

## ORIGINAL PAPER

# The application of near infrared (NIR) spectroscopy for the quantitative assessment of soil organic matter fraction in forests

Mateusz Kania<sup>✉</sup>, Dawid Kupka, Piotr Gruba

Department of Forest Ecology and Silviculture, Faculty of Forestry, University of Agriculture in Kraków, Al. 29 Listopada 46, 31-425 Kraków, Poland

## ABSTRACT

In this study we investigated if near infrared (NIR) spectroscopy can be effectively used to predict content of various fraction (LF – light fraction and MAF – mineral associated fractions) of soil organic matter, also in the context of their spatial distribution. Additionally, we used NIR spectroscopy to evaluate basic properties of forest soils. We analyzed 256 soil samples from the topsoil of plots in central Poland. Using laboratory techniques, we divided the soil samples into two soil carbon fractions: the light fraction (LF) and mineral-associated fraction (MAF). A calibration model was developed using the spectra from 171 soil samples and their corresponding measured values. The regression model was validated using 85 independent soil samples. Using this model, we estimated the following forest soil properties: carbon concentration in light fraction ( $C_{LF}$ ), carbon concentration in mineral-associated fraction ( $C_{MAF}$ ), ratio of the  $C_{LF}$  to the total carbon content of the soil sample ( $C_{LF}/C$ ), ratio of the  $C_{MAF}$  to the total carbon content of the soil sample ( $C_{MAF}/C$ ), the total concentration of carbon ( $C_t$ ) and nitrogen ( $N_t$ ), C:N ratio (CN), pH, the concentration of exchangeable base cations (BC) and cation exchange capacity (CEC). The best calibration results were obtained for  $C_{LF}$ ,  $C_{LF}/C$  and  $C_{MAF}/C$ . The largest adjusted coefficients of determination for validation were obtained for  $N_t$ , CN, BC and CEC. Model developed for the  $C_{LF}$  was characterized by inaccurate value prediction. The paper shows also the relationship between the optimum number of soil sample spectra and the absolute and relative measurement error. Comparison of the measured and predicted values show that NIR spectroscopy has potential for determining soil parameters. The statistical assessment and spatial distribution analysis of the modelled  $C_{MAF}$  demonstrated relatively good agreement with measured values. However, the model's assessment of the  $C_{LF}$  was less accurate. We conclude that NIR spectroscopy is most applicable for use in soil science to determine the parameters:  $C_t$ ,  $N_t$ , C/N, pH, CEC,  $C_{MAF}$  and  $C_{MAF}/C$ .

## KEY WORDS

forest soils, near infrared spectroscopy, soil analysis, soil fraction, soil organic matter

<sup>✉</sup>e-mail: [mateusz.kania@urk.edu.pl](mailto:mateusz.kania@urk.edu.pl)

Received: 18 November 2022; Revised: 19 December 2022; Accepted: 13 January 2023; Available online: 3 March 2023

 Open access

©2022 The Author(s). <http://creativecommons.org/licenses/by/4.0>

## Introduction

Near infrared (NIR) spectroscopy is a dynamic and innovative research technique based on the interpretation of the oscillation-rotation spectra of a given substance (Atkins, 2001; Viscarra Rossel *et al.*, 2006). NIR spectroscopy can act as an alternative to expensive and time-consuming laboratory measurements that require more complex sample preparation and analysis. The NIR technique is relatively simple to perform and can be carried out by persons after initial training. It is a non-invasive method that allows for the assessment of several soil properties through the analysis of a single spectrum that can be obtained within seconds (Ben-Dor and Banin, 1995; Islam *et al.*, 2003; Seema *et al.*, 2020). The use of this technique in forest soil science is in its infancy. However, studies show there is potential for NIR spectroscopy as a tool to facilitate the determination of soil properties (Kania and Gruba, 2016; Kania *et al.*, 2017).

The use of NIR spectroscopy to determine soil properties is subject to certain limitations. The results obtained are only a model estimate of the real values and are prone to error. Therefore, it is necessary to scan a large enough number of samples to obtain an average value close to the laboratory-measured value. Additionally, it is not known whether the model parameters are sensitive to the species composition of the forest stand, which will be important to understand when developing models for forest soils (Chodak *et al.*, 2007; Stenberg *et al.*, 2010; Ng *et al.*, 2020).

Soil organic matter (SOM) is an important source of plant nutrients in both agricultural and natural forest ecosystems. In addition, understanding the properties of SOM is useful when assessing the soil type, climate, vegetation and soil productivity (Loveland and Webb, 2003). SOM is a factor that influences the overall soil quality, and is an indicator of sustainable agricultural practices (Freixo *et al.*, 2002; Loveland and Webb, 2003).

The fractionation of SOM into its different organic components is considered an important step in understanding a soil's properties, as different organic fractions vary in their function in the soil ecosystem, chemical composition and circulation time (degree of stability). This process can also separate the primary and secondary organic complexes (Christensen, 2001). However, this is an expensive process requiring intensive labor from trained personnel and a properly equipped laboratory. Therefore, there is a need to develop alternative methods to classify SOM.

Forest soil research is largely based on assessing the properties of samples taken from a single soil cover, that are then considered representative of the wider soil surface. However, the chemical properties of forest soils can change over small distances. This is particularly evident in topsoil layer (Conyers and Davey, 1990; Gil, 1995). It is therefore necessary to determine the optimum number of samples needed to produce an average value that is accurate to a reasonable degree (typically 95% confidence) and that can be considered representative.

The aim of this study was to investigate if NIR spectroscopy can be effectively used to predict the content of various fraction (light fraction and mineral associated fractions) of soil organic matter, also in the context of their spatial distribution. Additionally, we used the NIR technique to assess basic properties of forest soils. The properties determined from the NIR spectra were compared with the analogous results obtained via chemometric methods, utilizing soil samples with known physicochemical properties.

## Material and methods

**STUDY SITE.** The study site for sample collection was located in the Suchedniów Forest District area in central Poland (Fig. 1).

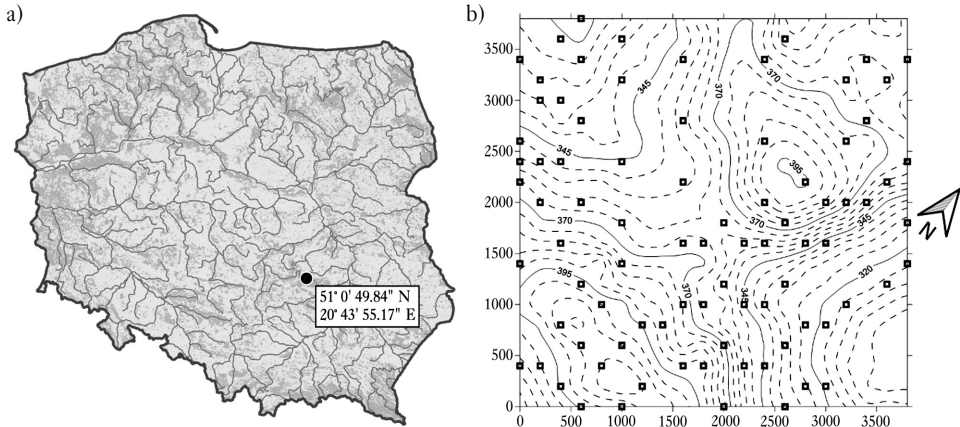


Fig. 1.

Research site: (a) locating in Poland, (b) hypsometric map

Sampling plots were established on soils under silver fir *Abies alba* Mill. and European beech *Fagus sylvatica* L. stands, as well as a mixture of Scots pine *Pinus sylvestris* L., European oak *Quercus robur* L. and common birch *Betula pendula* Roth stands. The site terrain is gently inclined to the NW. An elevation is ranging between 290 and 412 m above sea level. According to the WRB classification, local soils were represented by Dystric Cambisols, Haplic Luvisols and Albic Podzols (FAO, 2014). Soils are derived from Triassic sandstones and claystones, Quaternary sands and fluvio-glacial Quaternary sands. The Triassic sandstones, developed during the lower Triassic, were reddish in color, resulting from inclusions of iron-rich hematite. The major clay mineral component in the sandstones and claystones from the research site was kaolinite. The Quaternary sands were derived mostly from material derived from sandstone weathering, and are underlain by sandstones and claystones (Krajewski, 1955).

**SOIL SAMPLING AND LABORATORY ANALYSIS.** We analyzed 256 soil samples. Each sample was collected from the mineral topsoil (A or E) at the soil profile located at each plot. A detailed sampling scheme has been previously outlined by Gruba *et al.* (2015). The soil pH was measured by a potentiometric method using a combined electrode in a soil suspension in distilled water (1:5 mass-to-volume ratio), according to Buurman *et al.* (1996). Concentrations of exchangeable base cations ( $BC = Ca^{2+} + K^{+} + Mg^{2+} + Na^{+}$ ) and cation exchange capacity (CEC, the sum of base cations and total acidity) was measured by ICP (ICP-OES Thermo iCAP 6500 DUO, Thermo Fisher Scientific, Cambridge, U.K.). Fifteen g sub-samples were used to perform physical separation of SOM fractions. Fractionation procedure was described in details in Gruba *et al.* (2015). Briefly, the soil samples were divided into two carbon fractions: a light fraction (LF) and mineral-associated fraction (MAF). Separation of the SOM was carried out using a method designed for isolating the C content (Sohi *et al.*, 2001). Then, were measured soil property values, including the carbon concentration in light fraction ( $C_{LF}$ ), carbon concentration in mineral-associated fraction ( $C_{MAF}$ ), ratio of the  $C_{LF}$  to the total carbon content of the soil sample ( $C_{LF}/C$ ), ratio of the  $C_{MAF}$  to the total carbon content of the soil sample ( $C_{MAF}/C$ ), total concentration of carbon ( $C$ ) and nitrogen ( $N_t$ ), C:N ratio (CN) (using LECO CNS TruMac analyzer, Leco, St. Joseph, MI, USA).

**NIR SPECTROSCOPY ANALYSIS.** NIR spectroscopy was performed using a Fourier transform (FT) near infrared spectrometer (Antaris II FT-NIR; Thermo Fisher Scientific, Waltham, MA, USA).

Soil samples were placed in a glass tube (5 cm height, 1.9 cm diameter) in the spectrometer for analysis. Spectra were collected in the wavelength range of 1000 to 2500 nm. The spectrometer did not require any additional adjustment. Spectral analyses were performed using TQ Analyst 8 software (Thermo Fisher Scientific, Waltham, MA, USA). To calibrate the regression models for the soil properties of interest, we used the spectra from 171 soil samples. To assess the accuracy and reliability of these models, we performed a validation test using a subset (every third file) of the remaining 85 sample spectra that had not been used to calibrate the regression model. After multiple test calibrations, we selected the settings that yielded the highest adjusted coefficients of determination ( $R^2_{adj}$ ) for the soil properties. The level of statistical significance was defined at  $p < 0.05$ . The quality of the calibration and validation steps is expressed by the coefficient of determination of calibration ( $R^2C$ ) and coefficient of determination of validation ( $R^2V$ ). Statistical analyses were conducted using Statistica 13 software (Dell Inc., 2016).

The best results for most soil properties were obtained using the PLS (partial least square regression model and calculating the first derivative (1<sup>st</sup>) spectrum. Only for pH and BC was the second derivative (2<sup>nd</sup>) spectrum used. The spectral range was cut for  $C_t$ ,  $N_t$ , CN and CEC to obtain the best fit and to optimize the coefficients of calibration ( $R^2C$ ). For the models for the remaining properties, the full spectrum ranges gave the best results (Table 1).

## Results and discussion

GENERAL CHARACTERISTICS OF THE INVESTIGATED SOILS. The soil samples had varied characteristics, as evidenced by the range of values for the analyzed soil properties (Table 2). The pH of the soil samples ranged from acidic to slightly alkaline. The concentration of carbon in the light fraction was lower than the mineral-associated fraction. However, the percentage of carbon concentration in the light and mineral-associated fraction was relatively wide. The variability in the C:N ratios was also large, with a difference between the lowest and highest values of 39 units.

CALIBRATION RESULTS. The best calibration results, expressed by larger values of  $R^2C$ , were obtained for  $C_{LF}$ ,  $C_{LF}/C$  and  $C_{MAF}/C$  ( $0.90 \leq R^2C$ ). Reasonable calibration results were calculated

Table 1.

Optimum program settings (spectral range, derivative) and calibration coefficients ( $R^2C$ ) for selected soil properties

Soil properties <sup>a</sup>	Spectral range [cm <sup>-1</sup> ]	Derivative <sup>b</sup>	$R^2C$ <sup>c</sup>
$C_t$	4000-6975	1 <sup>st</sup>	0.87
$N_t$	4000-6975	1 <sup>st</sup>	0.84
CN	4000-5200	1 <sup>st</sup>	0.81
pH	4000-10000	2 <sup>nd</sup>	0.77
BC	4000-10000	2 <sup>nd</sup>	0.76
CEC	4000-6975	1 <sup>st</sup>	0.80
$C_{LF}$	4000-10000	1 <sup>st</sup>	0.90
$C_{LF}/C$	4000-10000	1 <sup>st</sup>	0.92
$C_{MAF}$	4000-10000	1 <sup>st</sup>	0.82
$C_{MAF}/C$	4000-10000	1 <sup>st</sup>	0.93

<sup>a</sup> $C_t$  – concentration of total carbon,  $N_t$  – concentration of total nitrogen; CN – carbon to nitrogen ratio; BC – concentration of exchangeable base cations ( $BC = Ca^{2+} + K^+ + Mg^{2+} + Na^+$ ); CEC – cation exchange capacity;  $C_{LF}$  – carbon concentration in light fraction;  $C_{LF}/C$  – carbon concentration in light fraction to total carbon concentration;  $C_{MAF}$  – carbon concentration in mineral-associated fraction;  $C_{MAF}/C$  – carbon concentration in mineral-associated fraction to total carbon concentration; <sup>b</sup>1<sup>st</sup> – first derivative; 2<sup>nd</sup> – second derivative; <sup>c</sup> $R^2C$  – coefficient of determination for calibration

for  $C_t$ ,  $N_t$ , CN and CEC and  $C_{MAF}$  ( $0.80 \leq R^2C < 0.90$ ). The lowest quality results ( $R^2C < 0.80$ ) were achieved for the BC and pH calibrations (Fig. 2, Table 1). In summary, the optimum correlation coefficients were achieved for the parameters assessing the percentage of the organic fraction of the total carbon, regardless of the fraction type. The regression models created during calibration were characterized by high accuracy. Low flexibility is a feature of over-fitted models. (Table 1). It should be noted that limited flexibility in the models may negatively impact the validation results.

**VALIDATION RESULTS.** The strength of the validation results is demonstrated by the  $R^2V$  coefficients (Table 3). The largest coefficients were obtained for  $N_t$ , CN, BC, CEC ( $0.60 \leq R^2V$ ). Medium strength coefficients were calculated for  $C_t$ ,  $C_{MAF}$  and  $C_{MAF}/C$  ( $0.4 \leq R^2V < 0.6$ ), and the weakest coefficients were derived for pH,  $C_{LF}$  and  $C_{LF}/C$  ( $R^2V < 0.4$ ). Weaker validation results are likely the result of a high accuracy obtained during the calibration stage. The differences between the measured and calculated values were not statistically significant, except for BC (Table 3).

The model developed for the carbon concentration in the light fraction is characterized by inaccurate value prediction, especially from the upper end of range, *i.e.*,  $>10 \text{ g}\cdot\text{kg}^{-1}$ . This inaccurate prediction is visible compared to  $C_{MAF}$ , C and N scatter plots. This is also confirmed by the lower  $R^2V$  coefficients obtained for  $C_{LF}$  and  $C_{LF}/C$  (Table 3).

The instances where our models performed relatively poorly, *i.e.*, for  $C_{LF}$ ,  $C_{LF}/C$  and BC, provides weight to the opinions of some researchers that the NIR technique is not yet sufficiently developed to replace laboratory methods (Ludwig *et al.*, 2002; Cozzolino and Morón, 2006). Conversely, Pietrzykowski and Chodak (2014) showed that NIR spectroscopy can be successfully used to analyze properties of soil organic matter such as the C content or C:N ratio. Similar conclusions can be drawn from the work by Cozzolino and Morón (2006). In agreement with this, in our study we developed reasonable models for predicting these parameters. For our samples, the estimated carbon concentrations were comparable to the measured values. Therefore, our study demonstrated that the NIR technique has the potential to estimate select properties of organic matter and to replace traditional laboratory methods in the future. However, further research is still needed to optimize the technique for all organic matter properties of interest (Cozzolino and Morón, 2006; Pietrzykowski and Chodak, 2014; Kania and Gruba, 2016).

**Table 2.**

Descriptive statistics of selected soil properties

Soil properties <sup>a</sup>	Unit	n <sup>b</sup>	Mean	Range
$C_t$	$\text{g}\cdot\text{kg}^{-1}$	256	2.4	(0.2-6.4)
$N_t$	%	256	0.13	(0.1-0.45)
CN	–	256	21	(11-50)
pH	–	247	4.3	(3.7-7.2)
BC	$\text{cmol}_{(+)}\cdot\text{kg}^{-1}$	256	0.98	(0.06-25.25)
CEC	$\text{cmol}_{(+)}\cdot\text{kg}^{-1}$	256	8.9	(1-26.8)
$C_{LF}$	$\text{g}\cdot\text{kg}^{-1}$	256	7.1	(0.7-33.9)
$C_{LF}/C$	%	256	34	(3-84)
$C_{MAF}$	$\text{g}\cdot\text{kg}^{-1}$	256	8.7	(0.1-28.2)
$C_{MAF}/C$	%	256	41	(0.17-87)

<sup>a</sup> $C_t$  – concentration of total carbon,  $N_t$  – concentration of total nitrogen; CN – carbon to nitrogen ratio; BC = concentration of exchangeable base cations ( $\text{BC} = \text{Ca}^{2+} + \text{K}^+ + \text{Mg}^{2+} + \text{Na}^+$ ); CEC – cation exchange capacity;  $C_{LF}$  – carbon concentration in light fraction;  $C_{LF}/C$  – carbon concentration in light fraction to total carbon concentration;  $C_{MAF}$  – carbon concentration in mineral-associated fraction;  $C_{MAF}/C$  – carbon concentration in mineral-associated fraction to total carbon concentration; <sup>b</sup>n – number of samples

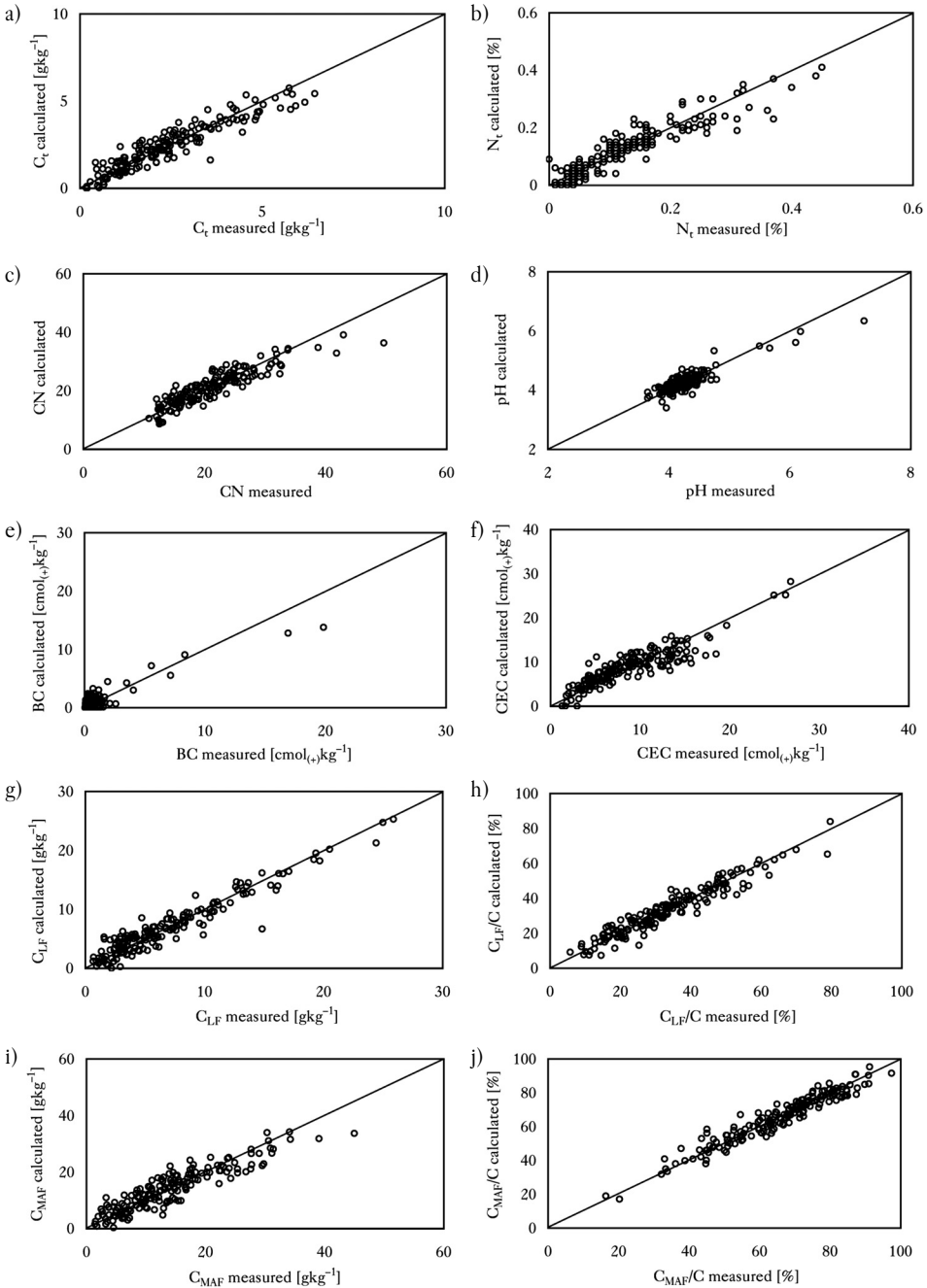


Fig. 2.

Calibration scatter plots for the measured vs calculated values for (a) concentration of total carbon ( $C_t$ ), (b) concentration of total nitrogen ( $N_t$ ), (c) carbon to nitrogen ratio (CN), (d) pH, (e) concentration of exchangeable base cations (BC), (f) cation exchange capacity (CEC), (g) carbon concentration in light fraction to total carbon concentration ( $C_{LF}$ ), (h) carbon concentration in light fraction to total carbon concentration ( $C_{LF}/C$ ), (i) carbon concentration in mineral-associated fraction ( $C_{MAF}$ ) and (j) carbon concentration in mineral-associated fraction to total carbon concentration ( $C_{MAF}/C$ )

**Table 3.**

Statistical results for NIR model validation

Soil properties <sup>a</sup>	R <sup>2</sup> V <sup>b</sup>	Measured values		Calculated values		p <sup>c</sup>
		mean (min-max) <sup>c</sup>	σ	mean (min-max)	σ <sup>d</sup>	
C <sub>t</sub>	0.53	2.5 (0.3-6.2)	1.42	2.4 (0.5-4.9)	0.96	NS
N <sub>t</sub>	0.63	0.14 (0.01-0.42)	0.10	0.13 (0.0-0.28)	0.07	NS
CN	0.60	20 (12-34)	4.71	20 (8-32)	4.61	NS
pH	0.11	4.2 (3.8-5.6)	0.27	4.2 (3.6-5.4)	0.26	NS
BC	0.67	0.9 (0.07-24.28)	2.82	1.18 (-1.5-14.67)	2.06	<0.001
CEC	0.72	8.8 (1.0-24.6)	4.69	8.4 (-2.1-21.3)	3.86	NS
C <sub>LF</sub>	0.33	7.7 (0.8-33.9)	6.05	6.9 (-0.4-19.4)	3.82	NS
C <sub>LF</sub> /C	0.31	35 (3-84)	14.69	33 (5-60)	11.88	NS
C <sub>MAF</sub>	0.58	13.5 (1.5-46.5)	7.94	13.6 (0.2-29.0)	6.72	NS
C <sub>MAF</sub> /C	0.40	67 (21-94)	13.99	65 (43-90)	11.54	NS

<sup>a</sup>C<sub>t</sub> – concentration of total carbon, N<sub>t</sub> – concentration of total nitrogen; CN – carbon to nitrogen ratio; BC – the concentration of exchangeable base cations (BC=Ca<sup>2+</sup>+K<sup>+</sup>+Mg<sup>2+</sup>+Na<sup>+</sup>); CEC – cation exchange capacity; C<sub>LF</sub> – carbon concentration in light fraction; C<sub>LF</sub>/C – carbon concentration in light fraction to total carbon concentration; C<sub>MAF</sub> – carbon concentration in mineral-associated fraction; C<sub>MAF</sub>/C – carbon concentration in mineral-associated fraction to total carbon concentration; <sup>b</sup>R<sup>2</sup>V – adjusted coefficient of determination for validation; <sup>c</sup>max – maximum values; min=minimum values; <sup>d</sup>σ – standard deviation; <sup>e</sup>p – significance level; NS – no significant differences p>0.05

RELATIONSHIP BETWEEN NUMBER OF SOIL SPECTRA AND MEASUREMENT ERROR. Collection of a single sample spectrum will not provide sufficiently comparable results to measured values. It is therefore necessary to scan multiple samples to obtain a value close to those derived from laboratory measurements (Ben-Dor and Banin, 1995; Chodak *et al.*, 2007). It is therefore important to estimate the number of scans required. If we consider the values for C<sub>LF</sub> and C<sub>MAF</sub> and the expected standard error (with 95% confidence), the optimum number of sample scans (*p*) can be calculated according to the formula by Kryszicki *et al.* (1995):

$$p = \left( \frac{2\sigma}{\Delta x} \right)^2$$

where:

σ – is the standard deviation

Δ*x* – is the assumed absolute (g·kg<sup>-1</sup>) or relative (%) measurement error.

The relationship between the optimum number of soil sample spectra and the absolute error for C<sub>LF</sub> and C<sub>MAF</sub> is shown in Figure 4. To estimate C<sub>MAF</sub> at the absolute error level of 2 g·kg<sup>-1</sup>, 45 spectra are required, whereas only 15 spectra are required for C<sub>LF</sub>. As the error increases, the difference between the optimum number of spectra for C<sub>LF</sub> and C<sub>MAF</sub> decreases. To obtain an accuracy of 3 g·kg<sup>-1</sup>, we need 20 C<sub>MAF</sub> spectra and 6 C<sub>LF</sub> spectra (a difference of 14). At the error level of 5 g·kg<sup>-1</sup>, the difference in optimum number of spectra is only 5. The differences between the optimum number of spectra required for the estimation of C<sub>LF</sub> and C<sub>MAF</sub> and the relative error rate (5-50%) are shown in Figure 5. The smaller the relative error, the more C<sub>LF</sub> spectra needed in relation to C<sub>MAF</sub>. To estimate values within 5% error, 5 and 23 spectra are needed for C<sub>MAF</sub> and C<sub>LF</sub> respectively. The smaller optimum number of C<sub>LF</sub> spectra required compared with C<sub>MAF</sub> is caused by a difference in their standard deviations. The carbon content in the light fraction was lower than in the mineral-associated fraction, resulting in a standard deviation almost twice as small (Table 3).

MAPPING. Spatial analysis followed by mapping of the laboratory-determined C<sub>LF</sub> values highlighted two hotspots of higher concentrations of this fraction (Fig. 6a). The hotspots are in the

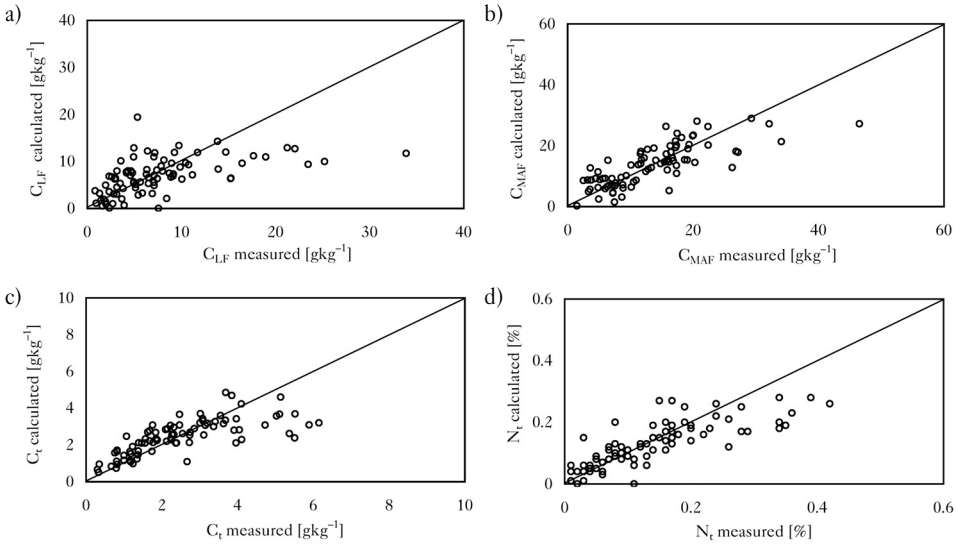


Fig. 3.

Scatter plots of the measured vs calculated values for the validation dataset derived for (a) carbon concentration in light fraction ( $C_{LF}$ ), (b) carbon concentration in mineral-associated fraction ( $C_{MAF}$ ), (c) total concentration of carbon ( $C_t$ ) and (d) total concentration of nitrogen ( $N_t$ )

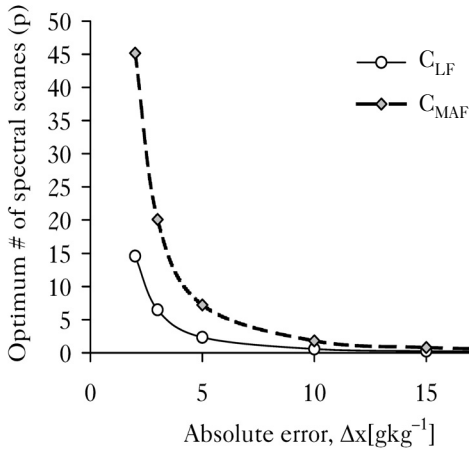


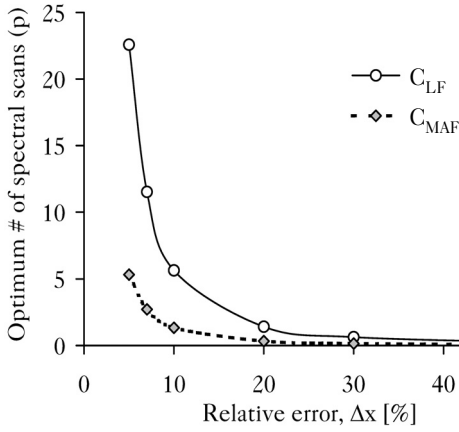
Fig. 4.

Relationship between the optimum number of soil sample spectra ( $p$ ) and the assumed absolute error  $\Delta x$  [ $\text{g}\cdot\text{kg}^{-1}$ ] for the carbon concentration in light fraction ( $C_{LF}$ ) and carbon concentration in mineral-associated fraction ( $C_{MAF}$ )

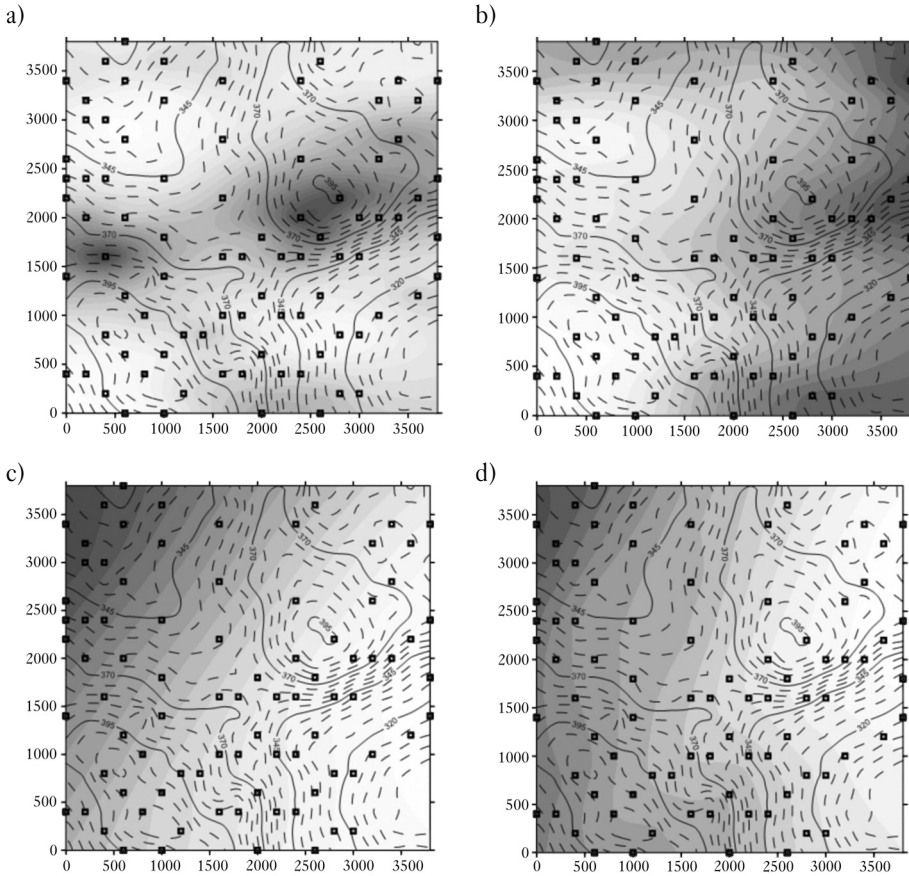
elevated areas (360-400 m above sea level) composed of soils derived from Triassic sandstones. The kriged values for observed  $C_{LF}$  range between 3.5 and 16  $\text{g}\cdot\text{kg}^{-1}$ . Interestingly, the spatial distribution for NIR-predicted  $C_{LF}$  values does not show the  $C_{LF}$  hotspots. The kriged  $C_{LF}$  values lie in the narrow range of 6 to 8  $\text{g}\cdot\text{kg}^{-1}$  (Fig. 6b). As was shown in Figure 3a, the NIR analysis failed to predict values above 10  $\text{g}\cdot\text{kg}^{-1}$ . It is therefore not surprising that the kriging analysis failed to reproduce the  $C_{LF}$  concentration hotspots visible in Figure 6a.

In the case of the spatial distributions for the observed and NIR-predicted  $C_{MAF}$  values, the trends are similar. The distributions show a gradual increase in  $C_{MAF}$  in a westerly direction (Fig. 6c, d). Additionally, the values overlap, i.e., 12-16  $\text{g}\cdot\text{kg}^{-1}$  and 10-18  $\text{g}\cdot\text{kg}^{-1}$  for the observed and predicted  $C_{MAF}$  values, respectively.





**Fig. 5.** Relationship between the optimum number of soil sample spectra (p) and the assumed relative error Δx [%] for the carbon concentration in light fraction (C<sub>LF</sub>) and carbon concentration in mineral-associated fraction (C<sub>MAF</sub>)



**Fig. 6.** Kriged maps for (a) observed carbon concentration in light fraction (C<sub>LF</sub>), (b) predicted carbon concentration in light fraction (C<sub>LF</sub>), (c) observed carbon concentration in mineral-associated fraction (C<sub>MAF</sub>) and (d) predicted carbon concentration in mineral-associated fraction (C<sub>MAF</sub>) in soils [g·kg<sup>-1</sup>]

## Conclusions

In this study we aimed to produce models that could predict properties of SOM with a high accuracy relative to laboratory-measured values. The calibrated models allowed us to predict values with a high accuracy. Satisfactory results were also produced during model validation. The statistical comparison of the measured and predicted values highlighted the applicability of the NIR technique for assessing soil parameters. However, the model developed for the carbon concentration in the light fraction ( $C_{LF}$ ) lacked sensitivity for predicting values at the upper end of the range, *i.e.*,  $>10 \text{ g}\cdot\text{kg}^{-1}$  of soil. This translated into an underestimation in the peak  $C_{LF}$  values when mapping the spatial distribution of this fraction in soils. This underestimation is also responsible for errors in the NIR prediction of the total soil carbon content ( $C_t$ ). However, the application of the NIR technique for quantitative assessment of the stabile  $C_{MAF}$  content produced relatively good results relative to measured values, in term of its statistical assessment and spatial distribution.

Additionally, our study demonstrated that to optimize the use of the NIR method for the prediction of soil properties, the number of optimum spectra required to achieve a desired level of accuracy in the estimated value should be taken into account during the sampling stage.

Overall, our study shows that with further work to optimize the technique, NIR spectroscopy has potential for use in soil science as a less labor-intensive alternative to laboratory techniques for determining parameters of SOM such as the  $C_t$ ,  $N_t$ , CN, pH, CEC,  $C_{MAF}$  and  $C_{MAF}/C$ . Based on average commercial prices, the cost and time of using the NIR technique is a few percent compared to laboratory analysis (not including the cost of the NIR spectrometer).

## Authors' contribution

M.K. – conceptual planning, performed the literature review, calibration, validation, model research, analysis and interpretation, manuscript planning and writing, data collection; D.K. – text writing and editing, analysis interpretation, figures preparation; P.G. – conceptual planning, analysis and interpretation, text writing-reviewing and editing, literature review.

## Conflicts of interest

The authors declare no conflicts of interest.

## Funding source

This research did not receive any specific funding.

## References

- Atkins, P.W., 2001. *Chemia fizyczna*. Warszawa: Wydawnictwo Naukowe PWN, 936 pp.
- 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 (2): 364-372. DOI: <https://doi.org/10.2136/sssaj1995.03615995005900020014x>.
- Buurman, P., van Lagen, B., Velthorst, E.J., 1996. *Manual for Soil and Water Analysis*. Leiden: Backhuys Publishers, 316 pp.
- Chodak, M., Niklińska, M., Beese, F., 2007. Near-infrared spectroscopy for analysis of chemical and microbiological properties of forest soil organic horizons in a heavy-metal-polluted area. *Biology and Fertility of Soils*, 44 (1): 171-180. DOI: <https://doi.org/10.1007/s00374-007-0192-z>.
- Christensen, B.T., 2001. Physical fractionation of soil and structural and functional complexity in organic matter turnover. *European Journal of Soil Science*, 52 (3): 345-353. DOI: <https://doi.org/10.1046/j.1365-2389.2001.00417.x>.
- Conyers, M.K., Davey, B.G., 1990. The variability of pH in acid soils of the southern highlands of New South Wales. *Soil Science*, 150 (4): 695-704.
- Cozzolino, D., Morón, A., 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil and Tillage Research*, 85 (1-2): 78-85. DOI: <https://doi.org/10.1016/j.still.2004.12.006>.

- FAO, 2014. World reference base for soil resources. Rome: World Soil Resources Report 106.
- Gil, W., 1995. Przestrzenne mikrozdóżnicowanie gleby leśnej – jego charakter i związki z drzewostanem. [Spatial microdifferentiation of forest soil – its character and links with tree stand (in Polish)]. *Sylwan*, 139 (4): 33-39.
- 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. DOI: <https://doi.org/10.1071/SR02137>.
- Kania, M., Gruba, P., 2016. Estimation of selected properties of forest soils using near-infrared spectroscopy (NIR). *Soil Science Annual*, 67 (1): 32-36. DOI: <https://doi.org/10.1515/ssa-2016-0005>.
- Kania, M., Gruba, P., Wiecheć, M., 2017. Zastosowanie techniki bliskiej podczerwieni do obliczania Siedliskowego Indeksu Glebowego. *Sylwan*, 161 (11): 935-939. DOI: <https://doi.org/10.26202/sywan.2017096>.
- Krajewski, R., 1955. Szczegółowa mapa geologiczna Polski w skali 1:50 000. Warszawa: Wydawnictwo Geologiczne.
- Krysicki, W., Bartos, J., Dyzka, W., Królikowska, K., Wasilewski, M., 1995. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część II. Statystyka matematyczna. Warszawa: Wydawnictwo Naukowe PWN, 331 pp.
- Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: A review. *Soil and Tillage Research*, 70 (1): 1-18. DOI: [https://doi.org/10.1016/S0167-1987\(02\)00139-3](https://doi.org/10.1016/S0167-1987(02)00139-3).
- 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 (1-2): 121-132. DOI: [https://doi.org/10.1016/S0378-1127\(02\)00467-X](https://doi.org/10.1016/S0378-1127(02)00467-X).
- Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., Minasny, B., 2020. Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia. *Geoderma Regional*, 22: e00319. DOI: <https://doi.org/10.1016/j.geodrs.2020.e00319>.
- Pietrzykowski, M., Chodak, M., 2014. Near infrared spectroscopy – A tool for chemical properties and organic matter assessment of afforested mine soils. *Ecological Engineering*, 62: 115-122. DOI: <https://doi.org/10.1016/j.ecoleng.2013.10.025>.
- Seema, Ghosh, A.K., Das, B.S., Reddy, N., 2020. Application of VIS-NIR spectroscopy for estimation of soil organic carbon using different spectral preprocessing techniques and multivariate methods in the middle Indo-Gangetic plains of India. *Geoderma Regional*, 23: e00349. DOI: <https://doi.org/10.1016/j.geodrs.2020.e00349>.
- Sohi, S.P., Mahieu, N., Arah, J.R.M., Powelson, D.S., Madari, B., Gaunt, J.L., 2001. A procedure for isolating soil organic matter fractions suitable for modeling. *Soil Science Society of America Journal*, 65 (4): 1121-1128. DOI: <https://doi.org/10.2136/sssaj2001.6541121x>.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107 (C): 163-215. DOI: [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7).
- Viscarra Rossel, R.A., McGlynn, R.N., McBratney, A.B., 2006. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 137 (1-2): 70-82. DOI: <https://doi.org/10.1016/j.geoderma.2006.07.004>.

## STRESZCZENIE

### Zastosowanie spektroskopii w bliskiej podczerwieni (NIR) do ilościowej oceny materii organicznej w glebach leśnych

W ostatnim czasie obserwuje się znaczny wzrost zainteresowania techniką bliskiej podczerwieni (NIR – near infrared), stanowiącej alternatywę dla kosztownych i pracochłonnych analiz laboratoryjnych. Niskie koszty badań wynikają m.in. z tego, że próbka gleby wymaga minimalnego stopnia przygotowania. Jest to metoda nieinwazyjna, pozwalająca na oszacowanie kilku właściwości badanej gleby na podstawie pojedynczego widma. Uzyskanie obrazu widma badanej gleby trwa kilkanaście sekund (Ben-Dor, Banin 1995; Islam i in. 2003; Viscarra Rossel i in. 2006). W badaniach gleboznawczych technika ta jest jeszcze stosunkowo rzadko wykorzystywana. Celem pracy było zbadanie, czy rodzaj frakcji glebowych wpływa na zastosowanie techniki NIR do ilościowej oceny materii organicznej, także w kontekście ich przestrzennego rozmieszczenia. Dodatkowo wykorzystano spektroskopię bliskiej podczerwieni do oznaczenia podstawowych właściwości gleb leśnych. Analizie poddano 256 próbek z wierzchniej warstwy gleb, z powierzchni zlokalizowanych w cen-

tralnej części Polski (ryc. 1). W wyniku prac laboratoryjnych oznaczono pH, zawartość kationów zasadowych (BC) oraz pojemność wymiany kationów (CEC). Każda z próbek gleby została podzielona na 2 odrębne frakcje glebowe: frakcję lekką (LF – light fraction) oraz powiązaną z frakcją mineralną (MAF – mineral associated fraction), zgodnie z Sohi i in. (2001). Następnie oznaczono właściwości gleb, tj.: zawartość węgla organicznego we frakcji lekkiej ( $C_{LF}$ ), zawartość węgla organicznego we frakcji powiązanej z frakcją mineralną ( $C_{MAF}$ ), stosunek  $C_{LF}$  do ogólnej zawartości węgla w próbce gleby ( $C_{LF}/C$ ), stosunek  $C_{MAF}$  do ogólnej zawartości węgla w próbce gleby ( $C_{MAF}/C$ ), całkowitą zawartość węgla ( $C_T$ ) i azotu ( $N_T$ ) oraz stosunek węgla do azotu (CN). Próbki gleb były zróżnicowane, a wartości poszczególnych właściwości gleb charakteryzowały się relatywnie dużą zmiennością (tab. 1). Do kalibracji modelu regresji wykorzystano widma 171 próbek gleb. W wyniku próbnych i wielokrotnych kalibracji wybrano ustawienia programu, które pozwoliły na opracowanie najdokładniejszych modeli. W celu sprawdzenia dokładności i wiarygodności opracowanych modeli przeprowadzono test walidacyjny na 85 wybranych schematycznie widmach, które nie zostały użyte wcześniej do kalibracji modelu. Przy użyciu opracowanych modeli kalibracyjnych oszacowano poszczególne właściwości gleb leśnych. Najlepsze wyniki kalibracji uzyskano dla  $C_{LF}$ ,  $C_{LF}/C$  i  $C_{MAF}/C$  (tab. 2; ryc. 2). Na etapie walidacji najlepsze rezultaty osiągnięto dla  $N_T$ , CN, BC i CEC, natomiast przeciętne dla  $C_T$ ,  $C_{MAF}$  oraz  $C_{MAF}/C$  (tab. 3). Opracowany model  $C_{LF}$  charakteryzował się niedokładną prognozą, głównie dla górnej granicy zakresu wartości (ryc. 3). W pracy przedstawiono zależność pomiędzy optymalną liczbą widm  $C_{LF}$  i  $C_{MAF}$  a błędem bezwzględnym (2-20 g·kg<sup>-1</sup>). Do uzyskania dokładności 3 g·kg<sup>-1</sup> potrzeba 20 widm  $C_{MAF}$  i 6 widm  $C_{LF}$  (różnica wyniosła 14 widm). Przy poziomie błędu 5 g·kg<sup>-1</sup> różnica optymalnej liczby widm wynosi już tylko 5 (ryc. 4). Na ryc. 5 pokazano różnice pomiędzy optymalną liczbą widm gleby potrzebnych do oszacowania  $C_{LF}$  i  $C_{MAF}$  a względnym poziomem błędu (5-50%). Im mniejszy zakres błędu, tym więcej jest potrzebnych widm  $C_{LF}$  w stosunku do  $C_{MAF}$ . Aby oszacować wartości w granicach błędu 5%, potrzebnych jest 5 widm  $C_{MAF}$  i 23  $C_{LF}$ . W pracy dokonano analizy rozkładu przestrzennego pomiędzy wartościami  $C_{LF}$  i  $C_{MAF}$  oznaczonymi i oszacowanymi. Mapowanie oznaczonych wartości  $C_{LF}$  wykazało 2 zagęszczone miejsca (360-400 m n.p.m.), natomiast analiza przestrzenna  $C_{LF}$  oznaczonych NIR nie wykazała takich miejsc (ryc. 6a, b). Rozkłady przestrzenne  $C_{MAF}$  wykazały podobną tendencję: stopniowy wzrost wartości w kierunku zachodnim (ryc. 6c, d). Porównanie zmierzonych i oszacowanych wartości pokazuje, że spektroskopia NIR ma potencjał określania parametrów gleby. Ocena statystyczna oraz analiza rozkładu przestrzennego modelowanego  $C_{MAF}$  wykazały stosunkowo dobrą zgodność z oznaczonymi laboratoryjnie wartościami. Ocena modelu  $C_{LF}$  była mniej dokładna. Wyniki badań pozwalają wysnuć wniosek, że spektroskopia NIR może być stosowana w gleboznawstwie do określania parametrów gleb, tj.:  $C_T$ ,  $N_T$ , CN, pH, CEC,  $C_{MAF}$  i  $C_{MAF}/C$ .