

<https://doi.org/10.37501/soilsa/186454>

Evaluation of spectral data based soil organic carbon content estimation models in VIS-NIR

Attila Nagy¹, Andrea Szabó^{1*}, Diana Quintín Escobar², János Tamás¹

¹Institute of Water and Environmental Management, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 138, Environment str., 4032 Debrecen, Hungary

²Proforest – Latin America Office, Cl 11 #100-121 Of. 203, Campestre Towers, Cali, Colombia

* Corresponding author: Andrea Szabó, szabo.andrea@agr.unideb.hu, ORCID iD: <https://orcid.org/0000-0002-6782-6793>

Abstract

Received: 2023-07-27

Accepted: 2024-03-25

Published online: 2024-03-25

Associated editor: B. Glina

Keywords:

Soil organic carbon

VIS-NIR

PCA

Spectral analysis

The precision agriculture is a key management tool for food security, requiring rapid and cost-effective field assessment methods to support agricultural decision-making. One solution is proximal sensors that collect spectral characteristics of soils. The objective of this study was to establish models for predicting soil organic carbon (SOC) content using VIS-NIR (400–2500 nm) spectroscopy as an alternative to destructive SOC measurement methods. For the modelling, 90 soil samples were collected from 0–0.20 m depth representing the most common soil types in the Northern Great Plain (Hungary), 60 soil samples were used for calibration, and 30 soil samples for model validation. The soil samples were evaluated both chemically and physically. The estimation models were fitted based on spectral indices, and the spectral bands used for indexing were identified by principal component analysis (PCA) of reflectance. Based on the PCA results, four SOC models were set up with moderately good coefficients of determination ($R^2=0.47-0.61$). The results demonstrated that VIS-NIR spectroscopy (especially NIR) based organic carbon content estimation models are suitable for rapid estimation of soil SOC%. This can reduce sampling costs by optimizing the number of samples to be sent to the laboratory and by identifying heterogeneous patches in the study area.

1. Introduction

Properly functioning agricultural systems are needed to meet the food needs of a globally growing population (Boscaro et al., 2018). To feed a growing world population sustainably, the production of accurate, up-to-date soil health information is essential to support agricultural development, environmental sustainability, and to mitigate the effect of climate change. One of the most important factors affecting agricultural production is the soil organic matter (SOM) content. Depletion of SOM can lead to soil degradation with implications for sustainable agricultural development (Tang et al., 2006), so it is important to investigate and estimate its values quickly and non-invasively. Traditional methods for estimating soil organic carbon (SOC) content are destructive. Walkley and Black (1934) method, Walkley and Black modified with external heating, Kalembsa and Jenkinson (1973), Tyurin method, and modified Tyurin method (Ponomariova and Plotnikova, 1980) are the most used conventional methods (Kononova, 1966; Filcheva, 2002) for the evaluation of organic carbon, used to quantify humus and its components. Most of these SOC monitoring techniques need a lot of time and resources since they rely heavily on rigorous sampling (Goidts and van Wesemael, 2007). Inventory and evaluation of the spatial and temporal changes in

SOC would therefore be considerably facilitated by the development of a reliable, accurate, and affordable technique to estimate SOC content. Visible and near-infrared reflectance spectroscopy (Vis-NIRS), in contrast to traditional testing methods, has developed into a rapid, non-destructive, and economical tool for determining a variety of soil parameters, including SOC content in the previous years (Ben-Dor et al., 1995; Stenberg et al., 2010; Nagy et al., 2007; Nagy and Tamás, 2009). Rapid and valuable SOC estimation methods are essential for precision agriculture to explore the spatial heterogeneity of SOC, which is a determinant factor of nutrient supply. The reflectance values can be used to predict the physical and chemical properties of soils using multivariate equations (Cecillon et al., 2009). The spectral reflectance of soil is influenced by several factors, including soil moisture, aggregation, texture, and surface roughness, all of which interact with spectral formations in complex ways (Ben-Dor et al., 2008; Kuang and Mouazen, 2013; Piekarczyk et al., 2016). NIRS calibration consists of four main steps (Ge et al., 2007). The first is the measurement, during which the concentration of the targeted soil components is determined in the laboratory and their spectroscopic reflectance is obtained. The second is pre-treatment, where the spectroscopic reflectance is pre-processed. The third is calibration, during which a subset of the samples is used to develop regression

models, and the fourth is validation, during which the remaining samples are used to assess the validity of the regression models for the remaining samples. NIRS calibrations that have not been validated with independent samples or that have not been used in the calibration procedure are generally less reliable (Brunet et al., 2007; Nduwamungu et al., 2009) and over-optimistic (Russell, 2003; Siebielec et al., 2004; Butkute and Slepeliene, 2006; Brunet et al., 2007; Terhoeven-Urselmans, 2008). Soil samples should be pre-treated before calibration of the NIRS model to increase the homogeneity of the sample and thus reduce interference from variations in particle size distribution (Couteaux et al., 2003). The 560 nm, 750 nm, 760 nm, 1000 nm, 1100 nm, 1600 nm, 1400 nm, 1700 nm, 1800 nm, 1900 nm, 2000 nm, 2200 nm, and 2400 nm VIS-NIR bands are relevant for SOC estimation in several studies (Laamrani et al., 2019). Other investigations have also demonstrated a considerable effect of water, a highly essential soil element, in the VIS-NIR region in bands approximately 1400–1900 nm (Viscarra et al., 2006; Viscarra and McBratney, 2008). Regression methods used with NIRS in soil analyses include multiple regression analysis (MRA), stepwise multiple linear regression (SMLR), multivariate adaptive regression splines (MARS), radial basis function networks (RBFN), principal component analyses (PCA), and partial least-squares regression (PLSR) and wavelet analysis is also promising (Ge et al., 2007). Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR) are two of the most widely used methods for quantifying SOC (Viscarra et al., 2006). They are multivariate statistical methods for meshing high-dimensional correlated variables by decreasing data variance to fewer new bands. This will determine which new tracks contain the most relevant information for further modelling and analysis (Liu et al., 2020). On the other hand, using spectral indices for less robust SOC modelling is also promising. Good results were found for indices based on specific SOC-sensitive spectra ranges of Vis and NIR ($R^2=0.80-0.81$) which gives comparable results with results of PLSR (Bartholomeus et al., 2008). However, this study used previously reported specific SOC-sensitive spectra ranges for index development, which might result in a considerable error in extrapolation beyond the SOC range in the training data.

Another main uncertainty in SOC modelling is that soil properties are very heterogeneous and diverse, and there is a lack of national or even internationally uniform standards for

spectral measurement methods, databases, and even laboratory measurement methods for SOC determination (Zhou and Zhou, 2009). In sensing techniques including Vis-NIR, to overcome site-specific constraints and to be useful for monitoring spatiotemporal SOC dynamics, regional spectral measurements, spectral libraries, and estimation models are required for more accurate estimation of SOC content (Viscarra and McBratney, 2008). Therefore, closing this information gap was the primary goal of the current study providing regionally accurate non-invasive SOC monitoring methods with a less robust modelling approach combining PCA results and spectral indexing. The objectives of this study were (I) to set soil spectral database at a regional scale (II) to integrate the use of PCA into spectral reflectance-based indices for soil organic carbon quantification to set and validate SOC estimation models based on regionally collected soil spectral data (III) to evaluate the performance of SOC estimation models using VIS and NIR spectroscopy performing the best at regional scale. The results can serve spectral-based SOC estimation solutions specific to the soils of the most important agricultural region in the Carpathian basin.

2. Materials and methods

2.1. Study area

The samples of this study were taken in the Northern Great Plain (Hungary) (Fig. 1). The study region is in an international watershed, the Tisza river's lowlands and it is the most important agricultural region in the Carpathian basin. During the growing season, the average annual daily temperature in Hungary's plain locations of the Tisza watershed ranges between 10°C–17.5°C. Despite having a humid continental climate with an average precipitation of 495 mm, the Great Plains region is plagued by water management problems. Floods, excess water, and droughts occur frequently even in the same year, often during the same vegetation period (Nagy et al., 2021). The soils of the lowland areas are very heterogeneous according to the WRB (IUSS Working Group WRB, 2022) with the following soil groups: Chernozems, Arenosols, Vertisols, Regosols, Solonetz, and Solonchak (Michéli et al., 2006).



Fig. 1. Location of the study area

2.2. Measurement method and data processing

A total of 90 soil samples were collected from 0–20 m² depth, of which 60 soil samples were used for calibration and 30 soil samples were used to validate the models. Regosols (15%), Chernozems (35%), Vertisols (15%), Arenosols (15%), Solonetz, and Solonchak (20%) soil samples were collected to cover the most common soil types occurring in the study area. The sample sites were randomly defined based on the soil map of Hungary (Mezősi, 2016). Samples were collected from arable land before the vegetation period, one sample is the mean of 5 samples, randomly collected from each (90) sample site. Samples for validation are selected randomly from each soil type. The soil samples were prepared for laboratory measurements made in accordance with Hungarian standards (MSZ). The soil sample was dried in a drying oven at a maximum temperature of 40°C and ground. Before grinding plant, residues were removed from the samples (MSZ-08-0206-1:1978). The pH_{H₂O} and pH_{KCl} was determined potentiometrically, the soil to solution ratio was 1:2.5. (MSZ-08-0206-2:197). The saturation percentage was determined by weighing 100 g of air-dry soil in a porcelain mortar and slowly saturated with water. The amount of water consumed (ml) gives the saturation percentage (MSZ-08-0205:1978). The amount of water-soluble salt was determined based on the conductivity and temperature data (MSZ-08-0206-2:1978). The content of calcium carbonates was tested with a Scheibler calcimeter (MSZ-08 0206/2-78). SOC content was measured using 5% K₂Cr₂O₇ and concentrated H₂SO₄ (MSZ-08-0210:1977). For the determination of ammonium lactate (AL) soluble P₂O₅ was 35.6–798 mg/kg and AL-soluble K₂O content extracting solution consisting of lactic acid, acetic acid, and ammonium acetate with a pH 3.70 was used for extraction and the soil suspension is filtered off after shaking. The P, K, content was measured using an ICP-OES apparatus (MSZ-20135:1999). For the determination of potassium chloride (KCl)-soluble NO₃-N in soils extracting solution with 1 M KCl concentration was used. Then after shaking, the soil suspension was filtered, and the nitrite-nitrate nitrogen was determined by spectrometry (MSZ-20135:1999). For statistical evaluation, the R studio agricolae package was used (Mendiburu, 2019). The normal distribution of the results was analysed using the Shapiro-Wilk test, and the Kruskal-Wallis test was used to test the homogeneity of the samples. Spectral measurements were performed in a controlled laboratory environment. Spectral profiles (reflectance) between 400–2500 nm was measured using two separate laboratory-scale spectrometers. The AvaSpec 2048 spectrometer in the 400–1000 nm wavelength range with a spectral resolution of 0.6 nm and the AvaSpec-NIR spectrometer in the 1000–2500 nm wavelength range with a 5 nm spectral resolution. The measurement process with both spectrometers was similar. The soil samples were scanned in a special proprietary sampling box using an Ava-Light-HAL halogen light source at high intensity. The sampling box provides a controlled, completely, dark scanning environment for spectral measurements. An equal distance of 5 mm between the sample and the sensor was used to ensure homogeneous measurements in three replicates. Three replicates of each sample were taken and averaged into one spectrum per sample. For accurate estimation of the

percentage of SOC, proper pretreatment of the soil samples for spectral analyses is important (Brunet et al., 2007; Ludwig et al., 2002; Van Waes et al., 2005; Croft et al., 2012). The soil samples were dried at 105°C, sieved and ground to a size fraction below 0.25 mm to exclude coarse and medium fine sand, thus obtaining a more homogeneous sample for analysis (Guillou et al., 2015; Nawar et al., 2016; Terra et al., 2015; Shi et al., 2015). Furthermore, the smaller the soil particle, the larger the surface area and the higher the spectral reflectance (Banninger and Fluhrer, 2004), which can enhance the detectability of soil changes in reflectance.

2.3. SOC model building and performance assessment

Statistical analysis of the results was performed using SPSS software. PCA with varimax rotation was used for data compression outliers detection and to study the patterns and internal structure within the whole data to identify wavelengths with the highest variation based on factor weights to identify SOC-sensitive wavelengths. PCA analyses were performed on log 1/R, data, where R represents reflectance. Varimax is an orthogonal rotation method used to generate incredibly high or significantly lower factor loadings, making it much easier to assign individual objects to a single factor (Allen, 2017). Besides PCA, the SD of the spectral features was also studied to select those wavelengths where the variations are the highest indicating the potential variability of SOC. After selecting SOC-sensitive wavelengths based on PCA and SD results in the 400–2500 nm range, spectral indices were set using the most sensitive, and the least sensitive wavelengths. Spectral indexing is a common practice in the monitoring of soils in remote sensing studies (Nagy et al., 2014; Babaeian et al., 2019; Mohamed et al., 2020; Béni et al., 2021). The simple linear regression method was used to generate a model to estimate SOC based on 60 samples. The coefficient of determination (R²) was used to compare the strength of the regression models. For validation 30 independent samples were used. The root means squared error (RMSE) of squared error was used to measure the accuracy of the estimation models:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

and the normalized root means square error (NRMSE):

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}}{(\bar{y})} \quad (2)$$

The Nash-Sutcliffe efficiency (NSE) (1970) (3) was also used to assess the accuracy of the estimation of forecast:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (3)$$

where:

\hat{y}_i : predicted yield data;

y_i : the observed yield data;

\bar{y} : the average yield;

n: the number of samples used for validation

Values of NSE range between 1.0 (which is a perfect fit) and $-\infty$. In this case, an efficiency of less than zero means that the observed yield's average value would have been a better predictor than the model.

3. Results and discussion

SOC was ranging between 17.2–43.5 g kg⁻¹ with an median of 30.6±6.76 g kg⁻¹, and the lower quartile was 25.15 g kg⁻¹, the upper quartile was 34.80 g kg⁻¹. The results suggest that the samples are homogeneous or similar with respect to the humus variable. The Kruskal-Wallis test found no statistically significant difference in the amount of humus in samples. This means that the amounts of humus measured in the respective samples are probably similar or close to each other. This gives the possibility to claim that the variability in terms of humus is homogeneous in the different samples. The Kruskal-Wallis chi-squared = 62, df = 58, p-value = 0.3355.

Physical and chemical parameters of soils were measured. The saturation percentage ranged from 26–48, so the samples included sand, sandy loam, loam, and clay loam. The numerical difference in the values of pH measured in 1 M KCl and H₂O (pH_{KCl} – pH_{H₂O}). The pH_{KCl} of the soil samples tested ranged from 4.57–7.40, so samples were found to be acidic to slightly alkaline. The pH_{H₂O} values ranged from 6.31–8.18. Calculations for the hydrogen ion concentration indicate that pH_{KCl} is $1.08 \cdot 10^{-6}$ – $3.32 \cdot 10^{-6}$ mol dm⁻³ and pH_{H₂O} is $1.13 \cdot 10^{-7}$ – $1.46 \cdot 10^{-8}$ mol dm⁻³. The total water-soluble salt concentration was low and slightly saline ranged from 0.02–0.08 (m/m)%. Calcium carbonate concentration ranged from 0.107–1.73 (m/m)%, thus the soil samples were weakly calcareous. AL-soluble P₂O₅ was 35.6–798 mg kg⁻¹ and AL-soluble K₂O was 144.4–1489 mg kg⁻¹. The KCl-soluble NO₃-N (nitrogen in nitrate form) ranged from 1.18–25.169 mg kg⁻¹ (Table 1). The chemical and physical parameters also show a high variability of the samples.

The reflectance profiles of dry soil samples were evaluated in the 400–2500 nm range. On average, the reflectance increases linearly with longer wavelengths. The reflectance of the soil samples varied between 9–14% at shorter wavelengths, reaching 34–39.5% at 1000 nm. The standard deviation curve of the reflectance profile took a parabolic shape. The minimum was at short wavelengths (400–430 nm) due to the high absorption of the soil. It then reaches a maximum at a relatively broad plateau in the 600–800 nm wavelength range and then decreases

again up to 1000 nm. The standard deviation curve of the reflectance profile has a parabola-like shape. The minimum was at short wavelengths (400–430 nm) due to the high absorption of the soil. It then reaches a maximum in the 600–800 nm wavelength range at a relatively broad plateau and then decreases again up to 1000 nm. Based on the SD, the largest variation of the reflectance was observed in the 650–750 nm range. In the 1000–2500 nm range, on average, an increase in reflectance is observed with longer wavelengths, reaching a plateau between 2150–2300 nm, followed by a slight decrease. However, there were three local decreases in reflectance at 1420 nm, 1930 nm, and 2210 nm, resulting in valleys in the reflectance curves. The first two deviations are typical of the water absorption bands of the soil spectrum (Bowers et al., 1965). Although for organic functional groups on the other side, there is a peak near 1930 nm, which may be masked by the strong influence of water in the 1900 nm band (Knadel et al., 2009). The dynamics of the standard deviation (SD) curves were very similar to the reflectance curves. The SD curve showed a minimum at the 1020–1050 nm wavelength (Fig. 2).

The PCA resulted in five main components. The first factor explained 93.2% of the total variance, with the remaining four components explaining 5.58% of the variance, for a total explained variance of 99.8%. Based on the factor weights of the first component, the largest variance of the reflectance was in the 580–600 nm range, so the average of the reflectance factors measured at these wavelengths was used as the numerator of the first spectral index (Fig. 3).

There were two minima in the factor weights, but because of the differences and variations identified earlier, the 960–970 nm range was considered less sensitive to SOC, and thus

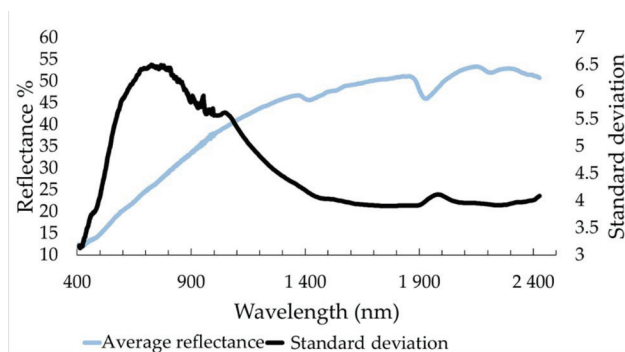


Fig. 2. Reflectance curves and standard deviation obtained in VIS-NIR region

Table 1
Mean physical and chemical properties of the soil samples

SOC%	Saturation percentage	pH _{KCl} * (mol dm ⁻³)	pH _{H₂O} * (mol dm ⁻³)	Calcium carbonate (m/m)%	Total water soluble salt (m/m)%	AL-soluble P ₂ O ₅ (mg kg ⁻¹)	AL-soluble K ₂ O (mg kg ⁻¹)	KCl-soluble NO ₃ -N (mg kg ⁻¹)
<2.5	38.58 ± 5.72	$1.08 \cdot 10^{-6}$	$1.13 \cdot 10^{-7}$	0.44 ± 0.50	0.05 ± 0.02	149.5 ± 59.58	557.8 ± 279.6	7.88 ± 5.23
2.5–3.5	41.67 ± 0.87	$9.49 \cdot 10^{-7}$	$1.33 \cdot 10^{-8}$	1.43 ± 0.25	0.03 ± 0.01	304.6 ± 167.3	768.1 ± 381.8	7.72 ± 4.84
3.5<	41.28 ± 1.70	$3.32 \cdot 10^{-6}$	$1.46 \cdot 10^{-8}$	1.29 ± 0.36	0.02 ± 0.01	339.6 ± 192.3	697.7 ± 247.2	9.97 ± 5.44

* calculations for the hydrogen ion concentration

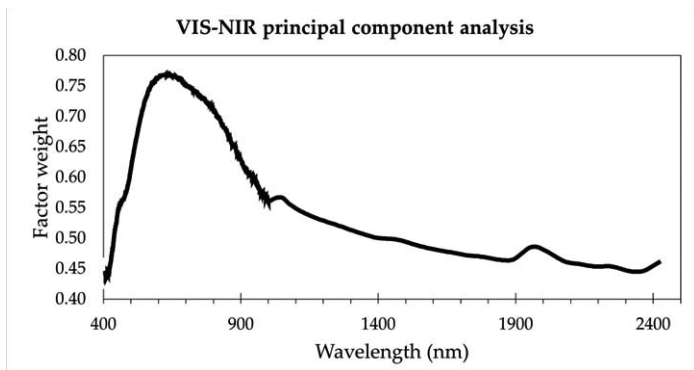


Fig. 3. Factor weights of the first component of the PCA result

we calculated the $Index_1 = \lambda_{580-600} / \lambda_{960-970}$. The first SOC model (SOC_{model1}) was based on linear regression with a moderately strong regression value $R^2=0.47$ ($p=0.000$). The first highlight of the PCA curves in the NIR range was observed at 1020–1040 nm, which was combined with the second wavelength difference in 1900–2100 nm. The two observed deviations were used to form the $Index_2 = \lambda_{1020-1040} / \lambda_{1900-2100}$ model. The other SOC model (SOC_{model2}) showed a strong correlation ($R^2=0.61$, $p=0.000$). The SOC_{model3} is defined by the factor weights of the starting and ending wavelength ranges in the NIR. Based on this, a difference was observed in the 1000–1010 nm range, and the minimum of the factor weights was found in the 2420–2500 nm wavelength range, which is less sensitive to the SOC. From the results obtained, the created $Index_3 = \lambda_{1000-1010} / \lambda_{2420-2500}$ had a strong regression value, $R^2=0.56$ ($p=0.000$). Based on the PCA curve results, SOC_{model4} was constructed from the first VIS range rising factor weight data and the third NIR range rising factor weight data using $Index_4 = \lambda_{640-660} / \lambda_{2200-2300}$. The SOC_{model4} was based on linear regression with a moderately strong regression value $R^2=0.49$ ($p=0.000$) (Table 2). Since the SOC content of samples are homogeneous, the potential reason for low regression is that Walkley and Black method underestimates the SOC content in the soil, which could affect the strength of regression (Meersmans et al., 2009).

SOC_{model1} , which performs also in the 400–1000 nm wavelength range, provided $RMSE=0.36$ SOC%, $NRMSE=11\%$, and

$NSE=0.56$. SOC values ranged from 1.98–3.81 SOC% with an average standard deviation of 3.02 ± 0.49 . Two models have been developed to predict SOC in the NIR wavelength range. SOC_{model2} has the strongest regression coefficient resulting from $RMSE=0.21$ SOC%, $NRMSE=6.30\%$, and $NSE=0.85$ supporting the best reliability of the model estimation. The predicted values varied from 1.72–4.35 SOC% with an average of 3.15 ± 0.69 . The use of SOC_{model3} performed $RMSE=0.29$ SOC%, $NRMSE=8.70\%$ and $NSE=0.72$. Estimated SOC values ranged from 2.05–4.22 SOC%, with a mean of 3.17 ± 0.68 . The estimation accuracy of the SOC_{model4} was $RMSE=0.29$ SOC%, $NRMSE=8.78\%$, and $NSE=0.72$. The predicted SOC values ranged from 2.18–3.83 SOC%, with an average SOC value of 3.10 ± 0.48 (Fig. 4).

In addition to the current study, previously developed models based on similar or different principles were collected to evaluate the accuracy of VIS-NIR spectroscopy in SOC determination. Measurements were made using RED and NIR reflectance in a study by Béni et al. (2021). The values were determined by their patented instrument and compared with SOC results for different soil types in the reference laboratory study. The calibration was derived from the NIR/RED index, where a strong correlation ($R^2=0.73$) was found between the values obtained with their patented detection method and the reference measurement data. On the other hand, there was no validation information. Ogrič et al. (2019) analyzed soil samples in the 400–2500 nm wavelength range in a study, but unlike the present study, they used Partial Least Square (PLS) to develop the estimation method and 1050°C dry combustion to determine the reference SOC data. Chang, et al. (2005) found a strong correlation between SOC and carbonate corrected dry combustion reference values in the 1100–2500 nm wavelength range with $R^2=0.88$. Islam, et al. (2003) followed a similar approach, then this research, analyzing soil samples in the 400–2500 nm range, and the reference SOC value was also similar to that determined by Walkley and Black, but applying a combination of Principal component regression, and found similar results. Morón and Cozzolino (2002) also investigated the estimation of SOC in the 400–2500 nm wavelength range with a strong correlation $R^2=0.74$ with the reference values. Kühnel and Bogner (2017) in their study used 350–2500 nm PCA and CHN (carbon,

Table 2
Statistics for the linear regression based SOC models

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
SOC_{model1}	Constant	6.093	.730		8.348	.000
	Index1	-6.125	1.363	-.684	-4.493	.000
SOC_{model2}	Constant	9.344	1.061		8.806	.000
	Index2	-8.167	1.327	-.782	-6.156	.000
SOC_{model3}	Constant	7.121	.800		8.904	.000
	Index3	-6.048	1.124	-.747	-5.381	.000
SOC_{model4}	Constant	4.954	0.464		10.681	.000
	Index4	-4.947	1.069	-.702	-4.630	.000

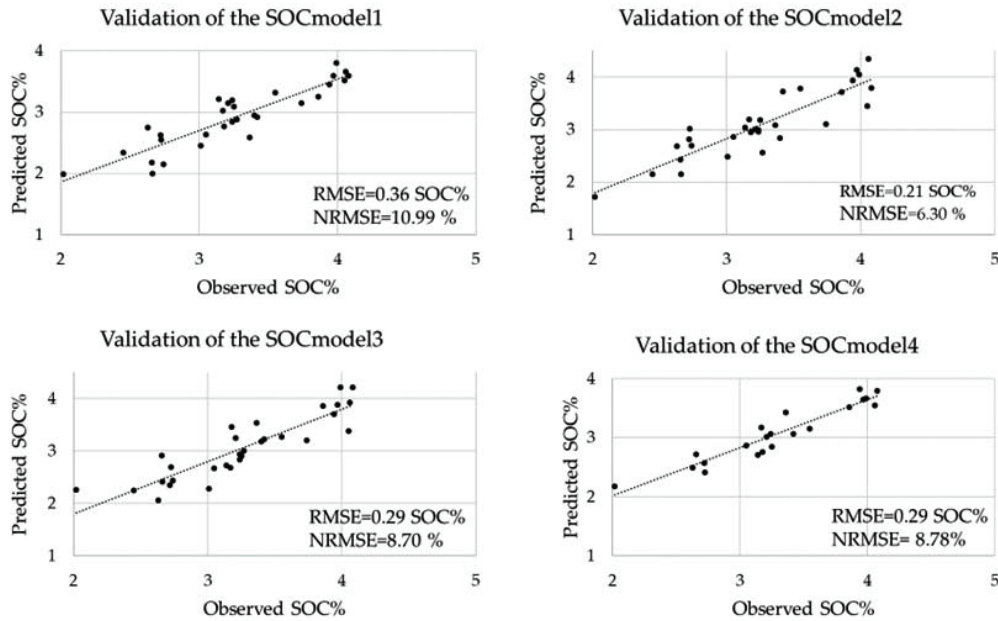


Fig. 4. The accuracy of the estimations based on validation of SOC models

hydrogen, nitrogen) analysis and found that the obtained SOC results gave a strong correlation $R^2=0.69$, which is almost identical to the result of the present study. In references, RMSE is between 0.25–0.64, in this study, it is 0.39, which proves, that the validation results of this study are also in correspondence with the studies described above and in Table 3.

5. Conclusions

Besides conventional analytical methods possess higher measurement accuracy for SOC, farmers still need reliable non-invasive, cost-effective, and rapid spectral-based methods to gain information on SOC. Considering the SOC content of

Table 3
Comparison of SOC estimation spectroscopy models

Predicted Method	Reference Method	λ – Range (nm)	Cal/ Val Samples	R^2	RMSE (SOC%)	Range of SOC content (g kg ⁻¹)	SD of SOC content (g kg ⁻¹)	Reference
PCA	SOC by Walkley and Black	400–2500	60/30	0.61	0.39	17.2–43.5	6.9	Present study
PLS	SOC by Walkley and Black	660–940	Total of 26	0.73	*	11.9–60.5	*	(Béni, et al., 2021)
PLS	SOC by dry combustion 1050°	400–2500	173/157	0.82	0.64	0.63–40.85	8.91	(Ogrič et al., 2019)
PCA	SOC by CHN-analyses	350–2500	29/41	0.69	0.5	10.1–90.6	10.5	(Kühnel and Bogner, 2017)
PLS	SOC by carbonate corrected dry combustion	1100–2500	161/83	0.88	0.38	0.5–40.8	10.4	(Chang, et al., 2005)
PCR	SOC by Walkley and Black	400–2500	121/40	0.81	0.35	0.6–49.5	8.2	(Islam, et al., 2003)
PLS	SOC by Walkley and Black	1100–2500	270/90	0.66	0.25	6.5–30	4.2	(Dunn, et al., 2002)
PLS	SOC by dichromatic oxidation	400–2500	177/139	0.74	0.5	10.3–68.5	15	(Morón & Cozzolino, 2002)

* not assessed

where

Cal/Val samples: calibration and validation samples

PCA: principal components analysis

PCR: principal components regression analysis

PLS: partial least square analysis.

the soil samples used for modeling the models are appropriate for the determination of SOC content of soils in a range of 17.2–43.5 g kg⁻¹. In this study four SOC% estimation models were developed based on spectral indices, supported by principal component analysis (PCA), resulted in moderately good coefficients of determination ($R^2=0.47\text{--}0.61$), indicating the feasibility of the method. Using the $\text{Index}_2 = \frac{\lambda_{1020-1040}}{\lambda_{1900-2100}}$ was performed the best in SOC estimation, resulting the strongest regression coefficient ($R^2=0.61$), and the outperforming from the other developed models with better performance indicators (i.e. RMSE=0.21 SOC%, NRMSE=6.30%, and NSE=0.85). Results show that non-destructive and chemical intensive spectral index-based models in VIS-NIR offer less robust alternatives to traditional SOC measurement methods. In conclusion, this study demonstrates that VIS-NIR spectroscopy, especially in the near infrared (NIR) range, can be a valuable tool for predicting soil organic carbon (SOC%) in the lowland region of Hungary.

Acknowledgments

This research was funded by TKP2021-NKTA-32 project. Project no. TKP2021-NKTA-32 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme. This research was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The research measurements were supported by University of Debrecen FAFSEM, Central Laboratory of Agricultural and Food Products.

References

Allen, M., 2017. The sage encyclopedia of communication research methods (Vols. 1–4). Thousand Oaks, CA: SAGE Publications, 1438–1440. <https://doi.org/10.4135/9781483381411>

Babaeian, E., Sidike, P., Newcomb, M.S., Maimaitijiang, M., White, S.A., Demieville, J., Ward, R.W., Sadeghi, M., Leb-Auer, D.S., Jones, S.B., Sagan, V., Tuller, M.A., 2019. New Optical Remote Sensing Technique for High-Resolution Mapping of Soil Moisture. *Frontiers in Big Data* 2, 37. <https://doi.org/10.3389/fdata.2019.00037>

Banninger, D., Fluhler, H., 2004. Modelling light scattering at soil surfaces. *IEEE T. Geoscience and Remote Sensing* 42(7), 1462–1471. <https://doi.org/10.1109/TGRS.2004.828190>

Bartholomeus, H., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W., Spaargaren, O., 2008. Spectral reflectance-based indices for soil organic carbon quantification. *Geoderma* 145, 28–36. <https://doi.org/10.1016/j.geoderma.2008.01.010>

Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* 59, 364–372. <https://doi.org/10.2136/sssaj1995.03615995005900020014x>

Ben-Dor, E., Heller, D., Chudnovsky, A., 2008. A novel method of classifying soil profiles in the field using optical means. *Soil Science Society of America Journal* 72(4), 1113–1123. <https://doi.org/10.2136/sssaj2006.0059>

Béni, Á., Juhász, E., Ragán, P., Rátonyi, T., Várbíró, G., Fekete I., 2021. Development of soil organic matter measurement system. *Soil and Water Research* 3, 1–6. <https://doi.org/10.17221/18/2021-SWR>

Boscaro, D., Pezzuolo, A., Sartori, L., Marinello, F., Mattioli, A., Bolzonella, D., Grigolato, S., 2018. Evaluation of the energy and

greenhouse gases impacts of grass harvested on riverbanks for feeding anaerobic digestion plants. *Journal of Cleaner Production* 172, 4099–4109. <https://doi.org/10.1016/j.jclepro.2017.02.060>

Bowers, S.A., Hanks, R.J., 1965. Reflection of radiant energy from soil. *Soil Science* 100, 130–138. <https://doi.org/10.1097/00010694-196508000-00009>

Brunet, D., Barthes, B.G., Chotte, J.L., Feller C., 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols, and Ultisols and from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma* 139, 106–117. <https://doi.org/10.1016/j.geoderma.2007.01.007>

Butkute, B., Slepeliene, A., 2006. Application of near infrared reflectance spectroscopy for the assessment of soil quality in a long-term pasture. *Communications in Soil Science and Plant Analysis* 37, 2389–2409. <https://doi.org/10.1080/00103620600819917>

Cecillon, L.C., Barthes, B.G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A., Burn, J.J. 2009. Assessment and monitoring of soil quality using Near-Infrared Reflectance Spectroscopy (NIRS). *European Journal of Soil Science* 60, 770–784. <https://doi.org/10.1111/j.1365-2389.2009.01178.x>

Chang, C.W., Laird, D.A., Hurburgh, C.R.J., 2005. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. *Soil Science* 170, 244–255. <https://doi.org/10.1097/01.ss.0000162289.40879.7b>

Couteaux, M.M., Berg, B., Rovira, P., 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. *Soil Biology and Biochemistry* 35, 1587–1600. <https://doi.org/10.1016/j.soilbio.2003.08.003>

Croft, H., Kuhn, N., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatiotemporal soil organic carbon dynamics in agricultural systems. *CATENA* 94, 64–74. <https://doi.org/10.1016/j.catena.2012.01.001>

Dunn, B.W., Batten, G.D., Beecher, H.G., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australia. *Animal Production Science* 42(5), 607–614. <https://doi.org/10.1071/EA01172>

Filcheva, E., Tsadilas, C.D., 2002. Influence of Clinoptilolite and Compost on Soil Properties. *Communications in Soil Science and Plant Analysis* 33, 595–607. <https://doi.org/10.1081/CSS-120002766>

Ge, Y., Morgan, C.L.S., Thomasson, J.A., Waiser, T., 2007. A new perspective to near-infrared reflectance spectroscopy: A wavelet approach. *Trans. ASABE* 2(50), 303–311. <https://doi.org/10.13031/2013.22394>

Goidts, E., van Wesemael, B., 2007. Regional assessment of soil organic carbon changes under agriculture in southern Belgium (1955–2005). *Geoderma* 141, 341–354. <https://doi.org/10.1016/j.geoderma.2007.06.013>

Guillou, F.L., Wetterlind, W., Viscarra Rosse, A.M., Hicks, W., Grundy, M., Tuomi, S., 2015. How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Research* 53(8), 913–921. <https://doi.org/10.1071/SR15019>

Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research* 41(6), 1101–1114. <https://doi.org/10.1071/SR02137>

IUSS Working Group WRB, 2022. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps. 4th edition. International Union of Soil Sciences (IUSS), Vienna, Austria

Kalembasa, S.J., Jenkinson, D.S., 1973. A comparative study of titrimetric and gravimetric methods for the determination of organic carbon in soil. *Journal of the Science of Food and Agriculture* 24, 1085–1090. <https://doi.org/10.1002/jsfa.2740240910>

Knadel, M., Greve, M.H., Thomsen, A., 2009. VIS/NIR mapping of TOC and extent of organic soils in the Nřrre Ł valley. *Nor-dic Association of Agricultural Scientists*, 7088.

Kononova, M.M., 1966. *Soil Organic Matter* Pergamon Press, 2nd edition, Oxford, 378. ISBN: 9781483185682

- Kuang, B., Mouazen, A.M., 2013. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. *Biosystems Engineering* 114(3), 249–258. <https://doi.org/10.1016/j.biosystemseng.2013.01.005>
- Kühnel, A., Bogner, C., 2017. In-situ prediction of soil organic carbon by vis-NIR spectroscopy: an efficient use of limited field data. *European Journal of Soil Science* 68, 689–702. <https://doi.org/10.1111/ejss.12448>
- Laamrani, A., Berg, A.A., Voroney, P., Feilhauer, H., Blackburn, L., March, M., Dao, P.D., He, Y., Martin, R.C., 2019. Ensemble Identification of Spectral Bands Related to Soil Organic Carbon Levels over an Agricultural Field in Southern Ontario, Canada. *Remote Sensing* 11, 1298. <https://doi.org/10.3390/rs11111298>
- Liu, R., Pan, Y., Bao, H., Liang, S., Jiang, Y., Tu, H., Nong, J., Huang, W., 2020. Variations in Soil Physico-Chemical Properties along Slope Position Gradient in Secondary Vegetation of the Hilly Region, Guilin, Southwest China. *Sustainability* 12, 1303. <https://doi.org/10.3390/su12041303>
- Ludwig, B., Khanna, P.K., Bauhus, J., Hopmans, P., 2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. *Forest Ecology and Management* 171, 121–132. [https://doi.org/10.1016/S0378-1127\(02\)00467-X](https://doi.org/10.1016/S0378-1127(02)00467-X)
- Meersmans, J., Van Wesemael, B., Van Molle, M., 2009. Determining soil organic carbon for agricultural soils: a comparison between the Walkley and Black and the dry combustion methods (north Belgium). *Soil Use and Management* 25, 346–353. <https://doi.org/10.1111/j.1475-2743.2009.00242.x>
- Mendiburu, F., 2019. *Agricolae: Statistical Procedures for Agricultural Research*. R Package Version 1.3-0. <https://CRAN.R-project.org/package=agricolae>
- Mezősi, G., 2016. Soils of Hungary. *The Physical Geography of Hungary, Akadémiai kiadó*, 165–174.
- Michéli, E., Fuchs, M., Hegymegi, P., Stefanovits, P., 2006. Classification of the Major Soils of Hungary and their Correlation with the World Reference Base for Soil Resources (WRB). *Agrokémia és talajtan*, 55.
- Mohamed, E.S., Ali, A., El-Shirbeny, M., Abutaleb, K., Shaddad, S.M., 2020. Mapping soil moisture and their correlation with crop pattern using remotely sensed data in arid region. *The Egyptian Journal of Remote Sensing and Space Science* 23, 347–353. <https://doi.org/10.1016/j.ejrs.2019.04.003>
- Morón, A., Cozzolino, D., 2002. Application of near Infrared Reflectance Spectroscopy for the Analysis of Organic C, Total N and pH in Soils of Uruguay. *Near Infrared Spectroscopy* 10, 215–221. <https://doi.org/10.1255/jnirs.338>
- Nagy, A., Tamás, J., 2009. Integrated airborne and field methods to characterize soil water regime. In: Ing., A. Celková (szerk.) *Transport of water, chemicals end energy in the soil-plant-atmosphere system*, 412–420.
- Nagy, A., Riczu, P., Gálya, B., Tamás, J., 2014. Spectral estimation of soil water content in visible and near infra-red range. *Eurasian Journal of Soil Science* 3(3), 163–171. <https://doi.org/10.18393/ejss.69645>
- Nagy, A., Szabó, A., Adeniyi, O.D., Tamás, J., 2021. Wheat Yield Forecasting for the Tisza River Catchment Using Landsat 8 NDVI and SAVI Time Series and Reported Crop Statistics. *Agronomy* 11, 652. <https://doi.org/10.3390/agronomy11040652>
- Nagy, A., Tamás, J., Burai, P., 2007. Application of advanced technologies for the detection of pollution migration. *Cereal Research Communications* 35, 805–809. <https://doi.org/10.1556/crc.35.2007.2.160>
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, Part I—A discussion of principles. *Journal of Hydrology* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research* 155, 510–522. <https://doi.org/10.1016/j.still.2015.07.021>
- Nduwamungu, C., Ziadi, N., Parent, L.É., Tremblay, G.F., Thurié's, L., 2009. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review. *Canadian Journal of Soil Science* 89(5), 531–541. <https://doi.org/10.4141/CJSS08076>
- Ogrič, M., Knadel, M., Kristianse, S.M., Peng, Y., De Jonge, L.W., Adhikari, K., Greve, M.H., 2019. Soil organic carbon predictions in Subarctic Greenland by visible-near infrared spectroscopy. *Antarctic, and Alpine Research* 51(1), 490–505. <https://doi.org/10.1080/15230430.2019.1679939>
- Piekarczyk, J., Kazmierowski, C., Krolewicz, S., Cierniewski, J., 2016. Effect of soil surface roughness on soil reflectance measured in laboratory and outdoor conditions. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing and Remote Sensing* 9(2), 827–834. <https://doi.org/10.1109/JSTARS.2015.2450775>
- Ponomariova, V.V., Plotnikova, T.A., 1980. *Humus and soil formation (Methods and results)*. Nauka, Leningrad Division.
- Russell, C.A., 2003. Sample preparation and prediction of soil organic matter properties by near infra-red reflectance spectroscopy. *Communications in Soil Science and Plant Analysis* 34, 1557–1572. <https://doi.org/10.1081/CSS-120021297>
- Shi, Z., Ji W., Viscarra Rossel, A.M., Chen, S., Zhou, Y., 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. *European Journal of Soil Science* 66(4), 679–687. <https://doi.org/10.1111/ejss.12272>
- Siebielec, G., McCarty, G.W., Stuczynski, T.I., Reeves, J.B. III., 2004. Near and mid infrared diffuse reflectance spectroscopy for measuring soil metal content. *Journal of Environmental Quality* 33, 2056–2069. <https://doi.org/10.2134/jeq2004.2056>
- Stenberg, B., Viscarra Rossel, A.M., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy* 107, 163–215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)
- Tang, H., Qiu, J., Van Ranst, E., Li, C., 2006. Estimations of soil organic carbon storage in cropland of China based on DNDC model. *Geoderma* 134, 200–206. <https://doi.org/10.1016/j.geoderma.2005.10.005>
- Terhoeven-Urselmans, T., Schmidt, H., Joergensen, R.G., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pretreatment. *Soil Biology and Biochemistry* 40, 1178–1188. <https://doi.org/10.1016/j.soilbio.2007.12.011>
- Terra, F.S., Dematte, J.A.M., Viscarra Rossel, A.M., 2015. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis-NIR and mid-IR reflectance data. *Geoderma* 255–256, 81–93. <https://doi.org/10.1016/j.geoderma.2015.04.017>
- Van Waes, C., Mestdagh, I., Lootens, P., Carlier, L., 2005. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. *The Journal of Agricultural Science* 143, 487–492. <https://doi.org/10.1017/S0021859605005630>
- Viscarra Rossel, R.A., McBratney, A.B., 2008. Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping, In: *Digital Soil Mapping with Limited Data* (eds. Hartemink et.al) Springer Science Business Media B.V., 165–172. https://doi.org/10.1007/978-1-4020-8592-5_13
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- Walkley, A., Black, L.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* 37, 29–38. <https://doi.org/10.1097/00010694-193401000-00003>
- Zhou, X.H., Zhou, D.W., 2009. Review of digital ground object spectral library. *Guang pu xue yu guang pu fen xi = Guang pu*, 29, 1616–1622. [https://doi.org/10.3964/j.issn.1000-0593\(2009\)06-1616-07](https://doi.org/10.3964/j.issn.1000-0593(2009)06-1616-07)