşahin, S., Yetgin, B., & Kutlu, B. (2008). Use of chroma-meter measured color parameters in estimating color related soil variables. *Communications in Soil Science and Plant Analysis*, *39*(6), 726-740.

Hartemink, A. E., & Minasny, B. (2014). Towards digital soil morphometrics. *Geoderma*, *230-231*, 305-317.

IBM Corp. Released (2015). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Koç, A., & İleri, O. (2016). Comparison of Cattle and Sheep Grazed Sub-Irrigated Rangelands Vegetation in Eskişehir Plain. *Journal of Field Crops Centrel Research Institute, 25*, 179-184.

Kirillova, N. P., Vodyanitskii, Y. N., & Sileva, T. M. (2015). Conversion of soil color parameters from the Munsell system to the CIE-L* a* b* system. *Eurasian Soil Science, 48*(5), 468-475.

McBratney, A. B., Stockmann, U., Angers, D. A., Minasny, B., & Field, D. J. (2014). *Challenges for soil organic carbon research.* In: Hartemink, A., McSweeney, K. (Eds.), Soil Carbon. Progress in Soil Science. Springer, Cham.

Moritsuka, N., Matsuoka, K., Katsura, K., Sano, S., & Yanai, J. (2014). Soil color analysis for statistically estimating total carbon. total nitrogen and active iron contents in Japanese agricultural soils. *Soil Science and Plant Nutrition*, *60*(4), 475-485.

Nishiyama K, Kimura, T., Isono, Y., & Inoue, Y. (2011). Color measurements of rocks and soils using colorimeters. *Journal of the Japan Society of Engineering Geology, 52,* 62–71 (in Japanese).

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Oruçoğlu, O. (2011). *Determination of environmental factors affecting 305-day milk yield of holstein cows by regression tree method.* Master Thesis, Süleyman Demirel University, Institute of Science, Isparta.

Post, D. F., Levine, S. J., Bryant, R. B. Mays, M. D., Batchily, A. K., Escadafal, R., & Huete, A.R. (1993). *Correlations between field and laboratory measurements of soil color*. In J. M. Bigham, & E. J. Ciolkosz (Eds.). *Soil Color*. Soil Science Society of America. Madison.

Rice, T. D., Nickerson, D., O'Neal, A. M., & Thorp, J. (1941). Preliminary color-standards and color names for soil. *Miscellaneous Publication, 425,* 1–12.

Sawada, K, Wakimoto, T., Hata, N., Taguchi, S., Tanaka, S., Tafu, M., & Kuramitz, H. (2013). The evaluation of forest fire severity and effect on soil organic matter based on the L*, a*, b* color reading system. *Analytical Methods, 5,* 2660–2665.

Schulze, D. G., Nagel, J. L., Van Scoyoc, G. E., Henderson, T. L., Baumgardner, M. F., & Stott, D. E. (1993). Significance of organic matter in determining soil colors. In J. M. Bigham, & E. J. Ciolkosz (Eds.). *Soil Color*. Soil Science Society of America. Madison.

Shen, Z. X., Cao, J. J. Zhang, X. Y., Arimoto, R., Ji, J. F., Balsam, W. L., Wang, Y. Q., Zhang, R. J., & Li, X. X. (2006). Spectroscopic analysis of iron-oxide minerals in aerosol particles from northern China. *Science of the Total Environment*, *367*, 899-907.

Simonson, R. W. (1993). *Soil color standards and terms for field use history of their development*. In J. M. Bigham, & E. J. Ciolkosz (Eds.). *Soil Color*. Soil Science Society of America. Madison.

Soil Survey Staff. (2010). *Soil Taxonomy*. 11th ed. USDA National Resources Conservation Services. Washington DC.

Soil Survey Staff. (2014). *Keys to Soil Taxonomy*. Twelfth Edition Edition, United States Department of Agriculture, Natural Resources Conservation Service, Washington DC.

Sönmez, N. K., Sönmez, S., Türkkan, H. R., Altın, R., & Çoşlu, M. (2020). Spatial variability analysis of electrical conductivity (ec) of irrigation water used in agricultural production: an example of Dalaman-Muğla. *The Journal of Faculty of Agriculture*, *15*(1), 12-26.

Şimşek, O., Altunbaş, S., Demirel, B. Ç., & Gözükara, G. (2020). Land evaluation studies on different soils developing on alluvial physiographies. *Mediterranean Agricultural Sciences*, *33*(1), 129-135.

Thwaites, R. (2002). *Color*. In: Lal. R. (ed.) Encyclopedia of Soil Science.. Marcel Dekkers. Inc

Torrent, J., Schwertmann, U., & Schulze, D. G. 1980. Iron oxide mineralogy of some soils of two river terrace sequences in Spain. *Geoderma*, *23,* 191-208.

Torrent, J., & Barrón, V. (1993). *Laboratory measurement of soil color: Theory and practice*. In J. M. Bigham, & E. J. Ciolkosz (Eds.). *Soil Color*. Soil Science Society of America. Madison.

Vodyanitskii, Y. N., & Kirillova, N. P. (2016). Application of the CIE-L* a* b* system to characterize soil color. *Eurasian Soil Science*, *49*(11), 1259-1268.

Viscarra Rossel, R. A., Minasny, B., Roudier, P., & McBratney, A. B. (2006). Colour space models for soil science. *Geoderma, 133,* 320-337.

Wu, C. W., Yang, Y., & Xia, J. X. (2017). A simple digital imaging method for estimating blacksoil organic matter under visible spectrum. *Archives Agronomy Soil Sciences*, *63*(10), 1346-1354.

Yılmaz, E., Çanakçı, M., Topakçı, M., Sönmez, S., Ağsaran, B., Alagöz, Z., Çıtak, S., & Uras, D. S. (2019). Effect of vineyard pruning residue application on soil aggregate formation, aggregate stability and carbon content in different aggregate sizes. *Catena*, *183*, 104219.