RESEARCH NOTE

Analysis of soil organic matter in tropical soils with near-infrared spectroscopy (NIRS) and chemometrics

Aarón Jarquín-Sánchez[1], Sergio Salgado-García[2], David J. Palma-López[2], and Wilder Camacho-Chiu[3]

[1]Laboratorio de Suelos, Colegio de la Frontera Sur Unidad Villahermosa. Carretera Villahermosa-Reforma km 15.5, Ranchería Guineo, sección II CP 86280, Villahermosa, Tabasco.
[2]Ciencia Ambiental, Colegio de Postgraduados–Campus Tabasco. Periférico Carlos A. Molina s/n km 3.5. H. Cárdenas, Tabasco. México.
[3]Departamento de Riego y Drenaje, Instituto Tecnológico de la Zona Olmeca. Villa Ocuitzapotlán, Centro, Tabasco.

Abstract

A. Jarquín-Sánchez, S. Salgado-García, D.J. Palma-López, and W. Camacho-Chiu. 2012. Analysis of organic matter in tropical soils with near-infrared spectroscopy (NIRS) and chemometrics. Cien. Inv. Agr. 39(2): 387-394. The objective of this study was to predict the concentration of soil organic matter (SOM) in tropical soils using near-infrared spectroscopy (NIRS) for samples measured within a polyethylene bag and without a bag. One hundred and fifty six soil samples from the humid tropics of Tabasco, Mexico with contrasting chemical characteristics were selected. The samples were dried, ground, and sieved through 2 mm and 5 mm screens, and their SOM contents were determined using the Walkley-Black method. The soil samples were packed in polyethylene bags, and SOM was measured directly with and without a bag using a quartz probe (FOSS 5000 model of NIRsystems, DK-3400 Hillerod,Denmark) for a range from 1100 to 2000 nm to obtain a prediction model of SOM. The model for determining SOM for bagged samples had good fit and explained 88% of the variation (0 to 10, 2% of SOM of samples). The model for determining SOM for bag-less samples was not efficient for predicting independent samples and therefore was discarded. The analysis by NIRS was reliable, more rapid, and easier for the determination of SOM in soil samples measured through a plastic bag.

Key words: Chemical analysis, tropical soil, optical fiber, model, organic matter prediction

Introduction

The organic fraction of soil regulates chemical processes occurring in this medium, influences its physical properties, and is the center of nearly all biological activities (Loveland and Webb, 2003).

The conventional analytic methods to determine the concentration of SOM are the analytic techniques of wet combustion (requiring acid digestion and titration) and loss on ignition (calcinations of the sample) (Peck, 1990). These methods, even when they are reliable, require 16 h to generate results and require high environmental impacts. Near-infrared spectroscopy (NIRS) has been utilized for predicting total concentrations and

the spatial variation of nitrogen and carbon (and other elements) in soils (Ludwing *et al.,* 2002; Odlare *et al.,* 2005; Shenk, 2004; Cozzolino and Moron, 2006).

Because NIR generates large data sets, chemometric methods must be applied for the interpretation of the results and the generation of the element prediction models of interest. For the application of NIRS in soil analysis, there are several technical aspects which must be elucidated (Martin *et al.,* 2003; Viscarra *et al.,* 2005; He *et al.,* 2007), such as reading soil samples using an optical fiber probe on soil samples stored in plastic bags (Jarquin *et al.,* 2011), where it is easier to take the reading without cleaning the probe, versus readings in which the probe is placed directly into the soil. Therefore, the objective of this study was to use NIRS to generate a model to predict SOM content quickly and reliably in the humid tropical soils of Tabasco, Mexico.

Materials and methods

Soils under study

To obtain a model allowing the assessment of SOM, 156 samples were selected from the collection of soils of the Colegio de Postgraduados, Campus Tabasco, obtained in October 2005. The samples belonged to the following units: Vertisol, Fluvisol, Cambisol, Gleysol, and Ultisol (Palma-López *et al.,* 2007), with a wide range of SOM contents. Composite samples of soil were taken from 0 to 30 cm depth. These soils differed in texture, pH, SOM, total nitrogen (tN), and cation exchange capacity (CEC). The maximum, minimum, mean and coefficient of variation for samples were calculated to ensure that the variation in the results did not influence model development. The soil samples were air dried at 26 °C in the shade, and then they were ground and sieved through a 5 mm diameter screen for the analyses of SOM (Walkley and Black) and tN (Kjeldahl) and 2 mm for the analyses of P-Olsen, K, Ca, Mg, Na,

CEC (ammonium acetate 1N pH 7), and texture (Bouyoucos). The analyses were conducted using the methods established in the Official Mexican Standard 021 (NOM, 2001).

Analysis using NIRS

The soil particle size for analysis by NIRS was 2 mm. The analysis of soil samples for SOM determination by NIRS was performed with a calibrated FOSS 5000 near-infrared spectrophotometer (NIR-systems) with a probe of integrated optical fiber, with a quartz window of 25 cm^2 and a range from 1100 to 2500 nm. The soil sample size was 200 g, and the plastic bag had dimensions of 20.8 cm x 9.9 cm (weight 1.5 g). The readings of the soil were taken both through the bag and without it. In the first case, the probe window was pressed against the bag to secure good contact with the soil. In the second case, the soil was placed on a glass plate and measured with the probe, again securing good contact with the soil. The probe was cleaned after each measurement. The interval of the useful range was from 1100 nm to 2000 nm, and the spectral range was averaged every 2 nm, generating 450 data points for each observation with and without a bag (the analysis took approximately 3 min).

ISI software 1.02 (WinISI II, 1999) was utilized for the chemometric data analysis, applying the following mathematical and statistical treatments: analysis by principal components (Wold *et al.,* 1987), application of first and second derivatives, standard normal variate (SNV), correction of multiplicative scatter correction (MSC), and detrending (DT). Minimum partial least squares (MPLS) was used to generate a model, and the analysis of prediction error (SEP) and error of calibration prediction (SEPC) were used as criteria for goodness of fit (Geladi, 2003; Miller and Miller, 2002). These mathematical treatments have different statistical parameters, which allow the selection of the best equation (Jarquín *et al.,* 2011). The treatments are expressed in a series of

numbers indicating which mathematical treatment should be used. The first number indicates the use of derivatives, which are utilized to find and correct noise problems in the spectrum. The second number indicates the use of GAP, representing the distance of the interval from the independent variable that separates the two averaged segments. The third number represents the use of Smooth, which indicates the smoothness obtained from each spectrum. The fourth number indicates the use of MSC, applied to the previous smoothing to improve the linearity (WinISI II, 1999).

The regression models were constructed using the SOM concentration obtained from the sample with the Walkley-Black method as the independent variable and the readings of the NIRS spectra as the dependent variable.

Each model was expressed as:

$$SOM\ (\%) = \beta_0 + \beta_1 X_{\lambda 1} + \beta_2 X_{\lambda 2} + \beta_3 X_{\lambda 3} + \beta_4 X_{\lambda 4} + \ldots\ldots\ldots + \beta_{450} X_{\lambda 450}$$

where β_0 is the ordinate to the origin, β_n are the coefficients of the calibration curve, $X_{\lambda n}$ are the wavelengths read every 2 nm, where the correlation coefficient of the components is the maximum in positive or negative form. For the present study, 450 β coefficients were found for each model.

The model was evaluated using cross validation; the program generates at least six cancellation groups of the calibration set and separates one to confirm the results or, in other words, to verify

the prediction. This operation is repeated as many times as the existing groups pass the calibration and the prediction group. The best model is the one presenting lower SEP and SEPC and higher R^2 (Geladi, 2003; Miller and Miller, 2002; Wold *et al.,* 2001). Furthermore, external validation was performed with a group of 26 independent soil samples having different physical and chemical characteristics, with the purpose of evaluating the precision of the model. These samples were analyzed under the same conditions as those used for constructing the model.

Results and discussion

Chemical and physical properties of tropical soils

The SOM content of the samples ranged from very low to very high (from 0.02 to 10% SOM). Soils with these characteristics and variability of color (data not presented), texture, and SOM content (Table 1) are the soils primarily used for agricultural purposes in the tropics.

The general statistics of pH, tN, P-Olsen, K, Ca, CEC, and sand, silt, and clay content are shown in Table 1. The variation coefficient was used to select the soils for establishing the mathematical model, and those that did not show differences were eliminated. This process served to establish a group of soils that had the greatest diversity of SOM contents, although they belonged to the same soil unit.

Table 1. Chemical and physical characteristics of the tropical soils utilized1.

Values	pH	TN	SOM	P-Olsen	Ke	Ca	Mg	Na	CEC	Sand	Silt	Clay
		(%)		(mg kg⁻¹)	Cmol(+) kg⁻¹					(%)		
Minimum	3.9	0.02	0.02	0.3	0.01	0.1	0.01	0.1	2.5	11	1	20
Maximum	8.2	0.50	10.0	41.4	2.76	20.4	16.4	30.3	36.3	55	34	86
Mean	5.3	0.12	2.90	6.3	0.23	3.0	2.4	1.3	12.3	31	14	55
CV	16.9	78.6	88.4	134.3	170.8	119.1	155.8	380.1	62.3	32.7	46.1	22.5

[1]156 soil samples; CV: Coefficient of variation; TN: Total nitrogen; SOM: Soil organic matter; P: Phosphorus-Olsen method; Ke: exchangeable Potassium; Ca: Calcium; Mg: Magnesium; Na: Sodium.

Description of the NIR spectra

The spectra for each sample collected through a bag (Figure 1) had smoother signals at 1400 and 1900 nm than the bag-less samples (Figure 2). The interval from 2002 to 2500 nm was eliminated under both experimental conditions because there was background noise in the signal. This proves that the zone from 1100 to 2000 nm has greater utility for establishing prediction models, as reported by Cozzolino and Moron (2006). The spectra of the samples read with and without a bag had signals in the wavelengths between 1100 and 1900 nm where the following molecules are located: N-N, C-N, O-N, S-N, C-O-O-N. Both cases show the same signals of these functional groups.

The spectra of the bagged soil samples showed lower reflectance than those of the bag-less samples, which may be attributed to the pressure on the sample applied by the probe during the reading. The depth reached by the infrared beam is different because the soil - with repeated placements of the probe - is compacted in different proportions depending on the particle type.

The variability diminishes with the use of the plastic bag as it maintains the order of the soil particles and reduces the variation in the spaces among them because of the pressure applied. The principal components analysis allowed the elimination of two bagged samples and seven bag-less samples whose variance did not contribute to improve the fit of models to predict the SOM content (Table 2).

*Obtaining and validating the mathematical models*

Evaluation of soil organic matter with the use of a bag. The standard mathematical treatment, MSC 2, 10, 10, 1 (multiplicative scatter correction), had a higher R^2 and the least standard error of calibration (SEC) and standard error of cross validation (SECV). Figure 3 depicts to the average coefficients spectrum generated with this model. The model was created with 124 samples (Table 3) in contrast to the 60 samples indicated as necessary (WinISI II, 1999); this reduction in samples was due to a cross validation which

**Table 2.** Principle components analysis for each soil group.

Type of analysis	Mathematical treatment	Explained variability (%)	Eliminated samples
Soil organic matter with bag	Standard MSC[1] 2, 10, 10, 1	99.69	2
Soil organic matter without bag	SNV only[2] 0, 0, 1, 1	99.99	7

[1]Standard multiplicative scatter correction.
[2]Standard normal variate only.
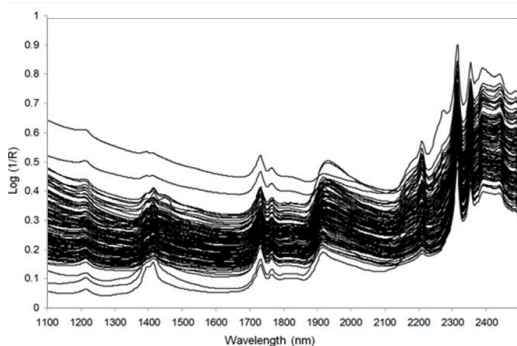
Figure 1. Spectrum of analyzed bagged soils.
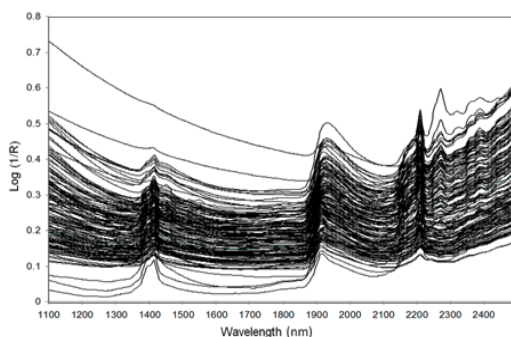
Figure 2. Spectrum of analyzed bag-less soils.

divided some samples into groups, identifying those with unusual performance or presenting differences (2½ times the standard error of the equation) between the laboratory data and the predictions. The SEC was lower than the SECV, and R^2 was 0.88, indicating that the model has good fit and predicts SOM in independent samples (Geladi, 2003).

*Evaluation of soil organic matter without a bag.* The correction of the spectrum was necessary, and thus the SNV mathematical treatment, only 2, 4, 4, 1 (SNV), was performed, which was sufficient correction for the model to present similarly low values of SEC and SECV (Table 3) with R^2 0.80; this value was higher than the R^2 of 0.75 at 0.78% obtained by Martin et al. (2003) to determine the SOM content of Orthic Black Chernozems soil. These results indicate that the model had good predictive abilities for predicting SOM contents for independent samples. Figure 4 presents the average SOM coefficient spectrum of this model.

The R^2 values obtained in this study for these two parameters are lower than those obtained by Ludwing *et al.* (2002), who reported an R^2 of 0.95 in calibration, with carbon levels of 1.4 to 16.8 % in soil samples.

External validation of models

Evaluation of oil organic matter with the use of a bag. To externally validate the generated model, 20 bagged and 26 bag-less soil samples were selected at random. These samples were sampled and evaluated similarly to those of the calibration group; the model chosen for each case was used, and SOM was calculated. Twenty and 26 soil samples were used to validate the model because a minimum of 10% of the total of the
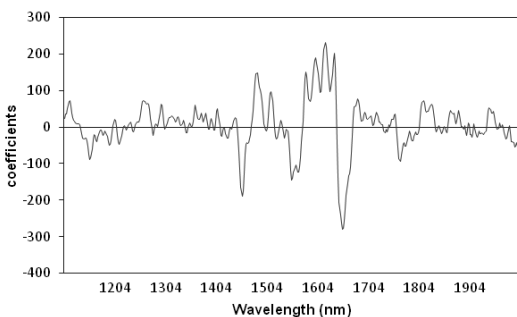
Figure 3. Coefficients obtained for the bagged SOM (soil organic matter) prediction model after mathematical treatment.
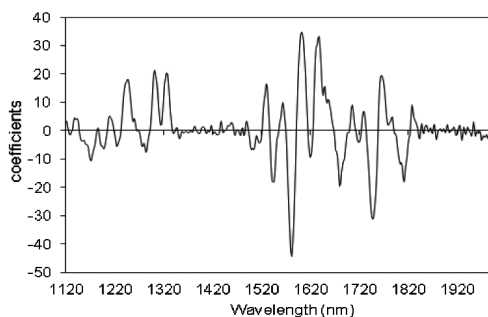
Figure 4. Coefficients obtained for the bag-less SOM (soil organic matter) prediction model after mathematical treatment.

Table 3. Statistics of calibration models for soil organic matter (SOM).

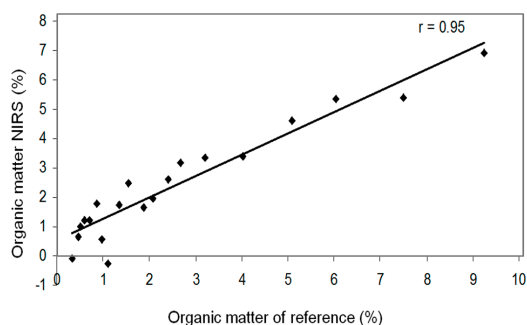| Type of analysis	Mathematical treatment	N	Average SOM (%)	Sd	R^2	Estimation SOM (%) Min	Estimation SOM (%) Max	SEC[3]	SECV[4]
SOM with bag	Standard MSC[1] 2, 10, 10, 1	124	2.83	2.45	0.88	0.00	10.18	0.71	0.86
SOM without bag	SNV only[2] 2, 4, 4, 1	120	2.75	2.25	0.80	0.00	9.50	1.00	1.28

[1]Standard multiplicative scatter correction.
[2]Standard normal variate only.
[3]Standard error of calibration.
[4]Standard error of cross validation.

samples was required to perform external validation; in this case, 18 and 19%, respectively, of the total bagged and bag-less samples were used. It is also possible to compare two populations with fewer than 30 data points using a Tukey test (Wold *et al.,* 2001).
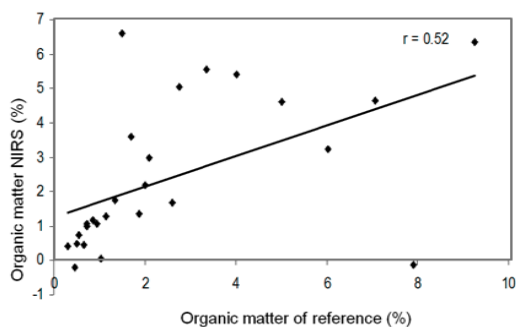
Figure 5 shows the external validation of the bagged SOM prediction model, according to the values obtained using the Walkley-Black method and NIRS. The correlation coefficient was 0.95, and the values of SEP (0.90) and SEPC (0.91) were low, which demonstrates that the model predicts the SOM contents with good approximation (Höskuldsson, 1996). The Tukey test proved that the values for SOM obtained by the two methods are not different ($P>0.05$). Based on the generated information, it can be concluded that analysis by NIRS is more reliable, quicker, and easier for SOM determination in bagged soil samples.

Evaluation of oil organic matter without a bag. The correlation between the NIRS and the Walkley-Black method was 0.50 (Figure 6); the low correlation is due to the high value of SEPC (2.27) and to difficulties in SOM determination in the external validation of some soil samples. This may be attributed to different situations, such as the effect of pressure on the sample by the optical fiber probe, as previously explained, or the nature of the soil itself, which may cause increasing errors according to the heterogeneity of the soil (Wold *et al.,* 2001); therefore, the model was discarded.

The SOM model for bagged samples showed good fit and explained 88% of the variation (0 to 10.2% of SOM). The SOM model for bag-less samples was not efficient to predict independent samples, so it was discarded. The NIRS analysis is reliable, faster and easier for the determination of SOM in bagged soil samples.

Figure 5. External validation of the NIRS (near-infrared spectroscopy) model for SOM (soil organic matter) with bag.

Figure 6. External validation of the NIRS (near-infrared spectroscopy) model for SOM (soil organic matter) without bag.

Resumen

A. Jarquín-Sánchez, S. Salgado-García, D.J. Palma-López y W. Camacho-Chiu. 2012. Análisis de materia orgánica en suelos tropicales por espectroscopia de infrarrojo cercano (NIRS) y quimiometría. Cien. Inv. Agr. 39(2): 387-394. El objetivo del presente trabajo fue predecir la concentración de materia orgánica del suelo (MOS) de regiones tropicales mediante espectroscopía de infrarrojo cercano (NIRS), en muestras medidas dentro de una bolsa de polietileno y sin ésta. Fueron seleccionadas 156 muestras de suelos del trópico húmedo de Tabasco, México, por sus características químicas contrastantes. Las muestras fueron secadas, molidas y tamizadas a través de una malla de 5 y 2 mm y su concentración de materia orgánica fue determinada por el método de Walkley Black. Las muestras de suelos se envasaron en bolsa

de polietileno y se midieron directamente con y sin bolsa con una sonda de cuarzo de un NIRS, modelo FOSS 5000 de NIRsystems, en un rango de 1.100 a 2.000 nm, para obtener un modelo de predicción del contenido de MOS en muestras dentro de la bolsa y fuera de ella. El modelo para determinar la MOS en muestras dentro de la bolsa presentó buen ajuste y explicó el 88% de la variación (0 a 10,2% de SOM). El modelo para materia orgánica para muestras fuera de la bolsa no resultó eficiente, para predecir muestras independientes, por lo que se descart**ó**. El análisis por NIRS es confiable, más rápido y sencillo en la determinación de materia orgánica en muestras de suelo con bolsa.

Palabras clave: Análisis químico, suelo tropical, fibra óptica, modelo, predicción de materia orgánica.

References

Cozzolino, D. and A. Morón. 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil and Tillage Research 85: 78-85.

Geladi, P. 2003. Chemometrics in spectrocopy. Part 1. classical chemometrics. Spectrochimica Acta part B 58: 767-782.

He, Y., M. Huang., A.G. Pereira., A.H. Gómez, and H. Song. 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. Computers and Electronics in Agric. 58: 144-153.

Höskuldsson, A. 1996. Prediction methods in science and technology. *In:* Hans-Henrik H. (ed.). Thor Publishing. Denmark. 404 pp.

Jarquín S.A., S.G. Salgado., D.J. Palma-López., W.Ch. Camacho, and A.P. Guerrero. 2011. Análisis de nitrógeno total en suelos tropicales por espectroscopía de infrarojo cercano (NIRS) y quimiometría. Agrociencia 45:653-662.

Loveland, P., and J. Webb. 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil and Tillage Research 70:1-18.

Ludwing, B., P.K. Khanna., J. Bauhus, and P. Hopmans. 2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. Forest Ecology and Management 171: 121-132.

Martin, P.D., D.F. Malley., G. Manning, and L. Fuller. 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. Can. J. Soil Sci. 82: 413–422.

Miller, N.J., and C.J. Miller. 2002. Estadística y quimiometría para química analítica. Edit. Pearson educación S. A. Madrid, España. 296 pp.

NOM-021-RECNAT-2001, Norma Oficial Mexicana (2001) que establece las especificaciones de fertilidad, salinidad y clasificación de suelos. Estudios, muestreo y análisis. NOM-021-RECNAT-2000. Emitida en el Diario Oficial el día Martes 31 de diciembre de 2002. México. 84 pp.

Odlare, M., K. Svensson, and M. Pell. 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. Geoderma 126: 193–202.

Palma-López, D.J., J.D. Cisneros., E.C. Moreno, and J.A. Rincon-Ramirez. 2007. Suelos de Tabasco su uso y manejo sustentable. Colegio de Postgraduados-ISPROTAB-Fundación Produce Tabasco, A.C. Villahermosa, Tabasco, México. 195 pp.

Peck, T.R. 1990. Soil Testing: Past, Present and Future. Cummun. In Soil Sci. Plant Anal. 21(13-16):1165-1186.

Shenk, J.S. 2004. Evaluation of the visible spectrum to measure plant and soil characteristics. *In*: A.M.C. Davies and A. Garrido-Varo (eds.). Near Infrared spectroscopy: procedigs of the 11th international conference 2003. Published by NIR publications, Chishester. p. 313-463.

Viscarra, R.R.A., D.J.J. Walvoort., J.L. McBratney, and J.O. Skjemstad. 2005. Visible, Near Infra-

red, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131: 59-75.

WinISI II.1999. Manual FOSS NIRsystems/TECA-TOR by Intrasoft International, LLC. Versión 1.02. 232 pp.

Wold, S., K. Esbensen, and P. Geladi. 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 2: 37–52.

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-Regresion: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58: 109-130.