DOI: 10.1255/jnirs.1219

A.M. de Souza *et al.*, *J. Near Infrared Spectrosc.* **24,** 293–303 (2016)
Received: 15 May 2015 ■ Revised: 4 March 2016 ■ Accepted: 12 May 2016 ■ Publication: 7 June 2016

JNIRS
JOURNAL
OF
NEAR
INFRARED
SPECTROSCOPY
Special Issue: Near Infrared Spectroscopy of Soil

Validation of the near infrared spectroscopy method for determining soil organic carbon by employing a proficiency assay for fertility laboratories

André Marcelo de Souza,[a]* Paulo Roberto Filgueiras,[b] Maurício Rizzato Coelho,[a] Ademir Fontana,[a] Thayane Christine Barbosa Winkler,[a] Patrícia Valderrama[c] and Ronei Jesus Poppi[b]

[a]Brazilian Agricultural Research Corporation (Embrapa Soils), 22460-000, Rio de Janeiro, RJ, Brazil. E-mail: andremarcelo.souza@embrapa.br

[b]Institute of Chemistry, University of Campinas (UNICAMP), PO Box 6154, 13084-971 Campinas, SP, Brazil

[c]Technological Federal University of Paraná (UTFPR), PO Box 271, 87301-899, Campo Mourão, PR, Brazil

Current research indicates that near infrared (NIR) spectroscopy is the most promising alternative technique for the determination of soil organic carbon (SOC) in laboratories worldwide and that it may serve as a total or partial replacement for the traditional wet chemistry methods. In this research, NIR spectroscopy was developed and validated as a method for the determination of SOC aiming to promote the use of NIR technology in the soil laboratories of Brazil as a routine analysis method. To this end, multivariate calibration models were constructed from a large number of soil samples (1490 samples) that encompassed the variability of Brazilian soils. These models were validated by submitting the concentrations of soil organic matter (SOM) as determined by NIR spectroscopy to the proficiency assay for fertility laboratories, which is coordinated by the Brazilian Agricultural Research Corporation (Embrapa Soils). The proposed methodology using NIR spectroscopy received the excellence index of quality, which gives a certificate seal issued by the interlaboratory programme. Therefore, the robustness of the NIR method was proved in a conclusive way by a proficiency test programme dedicated to evaluating the reference method for SOM determination employed by over a hundred soil laboratories in Brazil.

Keywords: soil carbon stock, figures of merit, PLS, routine soil analysis

Introduction

The continental dimensions of Brazil, the strong agricultural potential of the country and the lack of knowledge about Brazilian soils make the demand for research on soils a very strategic issue. Therefore, every year, millions of analyses of various attributes related to chemistry, fertility and classification of soil are performed in hundreds of laboratories throughout Brazil. Among all of these analyses, there is one that requires significant effort and time in soil laboratories: the determination of soil organic carbon (SOC). This attribute, among others, is traditionally used to evaluate soil fertility, and SOC is a diagnostic criterion that is used in the Brazilian System of Soil Classification.[1]

Analyses of SOC and seven other attributes related to soil fertility are performed by more than 500 Brazilian laboratories. Among these, 120 laboratories have the quality of their analysis attested and certified by a proficiency assay for fertility laboratories, which is coordinated by the Brazilian Agricultural Research Corporation (Embrapa Soils). Despite the large

ISSN: 0967-0335
doi: 10.1255/jnirs.1219

number of SOC analyses performed every year in Brazil, most laboratories still employ a wet method in which the soil organic matter (SOM) is oxidised with potassium dichromate and is then indirectly quantified by titrating the remaining ferrous ammonium sulphate with potassium dichromate.[2] This reference method has some limiting aspects, including (i) the uncertainties associated with the stages of digestion and titration, (ii) the high consumption of chemical reagents such as potassium dichromate, sulphuric acid and ferrous ammonium sulphate, (iii) the environmental impact due to the waste generated by this analysis and (iv) the time-consuming nature of the analysis.[2]

The limitations of the reference method that is currently employed by Brazilian soil laboratories and the need for a cheaper and faster alternative method have guided Embrapa to new paths in analytical chemistry. Current research has shown that the near infrared (NIR) spectroscopy methodology is the most promising alternative technique for routine soil analysis in total or partial replacement of traditional methods.[3–5] The growing interest in the application of NIR technology to soil analysis can be justified by the numerous and remarkable benefits that NIR spectroscopy offers over conventional analysis: (i) it is a non-destructive technique, (ii) it is free from undesirable residues and thus has minimal environmental impact, (iii) it is a low-cost technique and (iv) it is a rapid and innovative technique that requires little sample handling when combined with chemometrics. However, the main limitation of NIR spectroscopy is its dependence on chemometrics for data treatment. Because of this dependence on chemometrics and other factors, the NIR technique is not yet in wide use for routine soil analyses.

The innovative strategy of this work was to participate in the proficiency assay quality programme with the values of SOC determined by NIR spectroscopy rather than by the reference method and to obtain approval of the assay by earning a certificate of excellence. The main objective of this approach was to validate the calibration model for measuring SOC concentration using NIR spectra such that it can be employed in routine analysis.

Materials and methods

Acronyms and common symbols

Throughout this work, lowercase letters are used to represent scalar quantities, lowercase bold letters are used to represent vectors and uppercase bold letters are used to represent matrices. Italicised letters are used to represent dimensions of vectors and matrices.

Acronyms

RMSEC – root mean square error of calibration
RMSEP – root mean square error of prediction
RDP – ratio of performance to deviation
SD – standard deviation
SEP – standard error of prediction
LD – limit of detection
LQ – limit of quantification
MSEC – mean square error of calibration
MSECV – mean square error in cross validation
CV – coefficient of variation
EI – excellence index

Common symbols

n_{GL}– degrees of freedom
n_C – number of calibration samples
K – number of variables used in a calibration model
$\mathbf{y}_i$ – scalar reference value for the ith sample
$\hat{\mathbf{y}}_i$ – estimated y-value for ith sample based on a regression model
$\mathbf{b}_k$ – regression coefficient for kth variable based on a regression model
n – concentration level
m – number of replicates at each concentration level.
$S\hat{E}N$ – sensitivity
δ_X – estimation of noise level in the data
$V(PE_i)$ – variance of prediction error
φ_i – confidence limit
t – Student's t-value
T – scores of all calibration samples
$\mathbf{t}_i$ – score vector of a particular sample
α – significance level
A – number of latent variables
h_i– leverage
h_{limit} – leverage limit
I_C– number of calibration samples
J – number of spectral variables
$\mathbf{x}_{i,j}$ – absorbance value of sample i at wavelength j
$\hat{\mathbf{x}}_{i,j}$ – estimated value with A latent variables
ν – number of degrees of freedom
s(e) – standard deviation total residuals

Sample analysis

Complete profiles of the soils of the Brazilian territory (1490 samples) were selected from the Embrapa Soil collection, considering the environmental representativeness of the samples according to the Brazilian System of Soil Classification.[1]

The soil samples were dried at 40 °C for 48 hours, a rubber mallet was used to break clusters and the granulometry of samples was controlled by passing them through a sieve, with the samples ground to a size fraction <2 mm.[2]

The reference method employed was the dichromate oxidation method, where 0.5 g of soil sample was transferred to a 250 mL Erlenmeyer flask, to which was added 10 mL of 0.4 mol L^{-1} potassium dichromate solution. After this procedure, a test tube (25 mm in diameter and 250 mm in height) filled with water was put into the Erlenmeyer flask containing the soil and potassium dichromate solution. This combination formed a system condenser. The system was heated using an electric hot plate until the sample was boiling lightly for five minutes. After cooling, 80 mL of distilled water, 2 mL of orthophosphoric acid and 3 drops of diphenylamine indicator were added to the system. The system was titrated using a ferrous ammonium sulphate solution (0.1 mol L^{-1}) until the colour changed from blue to green. A white race titration was also performed. The SOC (g kg^{-1}) was calculated using the following equation:

$$\text{SOC} = (40 - \text{volume in mL used in the system titration}) \times f \times 0.6 \quad (1)$$

where f is 40/volume (in mL) used in the white race titration.[2]

In order to participate in the proficiency assay it was necessary to transform the values of SOC into SOM. The relationship between SOC and SOM is conventionally expressed by multiplying SOC by a conversion factor of 1.724.[2] This is based on the assumption that organic matter is 58% carbon, and this factor is known as the "van Bemmelen factor".[6] This conversion was made by all the laboratories that participated in the proficiency assay of soil fertility employed in this study.

Diffuse reflectance measurements were carried out using a Fourier transform spectrophotometer (Spectrum 100N; https://www.perkinelmer.com/lab-solutions/), equipped with a diffuse reflectance accessory (NIRA). The equipment and integrating sphere were equipped with DTGS/AsInGa detectors, respectively. The spectra were acquired in the range 4000–10,000 cm^{-1}, with a spectral resolution of 4 cm^{-1}, using 64 scans, in the percentage of reflectance mode.

Development of calibration and validation models

Spectral preprocessing was performed using the first derivative of Savitzky–Golay, with a second-order polynomial and a 15 point window.[7] Partial least squares regression (PLSR) models were developed with a selection of the most important NIR wavelengths. This variable selection was performed using the variable importance in projection (VIP) method.[8] The VIP method computes a score for each wavelength. VIP scores are a measure of the importance of each explanatory variable (i.e. wavelength). Because the average of the squared VIP scores equals 1, only influential wavelengths with a VIP score greater than 1 were kept in the model,[8,9] thus reducing the number of variables (from 6001 to 1033) used in the models. The subsets of calibration and validation data (2/3 to 1/3 for calibration and validation) were built by the Kennard and Stone algorithm.[10] Outliers were identified in the models based on data with extreme leverage, unmodelled residuals in spectral data[11] and unmodelled residuals in the dependent variable.[12]

Empirical local models were built aiming to achieve better predictive capability when compared with the global model. The SOC range was used to build the 11 PLSR local calibration models as follows (g kg^{-1}): (M0) $0.1 < y < 36.4$; (M1) $5 < y < 10$; (M2) $10 < y < 15$; (M3) $15 < y < 20$; (M4) $20 < y < 25$; (M5) $25 < y < 40$; (M6) $25 < y < 70$; (M7) $10 < y < 50$; (M8) $5 < y < 70$; (M9) $5 < y < 55$; (M10) $3 < y < 60$. Each PLSR model was tuned to minimise the root mean square error of cross validation (*RMSECV*) obtained by 10 samples in a contiguous block cross validation to maximise the determination coefficient value (which gives the predicting ability of the model) and the ratio of performance to deviation (*RPD*) obtained from each model.[13]

The performance of the PLS models was evaluated by determining figures of merit such as accuracy, sensitivity, analytical sensitivity, detection and quantification limits, adjustment and linearity (Table 1). The figures of merit such as accuracy, precision, sensitivity, analytical sensitivity, adjustment and confidence interval were calculated according to previous research.[14–17] The precision of the reference method was determined by the analysis of 23 concentration levels of soil samples covering the range of 0.1 to 43.8 g kg^{-1} of SOC, with six replicates at each level analysed by different analysts on different days.[18]

All models, routines and calculations used in this study were performed in Matlab R2013b (The MathWorks, Natick, MA, USA) and PLS_Toolbox 7.9 (R7.9.5; Eigenvector Research, Wenatchee, WA, USA).

Interlaboratory analysis of reference method

The database of the proficiency assay from 2013 was employed via simulation to evaluate the quality of the SOM analysis determined by the dichromate method (the reference method). In total, 111 laboratories were evaluated, and the statistical method and the calculation are described as follows. During the year, 12 (215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226) blind samples were analysed by each laboratory member of the programme, where two of the samples were triplicates. The laboratory then received 6 unique samples and 2 replicates, which resulted in 12 determinations of SOM. The resulting SOC concentration for each sample per laboratory was compiled in an Excel spreadsheet, and the calculation of mean, standard deviation(s), coefficient of determination (*CV*), minimum and maximum values was performed in three steps.

In the first step, all laboratories participated in the calculation, but in the next two steps, some criteria for penalties were employed based on the tolerance limits presented in Table 2. Every time that a laboratory was penalised because the SOC concentration of a sample was outside the tolerance limit, the laboratory received one asterisk. Because the calculation could be performed at most three times per sample, a laboratory could be penalised with up to three asterisks per sample because this sample was not a replicate sample. If the sample was a replicate sample, each asterisk was weighted by 3; thus, a laboratory could be penalised by up to nine asterisks per replicate sample.

Two important parameters for the quality evaluation of the laboratories (imprecision and inaccuracy) were computed employing the sample replicates and the number of asterisks (simple and weighted) received for each sample (Table 2). Because these two parameters are computed, they can be used to calculate the excellence index (a rank of quality) that will be employed to classify each laboratory in an excellence group. The laboratories classified in A and B excellence groups will be able to receive a certificate of quality (seal) from the quality programme.

Table 1. Figures of merit and outlier detection equations employed in this work.[11,14]

Figure of merit	Equation
Accuracy	$RMSEC=\sqrt{\frac{\sum_{i-1}^{n}\left(y_i-\hat{y}_i\right)^2}{n_c-k-1}}$ (2)
	$RMSEP=\sqrt{\frac{\sum_{i-1}^{n}\left(y_i-\hat{y}_i\right)^2}{n_v}}$ (3)
Determination coefficient	$R^2=1-\frac{\sum_{i=1}^{n}\frac{\left(y_i-\hat{y}_i\right)}{n-k-1}}{\sum_{i=1}^{n}\frac{\left(y_i-\overline{y}\right)^2}{n-1}}$ (4)
Ratio of performance to deviation	$RPD=\frac{SD}{SEP}$ (5)
Precision	$\text{Precision}=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\hat{y}_{ij}-\hat{y}_i\right)^2}{n\left(m-1\right)}}$ (6)
Sensitivity	$\mathrm{S\hat{E}N}=\frac{1}{\left\Vert b_k\right\Vert}$ (7)
Analytical sensitivity	$\gamma=\frac{\mathrm{S\hat{E}N}}{\delta_X}$ (8)
Limit of detection	$LD=3{,}3xb_k=3{,}3\delta_x\frac{1}{\mathrm{S\hat{E}N}}$ (9)
Limit of quantification	$LQ=10\delta_x b_k=10\delta_x\frac{1}{\mathrm{S\hat{E}N}}$ (10)
Confidence interval	$\mathrm{V}\left(PE_i\right)=\left(1+h_i+\frac{1}{n_c}\right)MSEC$ (11)
	$MSEC=\sum_{i=1}^{n_c}\frac{\left(y_i-\hat{y}_i\right)^2}{n_{GL}}$ (12)
	$\mathrm{n}_{\mathrm{GL}}=\mathrm{n}\left(1-\sqrt{\frac{MSEC}{MSECV}}\right)$ (13)
	$\varphi_i=t_{\frac{1-\alpha}{2n_{GL}}}\sqrt{V\left(PE_i\right)}$ (14)
Outlier detection	
Extreme leverage	$h_i=t_{A,i}^{T}\left(T_A^{T}T\right)^{-1}t_{A,i}$ $h_{\mathrm{limit}}=3\frac{A+1}{I_c}$ (15)
Unmodelled residuals in spectra	$s\left(e\right)^2=\frac{1}{I_cJ-J-A_{\max}\left(I_c,J\right)}\sum_{i=1}^{I_c}\left(\sum_{j=1}^{J}(x_{i,j}-\hat{x}_{i,j})^2\right)$ (16)
	$s\left(e\right)^2=\frac{1}{I_cJ-J-A_{\max}\left(I_c,J\right)}\sum_{j=1}^{J}(x_{i,j}-\hat{x}_{i,j})^2$ (17)
Unmodelled residuals in dependent variables	$RMSEC=\frac{1}{n-A-1}\sqrt{\sum_{i=1}^{n}\left(y_1-\hat{y}_1\right)^2}$ (18)
	$Erro\,abs\left(y_i-\hat{y}_i\right)=3\times RMSEC$ (19)

Table 2. Equations employed by the interlaboratory proficiency test to evaluate the quality of the reference method (oxidation of soil organic matter with dichromate) used by 114 laboratories for soil analysis in Brazil.[19]

Interlaboratory analysis	Equation	
Coefficient of variation (*CV*)	$CV(\%)=\frac{s}{\bar{x}}\times 100$	(20)
Tolerance limits	$CV > 50\%$ — interval = mean ± 1.0 s	(21)
	$20\% < CV < 50\%$ — interval = mean ± 1.5 s	(22)
	$CV < 20\%$ — interval = mean ± 2.0 s	(23)
Imprecision	$\text{Imprecision}=\frac{CV1+CV2}{2}$	(24)
Weighted asterisks	no = [(total of simple asterisks) × 1 + (total of weighted asterisks) × 3]	(25)
Inaccuracy	$\text{Inaccuracy}=\frac{\text{Number of weighted asterisks}}{\text{Number of determination}}\times 100$	(26)
Excellence index	$IE=100-\frac{(\text{inaccuracy}+\text{imprecision})}{2}$	(27)
Excellence groups	Group A: $IE > 90\%$	(28)
	Group B: $70 < IE < 90\%$	(29)
	Group C: $IE < 70\%$	(30)

Results and discussion

Calibration results

The NIR spectra of the soil samples are shown in Figure 1(A). No relevant information was removed by correcting the spectral baseline with the Savitsky–Golay algorithm, as shown in Figure 1(B). The spectral region between 10,000 cm^{-1} and 7500 cm^{-1} was removed because this region did not present relevant information. Figure 1(C) presents the average NIR spectra of the complete matrix of soil samples employed in this study.

Table 3 presents a compilation of several figures of merit and the number of outliers removed during the construction of all 10 calibration models built in this work. Also some information about the concentration range of SOC, *SD* and minimum, maximum and medium concentration is also presented in Table 3. First, the models can be more or less restrictive based on the range of SOC concentrations presented. The less restrictive models such as M0, M5 to M10 presented worse accuracy based on the values of *RMSEP* and *RMSEC* than the restrictive models such as M1, M2, M3 and M4 due to the narrow range of SOC concentrations. The explanation of this result is related to the lack of complexity and the narrow range of concentration of the restrictive models that results in more accuracy, i.e. the larger the range, the larger is the standard error of prediction.[5] Consequently, the same behaviour was observed based on the values of the coefficient of variation.

The accuracy index (*RPD*) and the R^2 values were always better for the models with the larger range of SOC concentrations (less restrictive models). Thus, these observations were very contradictory because larger values of *RPD* and R^2 indicated better fitting models; thus, the best performance was expected for the more restrictive models. A reasonable explanation of this observation relies on the fact that the relevance of these parameters is highly questionable because it depends on the range and distribution of the population,[5] respectively. This hypothesis was demonstrated here because the distribution of the SOC concentrations in the more restrictive models was more like a binomial distribution than a Gaussian distribution.

The number of latent variables in the multivariate calibration models varied from 5 to 20 (Table 3). High values of latent variables are characteristic of NIR-based modelling of soil samples according to the heterogeneity and complexity of this type of sample. Therefore, the accuracy of both the reference method and the NIR method was assessed at the level of intermediate precision, where samples over a large range of SOC concentrations representative of the calibration models were employed. The reference method presented an accuracy of 5.8%, whereas the accuracy of the NIR method ranged between 0.1% and 4.8%. These results demonstrate that the method based on NIR spectra was more accurate than the reference method for the determination of SOC. This result can be explained by the high repeatability of NIR spectrophotometers and by the errors in the reference method that arise from the steps of oxidation, digestion and titration with potassium dichromate.

The sensitivity (*SEN*), which is a fraction of the signal resulting in the addition of a unit of concentration of the properties of interest and is estimated by the inverse of the vector regression norm of the PLSR, showed values very close to zero (of the order of 10^{-5}). The results for sensitivity were not simple to judge because of the preprocessing adopted.[17] However, the analytical sensitivity (calculated by the relationship between the sensitivity and the instrumental noise) was more informative for comparison and for judging the sensi-

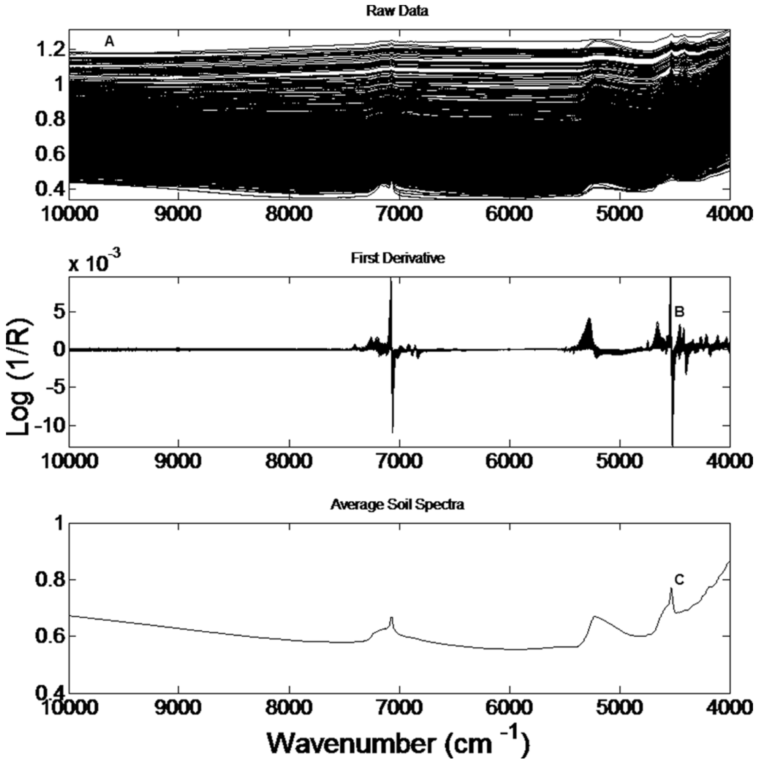

Figure 1. (A) NIR spectra of soil samples. (B) After preprocessing by the Savitzky–Golay first derivative with a 15 point window and second-order polynomial. (C) The average NIR spectra of the soil samples representative of the Brazilian territory.

tivity of a multivariate calibration method. The inverse of this parameter (analytical sensitivity^{-1}) permits one to establish a minimum concentration difference that is discernible by the analytical method in the range of concentrations where the minimum concentration difference was applied.

The model with the highest sensitivity was M4, which was capable of distinguishing a 0.1 g kg^{-1} concentration difference in SOC. Model M6 showed the lowest sensitivity because it could distinguish only 1.5 g kg^{-1} concentration differences between the soil samples.

The obtained *LD* and *LQ* values were adequate for the range of SOC concentrations employed in most calibration models, with the exception of model M0, because the lowest SOC concentration of this model was smaller than *LD* and *LQ*. The intervals of confidence found were also more suitable for those calibration models that had a larger range of SOC concentrations (less restrictive models).

The next section describes how these NIR spectrum calibration models built herein are employed to determine the SOC concentration of 12 samples of the proficiency assay programme. Finally, the models that are able to achieve an acceptable performance in the programme evaluation are considered as validated.

Validation results

Figure 2 shows the concentration profile of the SOM concentration (SOC concentration × 1724)[2] per sample calculated from the data evaluation of the proficiency assay programme of 2013.

To qualitatively estimate the representativeness of the proficiency assay samples, a principal components analysis (PCA) was conducted by joining the calibration and validation sets. Figure 3(A) shows the PC1 scores (61.5%) versus the PC2 scores (18.4%) for the calibration sets, validation sets and proficiency assay samples. The graph of the scores of the first two principal components demonstrated that the proficiency assay samples had a good representation on the calibration and validation datasets. However, the empty spaces on the right-hand side of PC1 indicated that the representativeness of the proficiency assay samples can be improved by inserting samples with higher SOM concentrations. The weights (or loadings) on PC1 (greater percentage of explained variance) shown in Figure 3(B) reveal the most important variable for the score configurations (Figure 2) that are characterised by CH bonding overtones of the aromatic groups (ν_1 C–H 7200 cm^{-1}, or ±1388 nm) and the combination bands of the fundamental transitions due to the aliphatic CH bond groups (ν_1 C–H 4600 cm^{-1}, or ±2173 nm). Assignments are as per Stenberg *et al.*[4]

Figure 4 shows the penalties for each calibration model in the proficiency assay sample evaluation: the number of simple asterisks [Figure 4(A)], the number of asterisks with repetition [Figure 4(B)] and the number of weighted asterisks [Figure 4(C)]. Some models were responsible

Table 3. Figures of merit, with outliers removed, that were used to evaluate the PLS calibration models.[14]

Figure of merit		**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**
Accuracy	*RMSEC*	2.7	0.8	0.8	0.8	1.0	2.0	5.0	2.4	2.6	2.5	2.3
	RMSEP	2.5	0.9	0.9	1.0	0.9	2.7	6.2	2.5	2.3	2.3	2.1
Coefficient of variation	*CV* (%)	36.4	12.6	7.3	5.5	4.0	9.2	15.6	16.1	21.4	21.5	23.6
RPD	*RPD*	2.1	1.1	1.1	1.1	1.0	0.9	1.6	1.4	1.9	1.9	2.2
Precision	(%)	0.5	0.4	0.3	0.3	0.1	0.2	2.1	1.1	0.6	0.6	0.5
Sensitivity	*SEN* (g kg^{-1})	4.1×10^{-5}	8.2×10^{-5}	1.5×10^{-4}	2.5×10^{-4}	7.2×10^{-4}	3.0×10^{-4}	3.4×10^{-5}	4.0×10^{-5}	4.7×10^{-5}	5.0×10^{-5}	5.1×10^{-5}
Analytical sensitivity	*SEN*(A) (g kg^{-1})	1.5	3.4	4.2	5.1	11.8	3.9	0.7	1.5	1.7	1.8	1.8
Inverse of analytical sensitivity	1/*SEN*(A) (kg g^{-1})	0.7	0.3	0.2	0.2	0.1	0.3	1.5	0.6	0.6	0.5	0.6
Limit of detection	*LD* (g kg^{-1})	2.1	0.9	0.7	0.6	0.3	0.8	4.4	1.9	1.8	1.6	1.7
Limit of quantification	*LQ* (g kg^{-1})	6.8	2.9	2.4	2.0	0.8	2.6	14.5	6.5	5.9	5.4	5.6
Adjustment	Coefficient of determination	0.83	0.65	0.61	0.69	0.67	0.70	0.86	0.75	0.77	0.80	0.83
Confidence interval	Maximum	5.5	1.7	1.8	1.8	2.3	5.1	12.7	5.2	5.6	5.1	4.7
	Minimum	5.3	1.5	1.6	1.6	2.0	4.2	10.3	4.8	5.1	4.9	4.5
Outlier analysis		**M0**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**
Latent variables	vl	20	20	13	9	6	5	9	17	17	17	18
Number of samples	cal	1986	484	260	138	81	36	54	528	1026	1026	1368
Number of samples	val	994	243	130	70	41	18	28	264	513	505	684
Outliers removed 1	cal	1772	405	217	116	70	32	47	463	908	898	1181
Outliers removed 1	val	936	210	106	49	26	2	16	246	481	470	632
Outliers removed 2	cal	1544	351	185	104	57	32	47	396	780	758	992
Outliers removed 2	val	871	188	73	32	19	2	16	222	441	430	575
Minimum *SOC* (g kg^{-1})	Min.	0.1	5.1	10.1	15.1	20.1	26.2	25.8	10.1	5.1	5.1	3.1
Maximum *SOC* (g kg^{-1})	Max.	36.4	9.9	14.9	19.9	24.9	37.5	66.5	34.4	34.4	34.4	34.4
Mean *SOC* (g kg^{-1})	Mean	7.7	7.1	12.2	17.3	22.0	31.4	36.4	16.0	11.7	11.6	9.5
Standard deviation	*SD*	3.6	1.0	1.0	1.2	1.4	2.7	4.3	3.3	3.7	3.7	3.5

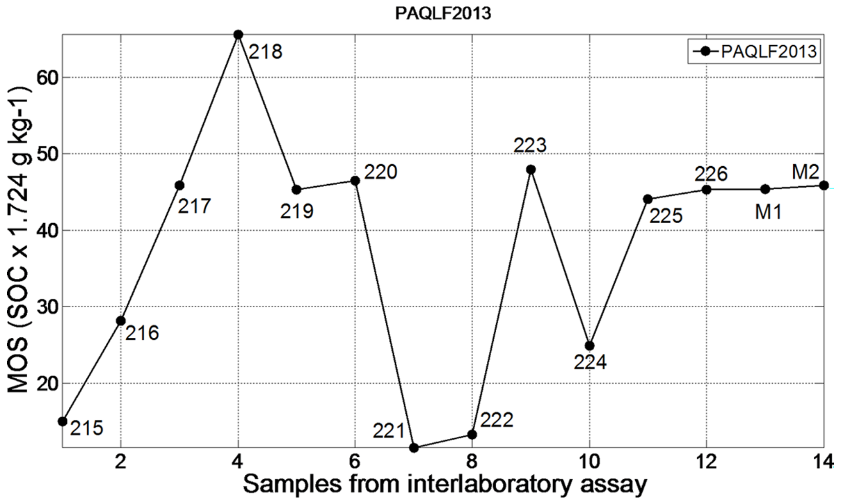

Figure 2. Average concentration of SOM for samples obtained from the results of 111 participating laboratories in the interlaboratory programme of 2013. The samples with repetitions (triplicate) are represented by M1 (217, 219 and 225) and M2 (220, 223 and 226).

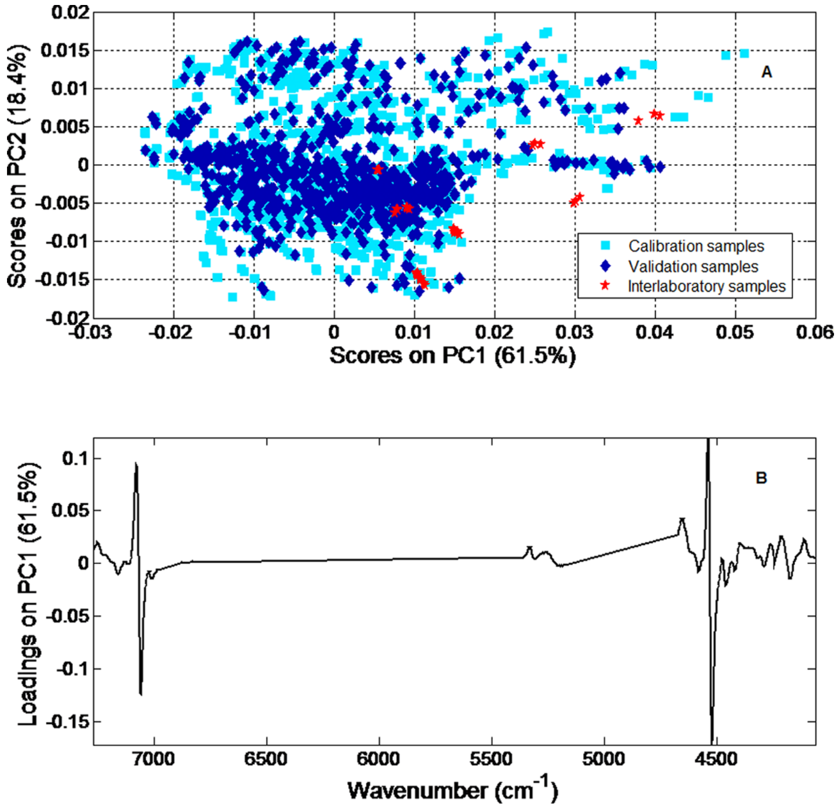

Figure 3. (A) PCA score plot of the first two principal components of the NIR spectra of calibration datasets, validation and 12 samples from the interlaboratory programme of 2013. (B) Loading graphics corresponding to the PCA plot.

for making the NIR calibration method outperform the reference method due to low penalties, where models M8 and M10 (Figure 5) presented the best performance. This observation can be supported by the diagnostic parameters presented in Table 4, where the best performances of models M8 and M10 were justified by their higher accuracy due to low penalties.

Conclusions

From the evaluation of the figures of merit, the calibration models constructed from the spectral database of representative samples of Brazilian soil types have a high potential for application. The results achieved in the validation step, which used the interlaboratory test to evaluate the reference

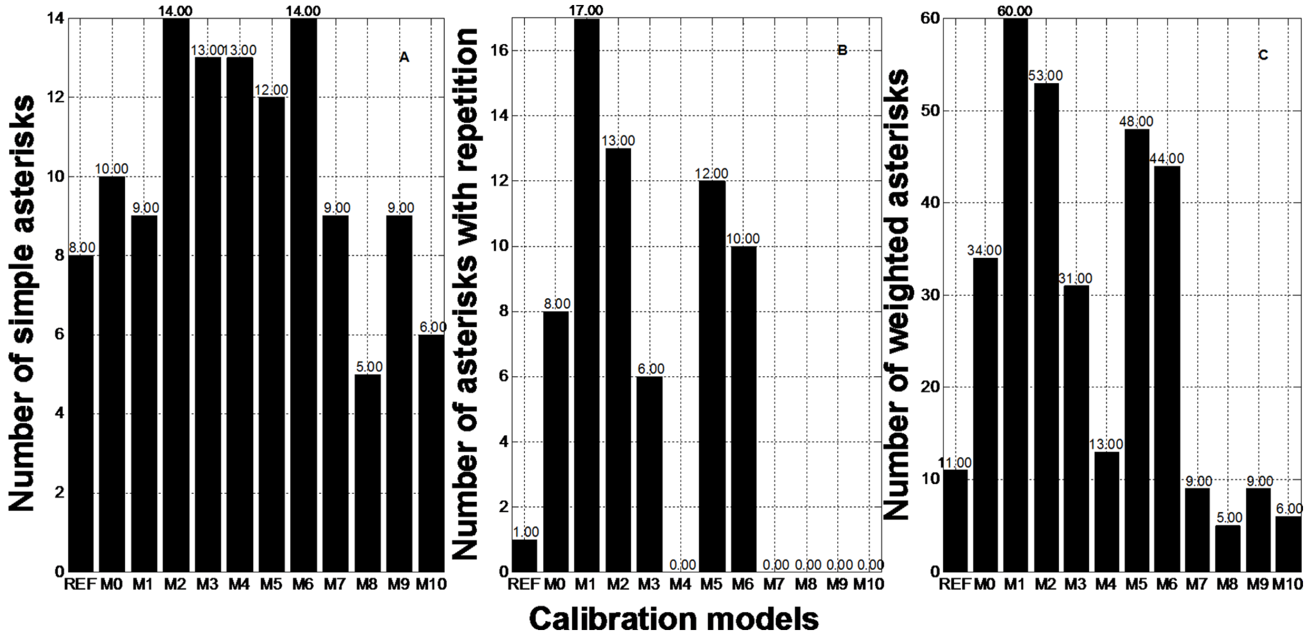

Figure 4. Bar graphs of the penalties for each calibration model submitted in the interlaboratory programme of 2013: (A) number of simple asterisks, (B) number of asterisks with repetition and C) number of weighted asterisks.

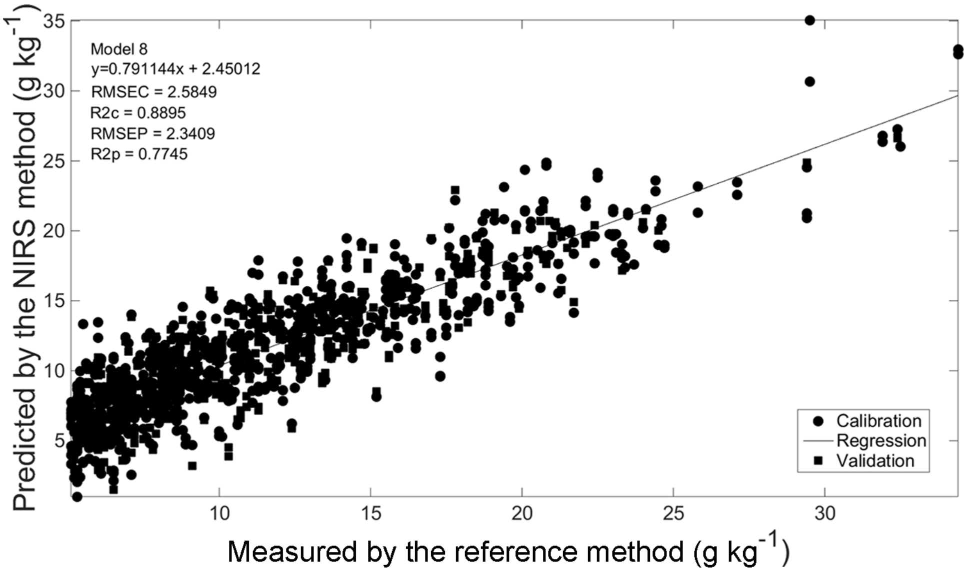

Figure 5. The concentrations of SOM measured by the reference method versus the concentrations of SOM predicted by NIR spectroscopy.

method that is employed by more than 100 laboratories in Brazil, show that the NIR spectroscopy method is a suitable replacement for the reference method and thus obtains approval in this proficiency programme. Furthermore, the calibration models were shown to provide even better results by creating local models to make a prediction for a specific sample. In addition, after implementing the NIR spectroscopy method in routine analysis, the spectral bank may grow daily and could reach even more significant levels of robustness.

Finally, the data treatment step is considered one of the bottlenecks in using a NIR spectroscopy method because the development of robust models is rather laborious and indispensable. Data analysis requires expertise in chemometrics to check that the best method is selected, and this fact hinders the implementation of NIR spectroscopy as a routine soil anal-

Table 4. Performance of each calibration model for predicting the SOM concentrations by NIR spectroscopy in samples from the interlaboratory programme of 2013 and the performance of the wet method (reference method).

Proficiency assay samples	Number of asterisks			Performance of calibration models				
Model	Simple asterisks	Asterisks with repetition	Weighted asterisks	Inaccuracy (%)	Imprecision (%)	Excellence index	Classification	Classification groups
Reference method	8	1	11	79	7	57	79	C
M0	10	8	34	243	2	–22	95	C
M1	9	17	60	429	1	–115	104	C
M2	14	13	53	379	1	–90	103	C
M3	13	6	31	221	1	–11	92	C
M4	13	0	13	93	0	53	80	C
M5	12	12	48	343	1	–72	103	C
M6	14	10	44	314	1	–58	102	C
M7	9	0	9	64	4	66	76	C
M8	5	0	5	36	2	81	63	B
M9	9	0	9	64	2	67	76	C
M10	6	0	6	43	2	77	67	B

ysis method. Those responsible for spectra-based analyses must have different technical skills compared with the skillset required by analysts traditionally employed in soil laboratories. Thus, issues such as the availability of professionals with knowledge of chemometrics and the costs involved in hiring and training these professionals should also be considered and evaluated when seeking to implement NIR spectroscopy as a routine soil analysis method in Brazil.

Acknowledgement

The authors acknowledge Embrapa for financial support (MP3 03.12.03.010.00.00).

References

1. H.G. dos Santos, P.K.T. Jacomine, L.H.C. dos Anjos, V.Á.V.L. de Oliveira, J.F. Lumbreras, M.R. Coelho, J.A. de Almeida, T.J.F. Cunha and J.B. de Oliveira, *Sistema Brasileiro de Classificação de Solos*, 3rd Edn. Embrapa, Brasilia – Distrito Federal (2013).
2. Embrapa, *Manual de Métodos de Análise de Solos*, 2nd Edn. Embrapa, Rio de Janeiro (2011).
3. J. Reeves, G. McCarty and T. Mimmo, "The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils", *Environ. Pollut.* **116,** S277 (2002). doi: http://dx.doi.org/10.1016/S0269-7491(01)00259-7
4. B. Stenberg, R.A.V. Rossel, A.M. Mouazen and J. Wetterlind, "Visible and near infrared spectroscopy in soil science", *Adv. Agronomy* **107,** 163 (2010). doi: http://dx.doi.org/10.1016/S0065-2113(10)07005-7
5. V. Bellon-Maurel and A. McBratney, "Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon in soils", *Soil Biol. Biochem.* **43,** 1398 (2011). doi: http://dx.doi.org/10.1016/j.soilbio.2011.02.019
6. D.W. Pribyl, "A critical review of the conventional SOC to SOM conversion factor", *Geoderma* **156,** 75 (2010). doi: http://dx.doi.org/10.1016/j.geoderma.2010.02.003
7. A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplifies least squares procedures", *Anal. Chem.* **36,** 1627 (1964). doi: http://dx.doi.org/10.1021/ac60214a047
8. I.-G. Chong and C.-H. Jun, "Performance of some variable selection methods when multicollinearity is present", *Chemom. Intell. Lab. Syst.* **78,** 103 (2005). doi: http://dx.doi.org/10.1016/j.chemolab.2004.12.011
9. L. Cécillon, N. Cassagne, S. Czarnes, R. Gros and J.-J. Brun, "Variable selection in near infrared spectra for the biological characterization of soil and earthworm ca sts", *Soil Biol. Biochem.* **40,** 1975 (2008). doi: http://dx.doi.org/10.1016/j.soilbio.2008.03.016
10. R.W. Kennard and L.A. Stone, "Computer aided design of experiments", *Technometrics* **11,** 137 (1969). doi: http://dx.doi.org/10.1080/00401706.1969.10490666
11. ASTM E1655-05. *Standard Practices for Infrared Multivariate Quantitative Analysis*. ASTM International, West Conshohocken, PA (2012).
12. T. Naes and H. Martens, "Multivariate calibration. II. Chemometric methods", *TrAC Trends Anal. Chem.* **3,** 266 (1984). doi: http://dx.doi.org/10.1016/0165-9936(84)80044-8
13. V. Bellon-Maurel, E. Fernandez-Ahumada, B. Palagos, J.-M. Roger and A. McBratney, "Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy", *TrAC Trends Anal. Chem.* **29,** 1073 (2010). doi: http://dx.doi.org/10.1016/j.trac.2010.05.006
14. P. Valderrama, J.W.B. Braga and R.J. Poppi, "Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy", *J. Agric. Food Chem.* **55,** 8331 (2007). doi: http://dx.doi.org/10.1021/jf071538s
15. P. Valderrama, J.W.B. Braga and R.J. Poppi, "Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy", *J. Braz. Chem. Soc.* **18,** 259 (2007). doi: http://dx.doi.org/10.1590/S0103-50532007000200003
16. P. Valderrama, J. Braga and R.J. Poppi, "Estado da arte de figuras de mérito em calibração multivariada", *Quim. Nova* **32,** 1278 (2009).
17. P. Valderrama, A.L. Romero, P.M. Imamura, I.R.S. Magalhães, P.S. Bonato and R.J. Poppi, "Quantification of chlorpheniramine maleate enantiomers by ultraviolet spectroscopy and chemometric methods", *Anal. Bioanal. Chem.* **397,** 181 (2010). doi: http://dx.doi.org/10.1007/s00216-010-3472-7
18. ASTM E177-10. *Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods*. ASTM International, West Conshohocken, PA (2010).
19. J.A. Quaggio, "Programa de laboratórios de análise de solo do Estadode São Paulo", *Bol. Inf. da Soc. Bras. Cienc. do Solo* **12,** 60 (1987).