Journal of the Optical Society of America B
OPTICAL PHYSICS

# Evaluation of univariate and multivariate calibration strategies for the direct determination of total carbon in soils by laser-induced breakdown spectroscopy: tutorial

Wesley Nascimento Guedes,[1] Diego Victor Babos,[1] Vinícius Câmara Costa,[2] Carla Pereira De Morais,[1] Vitor da Silveira Freitas,[1] Kleydson Stenio,[2] Alfredo Augusto Pereira Xavier,[2] Luís Carlos Leva Borduchi,[1] Paulino Ribeiro Villas-Boas,[1] and Débora Marcondes Bastos Pereira Milori[1,*]

[1]Embrapa Instrumentation, São Carlos, 13560-970, São Paulo, Brazil
[2]Agrorobótica, São Carlos, 13571-512, São Paulo, Brazil
*debora.milori@embrapa.br

Received 2 December 2022; revised 3 April 2023; accepted 4 April 2023; posted 12 April 2023; published 1 May 2023

**The correct choice of the calibration strategy method is a step that can ensure the accuracy of carbon determination in soils by laser-induced breakdown spectroscopy. In this paper, we evaluate and discuss the intrinsic characteristics of univariate and multivariate calibration strategies in overcoming soil matrix effects and their influence on the magnitude of calibration and prediction error values. The matrix-matching calibration with previous correction of Al spectral interference in the C I line at 193.03 nm was the best strategy of the 13 univariate models evaluated [root-mean-square error of calibration—(RMSEP) = 0.2 wt.% C]. In the multivariate evaluation of the 22 models obtained, the artificial neural network allowed obtaining a lower value of prediction error for C (RMSEP = 0.1 wt.% C) with very good accuracy (98%) for the set samples of validation. The observation of these results was justified considering the advantages and limitations of each calibration strategy explored.** 

https://doi.org/10.1364/JOSAB.482644

## 1. INTRODUCTION

Determining soil organic carbon (SOC) is essential due to its high storage potential. Soils can store carbon (C) amounts three times greater than the atmosphere and four times greater than plants [1,2] being very important for the $CO_2$ reduction in the atmosphere, minimizing the effects of global warming and climate change [2], increasing soil productivity, and moderating the water cycle [3]. Therefore, determining soil C content is the present and future focus of international negotiations and treaties related to global climate change.

For many years, the wet combustion method, Walkley–Black, based on the measurement of emitted $CO_2$ or sample weight changes, was the standard method used for measuring soil carbon [4]. Currently, dry combustion by an elemental analyzer is the official technique used to determine SOC [5,6] quantitatively. However, many other techniques have been used for soil C measurements, such as mid-infrared spectroscopy (MIRS) [7], near-infrared spectroscopy [8], inelastic neutron scattering (INS) [9], and laser-induced breakdown spectroscopy (LIBS) [5,10–12].

The methodology for improved agricultural land management [13], approved by Verra, recognized INS, LIBS, MIR, and visual-near infrared as emerging technologies to determine SOC content. However, LIBS has stood out in this scenario due to its inherent characteristics, such as minimal sample preparation, relative simplicity, high analytical frequency (on the order of seconds), multielement analysis, and relatively low maintenance cost [14]. Furthermore, the LIBS has the potential for miniaturization and application of portable instrumentation that can be used directly *in situ* for rapid measurement of soil samples [6,14].

LIBS is a versatile technique that enables multielement qualitative and quantitative analysis of solid, liquid, and gas samples. The interaction of the laser pulse with a solid sample, e.g., in the form of a pellet, leads to the ablation of micrograms of sample, rupture of chemical bonds in the material, and formation of a microplasma, whose initial temperature is close to 100,000 K [15,16]. The plasma formed is composed of several species (ions, atoms, molecules, and electrons) that collide with each other. The thermal energy of the plasma promotes the excitation of ions, atoms, and molecules present in the composition of

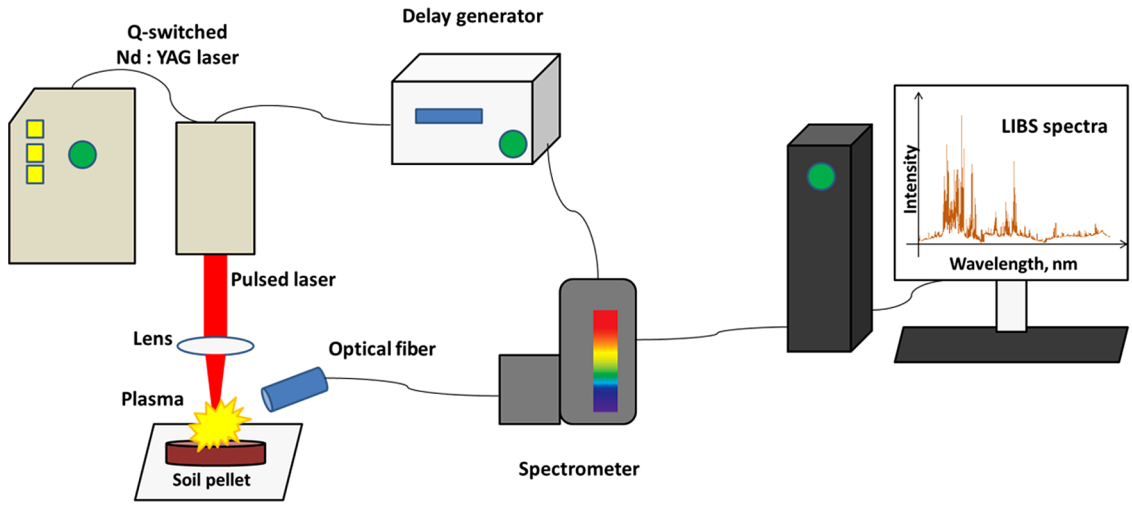


**Fig. 1.** Illustration of the main components of a LIBS system.

the sample. When returning to the ground state, each of these species emit its characteristic radiation, creating a fingerprint of the sample [17,18]. With the collection of this emitted radiation (with the cooling of the plasma formed (close to 10,000 K) and lower emission of continuum radiation), using an optical fiber and using a spectrometer, it is possible to obtain a LIBS emission spectrum, whose wavelengths refer to if the atomic–ionic-molecular emission characteristic of the elemental composition of the sample. The intensity and area of the lines (or bands, in the case of monitoring the emission of diatomic molecules) are proportional to the concentration of species present in the sample [15,16,19]. Using spectral pretreatments and appropriate calibration strategies, it is possible to obtain accurate and robust quantitative models and, thus, determine the elemental concentration [20].

A LIBS system basically consists of (i) a laser (such as a solid-state Nd:YAG laser), (ii) mirrors and lenses (which direct and focus the laser pulse on the sample), (iii) a sample holder, (iv) a plasma radiation collection system (fiber optics), (v) a spectrometer, a (vi) detector, as well as (vii) a delay generator (between the start of the laser pulse and the collection of the radiation emitted by the plasma). Figure 1 illustrates the main components of the LIBS system used in the analysis of soil samples.

The LIBS, when compared to other plasma-based spectroscopic elemental analysis techniques, such as inductively coupled plasma optical emission spectrometry (ICP OES), microwave-induced plasma optical emission spectrometry (MIP OES), and inductively coupled plasma mass spectrometry (ICP MS), has some advantages, such as (i) not requiring the solid sample to be converted into a solution to be introduced into the equipment, (ii) and, consequently, does not require oxidizing inorganic acids in the sample preparation step, (iii) minimal waste generation (the residue is the sample itself), and (iv) requires a smaller amount of sample for analysis (few micrograms of the sample). However, some limitations of the technique are mainly related to (i) matrix effects arising from the full sample analysis (which in some cases can severely compromise the accuracy of the determinations) [20], (ii) the difficulty of obtaining solid standards of calibration certificates for analysis of micrograms of sample, and (iii) high limits of detection (LOD) values of the method (which may preclude the use of LIBS in monitoring some elements in certain types of samples). However, these limitations can be circumvented using some strategies, such as the use of other reference techniques to obtain calibration standards, evaluation of univariate and multivariate calibration strategies, and use of double pulse systems and nanoparticles (to increase sensitivity and decrease LOD values).

Soil is a highly complex matrix due to several factors, such as moisture, physicochemical properties, soil organic matter, and inorganic C (in carbonate forms), being these some components that can insert uncertainties in C determination by LIBS [6]. Calibration curves built for C determination in soil samples by LIBS have been reported in the literature from emission intensity and/or area of the LIBS peak versus C content determined by an elemental analyzer [2]. However, due to above-mentioned matrix complexity, some univariate calibration strategies have been applied to improve the C determination, such as matrix-matching calibration (MMC), internal standardization (IS), standard addition, multienergy calibration, one-point calibration (OPC) [21,22]; and multivariate methods, such as principal component regression (PCR), partial least squares (PLS), support vector machine (SVM), multiple linear regression (MLR), and artificial neural networks (ANNs) [21,22].

Although many C emission lines are found in databases [23], those usually used to build calibration models for the C determination are at 193.03 nm and 247.86 nm [5,10,12,24]. This choice is because they are the strongest lines and have a greater emission probability (Einstein's coefficient). Still, when the analyses are performed under the air atmosphere, the line at 193.03 nm is attenuated by the $O_2$ absorption [15]. Furthermore, the C I emission lines at 193.03 nm and 247.86 nm are interfered by the Al II line at 193.04 nm and the Fe I line at 247.98 nm, respectively [5]. These interferences make it even more challenging to build calibration models to determine C in Brazilian soils as they have a high concentration of both elements. Another challenge for the development of

LIBS calibration models is the matrix effect, which is very accentuated in the LIBS technique, especially for complex samples, such as soils, which have varied compositions in terms of texture, granulometry, and chemical composition [25].

Thus, the objective of this paper is to present and compare several calibration strategies for determining C in soils via LIBS. It will be evaluated the linear univariate and multivariate models (MLR, PCR, and PLS). Also, the ANN will be used to minimize matrix effects, overcome spectral interference in the C lines, and mitigate self-absorption, aiming to improve sensitivity and precision. In addition, the advantages and limitations of each strategy will be discussed.

## 2. MATERIALS AND METHODS

### A. Sampling and Experimental Site

Soil samples were collected at EMBRAPA Pecuária Sudeste, a research center in São Carlos-SP, Brazil (GPS coordinates: −21.956094316858977 and −47.84542273769212). The study site includes four livestock managements and an adjacent native forest region. The soil collection was performed at eight depths from 0 cm to 100 cm with six replicates per management, totalizing 240 soil samples with C content ranging from 0.50 wt.% to 5.0 wt.%.

### B. Samples Characterizations and Pretreatments

The soil samples were sent to the laboratory and air dried until constant mass. First, the soil texture was determined by the reference method (Robinson pipette method). The purpose is to evaluate the distribution by size of particles smaller than 2 mm using Stokes' law, that is, the ratio between particle size and sedimentation rate [26]. Then, the soil samples were classified based on the United States of Agriculture (USDA) textural triangle (Fig. 2).

The textural triangle can verify an extensive range of sample textures. The collection field is located at a transition soil area between a red Latosol and red Alfisol. Beyond the texture characterization, the total C content was determined using dry combustion by 2400 CHNS elemental analyzer by Perkin-Elmer as described by Nicolodelli *et al*. [5].

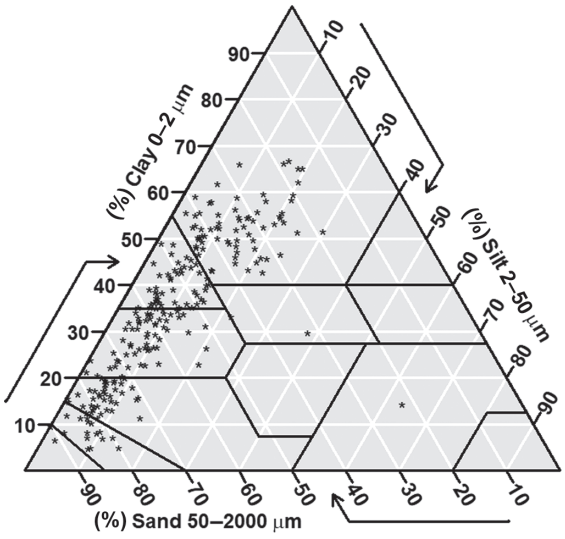


**Fig. 2.** Classification of all soils set by the USDA textural triangle.

The elemental quantification using the LIBS technique has a high dependency on the matrix, then the sample preparation is considered an essential step in the whole analysis process [6,10]. First, the samples at 2 mm of particle size were crushed with mortar and pestle, then passed through a 100 mesh (0.150 mm) sieve. Finally, soil pellets with radii of 1 cm and weight of 500 mg were prepared using a hydraulic press.

### C. LIBS Instrumentation and Measurements Conditions

The LIBS system used in this paper belongs to Embrapa Instrumentation, a research center located in São Carlos. It is a commercial LIBS system manufactured by Ocean Optics, LIBS2500+ model, equipped with a *Q*-switched Nd:YAG laser at 1064 nm produced by Quantel, Big Sky Laser Ultra 50 with a laser pulse of 50 mJ energy, 0.5 mm of laser spot, and duration of 8 ns. The LIBS2500+ has seven spectrometers gifted with a charge-coupled device (CCD) array. It provides spectral analysis across a wide 188–980 nm range at a resolution of approximately 0.1 nm resolution (full width half maximum). The acquisition conditions were delay time (*Q*-switch delay correlated) and integration time used were 10 μs and 2 ms, respectively.

For each sample, 60 spectra were acquired to reduce the variability among the shots (matrix heterogeneity). Each spectrum obtained corresponds to three shots accumulated, one cleaner shot and two shots generating an average spectrum.

In this paper, 20 spectra were collected for each soil sample (each spectrum obtained referring to the average of two individual spectra) in different regions of the pellet to get a representative analysis. This process was repeated three times, totaling 60 spectra per sample.

### D. Spectral Pretreatment

Spectral deconvolution was performed to minimize the inherent spectral interference of the C I line at 193.03 nm line by the Al lines [5,12]. Figure 2 shows a typical LIBS spectrum obtained by commercial LIBS 2500+. The C I line at 193.03 nm and the interferences of the Al lines are highlighted at the upper inset in Fig. 3.

Emission lines used to build all calibration models were C I at 193.03 nm, Al I at 193.58 nm, Al II at 198.99 nm, Si I at 212.41 nm, Mg II at 279.55 nm, Mg II at 280.27 nm, and Mg I at 285.21 nm. Furthermore, two different methods were used to remove outliers: the spectral angle mapper (SAM) and median absolute deviation (MAD) [27,28]. Both ways are explained by Stenio *et al*. [12]. In this paper, we used SAM and MAD to identify spectra called outliers, eliminating an average of two anomalous spectra per sample [27]. After the outlier's elimination, the remaining spectra were used to compose the average spectrum. In addition, spectral baseline correction was performed. Python programming language was used for the spectral pretreatments.

A MATLAB software routine (version 2010—MathWorks, Natick, MA, USA) was developed to calculate the area and

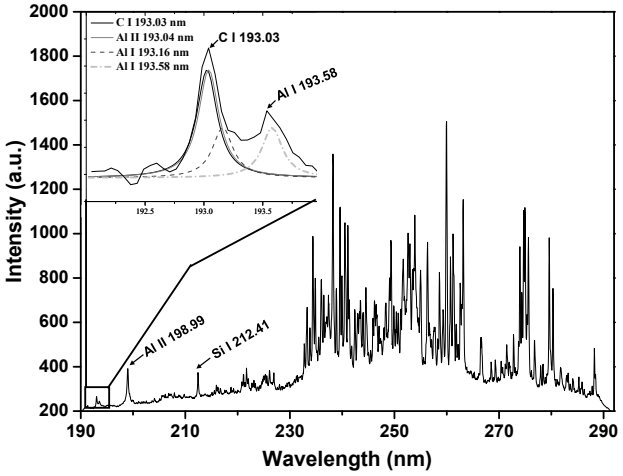


**Fig. 3.** LIBS spectral profile ranging 189 nm–293 nm with C I at the 193.03 nm peak highlighted at the top of the figure.

intensity of the C I line at 193.03 nm. Furthermore, eight spectral treatments (STs) to compensate for fluctuations in the analytical signal were evaluated [29], namely, ST 1 and ST 5: average and sum of the spectrum, respectively; ST 2 and ST 6: each individual spectrum is divided by its Euclidean norm. After this, the average (2) or sum (6) is calculated; ST 3 and ST 7: Each individual spectrum is divided by its area (sum of all signal intensities). After this, the average (3) or sum (7) is calculated; and ST 4 and ST 8: Each individual spectrum is divided by its maximum intensity value. After this, the average (4) or sum (8) is calculated.

The spectral treatment that provided the best analytical performance parameters for the models evaluated was ST 1 (average of the spectra). Among the spectral treatments evaluated, ST 1 provided satisfactory accuracy and precision in the C determinations. However, depending on the calibration strategy, using the C line's area or intensity allowed us to obtain a model with better predictive capacity.

### E. Proposition of Calibration Models

Univariate calibration approaches normally use information as line intensities, area, or intensity ($y$ axis) against the reference values ($x$ axis), determined by the elemental analyzer (CHNS) to obtain a regression. Thus, 13 models were calculated evaluating the strategies: (a) MMC: MMC is a simple and widely used strategy in proposing quantitative models using LIBS. This strategy uses a set of samples as calibration standards, previously analyzed by another technique to obtain a reference value for the C content in the soil (in this case, dry combustion with an elemental analyzer—CHNS). This method uses the approximation between the samples matrix to reduce signal oscillations caused by the sample surface variabilities [30]; (b) IS: In this case, an element present in the samples with known concentration is used to normalize the LIBS spectra; (c) inverse regression (IR): the least-squares regression is obtained using the intensity of the C content and other emission lines acquired by LIBS as the independent variable ($x$ axis) and reference values of the C content as the dependent variable ($y$ axis) [22]. IS with IR (IS–IR) and (e) MMC with spectral correction (SC): uses emission from the C I line at 193.03 nm, which is highly interfered with by emissions from the Al I line at 193.58 nm and the Al II line at 193.04 nm. However, considering that the interference is linear, it is possible to perform the deconvolution of C and Al areas and use a correction method proposed by Nicolodelli *et al*. [5] to remove the Al interference in the C line, according to Eq. (1),

$$A_{\mathrm{C}} = A_{\mathrm{C+Al}} - \alpha_1 A_{\mathrm{Al}}, \tag{1}$$

where $A_{\mathrm{C}}$ is the carbon area, $A_{\mathrm{C+Al}}$ is the interfered area of C + Al, $\alpha_1$ is a correction coefficient obtained by optimizing the best determination coefficient ($R^2$) of the calibration curve, and $A_{\mathrm{Al}}$ is the area of the Al (Al I or Al II) used as a reference. It is also possible to perform the correction using the two Al species where one more correction coefficient $\alpha_2$ is introduced, and two variables are optimized as presented in Eq. (2),

$$A_{\mathrm{C}} = A_{\mathrm{C+Al}} - \alpha_1 A_{\mathrm{Al\,I}} - \alpha_2 A_{\mathrm{I\,II}}. \tag{2}$$

After univariate approaches, four multivariate calibration strategies were evaluated: MLR [31], PCR [32], PLS [32], and ANNs [31,33].

The 3D plane model is a method for determining C in soils with wide textural variability using the LIBS spectrum [12]. In this case, the peak information was used, such as the intensity and area, to fit a regression model applying a 3D plane. First, in this approach, it was used on the C + Al peak at 193.03 nm and Al peak at 193.58 nm. It uses MLR for the case where there are only two variables: the intensity of a C emission and the intensity of the interference (which might be Al or Fe). Using that information makes it possible to generate a plane in space (3D) and, from its equation, make predictions with good recoveries and few errors.

PCR and PLS were also evaluated as multivariate strategies. Briefly, the PCR algorithm will transform the data matrix containing the independent variables ($X$) into a new calculated matrix with latent variables, which concentrates all the information regarding the samples in a small set of vectors called principal components. On the other hand, such as PCR, the PLS algorithm transforms the data matrix containing the independent variables into a more compact one with latent variables. Still, the difference is that the variables are calculated to best explain the information regarding dependent ($Y$) and independent ($X$) variables [32].

ANN is a logical-mathematical structure run with "machine learning" composed of several processing elements that seek to mimic the processing and functions of the brain. ANNs consist of a set of processing units called artificial neurons, a group of synapses characterized by a weight ($w$) with an individual value for each input. An adder, also known as a linear combiner, whose function is to add the input signals weighted by the respective synaptic weights of the neurons in the input layer. The bias increases or decreases the values obtained in the adder for the activation function; the activation function normalizes the permissible range of the values obtained by the adder and the output layer ($y$) that receives the activation function values [33]. ANN was processed in Python using machine learning from the Keras framework.

Accuracy results for all models (univariate and multivariate) were defined by recovery range between 80% and 120% according to Eq. (3),

$$\text{Recovery}(\%) = \frac{\hat{y} \cdot 100}{y_i}, \tag{3}$$

where $y_i$ is the C content determined by reference technique and $\hat{y}$ is the C content estimated by the prediction model.

Other parameters, such as Pearson's correlation coefficient ($r$), root-mean-square error of calibration (RMSEC) and root-mean-square error of prediction (RMSEP) were calculated to evaluate the quality of the generated models. The results were evaluated in terms of RMSEP, calculated from Eq. (4),

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}, \tag{4}$$

where $y_i$ is the C content determined by reference technique, $\hat{y}_i$ represents the C content estimated by prediction model, and $n$ is the total number of samples used in the prediction set.

## 3. RESULTS AND DISCUSSION

### A. Spectral Pretreatments: Contributing to the Accuracy and Precision of Measurements

Instrumental conditions can influence the acquisition of LIBS emission spectra from each sample. They must always be optimized considering the intrinsic characteristics of the set of soil samples that will be analyzed and the LIBS system used. In addition, matrix effects can significantly influence the acquisition of analytical signals from the analyzed sample. To minimize spectral fluctuations, some procedures must be used before and during the acquisition of spectra, such as standardization of granulometry, humidity, and pressure used in the preparation of the pellets of soil samples [22]; verification of the stability of the laser source (constant laser pulse energy), immobile optical parameters (the distance from mirrors, focusing lenses, and the collection optical fibers must be immobile during the entire spectra collection procedure), among other precautions.

However, even taking all these precautions when acquiring the data, some fluctuations in the emission spectra may occur. For this, spectral preprocessing should always be used, mainly for quantitative purposes, to minimize these instrumental fluctuations and some matrix effects in acquiring analytical information from the sample [34,35]. Thus, the identification and removal of anomalous spectra (outlier), spectral normalization, baseline correction, and correction of spectral interference are necessary preprocessing for the use of the area or intensity of a C emission line to be used in the proposition of calibration models.

Considering that the analyte is homogeneously distributed in the soil sample, it is possible to obtain reproducible emission spectra without significant fluctuations in the emission lines. However, in some regions of the sample pellet, a high concentration of C, called a "hotspot," can influence the spectral quality. In addition, the nonrecording of a spectrum or spectrum collected in a hotspot can generate spectra called outliers and must be eliminated from the spectra obtained. Some statistical tools can be used to overcome these situations and help to remove specific spectra, such as the Grubbs test, SAM [27], and MAD [12], among others. Another spectral pretreatment that can be applied and provided improvement in the results of calibration models was the baseline correction.

Besides, it is always necessary to evaluate different spectral pretreatments and kinds of signals because, depending on the intrinsic characteristics and complexity of the samples (for example, soil) and the analyte emission line, a spectral treatment can further the achievement of better statistical quality calibration models. To summarize, the performance of calibration models is closely related to the spectral pretreatment used.

### B. Univariate Calibration Approaches

Matrix-matching calibration-MMC and IS (both using least-squares regressions with direct regression or IR) were evaluated in univariate modeling. Thirteen models were obtained using these two strategies, differing in pretreatments and choice of emission lines in spectral normalization (for IS) as presented in Table 1.

In the MMC standards make the sample matrix compatible, and to be efficient it is necessary that the physicochemical properties of the soil samples are similar. Applying this strategy with direct regression, low accuracy in C determinations were obtained with high values of RMSEC (0.8 wt.% C) and RMSEP (1 wt.% C). Since the samples used in this paper presented differences in their compositions in texture (Fig. 2), depth of collection and soil management practices, low accuracy in C determinations is justified, demonstrating that MMC with direct regression was inefficient to overcome matrix effects.

However, in addition to the intrinsic characteristics of MMC, the choice of regression also influenced the accuracy of the determinations. Using MMC–IR RMSEC (0.5 wt.% C) and RMSEP (0.5 wt.% C) values were 38% and 50% lower, respectively, than those obtained using direct MMC regression (Table 1). These values show the importance of choosing the dependent and independent variables in the modeling used in the least-squares regression. Using IR (an independent variable as analytical signal measured by monitoring the C I line at 193.03 nm and a dependent variable as the C content obtained by the reference technique) enabled better accuracy. This indicates that the independent variable presents a smaller source of errors when compared to the C values obtained by the reference technique. Even taking care in homogenization, weighing the soil sample (approximately 10 mg) for analysis by dry combustion; the measurement of the analytical signal by LIBS (emission intensity) has a smaller source of errors. Thus, the MMC–IR allowed smaller errors in the determination when compared to the MMC with direct regression.

The C emission line at 193.03 nm is interfered with an Al II line at 193.04 nm and two Al I lines, one at 193.16 nm and another at 193.58 nm as can be seen in Fig. 3. Although the emission line at 193.58 nm can be obtained with a peak fit algorithm, the other two Al lines cannot be separated from the C I line at 193.03 nm. Knowing these spectral interferences, three MMC–SC models were built based in model proposed by Nicolodelli *et al*. [5], [Eq. (1)] along with the emission line of Al I line at 193.58 nm, an accuracy of 52% (recovery range

**Table 1. Univariate Calibration Models Generated from the Diverse Strategies**

| Parameter | MMC C 193 | MMC–IR C 193 | MMC–SC C 193+ Al 193 | MMC–SC C 193+ Al 198 | MMC–SC C 193+ Al 193+Al 198 | IS–C 193/ Al 193 | IS–IR C 193/Al 193 | IS–C 193/ Al 198 | IS–IR C 193/Al 198 | IS–C 193/ Si 212 | IS–IR C 193/Si 212 | IS–C 193/ Al 193+Si 212 | IS–IR C 193/ Al 193+Si 212 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Signal** | **Area** | **Area** | **Area** | **Area** | **Area** | **Intensity** | **Intensity** | **Intensity** | **Intensity** | **Intensity** | **Intensity** | **Intensity** | **Intensity** |
| $r$ calibration | 0.6182 | 0.6182 | 0.6368 | 0.8700 | 0.8962 | 0.4823 | 0.4823 | 0.8453 | 0.8453 | 0.5549 | 0.5549 | 0.8494 | 0.8494 |
| $R^2$ calibration | 0.3821 | 0.3821 | 0.4055 | 0.7569 | 0.8032 | 0.2326 | 0.2326 | 0.7145 | 0.7145 | 0.3079 | 0.3079 | 0.7215 | 0.7215 |
| RMSEC (%) | 0.8 | 0.5 | 0.5 | 0.3 | 0.3 | 1 | 0.6 | 0.4 | 0.4 | 1 | 0.5 | 0.4 | 0.3 |
| $r$ validation | 0.3100 | 0.3100 | 0.5200 | 0.7300 | 0.8300 | 0.4717 | 0.4717 | 0.8093 | 0.8093 | 0.3311 | 0.3311 | 0.6785 | 0.6785 |
| $R^2$ validation | 0.0961 | 0.0961 | 0.2704 | 0.5329 | 0.6889 | 0.2225 | 0.2225 | 0.6549 | 0.6549 | 0.1096 | 0.1096 | 0.4603 | 0.4603 |
| RMSEP (%) | 1 | 0.5 | 0.4 | 0.3 | 0.2 | 1 | 0.4 | 0.4 | 0.3 | 1 | 0.5 | 0.4 | 0.4 |
| Accuracy[a] | 8 (17%) | 20 (42%) | 25 (52%) | 32 (67%) | 39 (81%) | 6 (13%) | 17 (35%) | 23 (48%) | 31 (65%) | 8 (17%) | 22 (46%) | 14 (29%) | 28 (58%) |

[a]Recovery range values between 80% and 120% for the validation set (48 soil samples).

values between 80% and 120%) was obtained for 25 of 48 samples in the validation set. The employment of the Al II line at the 198.99 nm line, resulted in an accuracy improvement to 67% (32 samples) as shown in Table 1. The best model was generated with the two emissions lines Al I line at 193.58 nm and Al II line at 198.99 nm achieving $r_{val} = 0.83$, $R^2_{val} = 0.69$, RMSEP = 0.2 wt.% C and 81% accuracy.

Since then, the accuracy used the Al II emission line at 198.99 nm was greater than the Al I line at 193.58 nm, Al II is likely the most responsible for the interference, primarily due to the emission at 193.04 nm. The model improvement when using the two emissions lines is due to the fact that both Al species participate in the calculated area at 193.03 nm for C content predictions. It is possible to build a model more robustly and guarantee that the calculated peak area has a high correlation with the C content by eliminating the area contribution from both species [5]. However, the model for the interfered C I line with the Al I and II lines is still susceptible to matrix effects (line intensity dependent on the sample composition) and Al ionization degree (which changes the proportion of the Al I and Al II line intensities to the peak at 193.03 nm).

Another strategy widely used in calibration methods is IS. Using the IS, spectral normalization of the C I line at 193.03 nm can minimize matrix effects and signal bias due to instrumental fluctuations and laser-sample interaction. An excellent internal standard, in principle, should present a known concentration in all soil samples, have similar physicochemical properties, be affected by the same fluctuations during instrumental measurements, and be measured simultaneously with the analyte [22]. However, choosing the IS and sample preparation (addition of the internal standard to the sample via oxide or salt and subsequent homogenization) can be time consuming in analyzing solid samples. Thus, a strategy to speed up the analysis stage evaluated the use of major chemical elements in the composition of all soil samples. In this paper, Al and Si present in soil samples as oxides (aluminosilicates) were evaluated.

The profile of the lines evaluated as IS for the peak at C I 193.03 nm are shown in Fig. 3. Note that these are lines that are in a close spectral region (193 nm–212 nm) and that have different relative intensities. The excitation (Exc.) and ionization (Ion.) energies for the lines are C I at 193.03 nm (Exc.: 6.42 eV and Ion.: 11.26 eV), Al I at 193.58 nm (Exc.: 6.41 eV and Ion.: 5.98 eV), Al II at 198.99 nm (Exc.: 6.22 eV and Ion.: 18.82 eV), and Si I at 212.41 nm (Exc.: 5.83 eV and Ion.: 8.15 eV) (NIST Atomic Spectra Database Lines Data).

The profile and intensities of the three lines evaluated in the spectral normalization are different, and indeed the Al I line at 193.58 nm as it is partially superimposed on the C line requires more attention in the steps of baseline correction and calculation of the area or intensity. These pretreatments can significantly influence the quality of the normalization of the C I line at 193.03 nm. However, the Al II line at 198.99 and Si I line at 212.41 nm show good spectral and relative intensity profiles.

Among the eight models calculated using IS as a calibration strategy, all models from IR had lower calibration errors (0.3 wt.% C $\leq$ RMSEC $\leq$ 0.6 wt.% C) and were superior in predictive capacity (0.3 wt.% C $\leq$ RMSEP $\leq$ 0.5 wt.% C) when compared to direct regression (due to the same fact previously explained for MMC).

Although the Al I line at 193.58 nm presents excitation energy similar to the C I line at 193.03 nm, this model gave one of the most significant calibration errors for IS models. This result can be related to the spectral profile of the line, which may make it difficult to correct the baseline and accurately calculate the area or intensity of the signal. However, even using this line for spectral normalization, the RMSEP value (IS–IR: 0.4 wt.% C) was 20% lower than the value obtained in MMC–IR (0.5 wt.% C) without normalization, indicating an improvement in the model by using spectral normalization and its ability to improve the accuracy of determinations.

In addition to choosing the element to be used as an IS, selecting the element's line used in the normalization can also significantly influence the results, as in the case of Al. The Al atomic and ionic lines allowed to obtain models with different adjustments [$R^2 = 0.2326$ (IS–IR Al at I 193.58 nm) and $R^2 = 0.7145$ (IS–IR Al II at 198.99 nm)] and predictive capabilities (Table 1) for the same set of samples used in calibration and validation. The errors associated with the calibration and validation of models using the Al II line at 198.99 nm in the normalizations when compared to the Al I line at the 193.58 nm line were 33% and 25% lower, respectively.

Using the sum of the intensities of the Al II line at 198.99 and the Si I line at 212.41 nm for spectral normalization allowed us to obtain errors minor than 0.4 wt.% C. Furthermore, this strategy permitted evaluation of the contribution of these lines in minimizing the matrix effects since they are significant elements in the matrix of tropical soils.

Among all the internal standards evaluated (Al and Si), the Al I line at 198.99 nm ionic line showed the lowest errors associated with calibration (0.4 wt.% C) and prediction (0.3 wt.% C) of total C (65% of the samples in the validation set showed recovery range between 80% and 120%) (Table 1). The line has similar excitation and ionization energy values (thus, it can undergo fluctuations similar to the C in the formed plasma), excellent sensitivity, and spectral profile. Furthermore, as the C I line at the 193.03 nm line is interfered with Al emission lines, part of this interference is removed by normalizing the C line by this line. The choice of this line is then able to overcome matrix effects (such as spectral fluctuations and soil sample heterogeneity) and minimize spectral interference.

These results indicate the absolute need for and importance of the right element and emission line of this element to be used in spectral normalizations because they significantly influence the accuracy and the predictive capacity of the IS model.

### C. Multivariate Calibration Approaches

Several multivariate calibration strategies were evaluated, such as MLR, PCR, PLS, and ANN, aiming for total C determination. In addition, five MLR strategies were evaluated: MLR (1), using the peak intensity of C I line at 193.03 nm, plasma temperature (PT) index and self-absorption (SA) index as independent variables; MLR (2) using the same variables as MLR (1) and adding sand ($g\,kg^{-1}$); MLR (3) using peak intensity of the C I line at 193.03 nm, the Al II line at 198.99 nm, the Si I line at 212.41 nm, PT and SA indices; MLR (4) using the same variables as MLR (3) and adding sand ($g\,kg^{-1}$); and MLR (5) using the C + Al peak at 193.03 nm and the Al I peak at 193.58 nm on a 3D plane model proposed by Stenio *et al*. [12].

PT and SA indices were calculated using the ratios of Mg II at 279.55 nm/Mg II at 280.27 nm lines and Mg II at 280.27 nm/Mg I at 285.21 nm lines, respectively [36]. Both indices were included as independent variables in the MLR models to minimize their effects on the spectral profiles. As a result, the model's accuracy can increase significantly with a precise determination of the self-absorption coefficient [20].

In addition, the variable sand, which is a component of soil texture, was included in models aiming to bring information about the matrix since it inherently correlates with Si because most sand particles are derived from silica. Cousin *et al*. [37] also found correlations between Si and sand in soil analyses.

The parameters for each MLR model are shown in Table 2. Adding more independent variables to each MLR model significantly improved all the calculated parameters. The worst model (MLR 1) presented calculated $r$ (0.3574), $R^2$ (0.1277), RMSEP (0.5 wt.% C), and the best model (MLR 4) presented calculated $r$ (0.9289), $R^2$ (0.8628), and RMSEP (0.2 wt.% C).

The results show the accuracy within the validation set (48 soil samples) and how the model becomes more precise, following the order: MLR 1 = MLR 2 < MLR 5 < MLR 3 = MLR 4. For MLRs 1 and 2, only 19 samples were into recovery range between 80% to 120 %, whereas, the 3D model (MLR 5) presented 29 samples. In contrast, MLR 3 and MLR 4 showed the best accuracy (85%) with 41 of 48 soil samples into recovery range.

The reduction of RMSEP from 0.5 wt.% C in the worst model to 0.2 wt.% C in the best model, a prediction error 2.5 times lower, and an increase in an accuracy from 40% to 85% reflect the different approaches in each calibration model. For example, MLR 1 started with three variables, C I line at 193.03 nm, PT, and SA indices. In the MLR 2 model was added sand ($g\,kg^{-1}$). MLR 3 used five variables: C I at 193.03 nm, Al I at 193.58 nm, Al II at 198.99 nm, Si I at 212.41 nm lines, PT and SA. The best model (MLR 4) was built with all the previous variables: C I at 193.03 nm, Al I at 193.58 nm, Al II at 198.99 nm, Si I at 212.41 nm lines, sand ($g\,kg^{-1}$), and PT and SA indices were fitted into the model.

The MLR 5 was built with a different approach from the other MLR models. After removing outliers and performing baseline correction, the intensity values of the C + Al peak at 193.03 nm and the Al peak at 193.58 nm were calculated [12]. Then, the Linest [38] routine was used to estimate the plane's coefficients that best fit the data and with the adjusted values [12]. The correlation between the reference values predicted by LIBS for the calibration and validation sets were 0.89 and 0.79, respectively. The low values of RMSEC and RMSEP show the effectiveness of the 3D model for C prediction with an accuracy of 60% for soil samples with different textures as presented in Table 2.

For the presented models, adding Al and Si lines and sand ($g\,kg^{-1}$) as independent variables bring into the model's information about the matrix, resulting in more robust predictions and higher accuracy. The advantage of MLR models over univariate models is that in MLR, more emission lines can be used to improve calibration [39]. Still, one important limitation must be considered: The sample number must be greater than

**Table 2. Multivariate Calibration Models Obtained from MLR**

| Parameter | MLR (1) | MLR (2) | MLR (3) | MLR (4) | MLR (5) |
|---|---|---|---|---|---|
| Variables | C 193<br>PT index<br>SA index | C 193<br>PT index<br>SA index<br>sand | C 193<br>Al 198<br>Si 212<br>PT index<br>SA index | C 193<br>Al 198<br>Si 212<br>PT index<br>SA index<br>sand | C + Al peak at 193.03<br>+<br>Al I peak at 193.58 (3D) |
| $r$ calibration | 0.6340 | 0.6394 | 0.9182 | 0.9372 | 0.89 |
| $R^2$ calibration | 0.4019 | 0.4088 | 0.8430 | 0.8785 | 0.79 |
| RMSEC (%) | 0.5 | 0.5 | 0.3 | 0.2 | 0.3 |
| $r$ validation | 0.3574 | 0.3965 | 0.9042 | 0.9289 | 0.79 |
| $R^2$ validation | 0.1277 | 0.1572 | 0.8175 | 0.8628 | 0.62 |
| RMSEP (%) | 0.5 | 0.5 | 0.2 | 0.2 | 0.3 |
| Accuracy[a] | 19 (40%) | 19 (40%) | 41 (85%) | 41 (85%) | 29 (60%) |

[a]Recovery range values between 80% and 120% for the validation set (48 soil samples); PT; SA.

the number of variables [27]. These characteristics justify the better performance of the MLR 4 model and the lower value of RMSEC (0.2 wt.% C). Using the variables, which bring information from C, spectral interferences (Al), soil texture (Si and sand), and parameters of plasma information and matrix effects (PT and SA indices), this MLR model presented lower calibration error, among the MLR models compared, Table 2.

Following the MLR models, PCR and PLS models were calculated. The variables used to calculate all PCR and PLS models are presented in Table 3. The C I at 193.03 nm, Al I at 193.58 nm, Al II at 198.99, Si I at 211.41 nm lines, PT and SA indices, and texture [sand, clay, and silt (g kg$^{-1}$)] were used as independent variables for the proposition of the calibration models. Although the five PCR models showed lower accuracy (77%–79%) when compared with the best MLR model (MLR 4, 85% accuracy), the best PCR model (PCR 4) presented as an efficient method, achieving 79% accuracy with parameters $r_{val}$ (0.9374), $R_{val}$ (0.8787), RMSEP (0.2 wt.% C), and 38 of 48 samples into recovery range between 80% and 120% (Table 3), reflecting the importance of variable selection to modeling improvement.

Although the RMSEP values of the PCR and PLS models have not varied, the accuracy of the PLS models ranged from 79% to 90%, being superior to the PCR models as presented in Table 3. PLS 1, which uses all the aforementioned independent variables, is the accurate among the five proposed PLS models. The use of all the variables together, mainly the Si I line at 212.41 nm, which is a variable inherent to the sand can have produced an antagonistic effect, generating the model (PLS 1) with lower accuracy (79%). This observation is highlighted when the same variables (C I 193.03 nm, Al I 193.58 nm, and Al II at 198.99 nm lines) were used to build the PLS 5 adding only the texture was achieved the highest accuracy (90%) among all models PLS. The PLS 5 model showed the best results $r_{val}$ (0.9403), $R^2_{val}$ (0.8842), RMSEP (0.2 wt.% C), and 40 of 48 samples in the recovery range. Furthermore, a single PLS model can relate the independent variables (contained in the $X$ matrix) with one or more response variables ($Y$), overcoming the PCR modeling necessity to build one model for each ($Y$) response variable [32].

Finally, seven models from ANN were built using independent variables around the C, Al, and Si lines (C I at 193.03 nm, Al I at 193.58 nm, Al II at 198.99, and Si I at 211.41 nm), and PT and SA indices. ANN models showed better predictive ability in comparison with those multivariate calibration strategies (MLR, PCR, and PLS) evaluated.

Accuracy for all ANN models was equal or greater than 90% with RMSEP values very low, ranging from 0.2 wt.% C to 0.1 wt.% C. An overview of the ANN models for the validation set with 48 samples, evaluating the accuracy, follows the order ANN 7 = ANN 4 (90% accuracy) < ANN 3 = ANN 6 (92% accuracy) < ANN 1 = ANN 5 (94% accuracy) < ANN 2 (98% accuracy) as presented in Table 4.

A more specific observation on the results of the models, starting from ANN 1 to ANN 2, there is an improvement in the results and a reduction in the RMSEP values from 0.2 0.1 wt.% to 0.1 wt.% C when adding the variables of the Al line at II 198.99 nm, this being the best model (ANN 2). However, with the addition of the variables of the Si I line at 212.41 nm and PT and SA indices, there is no improvement in the results and an increase in RMSEP values. Therefore, it can be inferred that there is a synergism between the lines C I at 193.03 and the Al II at 198.99 nm for these data, improving the modeling with ANN. It was also observed in the univariate calibration approaches (model MMC with SC-C I at 193.03 + Al I at 193.58 + Al II at 198.99 nm lines), being the most accurate model (81%) among all univariate models.

The best model (ANN 2) was obtained using only independent variables, the C I line at 193.03 nm, the Al I line at 193.58 nm, and the Al II line at 198.99 nm, achieving 98% accuracy, i.e., 47 of 48 samples of the validation set into recovery range between 80% and 120%. The recoveries of the ANN 2 model presented really good values both calibration and validation sets. Table S1 in Supplement 1 shows subsets samples containing reference concentrations obtained by CHNS (reference method) and predicted by LIBS technique with its standard deviations and recovery values. The parameters obtained for this model (ANN 2) were $r_{val}$ (0.9669), $R^2_{val}$ (0.9349), and RMSEP (0.1 wt.% C). The C content was determined by LIBS versus

**Table 3. Multivariate Calibration Models Obtained from PCR and PLS**

| Parameter | PCR (1) | PCR (2) | PCR (3) | PCR (4) | PCR (5) | PLS (1) | PLS (2) | PLS (3) | PLS (4) | PLS (5) |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | C 193<br>Al 193<br>Al 198<br>Si 212<br>PT<br>SA<br>Texture | C 193<br>Al 193<br>Al 198<br>Texture | C 193<br>Al 193<br>Si 212 | C 193<br>Al 193<br>Al 198 | C 193<br>Al 193<br>Al 198<br>SA<br>PT | C 193<br>Al 193<br>Al 198<br>Si 212<br>PT<br>SA<br>Texture | C 193<br>Al 193<br>Al 198 | C 193<br>Al 193<br>Al 198<br>SA<br>PT | C 193<br>Al 193<br>Si 212 | C 193<br>Al 193<br>Al 198<br>Texture |
| $r$ calibration | 0.9399 | 0.9391 | 0.9113 | 0.9362 | 0.9392 | 0.9462 | 0.9428 | 0.9463 | 0.9222 | 0.9466 |
| $R^2$ calibration | 0.8834 | 0.8820 | 0.8305 | 0.8765 | 0.8821 | 0.8954 | 0.8889 | 0.8955 | 0.8504 | 0.8961 |
| RMSEC (%) | 0.2 | 0.2 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.3 | 0.2 |
| $r$ validation | 0.9231 | 0.9224 | 0.9088 | 0.9374 | 0.9310 | 0.9288 | 0.9371 | 0.9303 | 0.9282 | 0.9403 |
| $R^2$ validation | 0.8521 | 0.8508 | 0.8260 | 0.8787 | 0.8668 | 0.8628 | 0.8781 | 0.8656 | 0.8615 | 0.8842 |
| RMSEP (%) | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Accuracy[a] | 37 (77 %) | 37 (77 %) | 37 (77%) | 38 (79%) | 38 (79%) | 38 (79%) | 39 (81%) | 40 (83%) | 40 (83%) | 43 (90%) |

[a]Recovery range values between 80% and 120% for the validation set (48 soil samples); PT; SA; Texture = clay, sand, and silt.

**Table 4. Multivariate Calibration Models Obtained from ANN**

| Parameter | ANN (1) | ANN (2) | ANN (3) | ANN (4) | ANN (5) | ANN (6) | ANN (7) |
|---|---|---|---|---|---|---|---|
| Variables | C193<br>Al 193 | C 193<br>Al 193<br>Al 198 | C 193<br>Al 193<br>Si 212 | C 193<br>Al 193<br>Al 198<br>Si 212 | C 193<br>Al 193<br>Al 198<br>SA<br>PT | C 193<br>Al 193<br>Si 212<br>SA<br>PT | C 193<br>Al 193<br>Al 198<br>Si 212<br>SA<br>PT |
| $r$ calibration | 0.9621 | 0.9769 | 0.9790 | 0.9822 | 0.9783 | 0.9754 | 0.9835 |
| $R^2$ calibration | 0.9257 | 0.9542 | 0.9585 | 0.9648 | 0.9570 | 0.9514 | 0.9673 |
| RMSEC (%) | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.1 |
| $r$ validation | 0.9461 | 0.9669 | 0.9400 | 0.9474 | 0.9571 | 0.9553 | 0.9477 |
| $R^2$ validation | 0.8951 | 0.9349 | 0.8836 | 0.8976 | 0.9161 | 0.9127 | 0.8981 |
| RMSEP (%) | 0.1 | 0.1 | 0.2 | 0.2 | 0.1 | 0.1 | 0.2 |
| Accuracy[a] | 45 (94%) | 47 (98%) | 44 (92%) | 43 (90%) | 45 (94%) | 44 (92%) | 43 (90%) |

[a]Recovery range values between 80% and 120% for the validation set (48 soil samples); PT; SA.

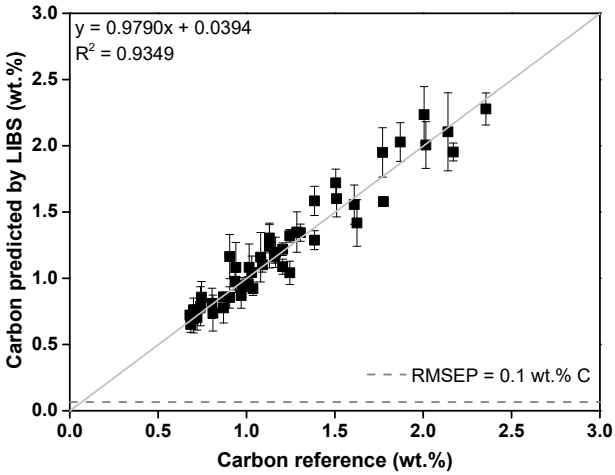


**Fig. 4.** Correlation between carbon content (wt.%) in soil predicted by LIBS and C values determined by reference method (elemental analyzer - CHNS) for best ANN model (ANN 2).

the reference values obtained by elemental analyzer—CHNS as shown in Fig. 4.

The superiority of ANNs over other multivariate calibration strategies can be explained due to the capacity of calibration in the presence of interferences, nonlinearly using several variables that can contribute overcoming matrix effects and that can explore more information from the analyzed samples [33,40].

Regression models are strongly related to the significance levels of the independent variables, and in ANN, all data are used with individual weights, making the models more accurate. Modeling with ANN makes it possible to accurately predict the total C content in soil samples analyzed directly by LIBS.

## 4. GUIDELINES FOR CALIBRATION STRATEGIES IN SOIL SAMPLES

The several calibration strategies presented in this tutorial demonstrated how the intrinsic characteristics of each modeling strategy to improve the soil carbon predictions. In Fig. 5, we show the advantages of each strategy, which can assist in the decision for calibration strategy employment. The univariate strategies (MMC and IS) simplify spectral treatment and modeling. Evaluating the most straightforward strategy (MMC) for C prediction using LIBS is based on the direct correlation

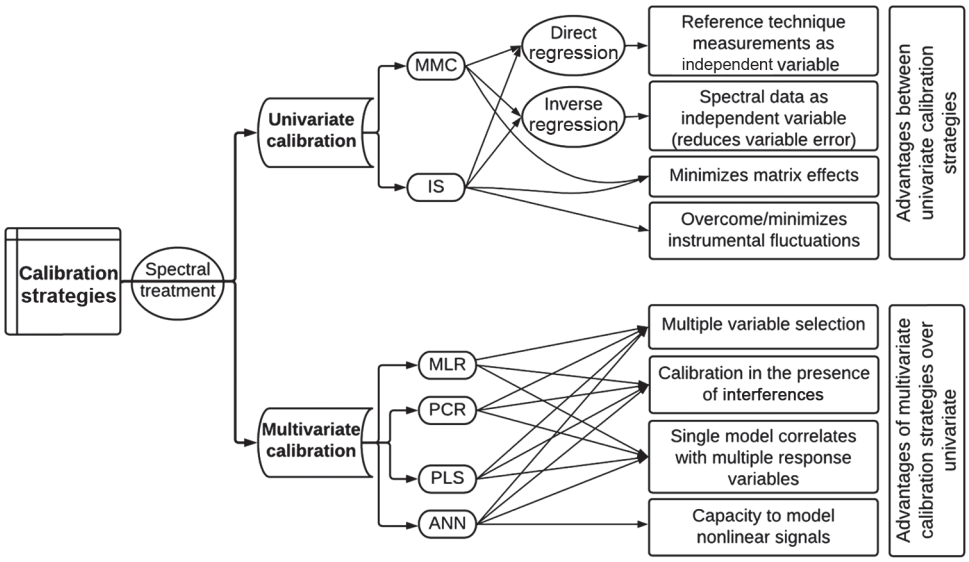


**Fig. 5.** Guidelines for choosing calibration strategies.

between soil C concentration obtained by the reference method and the analytical signal from LIBS spectra. However, MMC assumes that there are not significant matrix effects and that the samples have similar physicochemical properties, which is unlikely due to the complexity of the matrix (soil). The IS direct regression can partially correct matrix effects and instrumental fluctuations, an alternative to MMC. Still, a critical step is selecting an excellent IS to be utilized. Both univariate strategies are based on least-squares regression, and the adoption of dependent and independent variables can significantly affect the precision and accuracy of the model. The employment of IR can result in a better model, increasing accuracy due to using the analytical signal as an independent variable (minor error source). However, the univariate strategies need of spectral treatment to solve the spectral interference of Al I and Al II lines in the C I line at 193.03 nm (Fig. 3). In this context, the spectral correction can be applied to all calibration strategies (univariate and multivariate) to overcome this effect.

When the univariate strategies cannot sufficiently overcome matrix effects, we can employ multivariate calibrations once they utilize multiple variables, which is a significant advantage in improving modeling. Between the multivariate strategies evaluated, MLR presents the advantage of simplicity in modeling expertise. Still, one important limitation that needs to be observed is that the number of samples utilized on calibration needs to be higher than the number of selected variables. PCR and PLS strategies allow the selection of multiple variables selection, overcoming MLR limitations. Still, PLS modeling enables building a single model for multiple responses, an advantage over PCR calibration.

Even so, some datasets can be so complex that neither of the previously presented strategies can efficiently generate accurate models due to the nonlinearity of the dataset. Thus, the ANN must be adopted due to the ANN's capacity to assign individual weights to each variable added to the model calibration making it possible to explore more information from the data set and nonlinearly. However, there are also limitations for modeling applying the ANN, which require a large number of samples for calibration set and/or variables (the more information provided, the more robust the ANN model) and the analyst's great expertise.

## 5. CONCLUSION AND OUTLOOK

Several univariate and multivariate calibration strategies were applied and presented their advantages and limitations, allowing the user to decide which strategy can be used to quantify C in soil samples using the LIBS technique. Furthermore, it was demonstrated that more than one path could be chosen for soil C quantification; according to the matrix physical–chemical properties, the analyst decides which variables and calibration strategy can be used in the modeling.

Although the univariate strategies presented more limitations in soil C modeling and prediction, the MMC–SC strategy resulted in a high-accuracy model (81%). The best models were MLR 4, PLS 5, and ANN 2 with an accuracy of 85%, 90%, and 98%, respectively. These results demonstrated that the LIBS technique had great potential to be consolidated to quantify soil C.

In addition, other possibilities for modeling in more specific contexts can be considered, for example, calibration-free (CF-LIBS) [41] and OPC, not explored in this paper. In this case, there was no need to build multivariate models or calibration curves because the model was based on plasma parameters. Based on the local thermodynamic equilibrium, the concentration of one species can be determined using the relationships between the other species or by utilizing an element as an IS to simplify the computations. The model became more robust and less vulnerable to the matrix effect by considering the plasma parameters (temperature, electron density, and emitter density). By combining the model with the correction of the OPC method [42,43], there was an increase in the accuracy and precision of CF-LIBS even when the matrix effect was evident. This CF-LIBS model had been used efficiently in agrienvironmental

samples, such as soybean leaves [43] for nutrient quantification. However, more research was needed to develop a CF-LIBS data analysis procedure for measuring C and soil fertility; its application in soils had a great potential.

**Funding.** Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (382785/2022-3, 440226/2021-0); Fundação de Amparo à Pesquisa do Estado de São Paulo (2013/07276-1, 2022/05451-0).

**Acknowledgment.** The authors are grateful for the financial support provided by CNPq, Embrapa Instrumentation, National Laboratory of Agri-Photonics (LANAF) and Embrapa Soils for the laboratory and experimental field infrastructures that were crucial for this research development.

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** All data generated and analyzed are presented in the research presented.

**Supplemental document.** See Supplement 1 for supporting content.

## REFERENCES

1. E. C. Ferreira, E. J. Ferreira, P. R. Villas-Boas, G. S. Senesi, C. M. Carvalho, R. A. Romano, L. Martin-Neto, and D. M. B. P. Milori, "Novel estimation of the humification degree of soil organic matter by laser-induced breakdown spectroscopy," Spectrochim. Acta B **99**, 76–81 (2014).
2. G. S. Senesi and N. Senesi, "Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review," Anal. Chim. Acta **938**, 7–17 (2016).
3. G. Mishra, A. Sarkar, K. Giri, A. J. Nath, R. Lal, and R. Francaviglia, "Changes in soil carbon stocks under plantation systems and natural forests in Northeast India," Ecol. Modell. **446**, 109500 (2021).
4. A. Walkley, "An examination of methods for determining organic carbon and nitrogen in soils. (with one text-figure)," J. Agric. Sci. **25**, 598–609 (1935).
5. G. Nicolodelli, B. S. Marangoni, J. S. Cabral, P. R. Villas-Boas, G. S. Senesi, C. H. dos Santos, R. A. Romano, A. Segnini, Y. Lucas, C. R. Montes, and D. M. B. P. Milori, "Quantification of total carbon in soil using laser-induced breakdown spectroscopy: a method to correct interference lines," Appl. Opt. **53**, 2170–2176 (2014).
6. P. R. Villas-Boas, M. A. Franco, L. Martin-Neto, H. T. Gollany, and D. M. B. P. Milori, "Applications of laser-induced breakdown spectroscopy for soil characterization, part II: Review of elemental analysis and soil classification," Eur. J. Soil Sci. **71**, 805–818 (2020).
7. G. W. McCarty, J. B. Reeves, V. B. Reeves, R. F. Follett, and J. M. Kimble, "Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement," Soil Sci. Soc. Am. J. **66**, 640–646 (2002).
8. A. Segnini, A. Posadas, W. T. Lopes da Silva, D. M. B. P. Milori, C. Gavilan, L. Claessens, and R. Quiroz, "Quantifying soil carbon stocks and humification through spectroscopic methods: A scoping assessment in EMBU-Kenya," J. Environ. Manage. **234**, 476–483 (2019).
9. R. C. Izaurralde, C. W. Rice, L. Wielopolski, M. H. Ebinger, J. B. Reeves, A. M. Thomson, R. Harris, B. Francis, S. Mitra, A. G. Rappaport, J. D. Etchevers, K. D. Sayre, B. Govaerts, and G. W. McCarty, "Evaluation of three field-based methods for quantifying soil carbon," PLoS One **8**, e55560 (2013).
10. A. Segnini, A. Augusto Pereira Xavier, P. L. Otaviani-Junior, E. C. Ferreira, A. M. Watanabe, M. A. Sperança, G. Nicolodelli, P. R. Villas-Boas, P. P. Anchão Oliveira, and D. M. B. P. Milori, "Physical and chemical matrix effects in soil carbon quantification using laser-induced breakdown spectroscopy keywords soil C content, spectral correction, soil textural classes, field-based method, atomic emission," Am. J. Anal. Chem. **5**, 722–729 (2014).
11. R. S. Bricklemyer, D. J. Brown, P. J. Turk, and S. M. Clegg, "Improved intact soil-core carbon determination applying regression shrinkage and variable selection techniques to complete spectrum laser-induced breakdown spectroscopy (LIBS)," Appl. Spectrosc. **67**, 1185–1199 (2013).
12. K. Stenio, A. A. Pereira Xavier, C. P. De Morais, and D. M. B. P. Milori, "Carbon quantification in soils with different textures using laser-induced breakdown spectroscopy: spectral interference correction and use of a 3D plane model," Anal. Methods **14**, 4219–4229 (2022).
13. V. C. S. Methodology and S. Scope, "Vm0042 methodology for improved," (2021).
14. G. Nicolodelli, J. Cabral, C. R. Menegatti, B. Marangoni, and G. S. Senesi, "Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers," TrAC - Trends Anal. Chem. **115**, 70–82 (2019).
15. R. Noll, *Laser-Induced Breakdown Spectroscopy - Fundamentals and Applications* (Springer-Verlag, 2012).
16. A. Miziolek, V. Palleschi, and I. Schechter, *Laser-Induced Breakdown Spectroscopy (LIBS): Fundamentals and Applications* (Cambridge University, 2006).
17. D. A. Cremers and L. J. Radziemski, *Handbook of Laser-Induced Breakdown Spectroscopy* (Wiley, 2006).
18. S. Musazzi and U. Perini, *Laser-Induced Breakdown Spectroscopy Theory and Applications* (Springer-Verlag Berlin Heidelberg, 2014).
19. J. P. Singh and S. N. Thakur, *Laser-Induced Breakdown Spectroscopy* (Elsevier Science, 2007).
20. T. Takahashi and B. Thornton, "Quantitative methods for compensation of matrix effects and self-absorption in Laser Induced Breakdown Spectroscopy signals of solids," Spectrochim. Acta B **138**, 31–42 (2017).
21. T. L. Zhang, S. Wu, H.-S. Tang, K. Wang, Y.-X. Duan, and H. Li, "Progress of chemometrics in laser-induced breakdown spectroscopy analysis," Chin. J. Anal. Chem. **43**, 939–948 (2015).
22. V. C. Costa, D. V. Babos, J. P. Castro, D. F. Andrade, R. R. Gamela, R. C. Machado, M. A. Sperança, A. S. Araújo, J. A. Garcia, and E. R. Pereira-Filho, "Calibration strategies applied to laser-induced breakdown spectroscopy: A critical review of advances and challenges," J. Braz. Chem. Soc. **31**, 2439–2451 (2020).
23. A. Kramida, Y. Ralchenko, and J. Reader, "NIST Atomic Spectra Database," available at https://physics.nist.gov/asd, accessed 20 November 2022.
24. D. A. Cremers, M. H. Ebinger, D. D. Breshears, P. J. Unkefer, S. A. Kammerdiener, M. J. Ferris, K. M. Catlett, and J. R. Brown, "Measuring total soil carbon with laser-induced breakdown spectroscopy (LIBS)," J. Environ. Qual. **30**, 2202–2206 (2001).
25. E. C. Ferreira, D. M. B. P. Milori, E. J. Ferreira, R. M. Da Silva, and L. Martin-Neto, "Artificial neural network for Cu quantitative determination in soil using a portable Laser Induced Breakdown Spectroscopy system," Spectrochim. Acta B **63**, 1216–1220 (2008).
26. A. N. Beretta, A. V. Silbermann, L. Paladino, D. Torres, D. Bassahun, R. Musselli, and A. García-Lamohte, "Soil texture analyses using a hydrometer: modification of the Bouyoucos method," Cienc. Inv. Agr. **41**, 263–271 (2014).
27. B. S. Marangoni, K. S. G. Silva, G. Nicolodelli, G. S. Senesi, J. S. Cabral, P. R. Villas-Boas, C. S. Silva, P. C. Teixeira, A. R. A. Nogueira, V. M. Benites, and D. M. B. P. Milori, "Phosphorus quantification in fertilizers using laser induced breakdown spectroscopy (LIBS): a methodology of analysis to correct physical matrix effects," Anal. Methods **8**, 78–82 (2016).
28. N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," IEEE Trans. Geosci. Remote Sens. **42**, 1552–1565 (2004).
29. J. P. Castro and E. R. Pereira-Filho, "Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS)," J. Anal. At. Spectrom. **31**, 2005–2014 (2016).
30. D. F. Andrade, F. M. Fortunato, and E. R. Pereira-Filho, "Calibration strategies for determination of the In content in discarded liquid crystal displays (LCD) from mobile phones using laser-induced breakdown spectroscopy (LIBS)," Anal. Chim. Acta **1061**, 42–49 (2019).
31. T. Rajaee, S. Khani, and M. Ravansalar, "Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review," Chemom. Intell. Lab. Syst. **200**, 103978 (2020).

32. P. Yaroshchyk, D. L. Death, and S. J. Spencer, "Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS," J. Anal. At. Spectrom. **27**, 92–98 (2012).
33. L.-N. Li, X.-F. Liu, F. Yang, W.-M. Xu, J.-Y. Wang, and R. Shu, "A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis," Spectrochim. Acta B **180**, 106183 (2021).
34. F. M. V. Pereira, J. P. Castro, R. C. Machado, A. S. Araújo, D. F. Andrade, D. V. Babos, D. R. Beletti, E. R. Pereira-Filho, M. L. Mello, F. F. Hilario, J. A. Garcia, M. A. Speranca, R. R. Gamela, and V. C. Costa, *Laser-Induced Breakdown Spectroscopy (LIBS): Applications and Calibration Strategies*, 1st ed. (Editora Ibero-Americana de Educação, 2021).
35. W. N. Guedes and F. M. V. Pereira, "Classifying impurity ranges in raw sugarcane using laser-induced breakdown spectroscopy (LIBS) and sum fusion across a tuning parameter window," Microchem. J. **143**, 331–336 (2018).
36. S. Yao, Y. Shen, K. Yin, G. Pan, and J. Lu, "Rapidly measuring unburned carbon in fly ash using molecular cn by laser-induced breakdown spectroscopy," Energy Fuels **29**, 1257–1263 (2015).
37. A. Cousin, P. Y. Meslin, R. C. Wiens, *et al.*, "Compositions of coarse and fine particles in martian soils at gale: a window into the production of soils," Icarus **249**, 22–42 (2015).
38. "LibreOffice Linest function," available at https://wiki.documentfoundation.org/Documentation/Calc_Functions/LINEST, accessed 20 November 2022.
39. R. R. Gamela, V. C. Costa, M. A. Sperança, and E. R. Pereira-Filho, "Laser-induced breakdown spectroscopy (LIBS) and wavelength dispersive X-ray fluorescence (WDXRF) data fusion to predict the concentration of K, Mg and P in bean seed samples," Food Res. Int. **132**, 109037 (2020).
40. W. N. Guedes, L. J. dos Santos, É. R. Filletti, and F. M. V. Pereira, "Sugarcane stalk content prediction in the presence of a solid impurity using an artificial intelligence method focused on sugar manufacturing," Food Anal. Methods **13**, 140–144 (2020).
41. E. Tognoni, G. Cristoforetti, S. Legnaioli, and V. Palleschi, "Calibration-free laser-induced breakdown spectroscopy: state of the art," Spectrochim. Acta B **65**, 1–14 (2010).
42. L. C. L. Borduchi, D. M. B. P. Milori, and P. R. Villas-Boas, "One-point calibration of Saha-Boltzmann plot to improve accuracy and precision of quantitative analysis using laser-induced breakdown spectroscopy," Spectrochim. Acta B **160**, 105692 (2019).
43. L. C. L. Borduchi, D. M. B. P. Milori, M. C. Meyer, and P. R. Villas-Boas, "Reducing matrix effects on the quantification of Ca, Mg, and Fe in soybean leaf samples using calibration-free LIBS and one-point calibration," Spectrochim. Acta B **198**, 106561 (2022).