Geoderma 355 (2019) 113905

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques

Xu Xuebin[a,b], Du Changwen[a,b,*], Ma Fei[a], Shen Yazhen[a], Wu Ke[a,b], Liang Dong[a,b], Zhou Jianmin[a]

[a] The State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China
[b] University of Chinese Academy of Sciences, Yuquan Road 19, Beijing 100049, China

ARTICLE INFO

Editor: Alex McBratney

Keywords:
Laser-induced breakdown spectroscopy
FTIR-ATR spectroscopy
Spectral preprocessing
Data fusion

ABSTRACT

Spectroscopy is a useful method for soil monitoring because of its environmental friendliness, and its ability to produce rapid, nondestructive, simultaneous multi-element analysis. In this work, data fusion strategies for laser-induced breakdown spectroscopy (LIBS) and attenuated total reflectance Fourier-transform mid-infrared spectroscopy (FTIR-ATR), as well as a combination of multivariate calibration methods were investigated for prediction of soil organic matter (SOM) content. The root mean square error (RMSE) and residual prediction deviation (RPD) of the calibration and validation sets, systematic error, and residual assessment, were applied to evaluate the robustness and accuracy of these predictions. The results of a principal component analysis (PCA) indicated that baseline wander present in the spectral data could be effectively removed using morphological weighted penalized least squares (MPLS) and wavelet transform (WT) algorithms. The quantitative prediction ability of SOM content by a partial least squares regression (PLSR) model could be improved using principal component weighted mean (PCWM) and Euclidean distance weighted mean (EDWM) algorithms applied to parallel LIBS spectra. The prediction ability of SOM content was dramatically improved using mid-level data fusion based on the concatenation of latent variables of LIBS and FTIR-ATR spectra obtained by partial least squares algorithm. The considerable prediction accuracy and robustness were achieved using the PLSR model ($R_V^2 = 0.792$, $RMSE_V = 1.76\,g\,kg^{-1}$, and $RPD_V = 2.16$), the support vector regression (SVR) model ($R_V^2 = 0.811$, $RMSE_V = 1.68\,g\,kg^{-1}$, and $RPD_V = 2.27$), and the artificial neural network (ANN) model ($R_V^2 = 0.830$, $RMSE_V = 1.60\,g\,kg^{-1}$, and $RPD_V = 2.39$). The findings from this work suggest that the use of LIBS and FTIR-ATR spectra in combination with multivariate calibration can be a simple, fast, and nondestructive approach to monitor SOM. This strategy is potentially of great significance in the evaluation of soil fertility, the management of soil nutrients, and in guiding the agricultural production of precision agriculture.

1. Introduction

Soil organic matter (SOM), one of the key indicators of soil fertility and quality, is comprised of diverse and heterogeneous organic substances that can significantly influence the physical and biochemical reactions that occur in soils (Cambule et al., 2012; Dhillon et al., 2017; Luce et al., 2014). SOM content, considered as a good indicator of a healthy soil system, is related to a wide range of soil properties and processes, such as soil structure, water-retention capacity, nutrient cycling, and biological activity (de la Paz Jimenez et al., 2002; Lado et al., 2004; Oades, 1993; Sharma et al., 2017). SOM also plays an important role in impacting global climate change through influence on atmospheric carbon levels (Zhang et al., 2015).

Since SOM is a critical index for evaluating soil fertility and quality, there is an urgent need to conduct efficient and fast SOM characterization of arable soils, particularly with respect to the development of precision agriculture. Traditionally, the colorimetric method (Walkley and Black, 1934) and the dry combustion method (Nelson and Sommers, 1982) have been the primary tools for analyzing soil C content. Although these classical methods are effective, they are subject to well-known deficiencies such as significant time-consumption, high cost, destructive processes, and environmental unfriendliness. This is because the measurement usually involves tedious sample preparation which includes grinding, and wet or dry chemical procedures (Lu et al., 2014). A rapid, inexpensive, and effective detection technique is therefore desperately needed.

[*] Corresponding author at: The State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, China.
E-mail address: chwdu@issas.ac.cn (C. Du).

https://doi.org/10.1016/j.geoderma.2019.113905
Received 30 January 2019; Received in revised form 27 July 2019; Accepted 7 August 2019

Infrared spectroscopy has already been exploited for the quantitative determination of soil physical and chemical properties due to its numerous advantageous qualities. Compared to the classical approaches, this process is efficient, fast, non-destructive, and environmentally friendly (Jia et al., 2017; Tamburini et al., 2017; Xu et al., 2017). Attenuated total reflectance Fourier-transform mid-infrared spectroscopy (FTIR-ATR) is usually used to directly measure the surface properties of solid or thin film samples rather than their bulk properties. Shao et al. (2017) used FTIR-ATR for soil nitrate contents quantification, and they found that the second-order derivative of the nitrate characteristic bands (1270–1320 cm^{-1}) was proportional to the nitrate content, and an excellent correlation coefficient of 0.9751 was obtained, indicating that FTIR-ATR combined with second-order derivatives could be well suited for quantifying soil nitrates. Rial et al. (2016) used multivariate linear regression (MLR) to predict soil organic carbon (SOC) contents from selected FTIR-ATR bands as an independent proxy, and the model showed a good predictive performance of $R^2 = 0.88$ and RPD = 3.14, indicating that SOC can be effectively estimated from the identified spectral bands. Thus, FTIR-ATR can be an effective tool for the quantitative determination of SOM.

Laser-induced breakdown spectroscopy (LIBS) is a type of atomic emission spectroscopy, first reported in 1962 by Brech (1962). LIBS uses a laser beam of moderate power that is focused onto a sample to atomize the surface and generate luminous plasma (Rifai et al., 2017). Simultaneously, the light emitted by the plasma is recorded by a CCD (or ICCD) detector (Cremers and Radziemski, 2006). An increasing number of studies have been investigating the application of LIBS for soil content determination due in part to the technique's environmental friendliness, coupled with its ability to facilitate rapid, nondestructive, simultaneous multi-element analyses (El Haddad et al., 2014; Ferreira et al., 2008; Kim and Choi, 2018; Knadel et al., 2017; Meng et al., 2017; Rehan et al., 2018; Senesi and Senesi, 2016; Villas-Boas et al., 2016; Yongcheng et al., 2017; Zaytsev et al., 2018). For instance, Martin et al. (2003) used LIBS to measure the C and N content of soils with total C concentrations ranging from 0.16% to 4.3%. The LIBS C signal was shown to be highly correlated (coefficient of determination, $R^2 = 0.962$) with the organic carbon content measured by dry combustion using an elemental analyzer. However, many researchers have proposed that the LIBS spectral lines are affected not only by characteristics of the elemental emission lines, but also other interference information including the background spectrum, instrument noise, etc. (Tan et al., 2017; Yi et al., 2017). In addition, the "matrix effect" is a major factor limiting the application of LIBS for soil content determination. The "matrix effect" is caused by the changes in the emission line intensities of some elements in the samples when the physical properties and the chemical compositions of the matrix vary (Senesi, 2014). Due to the non-homogeneity of the soil system, the compositions (such as sand, clay, organic matter, etc.) are heterogeneously distributed in the soil samples at the micron level, which caused uneven ablation to the soil's surface at each paralleled shot and resulted in large differences between the paralleled LIBS spectra. Thus, the traditional arithmetic mean of the parallel spectra used to represent the characteristic spectra of soil samples is inappropriate for all soil samples. Unfortunately, there is generally an absence of literatures regarding appropriate algorithms for pre-processing of parallel LIBS spectra.

Data fusion of different spectroscopic data is considered as an effective strategy for improving prediction accuracy (Borràs et al., 2015). Data fusion techniques can be generally classified into three main groups (Silvestri et al., 2013): (a) low-level data fusion consisting of the simple concatenation of the separate spectroscopic data; (b) mid-level data fusion based on feature extraction or variable selection prior to multivariate analysis; (c) high-level or hierarchical data fusion is based on the concatenation of the scores extracted by means of multivariate projection techniques such as PCA, PLS, etc. or wavelet transform. To the best of our knowledge, data fusion of LIBS and FTIR-ATR spectra has not yet been investigated for SOM prediction.

In the present study, LIBS and FTIR-ATR spectroscopy combined with various chemometric algorithms (i.e., principal component analysis, partial least squares regression, support vector regression, and artificial neural network) were utilized to quantify the SOM content of soils. Several algorithms were proposed to evaluate the efficacy of the characteristic spectrum in the preprocessing of parallel LIBS spectra. Two levels of data fusion tactics (i.e., low-level and mid-level) were applied to improve prediction ability. The specific objectives of this study were: (i) to evaluate the preprocessing effectiveness of several algorithms for parallel LIBS spectra; (ii) to assess the potential of LIBS and FTIR-ATR spectroscopy for quantification of SOM content; (iii) to evaluate the efficacy of various data fusion methods and multivariate calibration models for the promotion of SOM prediction ability.

2. Material and methods

2.1. Soil sampling, pretreatment, and chemical analysis

One hundred and fifty-two topsoil (0–20 cm) samples and 152 subsoil (20–40 cm) samples were collected from Bayannur, Inner Mongolia, China, which covers about 1100 km^2. The detailed position information of the sampling sites was designed and recorded using a

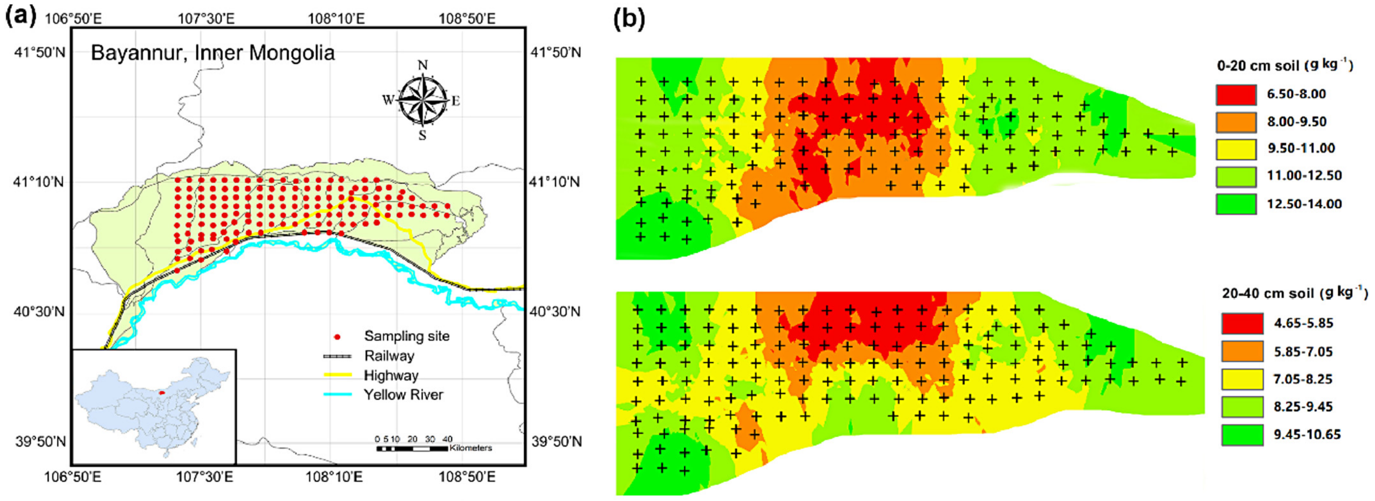

Fig. 1. (a) Locations of the studied area and the geographical distribution of soil sampling sites, (b) distribution of SOM contents of topsoils (0–20 cm) and subsoils (20–40 cm) in the studied area.

hand-held GPS device (Fig. 1a). The obtained soil samples were air-dried naturally at room temperature and crushed before being passed through a 2-mm mesh sieve to remove any plant roots and debris prior to analysis. Soil pH values were measured in water (1:2.5 soil/water ratio) using a pH meter (PHS-3BW, UK), with pH values ranging from 7.47 to 9.55. The SOC was chemically determined using a potassium dichromate oxidation colorimetry. Specifically, a 1.5 g soil sample was evenly mixed with 5.0 mL potassium dichromatic ($K_2Cr_2O_7$, AR) solution [c ($\frac{1}{6}$ $K_2Cr_2O_7$) = 0.8000 mol L^{-1}] and 5.0 mL of concentrated sulfuric acid in a glass tube. The mixture was placed in an incubator (DRP-9162, China) at 100 °C for 90 min and was diluted with pure water to 50 mL after cooling. After 24 h, supernatants were used for colorimetry with standards at 590 nm using a spectrophotometer (BioTek Epoch, USA). The SOC content was then multiplied by 1.724 to obtain the SOM content. The soil samples passed through a 0.25-mm mesh were palletized in dimensions of 1 cm diameter and 0.25 cm thickness with an applied pressure of ~ 55 MPa for the time duration of 2 min using a tablet machine (YP-2, China).

2.2. FTIR-ATR and LIBS spectra acquisition

FTIR-ATR spectra of the soil samples were acquired in the mid-infrared range (4000–400 cm^{-1}) at a 2.87 cm^{-1} resolution using an attenuated total reflectance infrared spectrophotometer (TRUDEFENDER FT, Thermo Scientific, USA). Approximately 2 g of air-dried and sieved soil (0.25 mm) was spread onto the FTIR-ATR crystal. Prior to the measurement of each soil sample, the FTIR-ATR crystal was cleaned, and background spectra were obtained for subsequent baseline correction.

LIBS spectra were obtained using a MobiLIBS system (IVEA, France) with the AnaLIBS control software. A fourth-harmonic Nd: YAG laser (Quantel, France) operated at 580 nm with a 5 ns pulse-duration was used as the ablation source. The frequency of the system was 20 Hz, and the delivery energy was 16 mJ. A lens with a focal length of 15 cm was used to focus the output of the laser onto the surface of the pelleted sample with a spot diameter of 50 μm. The emission line of the resulting plasma was transmitted from the light collector to the Mechelle 5000 Echelle spectrometer (Andor Technology, Ltd., Northern Ireland). The resolving power of this spectrometer was $\lambda/\Delta\lambda = 4000$. An intensified charge-coupled device camera (iStar, Andor Technology, Ltd., Northern Ireland) collected the diffracted light. The spectral range of this system was 200–1000 nm. The delay time and the gate width were set as 137 μs and 7.0 ms, respectively. The LIBS spectra of soil were collected at 3 depths through 3 × 3 cm windows cut. Thus a total of 27 parallel LIBS spectra of each soil sample were obtained.

2.3. Pre-treatment of spectra

Baseline correction for the raw spectra of the soil samples was performed using an algorithm of morphological weighted penalized least squares (MPLS), as previously described by (Li et al., 2013). The calculation procedures are as follows: (1) Obtaining the rough background profile using a mathematical morphological opening operation. (2) The rough background profile and the local minimum values are used as the input vector for penalized least squares to refine the background profile. (3) By subtracting the refined background profile from the original signal, a background corrected signal is calculated. After baseline correction, spectra of the soil samples were subsequently normalized and smoothed.

In the present study, we proposed four approaches for preprocessing parallel LIBS spectra: arithmetic mean (AM), principal component analysis (PCA), principal component weighted mean (PCWM), and Euclidean distance weighted mean (EDWM). Because the arithmetic mean is a common approach, only the other three methods will be introduced.

2.3.1. Principal component analysis (PCA)

PCA is a dimensionality reduction chemometrics technique that is able to reduce redundant information in a data set (Liang et al., 2017). PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs). For each soil sample, 27 parallel LIBS spectra were arranged into a p-by-27 matrix X with p variables (i.e., intensity values) and 27 measurements. The score matrix was obtained after principal component analysis. Rows of scores correspond to components, and columns correspond to observations. The first principal component was employed as the dimensionality reduced LIBS spectra for subsequent analysis.

2.3.2. Principal component weighted mean (PCWM)

According to the pre-experiment analysis, the first two principal components contained most of the information of the parallel LIBS spectra. Because all the 27 spectra lines are parallel spectra, the principal component scatterplots of these spectra lines should be located close to each other in principal component space if the influence of the background is ignored. In other words, if the principal component scatterplots of the parallel spectra are closer to the origin, the spectra have higher representativeness. Therefore, the principal component weighted mean (PCWM) was proposed. Firstly, the first two principal component scores (p_i^1 and p_i^2) of the ith parallel spectrum were acquired using PCA. Secondly, the distances (D_i) between principal component scores of the ith spectrum and the zero point (0, 0) (i.e., root-mean-square) were calculated (Eq. (1)). Thirdly, the weight coefficient (w_i) of the ith parallel spectrum was calculated by the proportions of reciprocal of D_i (Eq. (2)). Finally, the characteristic spectrum was then obtained by the average weighting method (Eq. (3)).

$$D_i = \sqrt{p_i^{12} + p_i^{22}} \tag{1}$$

$$w_i = \frac{1/D_i}{\sum_{i=1}^{n} 1/D_i} \tag{2}$$

$$\hat{X} = \sum_{i=1}^{n} w_i X_i \tag{3}$$

where n is the number of parallel spectra, $\hat{X}$ is the estimated characteristic spectrum, and X_i is the ith parallel spectrum.

2.3.3. Euclidean distance weighted mean (EDWM)

Similar to principal component weighted mean, the Euclidean distance (ED_i) between parallel spectra to arithmetic averaged spectrum was used to calculate the weight coefficient according to the following equation:

$$ED_i = \sqrt{\sum_{j=1}^{m} (s_i^j - s_m^j)^2} \tag{4}$$

$$w_i = \frac{1/ED_i}{\sum_{i=1}^{n} 1/ED_i} \tag{5}$$

where w_i is the EDWM weight coefficient of the ith parallel spectrum; s_i^j is the jth observed value of the ith spectrum (s_i); s_m^j is the jth observed value of the averaged spectrum; n is the number of parallel spectra; m is the number of the observed value of spectra. The characteristic spectrum was then obtained by weighting average according to Eq. 3.

2.4. Calibration models and data fusion methods

2.4.1. Prediction models

Partial least squares regression (PLSR) is one of the most commonly used chemometrics algorithm for regressing in the analysis of spectral data. This technique was used to correlate the spectral data to the chemical properties of the soil samples. In this study, ten-fold cross-validation was performed to obtain the optimal number of latent variables for better model construction while avoiding over-fitting (Reeves et al., 2002). PLSR model was used for evaluating the efficiency of preprocessing of parallel LIBS spectra and predicting the SOM content.

The Artificial neural network (ANN) model, has been demonstrated as a new approach for classification and prediction (Rezaei et al., 2014). The multi-layer feed-forward neural network (MLP), and the Levenberg– Marquardt (LM) back-propagation training algorithm are the most used neural network architecture and training algorithms, respectively because of their simplicity and good generalization capability (Rumelhart et al., 1986). In this study, the multi-layer feed-forward neural network consists of one input-layer, two hidden-layers, and one output-layer. The Levenberg–Marquardt algorithm, which provides a fast optimization, was used for network training.

Support vector regression (SVR) is a machine learning algorithm based on the statistical learning theory which seeks to maximize the ability to generalize using the structural risk minimization principle (Filgueiras et al., 2014b). Depending on the definition of this error function, there are two types of SVR models: (i) Epsilon-SVR (ε-SVR), which optimizes a model using the adjustable parameters epsilon (upper tolerance on prediction errors) and C (cost of prediction errors larger than epsilon), and (ii) Nu-SVR, which optimizes a model using the adjustable parameter Nu (0 V§ 1] which indicates a lower bound on the number of support vectors to use given as a fraction of the total calibration samples (Jimenez-Carvelo et al., 2017). Details of SVR theory and algorithms can be found elsewhere (Devos et al., 2009; Filgueiras et al., 2014b; Li et al., 2009).

2.4.2. Data fusion

In the present study, two levels of data fusion architectures (i.e., low-level and mid-level) were performed to enhance the quantitative ability of SOM content identification. In the low-level data fusion methodology, the whole preprocessed spectral matrices (LIBS and FTIR-ATR) were directly concatenated, and then the fused matrix was normalized before SOM prediction (Fig. 2a). In the mid-level approach, the variable selection process was dependently performed over the preprocessed spectral matrices before fusion, so that only the most relevant variables were fused. Principal component (PC) scores of preprocessed LIBS and FTIR-ATR spectra were obtained using PCA, and the results were subsequently concatenated as inputs data for SOM prediction (Fig. 2b). Similarly, latent variables of preprocessed LIBS and FTIR-ATR spectra were acquired by PLS and were subsequently concatenated as inputs data for SOM prediction (Fig. 2b).

2.4.3. Calibration and validation sets

To avoid the spatial pseudo replication and to make ensure that the validation set is independent, the sampling sites (total 152 sites) were randomly divided into a calibration set (114 sites, 75% of total sites) and validation set (38 sites, 25% of total sites). Thus, the calibration set contained 228 soil samples and the validation set contained 76 soil samples. The SOM contents ranged from 0.58 to 17.70 g kg^{-1} with a mean value of 9.05 g kg^{-1} in the calibration set and from 2.12 to 16.87 g kg^{-1} with a mean value of 9.21 g kg^{-1} in the validation set. The calibration set was used to build models and the validation set was employed to evaluate their predictive capability.

2.5. Statistical analysis

2.5.1. Robustness and accuracy assessment

For better evaluation of the robustness and accuracy of the models, the coefficients of determination (R^2), the root mean square error (*RMSE*), and the residual prediction deviation (*RPD*) of the calibration and validation sets were taken (Xing et al., 2016a). These evaluation parameters could be specifically defined by the following formulae:

$$RMSE = \sqrt{\frac{1}{N}\sum (x_i - y_i)^2} \tag{6}$$

$$RPD = SD/RMSE \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \overline{y_i})^2} \tag{8}$$

where x_i and y_i refer to the measured value and the corresponding estimated value respectively, $\overline{y_i}$ is the average of the estimated values, N denotes the number of samples, *SD* is the standard deviation of the measured values. High values of R^2 and *RPD*, along with a low *RMSE* value indicate a robust and accurate model. In soil science, models with *RPD* values < 1.4 are regarded as unacceptable for prediction; models with *RPD* values between 1.4 and 2 are supposed to have good prediction ability; and models with *RPD* values > 2 are considered as having excellent prediction ability.

The accuracy of the response obtained by the models was assessed using the F-test for variance according to the following formula (Filgueiras et al., 2014a; Oliveira et al., 2007; Pereira et al., 2008):

$$Fcalc = \frac{\mathrm{RMSE}_{V1}^2}{RMSE_{V2}^2} \tag{9}$$

where $\mathrm{RMSE}_{\mathrm{V1}} > \mathrm{RMSE}_{\mathrm{V2}}$. The Fcalc value was compared with the value of the Fisher-Snedecor distribution (F-statistic), with the degrees of freedom equal to the number of samples predicted, and a significance level of 5% ($\alpha = 0.05$). If the tabulated value of the F-statistic was less than Fcalc at the desired level of significance, then there was no statistical evidence of the homogeneity of the variances, and the method with $\mathrm{RMSE}_{\mathrm{V2}}$ presents better accuracy.

2.5.2. Systematic error assessment

Systematic errors may occur because of the inadequacy of the calibration models. Bias is the sum of the differences between the estimated values (y_i) and the measured values (x_i), and can be expressed as below (Filgueiras et al., 2014a; Steidle Neto et al., 2017b):

$$bias = \frac{\sum (y_i - x_i)}{n} \tag{10}$$

where n is the number of samples. A bias value close to zero indicates a low systematic error between the measured and predicted values (Steidle Neto et al., 2017a).

2.6. Software and codes

Data analysis was performed using MATLAB R2016a software (the MathWorks, Inc., Natick, MA, USA). Origin 2015 (Origin Lab Corporation, Northampton, MA, USA) was used for graphing. All Matlab codes for the present study are presented in Supplementary data.

3. Results

3.1. SOM contents, LIBS spectra, and FTIR-ATR spectra of soil samples

3.1.1. Descriptive statistics and distribution of SOM contents in the studied area

The SOM contents varied greatly among soil samples, wiht values ranging from 0.58 to 17.70 g kg^{-1} with a mean value of 9.10 g kg^{-1}. Wilding (1985) categorized coefficient of variation (CV) values into three classes with high (CV > 35%), moderate (15% < CV < 35%), and low variability (CV < 15%). The CV value of the SOM contents was 39.41%, suggesting that the SOM contents of the sampled soils were spatially varied within the studied area. The SOM distribution presented negative kurtosis (−0.56) with positive skewness (0.12), indicating a flat distribution concentrated at low values with relatively

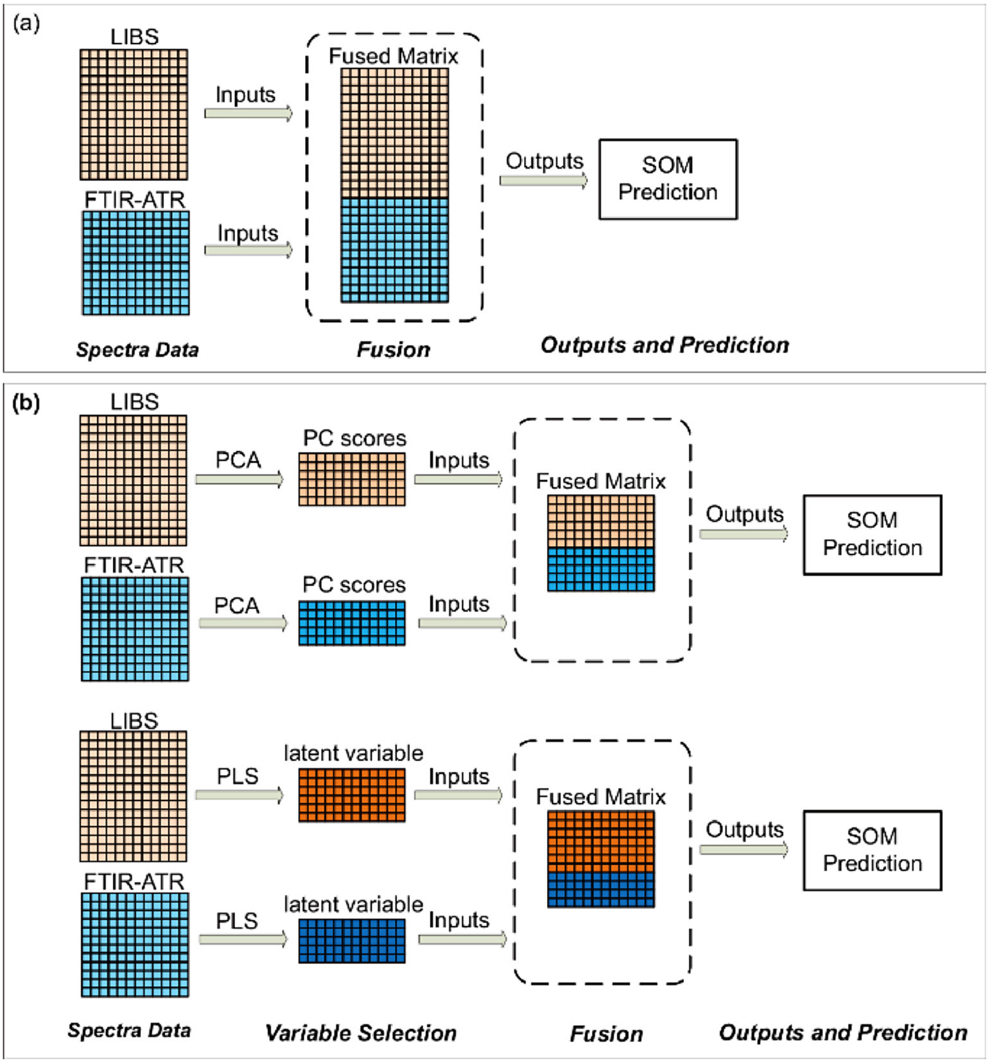

Fig. 2. Scheme of the data fusion process: (a) low-level data fusion and (b) mid-level data fusion.

few high values (Xu et al., 2018).

The distribution diagrams of the SOM contents of the top soils (0–20 cm) and the subsoils (20–40 cm) in the investigated area are shown in Fig. 1b. Obviously, the top soils (0–20 cm) had more SOM than the subsoils (20–40 cm). The middle part of the investigated area had less SOM than the marginal area both in the top soils and subsoils. The spatial variations of SOM in this area are likely caused by differences in environmental conditions and anthropogenic factors.

3.1.2. LIBS spectra of soil samples

The raw and preprocessed LIBS spectra in the wavelength range between 200 and 1000 nm were presented in Fig. 3a and b, respectively. The baseline wander of the LIBS spectra in the raw spectra could be observed, while these features were obviously eliminated after spectral preprocessing. The element characteristic lines were identified by comparing the experimental spectra with standard spectra from the NIST LIBS Database (Kramida, 2018) and from previous studies. The signals at 431.73, 656.825, and 746.83 nm were attributed to C II, H I, and N I emission lines, respectively (Harris et al., 2004; Juve et al., 2008; Khalil et al., 2017). Two O I emission lines were confirmed at 777.19 and 795.22 nm, respectively (Juve et al., 2008; Khalil et al., 2017). The continuous double peak at 742.43/744.32 nm was attributed to the S I emission line (Juve et al., 2008). Other mineral element peaks were also confirmed. Therefore, the results of the CHONS element identification in the soil LIBS spectra indicated that the SOM content can be characterized based on the LIBS spectra.

3.1.3. FTIR-ATR spectra of soil samples

Fig. 3c & d show the raw and preprocessed FTIR-ATR spectra of the soil samples in the wavenumber range from 4000 to 500 cm^{-1}. The noise and baseline shift in the spectra were already removed by spectral pretreatment. The intense vibrational bands in the region between 1760 and 800 cm^{-1} were likely attributed to C═C vibration of alkenes and aromatic compounds, which were centered at approximately 1640 cm^{-1}, and might also be caused by C═O vibration of amides and carboxylated groups (Sisouane et al., 2017; Smidt et al., 2008). Bands in the region between 1380 and 1050 cm^{-1} were assigned to the C—O stretch of polysaccharides, and of other groups such as alcohols, ether and esters (Dhillon et al., 2017; Janik et al., 2007a; Janik et al., 2007b; Spaccini and Piccolo, 2007). However, Haberhauer et al. (2000) indicated that these bands may overlap with Si—O stretching of silicate bands from mineral particles at 1050 cm^{-1}. Consequently, the bands below 1000 cm^{-1} were associated with a mixture of organic and inorganic compounds. In the present study, these functional groups of the SOM were included in the broadband between 1760 and 800 cm^{-1}.

3.2. Spectra pretreatment of spectra

3.2.1. PCA of raw and preprocessed spectra

The baseline wander of the spectra was visibly eliminated using

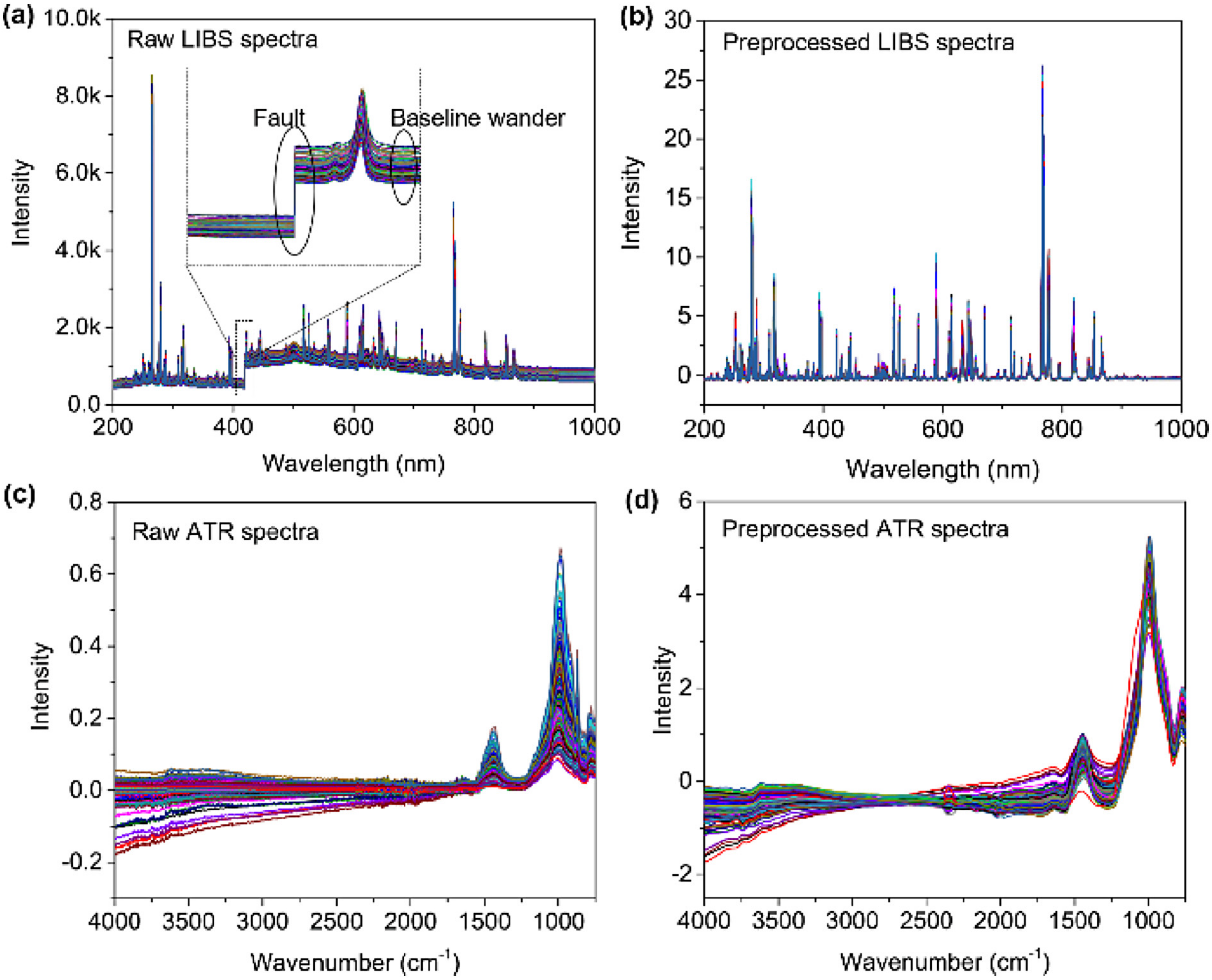

Fig. 3. LIBS and FTIR-ATR spectra of soil samples: (a) raw LIBS spectra, (b) preprocessed LIBS spectra, (c) raw FTIR-ATR spectra, and (d) preprocessed FTIR-ATR spectra.

various preprocessed approaches. To further understand the preprocessed efficiency, PCA was applied to evaluate the validity of the baseline correction on the LIBS spectra. The PCA results of one of the soil samples (S62) are presented in Fig. 4. Scatterplots between the first four scores of principal component analysis based on the original and baseline corrected spectra of 27 parallel samples are shown in Fig. 4a. The differences in the parallel samples were mainly reflected in the first four principal component directions since the first four principal components account for 90% of the total variance of the original spectra and 80% of the total variance of the baseline corrected spectra (Fig. 4b, c). It is important to highlight that the ranges of principal component scores of PC1 (δ1), PC2 (δ2), and PC3 (δ3) for the baseline corrected spectra were relatively lower in comparison with the original spectra (Fig. 4a), which indicated that MPLS could greatly reduce the variance originating from the background. Based on the PCA and spectral analysis, we conclude that it is the baseline wander of the spectra that causes large variations of the original spectra in the first four principal component directions from one spectrum to another. Evidently, these background variations among parallel spectra could be effectively removed using an MPLS algorithm.

PCA was also employed for assessing the effectiveness of the preprocessed methods on the FTIR-ATR spectra, and the results are shown in Fig. 5. The variances of the first two components increased from 39.54% and 3.38% of the total variance to 57.64% and 18.28% of the total variance, respectively, after spectral preprocessing. The first four principal components account for > 90% of the total variance for the preprocessed spectra which was higher than 68% of the total variance of the raw spectra (Fig. 5a). Furthermore, scatterplots of the first two principal components of the preprocessed spectra were uniformly located around the origin of the coordinates, compared with that of the raw spectra (Fig. 5b, c). This indicated that the calibration models would be well trained to predict the SOM content after spectral pretreatment.

3.2.2. PLSR of averaged LIBS spectra

The LIBS spectral average efficiency was assessed according to the statistic results using the PLSR model for SOM prediction (Table 1). In PLSR, the number of factors was optimized according to the minimum $RMSE_{CV}$ using cross-validation (Fig. S1, Supplementary data). For the PCA method, the $RMSE_{CV}$ (2.30 g kg^{-1}) and $RMSE_V$ (2.67 g kg^{-1}) of the PLSR model increased while R_V^2 (0.505) decreased compared to the AM method. This suggested that the PCA method was not suitable for preprocessing the parallel spectra and even reduced the representativeness of the spectra. Both the PLSR based on the EDWM and the PCWM methods achieved a better prediction capability with superior parameters ($RMSE_{CV}$ = 2.31 and 2.22, R_C^2 = 0.715 and 0.721, $RMSE_C$ = 1.88 and 1.86 g kg^{-1}, RPD_C = 1.88 and 1.90, R_V^2 = 0.540 and 0.581, $RMSE_V$ = 2.53 and 2.43 g kg^{-1}, and RPD_V = 1.51 and 1.57) compared with the AM method. As can be seen from the results, the EDWM and PCWM methods both improved the prediction ability of the SOM content based on the PLSR model, i.e., they were suitable for preprocessing LIBS parallel spectra. Because the PCWM method significantly improved both the prediction ability of SOM in calibration and validation sets, the PCWM method was employed for pretreatment

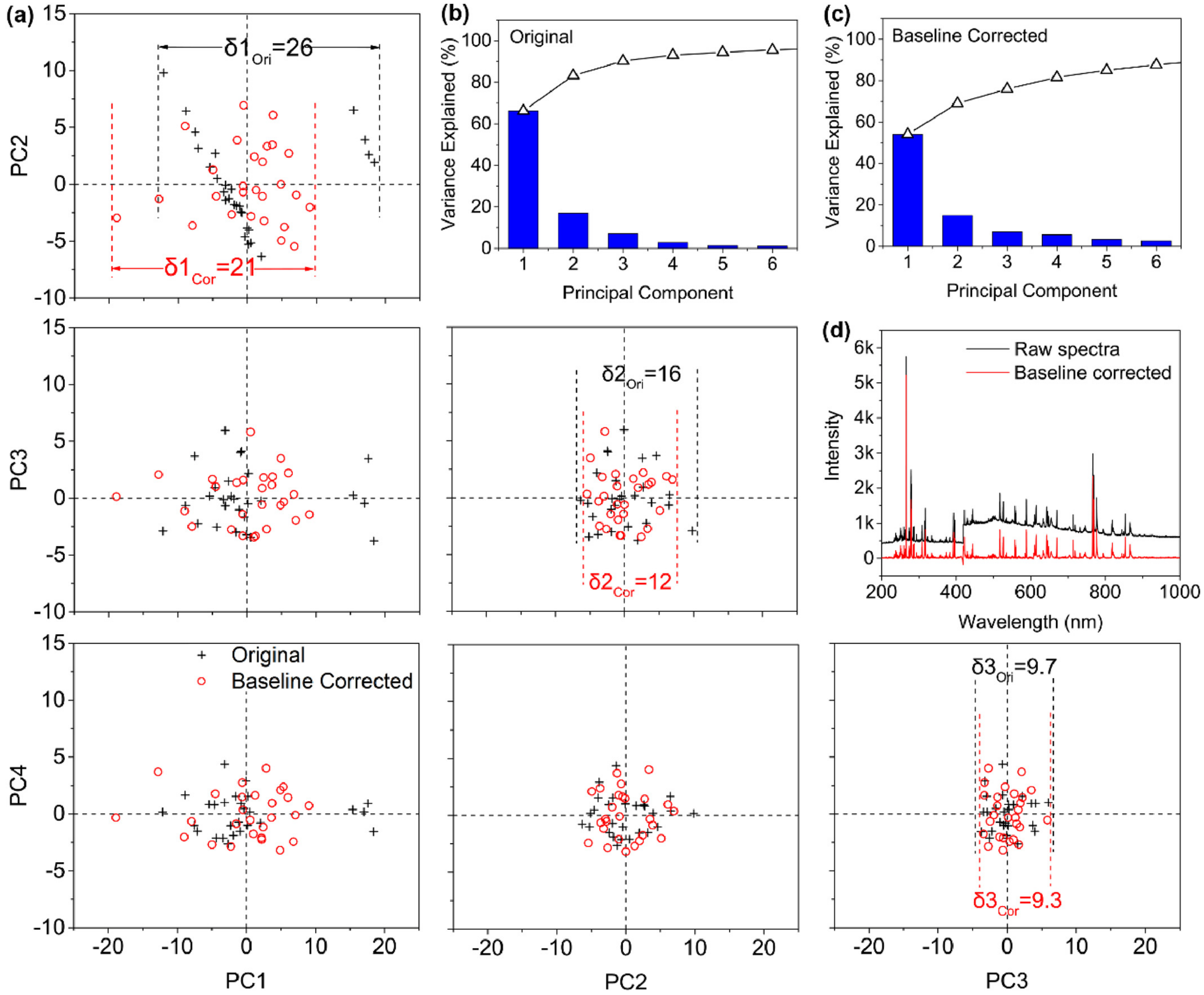

Fig. 4. Principal component analysis (PCA) results of raw and baseline-corrected LIBS spectra: (a) Scatterplot between the first four principal component analysis scores based on original and baseline corrected parallel spectra of sample, (b) variance explained change by first six principal components of original parallel spectra of sample, (c) variance explained change by first six principal components of baseline corrected parallel spectra of sample, and (d) baseline correction results of spectra using MPLS algorithm.

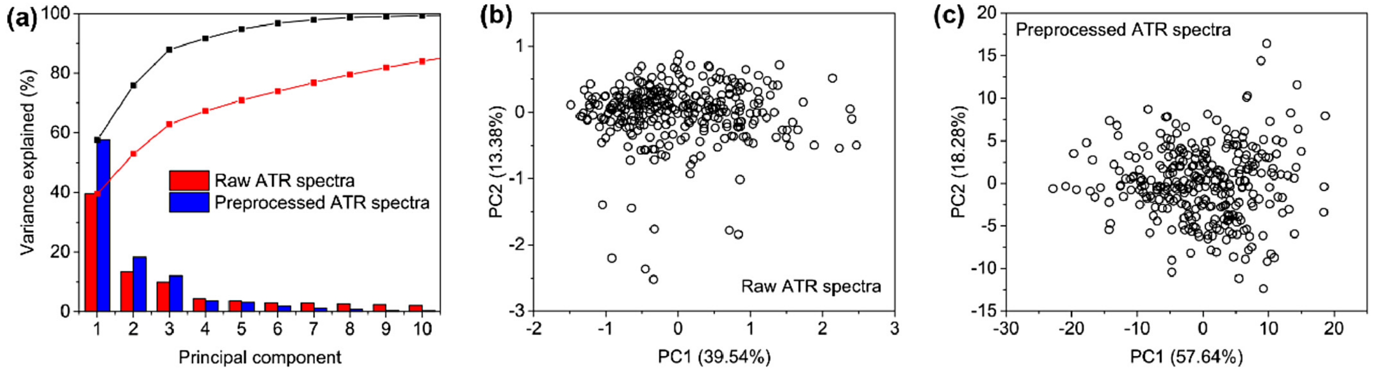

Fig. 5. Principal component analysis (PCA) results of raw and preprocessed FTIR-ATR spectra: (a) PCA variance explained by principal component, (b) scatterplots between the first two principal component analysis scores of raw FTIR-ATR spectra, and (c) scatterplots between the first two principal component analysis scores of preprocessed FTIR-ATR spectra.

Table 1
Statistics of the PLSR model used in the calibration and validation sets for the prediction of SOM in soil based on different pre-processing methods of parallel LIBS spectra.

Methods	NF	Calibration (228 samples)				Validation (76 samples)		
		$RMSE_{CV}$	R_C^2	$RMSE_C$	RPD_C	R_V^2	$RMSE_V$	RPD_V
AM	6	2.28	0.690	1.96	1.80	0.509	2.66	1.42
PCA	7	2.30	0.691	1.96	1.80	0.505	2.67	1.42
EDWM	8	2.31	0.715	1.88	1.88	0.540	2.53	1.51
PCWM	8	2.22	0.721	1.86	1.90	0.581	2.43	1.57

AM: arithmetic mean method for pre-processing of parallel LIBS spectra; PCA: principal component analysis used to dimensionality reduce the parallel LIBS spectra; EDWM: Euclidean distance as weight for weighted mean of parallel LIBS spectra; PCWM: principal component scores as weight for weighted mean of parallel LIBS spectra; NF: number of factors used in PLS regression; RMSE: the root mean square error ($g\,kg^{-1}$); RPD: the residual prediction deviation.

of parallel LIBS spectra.

3.3. Prediction of SOM

3.3.1. Comparison of various data fusion approaches

The prediction results of SOM based on various data fusion approaches were compared with those without data fusion models (Table 2 and Fig. 6). The number of factors was optimized according to the minimum $RMSE_{CV}$ using cross-validation (Fig. S2). The validation results were comparable to those from the calibration measurement, indicating that the models were robust and stable. The performance of the PLSR models for SOM prediction differed from the spectral the datasets as indicated by the RPD values (Xing et al., 2016a). For the FTIR-ATR spectral dataset, the RPD value in the validation set was < 1.4, which implied an unreliable prediction of SOM content. High $RMSE_V$ (2.89 $g\,kg^{-1}$) and low R_V^2 values of the validation set (below 0.500) also corresponded well to those low RPD values of the FTIR-ATR-PLSR model. For the LIBS spectral dataset, higher R^2 and RPD values were obtained compared to those of the FTIR-ATR spectra. This indicated that the soil LIBS spectra were more reliable and accurate than the FTIR-ATR spectra for predicting SOM content.

For the low-level data fusion (LLDF) strategy, higher R_C^2 and RPD_C values as well as lower $RMSE_{CV}$ and $RMSE_C$ values in the calibration set were observed compared with the FTIR-ATR-PLSR and LIBS-PLSR models, which indicated that a better calibration model was obtained with the LLDF strategy. However, the prediction accuracy (R_V^2 = 0.613, $RMSE_V$ = 2.44 $g\,kg^{-1}$, and RPD_V = 1.55) of the validation set was not promoted compared with the LIBS-PLSR model. These results suggested that there was over-fitting in the LLDF-PLSR model. For mid-level data fusion based on the concatenation of PCA scores (MLDF-PCs), the R^2 and RPD values of the MLDF-PCs-PLSR model were both higher than those of the FTIR-ATR-PLSR and the LIBS-PLSR models. The $RMSE_{CV}$, $RMSE_C$, and $RMSE_V$ of the MLDF-PCs-PLSR model were both lower than those of the FTIR-ATR-PLSR and LIBS-PLSR models. According to F-test, the $RMSE_V$ value of the LLDF-PCs-PLSR model showed significant ($P < 0.05$) differences compared with that of the FTIR-ATR-PLSR model (Table 3). These results implied that the MLDF-PCs strategy was able to improve the prediction ability of the SOM content whereas the LLDF strategy failed to do so. This was consistent with the observations from the scatterplots (Fig. 6a-d), where the scatters of the calibration set and the validation set in the MLDF-PCs-PLSR models were located closer to the 1:1 line than other models. For mid-level data fusion based on concatenating of latent variables (MLDF-LV), the MLDF-LV-PLSR model ($RMSE_{CV}$ = 1.48 $g\,kg^{-1}$, $RMSE_V$ = 1.76 $g\,kg^{-1}$, and RPD_V = 2.16) was the best method to predict SOM, and was considered to provide excellent prediction ability. The F-test of the $RMSE_V$ also indicated a significant ($P < 0.05$) improvement of the prediction accuracy (Table 3). Fig. 6e shows the splashes of the calibration set and validation set in the MLDF-LV-PLSR model located closer to the 1:1 line, which also suggested a precise prediction of SOM content.

3.3.2. Comparison of various calibration methods

In order to investigate the applicability of the MLDF-LV strategy for different models, latent variables obtained using the MLDF-LV strategy were calibrated and validated against the measured SOM content using SVR and ANN models, and the results were compared with the PLSR model. Three multivariate methods (including the PLSR model) provided different prediction accuracy for the SOM content. Scatterplots of the predicted vs. measured SOM from the soil samples for the different models are illustrated in Fig. 6e-g and the descriptive regression statistics are provided in Table 2. The SVR and ANN models provided better prediction accuracy (R_V^2 = 0.811, $RMSE_V$ = 1.68 $g\,kg^{-1}$, RPD_V = 2.27 for SVR and R_V^2 = 0.830, $RMSE_V$ = 1.60 $g\,kg^{-1}$, RPD_V = 2.39 for ANN) than the PLSR model. However, the $RMSE_V$ showed no significant differences among the three models according to the F-test (Table 3). All three models presented a relatively high RPD value > 2.0 for both calibration and validation sets and had excellent prediction ability, which were comparable and often better than those obtained by other techniques (Table 4).

3.4. Systematic error and residual assessment

Systematic errors are biases in the measurement which lead to situations where the mean of many separate measurements differs significantly from the actual value of the measured attribute, this may originate from imperfect calibration. A bias value close to zero indicates low systematic error between the measured and the predicted values.

Table 2
Statistics of the PLSR, SVR, and ANN models used in the calibration and validation sets for the prediction of SOM content in soil based on different data fusion strategies.

Datasets for modeling	Model	NF	Calibration (228 samples)				Validation (76 samples)			Bias
			$RMSE_{CV}$	R_C^2	$RMSE_C$	RPD_C	R_V^2	$RMSE_V$	RPD_V	
FTIR-ATR	PLSR	12	2.48	0.662	2.05	1.72	0.487	2.89	1.35	0.01
LIBS	PLSR	8	2.22	0.721	1.88	1.88	0.581	2.43	1.57	0.16
LLDF	PLSR	13	1.92	0.858	1.33	2.66	0.613	2.44	1.55	0.16
MLDF-PCs	PLSR	10	2.13	0.781	1.65	2.14	0.630	2.35	1.61	0.15
MLDF-LVs	PLSR	8	1.48	0.862	1.31	2.70	0.792	1.76	2.16	0.16
MLDF-LVs	SVR	–	–	0.854	1.39	2.54	0.811	1.68	2.27	−0.09
MLDF-LVs	ANN	–	–	0.860	1.34	2.64	0.830	1.60	2.39	−0.12

LLDF: low-level data fusion based on concatenation of full LIBS spectra and full FTIR-ATR spectra; MLDF-PCA: mid-level data fusion based on concatenation of PCA scores of LIBS spectra and ATR spectra; MLDF-LV: mid-level data fusion based on concatenation of latent variables of LIBS spectra and ATR spectra; PLSR: partial least squares regression; SVR: support vector regression; ANN: artificial neural network; NF: number of factors used in regression; RMSE: the root mean square error ($g\,kg^{-1}$); RPD: the residual prediction deviation; bias ($g\,kg^{-1}$).

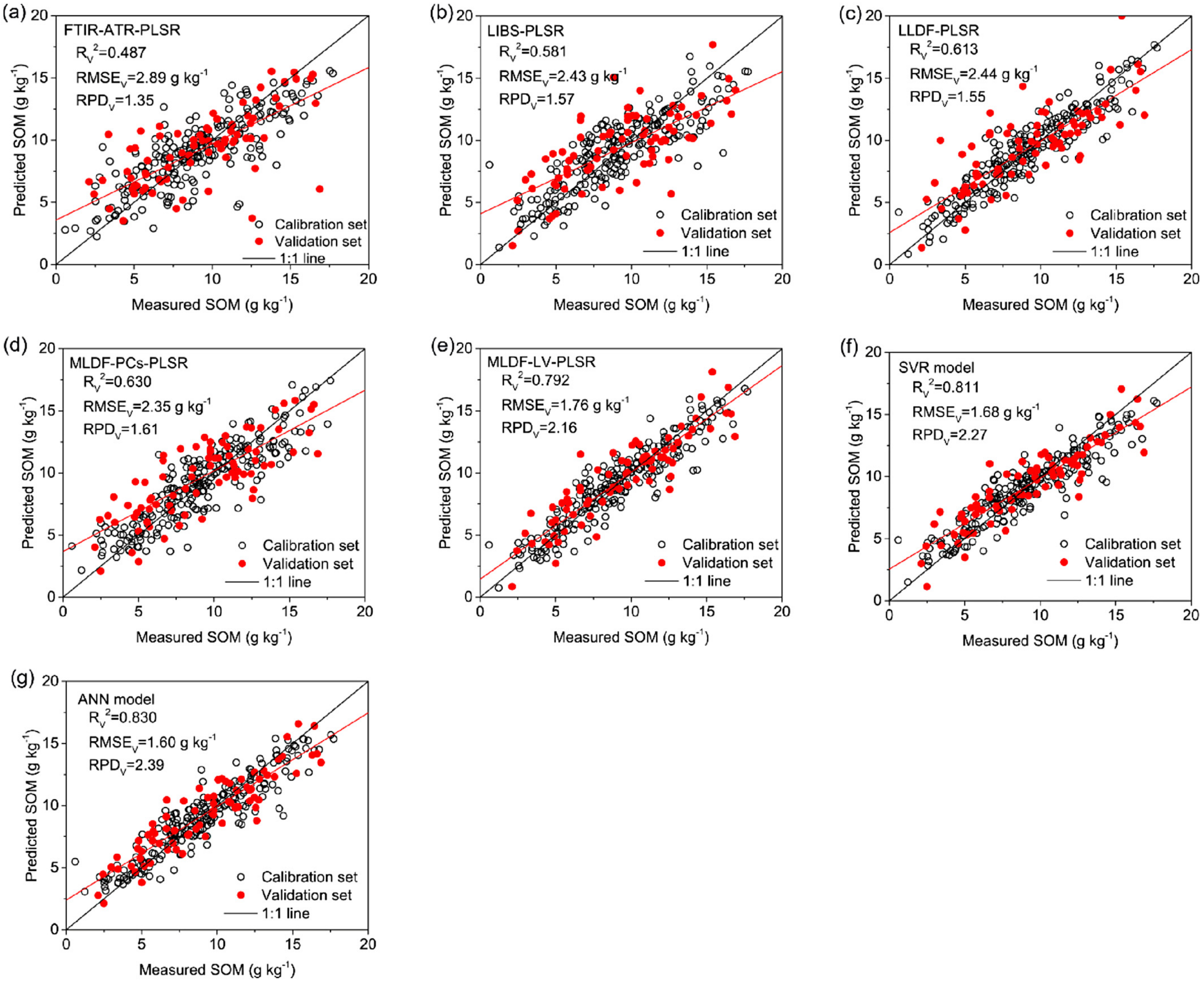

Fig. 6. (a-e) Scatterplots of measured values vs the predicted values of SOM contents based on FTIR-ATR spectra (a), LIBS spectra (b), low-level data fusion (c), mid-level data fusion of scores of PCA (d), and mid-level data fusion of latent variables of PLS (e), (f, g) scatterplots of measured values vs the predicted values of SOM content based on mid-level data fusion of latent variables by using SVR (f) and ANN (g) models.

Table 3
Accuracy assessment of SOM prediction models using F-test of $RMSE_V$ of various models.

		FTIR-ATR-PLSR	LIBS-PLSR	LLDF-PLSR	MLDF-PCs-PLSR	MLDF-LV-PLSR	MLDF-LV-SVR	MLDF-LV-ANN
	$RMSE_V$	2.89	2.43	2.44	2.35	1.76	1.68	1.60
FTIR-ATR-PLSR	2.89							
LIBS-PLSR	2.43	1.414						
LLDF-PLSR	2.44	1.403	1.008					
MLDF-PCs-PLSR	2.35	1.512*	1.069	1.078				
MLDF-LV-PLSR	1.76	2.696*	1.906*	1.922*	1.783*			
MLDF-LV-SVR	1.68	2.959*	2.092*	2.109*	1.957*	1.098		
MLDF-LV-ANN	1.60	3.263*	2.307*	2.326*	2.157*	1.210	1.103	

The values in table (Fcalc) are compared with the value of the Fisher–Snedecor distribution (F-statistic, 1.462), with degrees of freedom equal to the number of samples predicted ($\nu 1 = \nu 2 = 76$) and a significance level of 5% ($\alpha = 0.05$); if the Fcalc value is greater than F-statistic value (1.462), there is no statistical evidence of the homogeneity of the variances, and the model with lower $RMSE_V$ presents better accuracy (footnoted with *).

The biases between these two groups of values are given in Table 2. According to the absolute value, the bias in descending order was given as LIBS-PLSR model, LLDF-PLSR model, MLDF-LV-PLSR model, MLDF-PCs-PLSR model, MLDF-LV-ANN model, MLDF-LV-SVR model, and FTIR-ATR-PLSR model. Although the FTIR-ATR-PLSR model showed the lowest systematic error, it was still not suitable for SOM prediction because of the low prediction accuracy. However, the MLDF-LV-PLSR, MLDF-LV-SVR, and MLDF-LV-ANN models showed low systematic errors with high prediction accuracies, indicating that these models possessed potential for SOM prediction.

Fig. 7 presents the results of the calibration and validation residuals with the measured SOM content. In the MLDF-PCs-PLSR, MLDF-LV-

Table 4
Comparison of various chemometrics and spectral techniques for SOM content prediction in previous studies.

Soil type	N	SOM ($g\ kg^{-1}$)	Spectra	Multivariate calibration	Parameters			Reference
					PMSE$_V$ ($g\ kg^{-1}$)	R_V^2	RPD$_V$	
Black soil, fluvo-aquic soil, paddy soil, red soil	194	5.74–64.89	FTIR-PAS and Raman	PLSR	6.74	0.81	2.18	(Xing et al., 2016a)
Black soil, fluvo-aquic soil, paddy soil, red soil	200	5.74–97.41	Raman	PLSR	8.16	0.74	1.92	(Xing et al., 2016b)
Paddy soil	933	3.95–45.69	FTIR-PAS	Self-Adaptive PLSR	1.65	0.9293	3.18	(Ma et al., 2016)
Fluvo-aquic soil	56	6.23–13.73	FTIR-PAS	PLSR	0.99	0.9238	–	(Du et al., 2009)
Red soil and paddy soil	585	3.63–68.31	Vis-NIR	SVMR	4.87	0.88	2.84	(Xu et al., 2018)
Brown soil	304	0.58–17.7	LIBS and FTIR-ATR	PLSR (data fusion)	1.76	0.792	2.16	This study
				SVR (data fusion)	1.68	0.811	2.27	
				ANN (data fusion)	1.60	0.830	2.39	

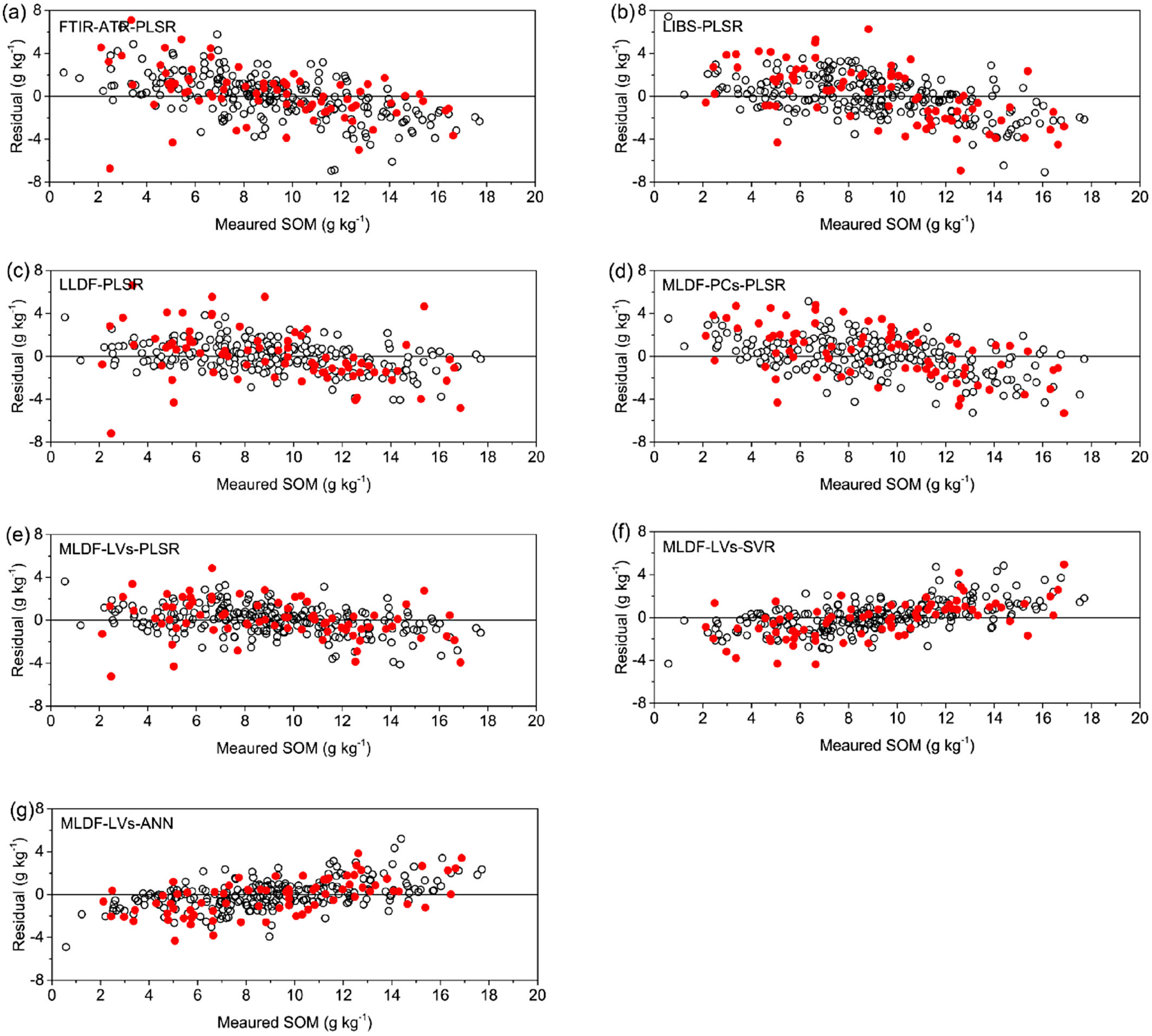

Fig. 7. Residuals for the quantification of SOM content from soil samples under various data fusion approaches and regression models: (a) PLSR model based on FTIR-ATR spectra, (b) PLSR model based on LIBS spectra, (c) PLSR model based on low-level data fusion, (d) PLSR model based on mid-level data fusion of scores of PCA, (e) PLSR model based on mid-level data fusion of latent variables of PLS, (f) SVR model based on mid-level data fusion of latent variables of PLS, and (g) ANN model based on mid-level data fusion of latent variables of PLS.

PLSR, MLDF-LV-SVR, and MLDF-LV-ANN models, the residuals were close to the horizontal line, which also indicated the excellent prediction accuracy of those models. Trends in the residuals for the FTIR-ATR-PLSR, LIBS-PLSR, LLDF-PLSR, MLDF-PCs-PLSR, and MLDF-LV-SVR models were observed (Fig. 7). In the FTIR-ATR-PLSR and LIBS-PLSR model, when the measured SOM values were below 4 g kg^{-1} or above 14 g kg^{-1}, the calibration and validation residuals increased, and the prediction accuracy decreased.

4. Discussion

4.1. Spectral pre-processing

Baseline correction approaches are usually used to remove background noises for Raman spectra and chromatograms (Baek et al., 2015; Fu et al., 2016; Hu et al., 2007). However, there is generally an absence of literatures discussing the use of baseline correction for LIBS spectra. In the present study, we employed the MPLS algorithm to remove noise and baseline shift in the LIBS and FTIR-ATR spectra. The resolution of the LIBS and FTIR-ATR spectra can be quickly improved using the MPLS method, thereby decreasing the large variations in parallel spectra according to the principal component analysis. The spot diameter of the laser employed in this study was 50 μm. Because the compositions of soil are heterogeneously distributed in the soil samples at the micron level, uneven ablation on the soil's surface at each paralleled shot was observed. This resulted in large differences between the paralleled LIBS spectra. The intensities of each element in the parallel LIBS spectra were significantly different in this study. This greatly limited the use of averaged LIBS spectra to reveal soil composition information. There are very few studies focusing on the development of superior algorithms to resolve this issue. In our work, three approaches, i.e., principal component analysis (PCA), principal component weighted mean (PCWM), and Euclidean distance weighted mean (EDWM), were proposed and compared with the arithmetic mean (AM) method. Optimal typical LIBS spectra of the soil samples were obtained. The prediction abilities of the SOM content by the PLS model was significantly improved after applying the PCMW and the EDMW algorithms to parallel LIBS spectra, which are very promising. In theory, more advanced parallel data integration techniques could be developed to remove errors caused by soil heterogeneity and to improve the prediction ability of the LIBS spectra.

4.2. Data fusion and multivariate techniques

A combination of multiplex spectroscopic techniques using data fusion strategies in soil property prediction is another effective approach for improving the prediction ability. According to the results of the calibration and statistical analysis, mid-level data fusion of the LIBS and FTIR-ATR spectra based on the latent variables of PLS could dramatically improve the SOM prediction accuracy. LIBS spectra canprovide information on the elemental content of soil but does not provide compositional information. The prediction accuracy could be affected by other components (e.g., CO_3^{2-}, HCO_3^-, NO_3^-, and H_2O) which had the same elements with SOM when using the separate LIBS spectra coupled with multivariate techniques for SOM prediction. However, the FTIR-ATR spectra, as molecular spectroscopy technique, reveal structural information about the components of soil. Therefore, a combination of LIBS and FTIR-ATR spectra could eliminate interference and provide supplementary information which is lacking in the individual data. The SOM prediction ability was thereby improved. Theoretically, more accurate models with fused data can be developed to exclude strong interferences caused by each spectroscopic approach when implementing more advanced data integration techniques (e.g., high-level data fusion).Nonetheless, care should be taken when using this approach since the data fusion may introduce the accumulation of prediction errors of the separate spectroscopic tools.

The combination of spectroscopic method with chemometrics is a novel, fast, and nondestructive approach for composition determination and material identification. In the present study, the PLSR, SVR, and ANN models were chosen for SOM prediction, and their prediction abilities were compared using statistical analysis. PLSR, SVR, and ANN models showed high prediction ability for SOM according to the $RMSE_V$, R_V^2, and RPD_V values. The preprocessing and data fusion of the LIBS and FTIR-ATR spectra obviously improved prediction ability. Using chemometrics based on the data fusion of LIBS and FTIR-ATR spectra offer a simple, fast, and nondestructive approach for monitoring SOM content, and it is of great significance in the evaluation of soil fertility, the management of soil nutrient, and the guidance of agricultural production.

4.3. Prediction influenced by soil spatial variation

The spatial structure of calibration and validation samples are of great importance (Brown et al., 2005). Generally, the prediction model shows more stability and effectiveness when applied to similar soils in the same physiographic region than heterogeneous soils (Brown et al., 2005). In this study, the variations of soil properties caused by different sampling sites were greater than those caused by different depths. Thus, the correlation of SOM contents between the 0–20 cm deeep soils and 20–40 cm deep soils was investigated. As shown in Fig. S3, the SOM contents of 0–20 cm deep soils showed a linear relationship with those of 20–40 cm deep soils (with $r^2 = 0.452$). Therefore, the spatial pseudo replication may occur if different depths of each profile are included in calibration and validation sets. The subset partition without regard to the depths resulted in higher prediction accuracy than that which considered the depths, because of spatial pseudo replication (Fig. S4). Thus, overestimation of prediction accuracy caused by the spatial pseudo replication should be avoided in future studies.

5. Conclusions

In the present study, various spectral preprocessing methods, data fusion strategies, and chemometrics algorithms were performed to improve the quantitative prediction ability of SOM content. The results indicated that the soil LIBS and FTIR-ATR spectral baseline wander could be removed by applying the MPLS and wavelet transform algorithm. The characteristic LIBS spectra obtained using the PCWM, and EDWM algorithms presented lower errors and higher quantitative prediction ability of SOM in the PLSR model than using the AM and PCA algorithms. The quantitative prediction ability of SOM could be dramatically improved by using mid-level data fusion based on feature extraction of the PLSR model (MLDF-LV). Moreover, the SVR and ANN models also showed excellent prediction accuracy and robustness for SOM prediction according to systematic error and residual assessment. The findings from this study indicate that the application of soil LIBS and FTIR-ATR spectroscopy combined with chemometrics can result in an affordable and efficient procedure for the determination of SOM content and facilitate improved soil fertility monitoring.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (41671238), the National Basic Research Program of China (No. 2015CB150403), the “STS” Project from Chinese Academy of Sciences (KFJ-PTXM-003, KFJ-STS-QYZX-047), and the Jiangsu Demonstration Project in Modern Agriculture (BE2017388).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2019.113905.

References

Baek, S.J., Park, A., Ahn, Y.J., Choo, J., 2015. Baseline correction using asymmetrically reweighted penalized least squares smoothing. Analyst 140 (1), 250–257.

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., Busto, O., 2015. Data fusion methodologies for food and beverage authentication and quality assessment – a review. Anal. Chim. Acta 891, 1–14.

Brech, F.a.C., L., 1962. Optical microemission stimulated by a ruby laser. Appl. Spectrosc. 16, 59–64.

Brown, D.J., Bricklemyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129 (3), 251–267.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2012. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. Geoderma 183-184, 41–48.

Cremers, D.A., Radziemski, L.J., 2006. Handbook of Laser-Induced Breakdown Spectroscopy. Wiley, New York.

Devos, O., Ruckebusch, C., Durand, A., Duponchel, L., Huvenne, J.P., 2009. Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation. Chemometr. Intell. Lab. 96 (1), 27–33.

Dhillon, G.S., Gillespie, A., Peak, D., Van Rees, K.C.J., 2017. Spectroscopic investigation of soil organic matter composition for shelterbelt agroforestry systems. Geoderma 298, 1–13.

Du, C.W., Zhou, J.M., Wang, H.Y., Chen, X.Q., Zhu, A.N., Zhang, J.B., 2009. Determination of soil properties using Fourier transform mid-infrared photoacoustic spectroscopy. Vib. Spectrosc. 49 (1), 32–37.

El Haddad, J., Bruyere, D., Ismael, A., Gallou, G., Laperche, V., Michel, K., Canioni, L., Bousquet, B., 2014. Application of a series of artificial neural networks to on-site quantitative analysis of lead into real soil samples by laser induced breakdown spectroscopy. Spectrochim. Acta B 97, 57–64.

Ferreira, E.C., Milori, D., Ferreira, E.J., Da Silva, R.M., Martin-Neto, L., 2008. Artificial neural network for cu quantitative determination in soil using a portable laser induced breakdown spectroscopy system. Spectrochim. Acta B 63 (10), 1216–1220.

Filgueiras, P.R., Alves, J.C.L., Sad, C.M.S., Castro, E.V.R., Dias, J.C.M., Poppi, R.J., 2014a. Evaluation of trends in residuals of multivariate calibration models by permutation test. Chemometr. Intell. Lab. 133, 33–41.

Filgueiras, P.R., Sad, C.M.S., Loureiro, A.R., Santos, M.F.P., Castro, E.V.R., Dias, J.C.M., Poppi, R.J., 2014b. Determination of API gravity, kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. Fuel 116, 123–130.

Fu, H.Y., Li, H.D., Yu, Y.J., Wang, B., Lu, P., Cui, H.P., Liu, P.P., She, Y.B., 2016. Simple automatic strategy for background drift correction in chromatographic data analysis. J. Chromatogr. A 1449, 89–99.

Haberhauer, G., Feigl, B., Gerzabek, M.H., Cerri, C., 2000. FT-IR spectroscopy of organic matter in tropical soils: changes induced through deforestation. Appl. Spectrosc. 54 (2), 221–224.

Harris, R.D., Cremers, D.A., Ebinger, M.H., Bluhm, B.K., 2004. Determination of nitrogen in sand using laser-induced breakdown spectroscopy. Appl. Spectrosc. 58 (7), 770–775.

Hu, Y.G., Jiang, T., Shen, A.G., Li, W., Wang, X.P., Hu, J.M., 2007. A background elimination method based on wavelet transform for Raman spectra. Chemometr. Intell. Lab. 85 (1), 94–101.

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M., Rawson, A., 2007a. Rapid prediction of soil water retention using mid infrared spectroscopy. Soil Sci. Soc. Am. J. 71 (2), 507–514.

Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007b. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Aust. J. Soil Res. 45 (2), 73–81.

Jia, S.Y., Li, H.Y., Wang, Y.J., Tong, R.Y., Li, Q., 2017. Hyperspectral imaging analysis for the classification of soil types and the determination of soil total nitrogen. Sensors 17 (10), 2252.

Jimenez-Carvelo, A.M., Gonzalez-Casado, A., Cuadros-Rodriguez, L., 2017. A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction. Talanta 164, 540–547.

Juve, V., Portelli, R., Boueri, M., Baudelet, M., Yu, J., 2008. Space-resolved analysis of trace elements in fresh vegetables using ultraviolet nanosecond laser-induced breakdown spectroscopy. Spectrochim. Acta B 63 (10), 1047–1053.

Khalil, A.A.I., Morsy, M.A., El-Deen, H.Z., 2017. Development of double-pulse lasers ablation system and electron paramagnetic resonance spectroscopy for direct spectral analysis of manganese doped PVA polymer. Opt. Laser Technol. 96, 227–237.

Kim, E.-A., Choi, J.H., 2018. Changes in the mineral element compositions of soil colloidal matter caused by a controlled freeze-thaw event. Geoderma 318, 160–166.

Knadel, M., Gislum, R., Hermansen, C., Peng, Y., Moldrup, P., de Jonge, L.W., Greve, M.H., 2017. Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination. Biosyst. Eng. 156, 157–172.

Kramida, A., Ralchenko, Y., Reader, J., NIST ASD Team, 2018. 2018. NIST Atomic Spectra Database (ver. 5.5.6), [Online]. National Institute of Standards and Technology, Gaithersburg, MD.

Lado, M., Paz, A., Ben-Hur, M., 2004. Organic matter and aggregate-size interactions in saturated hydraulic conductivity. Soil Sci. Soc. Am. J. 68 (1), 234–242.

Li, H.D., Liang, Y.Z., Xu, Q.S., 2009. Support vector machines and its applications in chemistry. Chemometr. Intell. Lab. 95 (2), 188–198.

Li, Z., Zhan, D.J., Wang, J.J., Huang, J., Xu, Q.S., Zhang, Z.M., Zheng, Y.B., Liang, Y.Z., Wang, H., 2013. Morphological weighted penalized least squares for background correction. Analyst 138 (16), 4483–4492.

Liang, D., Du, C.W., Ma, F., Shen, Y.Z., Wu, K., Zhou, J.M., 2017. Characterization of nano Fe-III-tannic acid modified polyacrylate in controlled-release coated urea by Fourier transform infrared photoacoustic spectroscopy and laser-induced breakdown spectroscopy. Polym.Test. 64, 101–108.

Lu, Y.Z., Du, C.W., Yu, C.B., Zhou, J.M., 2014. Fast and nondestructive determination of protein content in rapeseeds (Brassica napus L.) using Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS). J. Sci. Food Agr. 94 (11), 2239–2245.

Luce, M.S., Ziadi, N., Zebarth, B.J., Grant, C.A., Tremblay, G.F., Gregorich, E.G., 2014. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. Geoderma 232-234, 449–458.

Ma, F., Du, C., Zhou, J., 2016. A self-adaptive model for the prediction of soil organic matter using mid-infrared photoacoustic spectroscopy. Soil Sci. Soc. Am. J. 80 (1), 238–246.

Martin, M.Z., Wullschleger, S.D., Garten, C.T., Palumbo, A.V., 2003. Laser-induced breakdown spectroscopy for the environmental determination of total carbon and nitrogen in soils. Appl. Opt. 42 (12), 2072–2077.

Meng, D.S., Zhao, N.J., Ma, M.J., Fang, L., Gu, Y.H., Jia, Y., Liu, J.G., Liu, W.Q., 2017. Application of a mobile laser-induced breakdown spectroscopy system to detect heavy metal elements in soil. Appl. Opt. 56 (18), 5204–5210.

Nelson, D.W., Sommers, L.E., 1982. Total carbon, organic carbon and organic matter. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis. American Society of Agronomy and Soil Science Society of American, Madison, pp. 552–553.

Oades, J.M., 1993. The role of biology in the formation, stabilization and degradation of soil structure A2 - Brussaard, L. In: Kooistra, M.J. (Ed.), Soil Structure/Soil Biota Interrelationships. Elsevier, Amsterdam, pp. 377–400.

Oliveira, F.C.C., Brandão, C.R.R., Ramalho, H.F., da Costa, L.A.F., Suarez, P.A.Z., Rubim, J.C., 2007. Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. Anal. Chim. Acta 587 (2), 194–199.

de la Paz Jimenez, M., de la Horra, A., Pruzzo, L., Palma, M.R., 2002. Soil quality: a new index based on microbiological and biochemical parameters. Biol. Fert. Soils 35 (4), 302–306.

Pereira, C.F., Pimentel, M.F., Galvão, R.K.H., Honorato, F.A., Stragevitch, L., Martins, M.N., 2008. A comparative study of calibration transfer methods for determination of gasoline quality parameters in three different near infrared spectrometers. Anal. Chim. Acta 611 (1), 41–47.

Reeves, J., McCarty, G., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environ. Pollut. 116, S277–S284.

Rehan, I., Gondal, M.A., Rehan, K., 2018. Determination of lead content in drilling fueled soil using laser induced spectral analysis and its cross validation using ICP/OES method. Talanta 182, 443–449.

Rezaei, F., Karimi, P., Tavassoli, S.H., 2014. Effect of self-absorption correction on LIBS measurements by calibration curve and artificial neural network. Appl. Phys. B Lasers Opt. 114 (4), 591–600.

Rial, M., Cortizas, A.M., Rodríguez-Lado, L., 2016. Mapping soil organic carbon content using spectroscopic and environmental data: a case study in acidic soils from NW Spain. Sci. Total Environ. 539, 26–35.

Rifai, K., Laflamme, M., Constantin, M., Vidal, F., Sabsabi, M., Blouin, A., Bouchard, P., Fytas, K., Castello, M., Kamwa, B.N., 2017. Analysis of gold in rock samples using laser-induced breakdown spectroscopy: matrix and heterogeneity effects. Spectrochim. Acta B 134, 33–41.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: E.R. David, L.M. James, C.P.R. Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. vol. 1. MIT Press, pp. 318–362.

Senesi, G.S., 2014. Laser-Induced Breakdown Spectroscopy (LIBS) applied to terrestrial and extraterrestrial analogue geomaterials with emphasis to minerals and rocks. Earth-Sci. Rev. 139, 231–267.

Senesi, G.S., Senesi, N., 2016. Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. Anal. Chim. Acta 938, 7–17.

Shao, Y.Q., Du, C.W., Shen, Y.Z., Ma, F., Zhou, J.M., 2017. Evaluation of net nitrification rates in paddy soil using mid-infrared attenuated total reflectance spectroscopy. Anal. Methods 9 (5), 748–755.

Sharma, P., Laor, Y., Raviv, M., Medina, S., Saadi, I., Krasnovsky, A., Vager, M., Levy, G.J., Bar-Tal, A., Borisover, M., 2017. Compositional characteristics of organic matter and its water-extractable components across a profile of organically managed soil. Geoderma 286, 73–82.

Silvestri, M., Bertacchini, L., Durante, C., Marchetti, A., Salvatore, E., Cocchi, M., 2013. Application of data fusion techniques to direct geographical traceability indicators. Anal. Chim. Acta 769, 1–9.

Sisouane, M., Cascant, M.M., Tahiri, S., Garrigues, S., El Krati, M., Boutchich, G.E.K., Cervera, M.L., de la Guardia, M., 2017. Prediction of organic carbon and total nitrogen contents in organic wastes and their composts by infrared spectroscopy and partial least square regression. Talanta 167, 352–358.

Smidt, E., Meissl, K., Schwanninger, M., Lechner, P., 2008. Classification of waste materials using Fourier transform infrared spectroscopy and soft independent modeling of class analogy. Waste Manag. 28 (10), 1699–1710.

Spaccini, R., Piccolo, A., 2007. Molecular characterization of compost at increasing stages of maturity. 1. Chemical fractionation and infrared spectroscopy. J. Sci. Food Agr. 55 (6), 2293–2302.

Steidle Neto, A.J., Lopes, D.C., Pinto, F.A.C., Zolnier, S., 2017a. Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. Biosyst. Eng. 155, 124–133.

Steidle Neto, A.J., Toledo, J.V., Zolnier, S., Lopes, D.d.C., Pires, C.V., Silva, T.G.F.d., 2017b. Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance. Biosyst. Eng. 156, 17–26.

Tamburini, E., Vincenzi, F., Costa, S., Mantovi, P., Pedrini, P., Castaldelli, G., 2017. Effects of moisture and particle size on quantitative determination of total organic carbon (TOC) in soils using near-infrared spectroscopy. Sensors 17 (10).

Tan, B., Huang, M., Zhu, Q.B., Guo, Y.M., Qin, J.W., 2017. Detection and correction of laser induced breakdown spectroscopy spectral background based on spline interpolation method. Spectrochim. Acta B 138, 64–71.

Villas-Boas, P.R., Romano, R.A., Franco, M.A.D., Ferreira, E.C., Ferreira, E.J., Crestana, S., Milori, D., 2016. Laser-induced breakdown spectroscopy to determine soil texture: a fast analytical technique. Geoderma 263, 195–202.

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37 (1), 29–38.

Wilding, L.P., 1985. Spatial variability: Its documentation, accommodation and implication to soil survey, Proceedings of a Workshop of ISSS and the SSSA, Les Vegas USA, pp. 166–187.

Xing, Z., Du, C.W., Tian, K., Ma, F., Shen, Y.Z., Zhou, J.M., 2016a. Application of FTIR-PAS and Raman spectroscopies for the determination of organic matter in farmland soils. Talanta 158, 262–269.

Xing, Z., Du, C.W., Zeng, Y., Ma, F., Zhou, J.M., 2016b. Characterizing typical farmland soils in China using Raman spectroscopy. Geoderma 268, 147–155.

Xu, S.X., Zhao, Y.C., Wang, M.Y., Shi, X.Z., 2017. Determination of rice root density from Vis-NIR spectroscopy by support vector machine regression and spectral variable selection techniques. Catena 157, 12–23.

Xu, S.X., Zhao, Y.C., Wang, M.i., Shi, X.Z., 2018. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. Geoderma 310, 29–43.

Yi, C., Lv, Y., Xiao, H., Ke, K., Yu, X., 2017. A novel baseline correction method using convex optimization framework in laser-induced breakdown spectroscopy quantitative analysis. Spectrochim. Acta B 138, 72–80.

Yongcheng, J., Wen, S., Baohua, Z., Dong, L., 2017. Quantitative analysis of magnesium in soil by laser-induced breakdown spectroscopy coupled with nonlinear multivariate calibration. J. Appl. Spectrosc. 84 (4), 731–737.

Zaytsev, S.M., Krylov, I.N., Popov, A.M., Zorov, N.B., Labutin, T.A., 2018. Accuracy enhancement of a multivariate calibration for lead determination in soils by laser induced breakdown spectroscopy. Spectrochim. Acta B 140, 65–72.

Zhang, K.R., Dang, H.S., Zhang, Q.F., Cheng, X.L., 2015. Soil carbon dynamics following land-use change varied with temperature and precipitation gradients: evidence from stable isotopes. Glob. Chang. Biol. 21 (7), 2762–2772.