

Article

Hyperspectral Estimation of Soil Organic Matter Content using Different Spectral Preprocessing Techniques and PLSR Method

Lanzhi Shen ^{1,2}, Maofang Gao ^{1,*}, Jingwen Yan ², Zhao-Liang Li ¹, Pei Leng ¹, Qiang Yang ³ and Si-Bo Duan ¹

¹ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; 17lzshen@stu.edu.cn (L.S.), lizhaoliang@caas.cn (Z.-L.L.), lengpei@caas.cn (P.L.), duansibo@caas.cn (S.-B.D.)

² Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou University, Shantou 515063, China; jwyan@stu.edu.cn

³ College of Engineering, Nanjing Forestry University, Nanjing 210037, China; qiangyang@njfu.edu.cn

* Correspondence: gaomaofang@caas.cn

Received: 1 March 2020; Accepted: 6 April 2020; Published: 8 April 2020

Abstract: Soil organic matter (SOM) is the main source of soil nutrients, which are essential for the growth and development of agricultural crops. Hyperspectral remote sensing is one of the most efficient ways of estimating the SOM content. Visible, near infrared, and mid-infrared reflectance spectroscopy, combined with the partial least squares regression (PLSR) method is considered to be an effective way of determining soil properties. In this study, we used 54 different spectral pretreatments to preprocess soil spectral data. These spectral pretreatments were composed of three denoising methods, six data transformations, and three dimensionality reduction methods. The three denoising methods included no denoising (ND), Savitzky–Golay denoising (SGD), and wavelet packet denoising (WPD). The six data transformations included original spectral data, R; reciprocal, 1/R; logarithmic, log(R); reciprocal logarithmic, log(1/R); first derivative, R'; and first derivative of reciprocal, (1/R)'. The three dimensionality reduction methods included no dimensionality reduction (NDR), sensitive waveband dimensionality reduction (SWDR), and principal component analysis (PCA) dimensionality reduction (PCADR). The processed spectra were then employed to construct PLSR models for predicting the SOM content. The main results were as follows—(1) the wavelet packet denoising (WPD)-R' and WPD-(1/R)' data showed stronger correlations with the SOM content. Furthermore, these methods could effectively limit the correlation between the adjacent bands and, thus, prevent “overfitting”. (2) Of the 54 pretreatments investigated, WPD-(1/R)'-PCADR yielded the model with the highest accuracy and stability. (3) For the same denoising method and spectral transformation data, the accuracy of the SOM content estimation model based on SWDR was higher than that of the model based on NDR. Furthermore, the accuracy in the case of PCADR was higher than that for SWDR. (4) Dimensionality reduction was effective in preventing data overfitting. (5) The quality of the spectral data could be improved and the accuracy of the SOM content estimation model could be enhanced effectively, by using some appropriate preprocessing methods (one combining WPD and PCADR in this study).

Keywords: soil organic matter; partial least squares regression; wavelet packet denoising; principal component analysis

1. Introduction

Soil organic matter (SOM) is an important indicator of soil fertility [1]. It is rich in a *variety* of organic acids and humic acids, which have a certain ability to dissolve soil mineral and can promote

the absorption of nutrients. SOM not only provides the nutrients necessary for crops and improves soil physical structure but also helps with water and fertilizer retention [2]. Therefore, it is of great significance to be able to estimate the SOM content rapidly and accurately, in order to increase grain production and aid the sustainable development of agriculture. However, conventional SOM content estimation methods are costly, time consuming, and laborious and, thus, do not meet the needs of the current production management. Fortunately, the development of hyperspectral remote sensing technology has yielded several new methods for soil analysis. One can readily obtain large amounts of remote sensing data through satellites, radar and field and laboratory spectrometers. Soil spectral information has been used to infer soil properties by many researchers [3–8]. In addition, modeling methods are increasingly being used in SOM hyperspectral analysis, and the accuracy of most of these models is high. Some researchers used ground spectral information to analyze soil properties. Alexakis et al. estimated the soil parameters related to soil erosion using integrated satellite remote sensing data, artificial neural networks, and field spectroscopy and GIS data [3]. Kawamura et al. suggested that the soil oxalate-extractable P content can be predicted using visible-near infrared (NIR) spectroscopy [4]. They also stated that the genetic algorithm–partial least squares (GA–PLS) regression method is suitable to find the optimal bands for the PLS regression, contributing to a better predictive ability. Gholizadeh et al. evaluated the potential of the new datamining engine, PARACUDA-II, by comparing its performance in predicting the content of oxidizable carbon in soil, against that of other common datamining algorithms [5]. Kopáčková et al. evaluated the performance of a datamining engine (PARACUDA) in predicting various soil attributes, using reflectance data corresponding to the visible and thermal infrared regions [6]. Rossel et al. assessed various soil properties simultaneously, by using visible, NIR, mid-infrared (MIR), and combined diffuse reflectance spectroscopies [7]. Others have analyzed soil properties using satellite hyperspectral imagery data. Peón et al. predicted the organic carbon content of topsoil using airborne and satellite hyperspectral imagery [8]. Soil spectral information is not only related to the characteristics of the chemical components of the soil, such as its SOM content, iron oxide content, and soil moisture but also to the physical properties of the soil such as its particle size, density, and surface roughness [9]. Due to the limitations of the measuring instruments, methods, and environment, the spectral reflectance data of soil usually contain noise. Therefore, researchers have used some types of data preprocessing methods as they analyzed soil spectral information. Liu et al. applied several spectral data pretreatments during sample selection to construct models for predicting the SOM content using visible and NIR spectroscopy [10]. Zhang et al. constructed a SOM estimation model based on the PLS regression (PLSR) method, using neural networks and spectral data subjected to four transformations (first-order differential, FDR; second-order differential, SDR; continuum removal, CR; continuous wavelet transform, CWT) [11]. Vohland et al. used different methods to select the spectral variables for improving model accuracy and assessing the indicators of arable soil quality [12]. Moreover, hyperspectral data have the characteristics of more bands, large amount of data, and data redundancies. These factors increase both the workload and the complexity of data processing and modeling. Therefore, for the sake of improving the accuracy of the models, it is very important to select the appropriate preprocessing methods (for the hyperspectral data) before modeling, such as denoising, dimensionality reduction, and data form transformation. Most of the studies described above had performed some types of data preprocessing before modeling. There are mainly three kinds of spectral data preprocessing methods—denoising, data transformations and dimensionality reduction. Denoising methods include wavelet packet denoising, SG filtering denoising, etc. Data transformations include $1/R$, R' , etc. [10]. Dimensionality reduction methods include PCA dimensionality reduction, continuum-removal, etc. It has been proved that denoising can reduce the noise in the spectral data [11]. Some data transformations can improve the correlation between some bands and soil parameters or vegetation parameters, especially the first derivative [11] and dimensionality reduction can reduce the data redundancy [13]. However, for assessing the SOM content based on the soil spectra, different researchers deal with soil samples in different ways, such as grounding or sieving the soil. Furthermore, the pretreatments performed on the spectral data are also different,

making it hard to compare the obtained results. The pretreatment methods (e.g., SG) selected in this paper are common and are considered to be “effective” by many researchers. The combinations of these single pretreatment methods were compared and used for spectrum data processing.

The aim of this study was as follows—(i) to estimate the practicability of using visible, NIR, and MIR spectroscopy for assessing the SOM content; (ii) to elucidate the correlation between the SOM content and the different processing data of soil spectra; (iii) to predict the SOM content based on spectral data subjected to different pretreatments, compare the SOM estimation results for the different spectral data pretreatments, and select the most effective preprocessing method for predicting the SOM content based on the PLSR approach; and (iv) to explore whether denoising can reduce correlation between adjacent bands. The overall aim was to improve the accuracy of hyperspectral SOM estimation approaches.

2. Materials and Methods

2.1. Study Area and Soil Sample Collection

The research area was Yitong County, Jilin Province, China (Figure 1). This county is located in the south-central part of Jilin Province, at an east longitude of 124°49′–125°46′ and a north latitude of 43°3′–43°38′. The samples were collected from 21 April to 23 April 2017. The soil sampling points were selected such that they lay on the grid with dimensions of 1 × 1 km (Figure 1c), and the sampling depth was 0–5 cm. For each grid, only one sample was collected. The land type of the samples was corn-cultivated land. The survey area falls in the black soil region, and the soil types included meadow soil, black soil, white soil, and paddy soil, according to the Chinese genetic soil classification system. A collector of diameter 10 cm and length 5 cm was used to vertically remove undisturbed soil samples at the collection points. The extracted samples were deposited in large aluminum boxes of diameter 10 cm and length 5 cm, such that their original structure was maintained for the indoor spectral measurements. A total of 213 soil samples were collected. We analyzed the SOM content (%) of the soil samples in the laboratory using the Walkley–Black [14], after the spectral measurements. The conversion factor to calculate the SOM from the SOC content was 1.724.

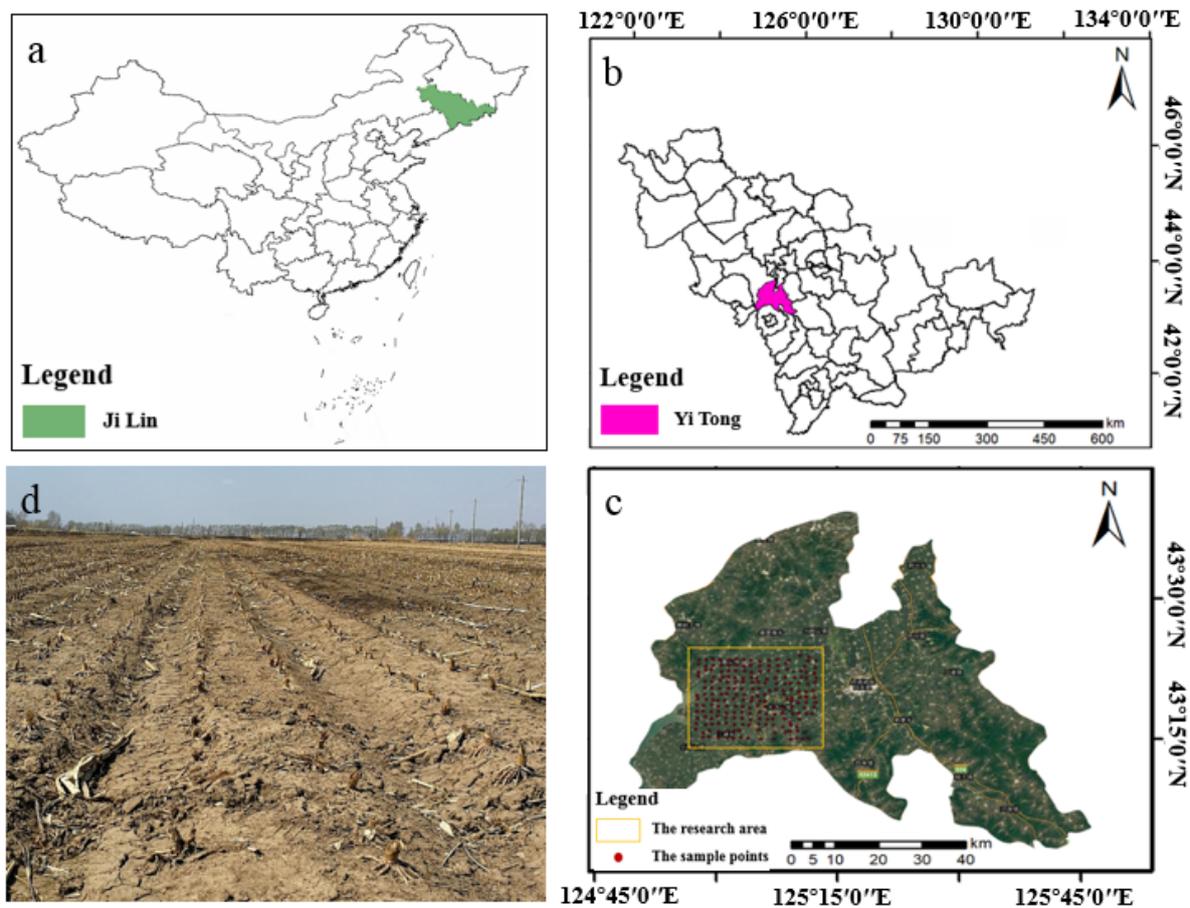


Figure 1. Maps of the study area (a: a map of China; b: a map of Ji Lin Province; c: The research area and sampling points distribution; d: landscape photograph by M.F. Gao).

2.2. Spectral Measurements

Analytical Spectral Devices (ASD) FieldSpec 4 High-Res spectroradiometer was used to perform the indoor spectral measurements on the undisturbed soil samples, which were stored in aluminum boxes and were not pretreated (Figure 2). The surface of the soil samples was also not processed. The spectral measurements were performed at the same day of soil sample collection. The ASD spectroradiometer, whose spectral resolutions was 3 nm @700 nm and 8 nm @1400/2100 nm, had a wavelength range of 350–2500 nm. During the measurements, a 50 W halogen lamp was placed beside the soil sample, such that the incident angle of the light source was 60° (zenith angle of 30°), and the distance between the lamp and the soil sample was 30 cm. The aluminum box was surrounded by black flannel. The probe was placed 15 cm vertically from the soil sample. Each soil sample was measured four times—the aluminum box was rotated by 90° after each measurement for a total of three rotations. A total of 10 spectral curves were collected automatically during each measurement and the arithmetic average of the curves was used as the spectral data. Standard whiteboard calibration was performed before each measurement.



Figure 2. Chart of indoor spectral measurements performed on soil samples.

The software ViewSpec Pro™, produced by ASD, was used to modify the breakpoint of the original indoor spectral data (GAP window of 5×5) and obtain the average of the spectra. The spectral resolution was set to 1 nm, for a total of 2151 bands. The noise caused by the instability of the equipment at 350–400 nm was removed. In order to reduce data redundancy, the indoor spectral data were resampled in the same manner as HyMap airborne hyperspectral images (acquired between 30 April and 1 May 2017) provided by a second-level project unit (spectral resolution for 400–905 nm was 15 nm while that for 880–2500 nm was 18 nm). Finally, each spectrum had 135 bands. A total of 213 indoor spectral curves were obtained.

2.3. Description of Sample Set

It has been reported that the spectral reflectance of soil decreases with an increase in its SOM content [15,16]. In this study, in view of the sample quality, 15 samples with abnormal data in which the soil surface did not strictly maintain the original shape, were not included. Additionally, their spectral curves were significantly different from the others. The remaining 198 samples were grouped into two categories in a ratio of 4:1, which were used to develop and validate the model. The 198 samples were sorted in the ascending order of their SOM content. Starting from the fifth sample, one sample was selected every four samples and allocated to the validation dataset, which contained a total of 40 samples. The remaining 158 samples were used as the training dataset for the model. The statistical information regarding the SOM contents of the various datasets is shown in Table 1.

Table 1. Descriptive Statistics of Various Sample sets.

Sample set	No. of samples	SOM (%)			
		Max	Min	Ave	Std
Total samples	198	4.254	1.150	2.203	0.495
Training dataset	158	4.254	1.458	2.212	0.499
Validation dataset	40	3.589	1.150	2.170	0.486

Soil organic matter (SOM); Maximum (Max); Minimum (Min); Average (Ave); Standard Deviation (Std).

2.4. Preprocessing Methods

In this study, the original soil spectral data were preprocessed using different methods (i.e., different denoising methods, different data transformations, and different dimensionality reduction

approaches), in order to obtain an accurate model for the SOM content estimation based on the PLSR method. Three different denoising methods were used, namely, no denoising (ND), Savitzky–Golay denoising (SGD) [17], and wavelet packet denoising (WPD). Further, six different data transformations were employed. These included using the original spectral data, R ; its logarithm, $\log(R)$; its first derivative, R' ; its reciprocal, $1/R$; the logarithm of its reciprocal, $\log(1/R)$, and the first derivative of its reciprocal, $(1/R)'$. Finally, three different dimensionality reduction approaches were used. These were, no dimensionality reduction (NDR), sensitive waveband dimensionality reduction (SWDR), and principal component analysis (PCA) dimensionality reduction (PCADR) (see Table 2). All the algorithms involved were implemented in the Python programming language (version 3.7).

Table 2. Pretreatment methods used in this study.

Denoising methods	ND, SGD, WPD
Data transformations	R , $1/R$, $\log(R)$, $\log(1/R)$, R' , $(1/R)'$
Dimensionality reduction methods	NDR, SWDR, PCADR

Descriptions of some commonly used preprocessing methods are shown below.

2.4.1. Savitzky–Golay Denoising

The SG smoothing filter is a popular filter for pretreating soil spectra [18]. It is a low-pass filter used to smooth the spectra by eliminating all high-frequency noise, while allowing the low-frequency signals to pass [19]. Further, SG filtering is based on the least squares fitting of a curve local polynomial and uses the weighted-average algorithm for the moving window. However, its weighting coefficient is not a simple constant window and is obtained by fitting the least squares of a given higher-order polynomial in a sliding window [20]. Its underlying idea is to make the reconstructed curve approximate the upper envelope of the original curve, gradually, through iteration [21]. Smooth filtering-based denoising using the SG method could improve the smoothness of the spectrum and reduce noise interference. The expression for SG filtering is as follows:

$$Y_j^* = \frac{\sum_{i=-m}^m C_i Y_j}{N} \quad (1)$$

where Y_j^* is the reconstructed spectral data, C_i is the filtering coefficient, Y_j is the original spectral data, N is the number of datapoints in the sliding window ($N = 2m+1$), and $2m+1$ is the window width. In practical applications, SG filtering requires two parameters—the filter window width and the order of the polynomial for the smooth fitting process. The filter window width can affect the smoothing results, in that the higher the window width, the smoother the resulting spectrum. The order of the fitting polynomial also affects the filtering results [22]. The higher the order, the smoother the fit. In this study, the size of the filter window was set to 21, and the order of the fitting polynomial was taken to be 2.

2.4.2. Wavelet Packet Denoising

Daubechies and others have shown that wavelet packets take into account both high-frequency and low-frequency components of the signal and can effectively extract useful information in each frequency band. As a result, the denoising effect is strong [23,24]. When using wavelet packets to denoise a signal, the choice of the wavelet basis function and the number of layers of the signal are of particular importance. WPD decomposes the original signal into high-frequency and low-frequency signals. The high-frequency signal includes the noise information while the low-frequency signal is an approximation of the original signal. In this study, the db2 wavelet basis function was used to decompose the two-layer wavelet packet, while the soft threshold function was used to denoise the high-frequency signal node, d , in the leaf layer, after signal decomposition. The signal was then reconstructed after threshold denoising. The threshold determination formula was as follows:

$$\text{tar} = \sigma \sqrt{2 \log_e N} \quad (2)$$

where σ is the median of the absolute value of all the coefficients in the high-frequency signal, d , divided by 0.6745 and N is the number of datapoints in d . The threshold proposed by Donoho was considered to be the maximum noise value. Further, $\text{tar}/2$ was taken to be the denoising threshold for the noise signal in this study.

2.4.3. Mathematical Transformations of Spectral Reflectance Data

As mentioned above, six different mathematical transformations were performed on the spectral data— R , $1/R$, $\log(R)$, $\log(1/R)$, R' , and $(1/R)'$. Since the spectrometer collected discrete data, R' was calculated using the following equation:

$$R'(\lambda_i) = \frac{R(\lambda_{i+1}) - R(\lambda_i)}{\lambda_{i+1} - \lambda_i} \quad (3)$$

where $R'(\lambda_i)$ is the first derivative of the reflectance at band λ_i , $R(\lambda_{i+1})$ is the reflectance at band λ_{i+1} , and $R(\lambda_i)$ is the reflectance at band λ_i .

2.4.4. PCA Dimensionality Reduction

PCA is a commonly used dimensionality reduction method that has been employed widely in hyperspectral remote sensing. Extracting meaningful features (or components) from multidimensional data is typically done using the canonical PCA [25,26]. The purpose of the PCA transformation is to determine the set of the optimal unit orthogonal vector bases (i.e., principal components) through a linear transformation, and to minimize the error of the mean square deviation of the original sample through a linear combination [13]. In PCA, data were transformed from the original coordinate system to a new one, and the choice of the new coordinate system was determined by the data itself. The first new coordinate axis was along the direction with the largest variance in the original data, while the second new coordinate axis was along the direction with the largest variance orthogonal to the first coordinate axis, and so on. Most variances were accounted for in the first few new coordinate axes; finally, one had to choose several coordinate axes. In other words, data were reduced to several dimensions.

2.4.5. Sensitive Band Dimensionality Reduction

The original spectra were first subjected to denoising and then to a data transformation before being processed for dimensionality reduction. The correlation between the preprocessed spectral data and the SOM content was determined, and the wavebands for which the coefficient of determination, r^2 (square of the correlation coefficient, r), was greater than or equal to 0.25 were selected as the sensitive wavebands for each spectral curve. The expression for calculating the correlation coefficient, r , was as follows:

$$r_i = \frac{\text{cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}} = \frac{\sum_{n=1}^N (x_{ni} - \bar{x}_i)(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (x_{ni} - \bar{x}_i)^2} \sqrt{\sum_{n=1}^N (y_n - \bar{y})^2}} \quad (4)$$

where r_i is the correlation coefficient between the spectral data of the i th band and the SOM content, x_{ni} is the corresponding spectral data value of the i th band of the n th sample, \bar{x}_i is the average value of the corresponding spectral data of the i th band, y_n is the SOM content of the n th sample, and \bar{y} is the average value of the SOM contents of all the samples.

2.5. Partial Least Squares Regression Method

PLSR is a popular technique used to correlate soil spectra with the SOM content [27,28]. The method combines the characteristics of multiple linear regression analysis, canonical correlation analysis, and principal component analysis, so that it not only provides a suitable regression model, but also expresses information more comprehensively. It is underpinned by the assumption that the dependent variable can be estimated via a linear combination of explanatory variables [29]. It provides a many-to-many linear regression modeling method, especially in this case, where the number of two sets of variables which had multiple correlations was large and the number of the sample size was small, the model established by PLSR had advantages that traditional classical regression analysis did not have. When solving many-to-many linear regression problem, multiple linear regression leads to overfitting due to the correlation between independent variables. While the PLSR method would find some new variables that are linearly independent, to replace the original independent variables that can maximize the difference between independent variables.

2.6. Metrics for Evaluating Model Performance

The parameters used in this study for evaluating the model performance included the coefficient of determination for the training set, R^2_T ; the coefficient of determination for the validation set, R^2_V ; the root-mean-square error of the training set ($RMSE_T$); the root-mean-square error of the validation set ($RMSE_V$), and the ratio of performance-to-deviation (RPD). The larger the R^2 value, the greater the accuracy of the model. On the other hand, the $RMSE_T$ and $RMSE_V$ values would have to be as small as possible. Furthermore, the more similar they were, the higher would be the estimation accuracy and stability of the model. Finally, the range of the RPD values could generally be divided into three categories—when RPD was equal to or more than 2.0, the model was suitable for estimating the SOM content from hyperspectral data; when RPD was less than 2.0, the reliability of the model could be improved by fine-tuning the model; and finally, when RPD was equal to or less than 1.4, the model was unreliable [30].

$$RMSE_T = \sqrt{\frac{\sum_{i=1}^n (y_{it} - y_{ip})^2}{n}} \quad (5)$$

where y_{it} is the true value of the SOM content of the i th sample, y_{ip} is the estimated value of the SOM content of the i th sample, and n is the number of samples in the training set. Equation (5) is the expression for calculating $RMSE_T$. Furthermore, the expression for calculating $RMSE_V$ was similar to that for $RMSE_T$.

Finally, Eq. (6) gave the expression for calculating RPD:

$$RPD = SD / RMSE_V \quad (6)$$

where SD is the standard deviation of the SOM contents for the samples in the validation set.

3. Results and Discussion

3.1. Correlation between SOM Content and Reflectance Data Subjected to Different Pretreatments

Correlation analysis is a classical and reliable method for analyzing the correlativity between independent and dependent variables [31]. We performed a correlation analysis between the SOM content and the reflectance spectra, subjected to different pretreatments. A total of eighteen different pretreatments, including three denoising methods (ND, SGD, and WPD) and six spectral data transformations (R, 1/R, log(R), log(1/R), R', and (1/R')) were used. The 18 pretreatments are listed in Table 3.

Table 3. Different Combinations of Pretreatments used in this study.

Pretreatment Methods Used	Denoising Methods		
	ND	SGD	WPD

	R	ND-R	SGD-R	WPD-R
	1/R	ND-1/R	SGD-1/R	WPD-1/R
Data	log(R)	ND-log(R)	SGD-log(R)	WPD-log(R)
transformations	log(1/R)	ND-log(1/R)	SGD-log(1/R)	WPD-log(1/R)
performed on	R'	ND-R'	SGD-R'	WPD-R'
spectral data	(1/R)'	ND-(1/R)'	SGD-(1/R)'	WPD-(1/R)'

The degree of correlation was expressed by the Pearson coefficient, r . Figure 3 shows the changes in the correlation coefficient with the wavelength. It can be seen from Figure 3 (a), Figure 3 (b), and Figure 3 (c), that irrespective of the denoising method used (ND, SGD, or WPD), there was a strong negative correlation between the SOM content and R in the range of 400–2500 nm. The curves for the correlation coefficients were almost smooth and horizontal. Furthermore, the fact that the variations in the correlation coefficient were small meant that the correlation coefficients corresponding to the adjacent bands were similar. Although the trend in the curves for the correlation coefficient for the SOM content and $\log(R)$ was similar to that in the case of the SOM content and R, the former curves were higher over the entire investigated bandwidth range. Furthermore, the curve for the correlation coefficient for the SOM content and $1/R$ was almost a mirror image of the curve for the SOM content and R, exhibiting a similar trend about the x-axis in the positive quadrant. In addition, the correlation (represented by the absolute value of the correlation coefficient) between the SOM content and $1/R$ was stronger in the visible band but weaker in the NIR and MIR band. Next, the correlation between the SOM content and $\log(1/R)$ was basically the same as that between the SOM content and $1/R$; however, the former curve was a little higher over the entire wavelength range. In contrast, the curve for the correlation between the SOM content and the first derivative of R showed significant variations over the entire wavelength region. More precisely, it exhibited approximately the same trend as the other correlation curves for wavelengths smaller than 1300 nm but then rose and fell sharply. In addition, the correlation between the SOM content and R' in the range of 400–640 nm was stronger than those between the SOM content and the other data. The same was also true for the correlation between the SOM content and $(1/R)'$ in the range of 840–1300 nm. Thus, it was likely that they had the possibility to become a sensitive correlation. At the same time, it can be seen from Figure 3 that regardless of the denoising method used (ND, SGD, or WPD), the bands with high correlation (where the correlation coefficient was greater than 0.5) were concentrated in the range of 610–1300 nm after R, $1/R$, $\log(R)$, and $\log(1/R)$. However, the bands with high correlation were concentrated in the range of 466–640 nm, 665 nm, and 704–767 nm, after ND- R' ; 716–1300 nm after ND- $(1/R)'$; 466–494 nm, 523–640 nm, 678 nm, 704–729 nm, and 767–780 nm after WPD- R' ; and 704 nm, 730 nm, 755–1300 nm, and 1341 nm after WPD- $(1/R)'$.

In conclusion, there was a strong correlation between the SOM content and R. $\log(R)$ would improve the correlation with SOM, $1/R$ would improve the correlation between the visible band range and SOM, and the R' and $(1/R)'$ would increase the correlation of some bands, and these bands were somewhat scattered.

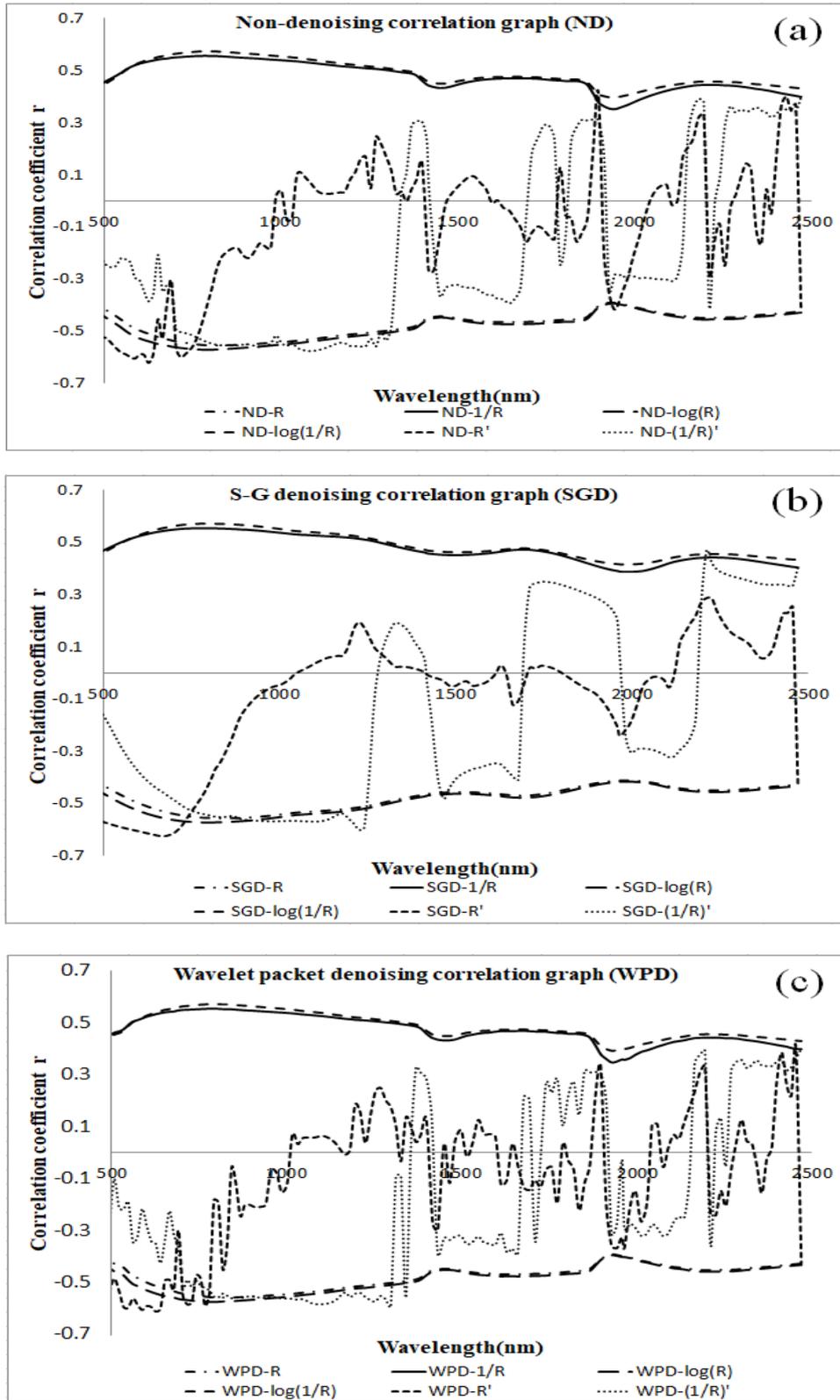


Figure 3. Curves for correlations between SOM content and spectral data subjected to six different transformations under (a) ND, (b) SGD, and (c) WPD at wavelengths of 400–2500 nm.

There were also some differences between Figure 3 (a), Figure 3 (b), and Figure 3 (c). It can be seen from Figure 3 (b) and Figure 3 (c) that the correlation curves between the SOM content and R, 1/R, log(R), and log(1/R) were basically the same after SGD or WPD. However, after SGD, the correlation curves between the SOM content, and R' and (1/R)' were smoother than those after ND.

Furthermore, after WPD, the correlation curves between the SOM content and R' and $(1/R)'$ showed more variations than those after ND. Moreover, the greater the variations in the curves, the more dispersed the band distribution at higher correlation coefficient values; this was beneficial for the subsequent dimensionality reduction operation, as it helped to effectively reduce the correlation between the bands and prevent the problem of “overfitting” caused by the strong correlation between them.

3.2. Determination of Optimal Parameter Value for PCA

Data redundancy is a disadvantage in hyperspectral data, and the key to the processing of hyperspectral data is the extraction of useful information present in large datasets [32]. The PCA method was selected to reduce the dimensions of the remote sensing data used in this study. Figure 4 shows the curve of the variations in the model accuracy, with the data dimension after dimensionality reduction. When the dimension was 25, both the RPD value and the R^2_v value of the model were maximized. At this time, the model showed the highest accuracy. Thus, 25 was chosen as the optimal dimension for PCA-based dimensionality reduction.

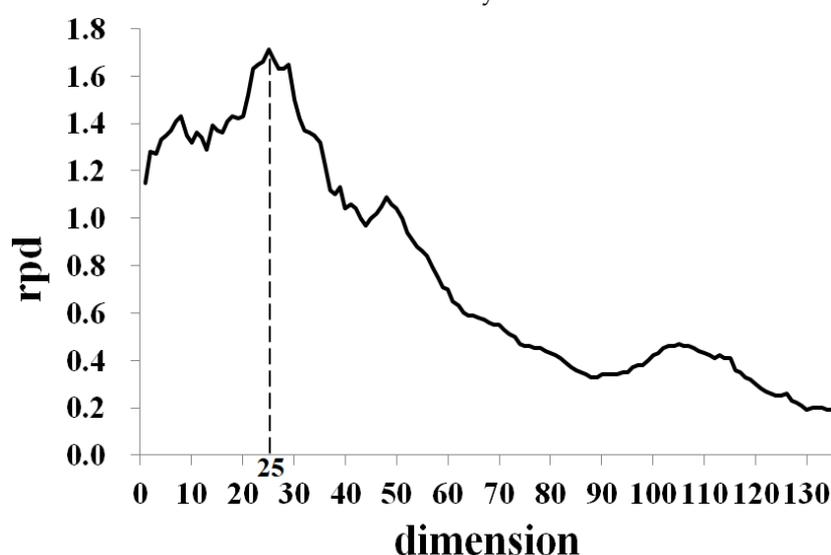


Figure 4. Model accuracy for different dimensions during dimensionality reduction.

3.3. Accuracy Analysis of the Hyperspectral Estimation Model of the SOM Content based on PLSR

The SOM content estimation model results based on the PLSR model in 54 different spectral pretreatment methods listed in Table 4.

Table 4. SOM Content Estimation Model Results Based on the PLSR Model in 54 Different Spectral Pretreatment Methods.

Pretreatment Method	RMSE _T (%)	RMSE _V (%)	R ² _T	R ² _V	RPD	
R	NDR	0.150	1.200	0.921	0.007	0.400
	SWDR	0.211	0.426	0.831	0.465	1.125
	PCAD	0.320	0.350	0.595	0.512	1.370
ND	NDR	0.113	0.934	0.950	0.229	0.660
	SWDR	0.269	0.605	0.718	0.508	0.793
	PCAD	0.338	0.326	0.544	0.589	1.472
log(R)	NDR	0.128	0.903	0.946	0.054	0.531
	SWDR	0.266	0.420	0.723	0.556	1.141
	PCAD	0.325	0.301	0.582	0.640	1.591

		R					
		NDR	0.128	0.903	0.946	0.054	0.531
	log(1/R)	SWDR	0.266	0.420	0.723	0.556	1.141
		PCAD					
		R	0.325	0.301	0.582	0.640	1.591
		NDR	0.150	0.200	0.921	0.007	0.400
	R'	SWDR	0.335	0.417	0.553	0.323	1.150
		PCAD					
		R	0.317	0.338	0.602	0.533	1.419
		NDR	0.114	0.856	0.960	0.129	0.560
	(1/R)'	SWDR	0.293	0.568	0.660	0.400	0.844
		PCAD					
		R	0.330	0.331	0.568	0.555	1.449
		NDR	0.208	0.817	0.836	0.110	0.586
	R	SWDR	0.347	0.391	0.519	0.389	1.227
		PCAD					
		R	0.285	0.348	0.680	0.538	1.377
		NDR	0.141	1.031	0.932	0.147	0.465
	1/R	SWDR	0.386	0.424	0.402	0.289	1.130
		PCAD					
		R	0.341	0.311	0.537	0.627	1.542
		NDR	0.148	1.014	0.923	0.127	0.473
	log(R)	SWDR	0.304	0.354	0.633	0.526	1.352
		PCAD					
		R	0.344	0.321	0.528	0.585	1.494
		NDR	0.148	1.014	0.923	0.127	0.473
	log(1/R)	SWDR	0.304	0.354	0.633	0.526	1.352
		PCAD					
		R	0.344	0.321	0.528	0.585	1.494
		NDR	0.204	0.710	0.842	0.176	0.675
	R'	SWDR	0.351	0.343	0.507	0.532	1.398
		PCAD					
		R	0.337	0.344	0.548	0.516	1.392
		NDR	0.141	1.032	0.932	0.147	0.465
	(1/R)'	SWDR	0.336	0.436	0.551	0.396	1.101
		PCAD					
		R	0.309	0.334	0.623	0.593	1.434
		NDR	0.211	0.834	0.831	0.111	0.575
	R	SWDR	0.320	0.404	0.593	0.362	1.184
		PCAD					
		R	0.271	0.346	0.713	0.563	1.386
		NDR	0.114	0.856	0.960	0.129	0.560
	1/R	SWDR	0.243	0.614	0.770	0.371	0.781
		PCAD					
		R	0.236	0.287	0.785	0.690	1.668
		NDR	0.121	1.384	0.953	0.038	0.346
	log(R)	SWDR	0.249	0.581	0.759	0.0279	0.825
		PCAD					
		R	0.239	0.306	0.780	0.642	1.569
		NDR	0.121	1.384	0.953	0.038	0.346
	log(1/R)	SWDR	0.249	0.581	0.759	0.0279	0.825

R'	PCAD	0.239	0.306	0.780	0.642	1.569
	R					
	NDR	0.211	2.885	0.831	0.0001	0.166
(1/R)'	SWDR	0.299	0.343	0.646	0.530	1.399
	PCAD	0.299	0.343	0.646	0.530	1.399
	R					
(1/R)'	NDR	0.108	2.421	0.965	0.0004	0.198
	SWDR	0.295	0.451	0.657	0.303	1.064
	PCAD	0.241	0.280	0.775	0.713	1.712
	R					

The analysis of some of these results with significant regularity are described below.

3.3.1. Comparison of the Modeling Results based on Original Data and Data Obtained after Effective Pretreatment

With respect to the SOM estimation model based on PLSR, the best performance was observed after the pretreatment WPD-(1/R)'-PCADR. Table 5 shows the PLSR-based estimation accuracy of the SOM content when using the untreated data and those subjected to the pretreatment WPD-(1/R)'-PCADR. Figure 5 shows the scatterplot of the SOM content estimation results based on PLSR, when unprocessed data were used [33]. The corresponding values of $RMSE_v$, R^2_v , and RPD were 1.200, 0.007, and 0.400 respectively. Figure 6 shows the scatter plot of the SOM content estimation results based on PLSR after the pretreatment WPD-(1/R)'-PCADR. The corresponding values of $RMSE_v$, R^2_v , and RPD were 0.280, 0.713, and 1.712, respectively. Thus, the R^2_v and RPD values for the latter case (i.e., after the pretreatment WPD-(1/R)'-PCADR) were improved by 0.706 and 0.312, respectively. When the PLSR model was used on the untreated data, the estimation results for the training and validation sets were significantly different.

Table 5. SOM Content Estimation Model Results Based on PLSR Model in the Case of Untreated Data and Those Subjected to Optimal Pretreatment.

Pretreatment Method	$RMSE_T$ (%)	$RMSE_v$ (%)	R^2_T	R^2_v	RPD
No pretreatment	0.150	1.200	0.921	0.007	0.400
WPD-(1/R)'-PCADR	0.241	0.280	0.775	0.713	1.712

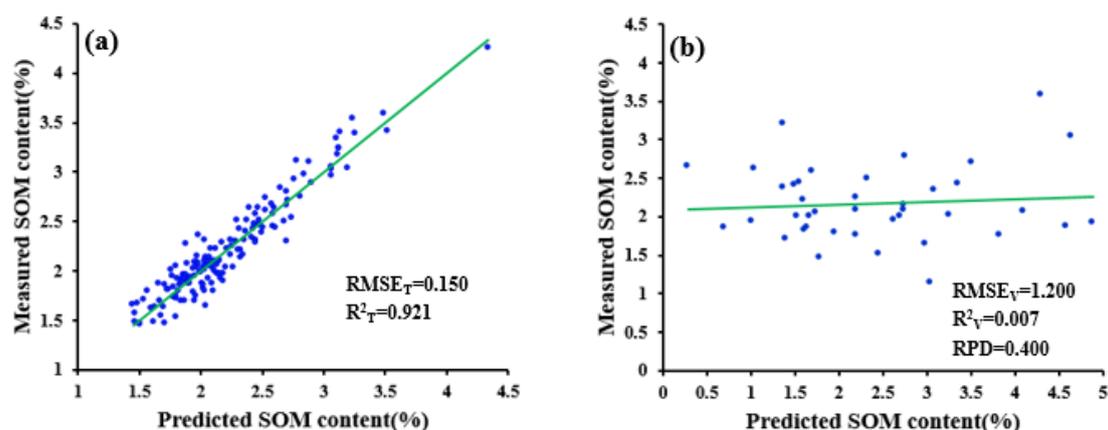


Figure 5. SOM content estimation results based on PLSR model for (a) training and (b) validation sets of untreated data.

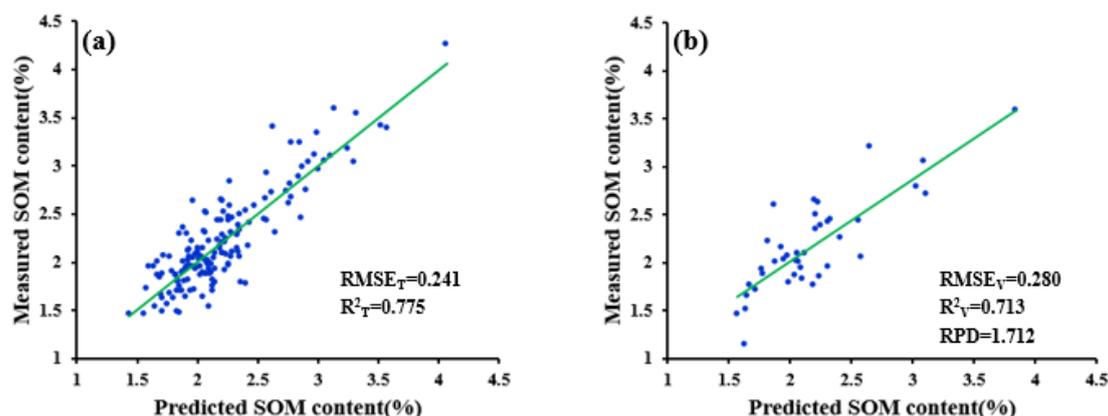


Figure 6. SOM content estimation results based on PLSR for (a) training and (b) validation sets of data subjected to WPD-(1/R)-PCADR pretreatment.

3.3.2. Comparison of Modeling Results for Different Dimensionality Reduction Methods

It was observed that, for the same denoising method and the same spectral data transformation, the dimensionality reduction performance of SWDR was better than that of NDR. Furthermore, the dimensionality reduction performance of PCADR was better than that of SWDR. Table 6 compares the SOM content estimation results for the spectral data subjected to the 1/R transformation, the WPD method, and different dimensionality reduction treatments. Figure 7 shows the scatterplot of the SOM content estimation results for WPD-1/R-NDR. Figure 8 shows the scatterplot for WPD-1/R-SWDR, and Figure 9 shows the scatterplot for WPD-1/R-PCADR. Here, it can be seen clearly that the accuracy of the SOM estimation model based on PLSR was higher when the data were subjected to dimensionality reduction using the SWDR, as compared to the NDR method; the R^2_v and RPD values using the SWDR had increased by 0.380 and 0.626, respectively, than the value using the NDR method. The same was also true for the PCADR method, as compared to the SWDR method; the R^2_v and RPD values using the PCADR increased by 0.181 and 0.482, respectively, than the value using the SWDR.

In addition, it could be seen that when dimensionality reduction was not performed, the model training resulted in significant overfitting, with the difference between the relative analysis errors for the training and validation sets being 0.742. On the other hand, after the dimensionality reduction of the sensitive bands, the problem of overfitting was resolved, and the difference in the errors of the two datasets reduced to 0.113. Furthermore, after dimensionality reduction using PCA, overfitting was eliminated completely, and the error difference was only 0.051. These results confirmed that dimensionality reduction could effectively prevent overfitting and improve the accuracy and stability of the model. The results for the other denoising methods and spectral data transformations were similar and hence were not included.

Table 6. Comparison of SOM Content Estimation Results for Spectral Data Subjected to 1/R Transformation, WPD Denoising, and Different Dimensionality Reduction Methods.

Pretreatment Method	RMSE _T (%)	RMSE _V (%)	R ² _T	R ² _V	RPD
WPD-1/R-NDR	0.114	0.856	0.960	0.129	0.560
WPD-1/R-SWDR	0.291	0.404	0.666	0.509	1.186
WPD-1/R-PCADR	0.236	0.287	0.785	0.690	1.668

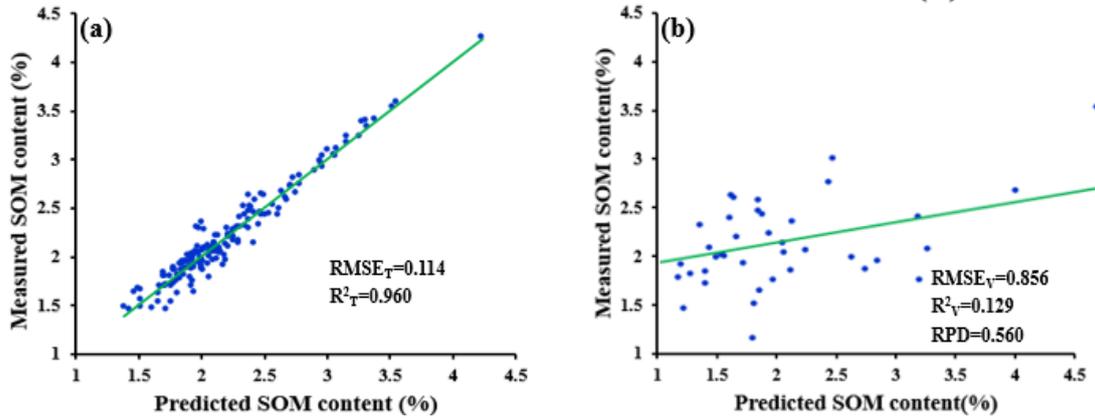


Figure 7. SOM content estimation results based on PLSR for (a) training and (b) validation set for data subjected to WPD-1/R-NDR pretreatment.

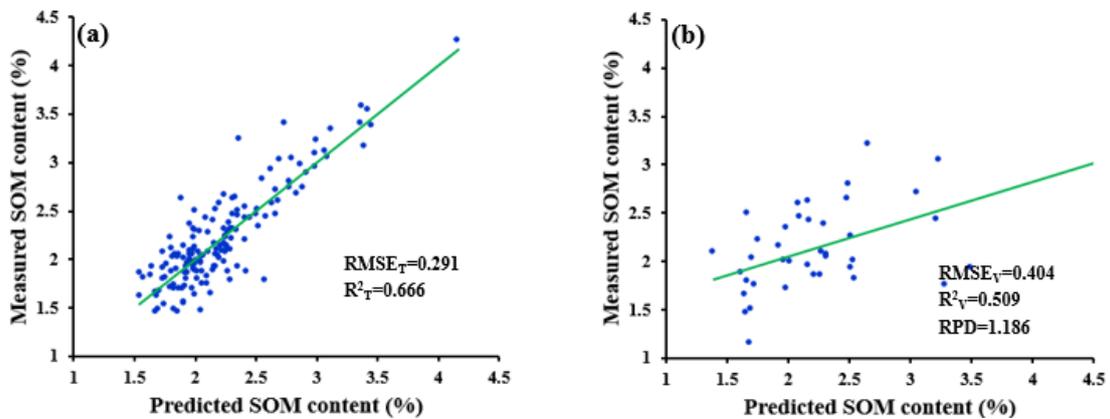


Figure 8. SOM content estimation results based on PLSR for (a) training and (b) validation set for data subjected to WPD-1/R-SWDR pretreatment.

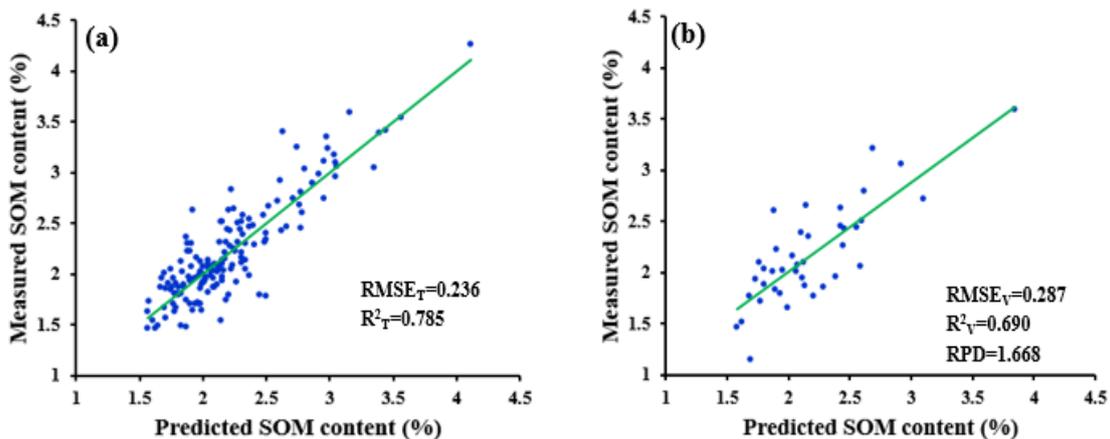


Figure 9. SOM content estimation results based on PLSR for (a) training and (b) validation set for data subjected to WPD-1/R-PCADR pretreatment.

3.3.3. Comparison of Modeling Results for Different Denoising Methods

Next, some of the data subjected to different denoising methods were treated using the same spectral transformations and dimensionality reduction methods. This improved the estimation accuracy of the model (see Table 7). For example, in the case of the SWDR process, when the R spectral data were denoised using the SG method, the RPD value of the model was 0.102 higher than

that without denoising. Furthermore, when the R data were denoised by the WPD method, the RPD was 0.059 higher than that without denoising. Figure 10 shows the scatterplot of the SOM content estimation results for ND-R-SWDR, Figure 11 shows the scatterplot for SGD-R-SWDR, and Figure 12 shows the scatterplot for WPD-R-SWDR.

Table 7. Comparison of the SOM Content Estimation Results for the R Spectral Data Subjected to SWDR and the Different Denoising Methods.

Pretreatment Method	RMSE _T (%)	RMSE _V (%)	R ² _T	R ² _V	RPD
ND-R-SWDR	0.211	0.426	0.831	0.465	1.125
SGD-R-SWDR	0.347	0.391	0.519	0.389	1.227
WPD-R-SWDR	0.320	0.404	0.593	0.362	1.184

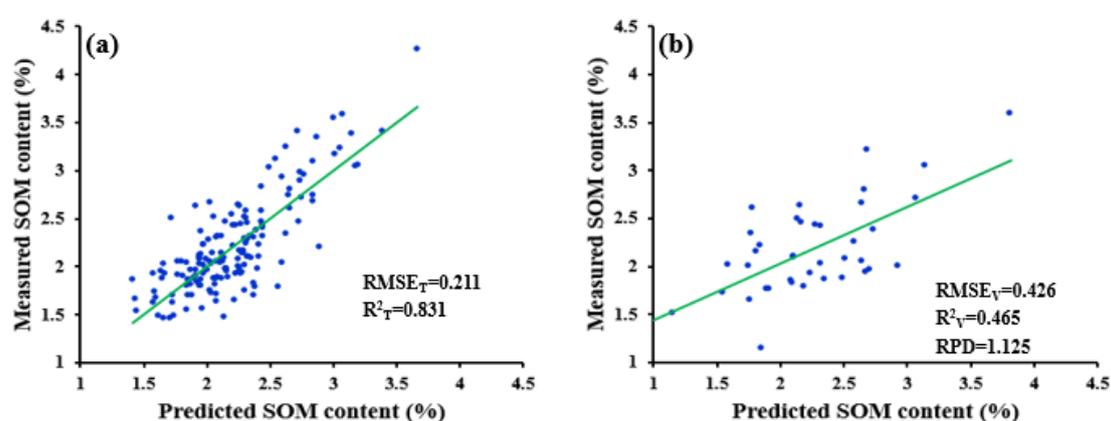


Figure 10. SOM content estimation results based on PLSR for the (a) training and (b) validation set for data subjected to the ND-R-SWDR pretreatment.

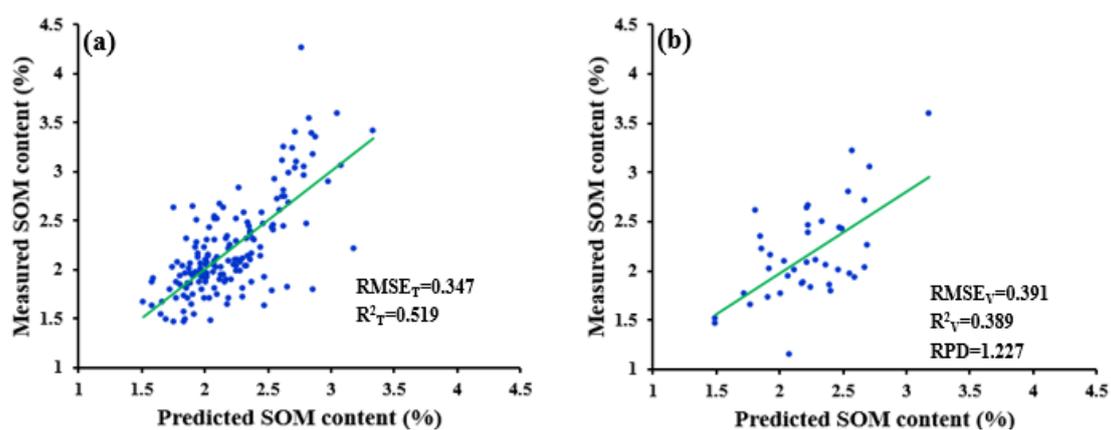


Figure 11. SOM content estimation results based on PLSR for the (a) training and (b) validation set for data subjected to SGD-R-SWDR pretreatment.

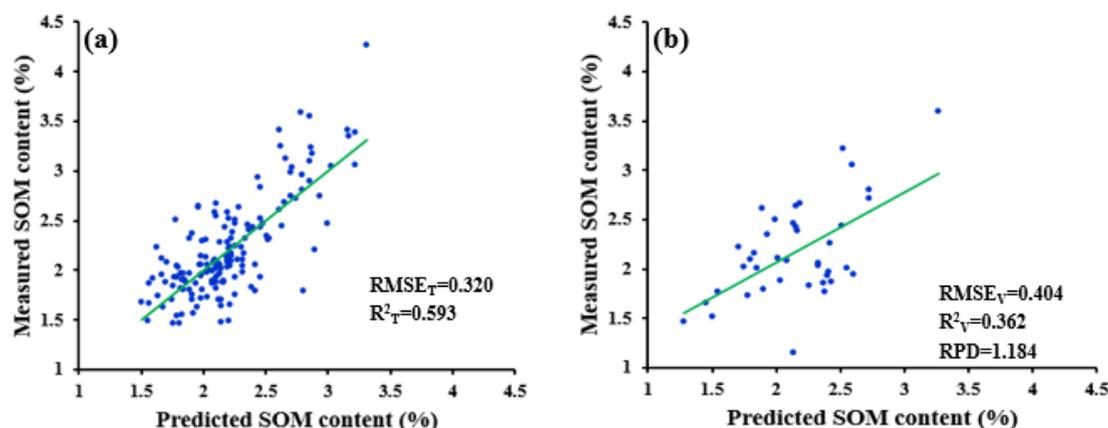


Figure 12. SOM content estimation results based on PLSR for the (a) training and (b) validation set for data subjected to WPD-R-SWDR pretreatment.

3.3.4. Discussion of Different Preprocessing Techniques for Soil Hyperspectral Data

It can be seen from Table 4 that, for the same denoising method and same spectral data transformation, the dimensionality reduction performance of SWDR and PCADA were better than that of NDR. However, the estimation accuracy was relatively low when the data were subjected to any of the spectral transformations and the SWDR method, with the RPD value being mostly less than 1.4. In contrast, the estimation accuracy improved greatly when the PCADR method was used. The RPD value was basically greater than 1.4 in this case, with the R^2_v value being as high as 0.713. The low estimation accuracy of the SWDR was probably because the sensitive bands selected were not the optimal bands for soil spectrum processing. In this study, the wavebands with high correlation coefficients ($r > 0.5$) were selected as the sensitive bands, and most of these lay in the visible region (Figure 3). In the visible-wavelength range of 400–700 nm, the variations in the spectral reflectance of soil are closely related to the presence of SOM and minerals such as iron oxides [34]. Rossel et al. showed that the reflectance spectrum at 410 nm is related to the SOM content [11]. In addition, several studies have reported that SOM can reduce the spectral reflectance in the visible region and that the SOM content is strongly correlated to the reflectance in the 550–680 nm range [35–37]. Mouazen et al. found that the reflectance of soil in this wavelength range is related to the soil color, which is determined by the electronic transitions [38]. In the NIR and MIR region, the peaks at wavelengths of 1853, 1000, and 2412 nm are mainly caused by the absorbance of the O-H bonds of the free moisture in the soil, as well as the absorbance of the other O-H groups existing in the soil, such as the clay minerals [39–41]. Thus, the sensitive bands selected based on only the Pearson coefficient might not be truly representative of the actual reflectance data. Moreover, we found that a high number of bands had r values of more than 0.5 did not result in a better model. Thus, the appropriate sensitive bands related to SOM should be selected with care, in order to be able to estimate the SOM content with accuracy, as this would minimize the errors caused by data redundancy.

Additionally, for the same spectral transformation and dimensionality reduction methods, the advantages of the SGD method were not apparent. The accuracy of the SOM estimation model used by SGD might be higher than that used by ND (e.g., SGD-R-SWDR), but some of others might be lower than that used by ND (e.g., SGD-log(R)-PCADR). The SGD method did not improve the correlation between the SG spectra and the SOM content (Figure 3). Barnes et al. have suggested that determining the appropriate smoothing window size was essential for processing spectral data [42]. The low correlation coefficient in the case of the data subjected to SG preprocessing might be attributable to the over-smoothing of the data, which probably resulted in information loss [43].

In addition, when WPD and PCADR are used simultaneously, the advantages of derivative method are obvious. Oldham et al. and Li et al. reported that derivatives are not only a powerful tool for analyzing spectral data but also help overcome several collinearity problems [44,45]. The derivative method strongly affects the local peaks in the spectrum. Thus, it could be used to enhance

the sensitivity of the analysis and the spectral resolution. To a certain degree, it also helps in removing noise. As is known, the first derivative (FD) and second derivative (SD) indicate the slope and change in the slope, respectively, of the reflectance spectrum. The peak absorption of the SD spectrum is greater than that of the FD spectrum, with the reflectance value in the former being lower. Although the SD can separate a greater number of absorption peaks, it can also introduce noise and might cause errors. The SD and FD lead to significant changes in the spectrum, resulting in sharp peaks. Fractional derivatives can limit the extent of the changes in the spectrum and ensure that the shape characteristics of the original spectrum are preserved. Thus, they are more advantageous than full derivatives (FD and SD) [32]. To extend the order to non-integers, fractional derivatives might be used to provide more useful information from remote sensing, which could add more detail to the spectra than whole derivatives. In this study, only the FD was used to estimate the SOM content. In future, however, we plan to use other higher-order full and fractional derivatives to preprocess the spectral data.

4. Conclusions

We collected 213 soil samples from Yitong County, Jilin Province, China, measured their spectral data, subjected the data to different preprocessing treatments, and subsequently used them to determine the SOM content using the PLSR method. This was done with the aim of establishing an accurate and efficient model for predicting the SOM content, based on hyperspectral reflectance data. The conclusions of the study could be summarized as follows. (1) After the WPD-R' and WPD-(1/R)' pretreatments, the bands with stronger correlations with the SOM content became more dispersed; this was beneficial to the subsequent dimensionality reduction operation, as it effectively reduced the correlation between the adjacent bands and prevented "overfitting." (2) In the Yitong area of Jilin Province, the WPD-(1/R)'-PCADR pretreatment of the 54 different pretreatments investigated in this study yielded the model with the highest accuracy for estimating the SOM content. (3) For the same denoising method and spectral data transformation method, the accuracy of the SOM estimation model based on PLSR was higher than when the data were subjected to dimensionality reduction using the SWDR, as compared to the NDR method, the R^2_v and RPD values using the SWDR were increased by 0.380 and 0.626, respectively, than the value using the NDR method. The same was also true for the PCADR method, as compared to the SWDR method, the R^2_v and RPD values using the PCADR increased by 0.181 and 0.482, respectively, than the value using the SWDR. (4) Dimensionality reduction was effective in preventing data overfitting. (5) The quality of the spectral data could be improved and the accuracy and stability of the SOM content estimation model could be enhanced effectively using appropriate preprocessing methods (a combination of WPD and PCADR in this case). For example, the RPD of the model based on the data preprocessed using WPD-(1/R)'-PCADR was higher than that of the model based on untreated data by 1.312.

Author Contributions: writing—original draft, L.S.; methodology, M.G.; writing—review and editing, M.G.; supervision, J.Y.; project administration, Z.-L.L.; Formal analysis, P.L.; investigation, Q.Y.; resources, S.-B.D.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41921001, 41871282.

Acknowledgments: The authors would like to thank their colleagues and students from the Institute of Agricultural Resources and Regional Planning, Beijing, China, for collecting and processing the soil data in Jilin province.

Conflicts of Interest: The authors declared that there is no conflicts of interest to this paper.

References

1. Munson, S.A.; Carey, A.E. Organic matter sources and transport in an agricultural dominated temperate watershed. *Appl. Geochem.* **2004**, *19*, 1111–1121.
2. Dou, S. *Soil Organic Matter*; Science Press: Beijing, China, 2010.

3. Alexakis, D.; Tapoglou, E.; Vozinaki, A.E.; Tsanis, I.K. Integrated Use of Satellite Remote Sensing, Artificial Neural Networks, Field Spectroscopy, and GIS in Estimating Crucial Soil Parameters in Terms of Soil Erosion. *Remote Sens.* **2019**, *11*, 1106.
4. Kawamura, K.; Tsujimoto, Y.; Nishigaki, T.; Andriamananjara, A.; Rabenarivo, M.; Asai, H.; Razafimbelo, T. Laboratory Visible and Near-Infrared Spectroscopy with Genetic Algorithm-Based Partial Least Squares Regression for Assessing the Soil Phosphorus Content of Upland and Lowland Rice Fields in Madagascar. *Remote Sens.* **2019**, *11*, 506.
5. Gholizadeh, A.; Saberioon, M.; Carmon, N.; Boruvka, L.; Ben-Dor, E. Examining the Performance of PARACUDA-II Data-Mining Engine versus Selected Techniques to Model Soil Carbon from Reflectance Spectra. *Remote Sens.* **2018**, *10*, 8.
6. Kopacková, V.; Eyal, B.D.; Nimrod, C.; Notesco, G. Modelling Diverse Soil Attributes with Visible to Longwave Infrared Spectroscopy Using PLSR Employed by an Automatic Modelling Engine. *Remote Sens.* **2017**, *9*, 134.
7. Rossel, R.A.V.; Walvoort, D.J.J.; Mcbratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75.
8. Peón, J.; Carmen, R.; Fernández, S.; Calleja, J. F.; De Miguel, E.; Carretero, L. Prediction of Topsoil Organic Carbon Using Airborne and Satellite Hyperspectral Imagery. *Remote Sens.* **2017**, *9*, 1211.
9. Liu, X.M. Near infrared diffuse reflectance spectra detection of soil organic matter and available N. *J. Chin. Agric. Mech.* **2013**, *34*, 202–206.
10. Liu, Y.; Liu, Y.L.; Chen, Y.Y.; Zhang, Y.; Shi, T.; Wang, J.; Fei, T. The Influence of Spectral Pretreatment on the Selection of Representative Calibration Samples for Soil Organic Matter Estimation Using Vis-NIR Reflectance Spectroscopy. *Remote Sens.* **2019**, *11*.
11. Zhang, S.; Shen, Q.; Nie, C.; Huang, Y.; Wang, J.; Hu, Q.; Chen, Y. Hyperspectral inversion of heavy metal content in reclaimed soil from a mining wasteland based on different spectral transformation and modeling methods. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *211*, 393–400.
12. Vohland, M.; Ludwig, M.; Thiele-Bruhn, S.; Ludwig, B. Quantification of Soil Properties with Hyperspectral Data: Selecting Spectral Variables with Different Methods to Improve Accuracies and Analyze Prediction Mechanisms. *Remote Sens.* **2017**, *9*, 1103.
13. Dong, C.W. Face Recognition Based on PCA and SVM Algorithm. *Radio Telev. Inf.* **2018**, *10*, 107–110.
14. Sahrawat, K.L. Simple modification of the Walkley-Black method for simultaneous determination of organic carbon and potentially mineralizable nitrogen in tropical rice soils. *Plant Soil.* **1982**, *69*, 73–77.
15. Feng, Y.S.; Wu, P.X.; Liu, Y.J.; Zhou B.J.; Ma, J. The Study of The Soil Spectral Characteristics. *J. Jilin Agric. Univ.* **1989**, *11*, 72–76.
16. Peng, J.; Zhang, Y.Z.; Zhou, Q. Spectral Characteristics of Soils in Hunan Province as Affected by Removal of Soil Organic Matter. *Soils* **2006**, *38*, 453–458.
17. Steinier, J.; Termonia, Y.; Deltour, J. Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.* **1972**, *44*, 1906–1909.
18. Askari, M.S.; Cui, J.F.; O'Rourke, S.M.; Holden, N.M. Evaluation of soil structural quality using VIS–NIR spectra. *Soil Tillage Res.* **2015**, *146*, 108–117.
19. Hook, J. Smoothing non-smooth systems with low-pass filters. *Physica D Nonlinear Phenom.* **2014**, *269*, 76–85.
20. Huang, Y.H.; Wang, J.H.; Jiang, D.; Zhou, Q. Reconstruction of MODIS-EVI Time-Series Data with S-G Filter. *Geomat. Inf. Sci. Wuhan Univ.* **2009**, *34*, 1440–1443.
21. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
22. Yang, H.; Shen, R.P.; Wu, L.Y.; Li, M Temporal and Spatial Analysis of Remotely Sensed Vegetation Coverage Changes in Jiangxi Province Based on S-G Filter. *Sci. Technol. Eng.* **2014**, *14*, 101–106.
23. Kong, L.J. *Matlab Wavelet Analysis Super Learning Manual*; The People's Posts and Telecommunications Press: Beijing, China, 2014.
24. Virmani, J.; Kumar, V.; Kalar, N.; Khandelwal, N. SVM-Based Characterization of Liver Ultrasound Images Using Wave Packet Texture Descriptors. *J. Digit. Imaging* **2012**, *26*, 530–543.
25. Rivera-Cacedo, J.P.; Jochem, V.; Jordi, M.; Camps-Valls, G.; Moreno, J. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 88–101.

26. Pearson, K. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **1901**, *2*, 559–572.
27. Goodarzi, M.; Sharma, S.; Ramon, H.; Saeys, W. Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. *TrAC Trends Anal. Chem.* **2015**, *67*, 147–158.
28. Giacomo, D.R.; Stefania, D.Z. A multivariate regression model for detection of fumonisins content in maize from near infrared spectra. *Food Chem.* **2013**, *141*, 4289–4294.
29. Wang, F.H.; Gao, J.; Zha, Y. Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 73–84.
30. Chang, C.W.; Laird, A.D.; Mausbach, M.J.; Hurburgh, C.R. Near infrared reflectance spectroscopy: Principal components regression analysis of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490.
31. Qiao, X.X.; Wang, C.; Feng, M.C.; Yang, W.D.; Ding, G.W.; Sun, H.; Shi, C.C. Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques. *Spectrosc. Lett.* **2017**, *50*, 156–163.
32. Wang, X.; Zhang, F.; Kung, H.T.; Johnson, V.C. New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sens. Environ.* **2018**, *218*, 104–118.
33. Pineiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* **2008**, *216*, 316–322.
34. Baumgardner, M.F. Reflectance properties of soils. *Adv. Agron.* **1985**, *38*, 1–44.
35. Zheng, G.H.; Ryu, D.; Jiao, C.Q. Estimation of Organic Matter Content in Coastal Soil Using Reflectance Spectroscopy. *Pedosphere* **2016**, *26*, 130–136.
36. Wang, J.; He, T.; Lv, C.Y.; Chen, Y.; Jian, W. Mapping soil organic matter based on land degradation spectral response units using Hyperion images. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S171–S180 (Suppl. 2).
37. Nocita, M.; Stevens, A.; Toth, G.; Panagos, P.; van Wesemael, B.; Montanarella, L. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biol. Biochem.* **2014**, *68*, 337–347.
38. Mouazen, A.M.; Maleki, M.R.; Baerdemaeker, J.D.; Ramon, H. On-line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Tillage Res.* **2007**, *93*, 13–27.
39. Luce, M.S.; Ziadi, N.; Zebarth, B.J.; Grant, C.A.; Tremblay, G.F.; Gregorich, E.G. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. *Geoderma* **2014**, *232–234*, 449–458.
40. Ben-Dor, E.; Banin, A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci. Soc. Am. J.* **1995**, *159*, 259–270.
41. Chabrilat, S.; Goetz, A.F.H.; Krosley, L.; Olsen, H.W. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sens. Environ.* **2002**, *82*, 431–445.
42. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777.
43. Chen, H.Z.; Song, Q.Q.; Tang, G.Q.; Feng, Q.X.; Lin L. The Combined Optimization of Savitzky-Golay Smoothing and Multiplicative Scatter Correction for FT-NIR PLS Models. *ISRN Spectrosc.* **2013**, 1–9. (doi: 10.1155/2013/642190)
44. Oldham, K.B.; Spanier, J. The fractional calculus. *Math. Gazette.* **1974**, *56*, 396–400.
45. Li, B.; Xie, W. Adaptive fractional differential approach and its application to medical image enhancement. *Comput. Electr. Eng.* **2015**, *45*, 324–335.

