



## Article

# Prediction of Soil Organic Matter by VIS–NIR Spectroscopy Using Normalized Soil Moisture Index as a Proxy of Soil Moisture

Yongsheng Hong <sup>1,2,3</sup>, Lei Yu <sup>4,5</sup>, Yiyun Chen <sup>1,2,3,\*</sup> , Yanfang Liu <sup>1,6,\*</sup>, Yaolin Liu <sup>1,6</sup>, Yi Liu <sup>1,2,3</sup> and Hang Cheng <sup>1,2,3</sup>

<sup>1</sup> School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China; hys@whu.edu.cn (Y.H.); yaolin6100@sina.com (Y.L.); liuyi2010@whu.edu.cn (Y.L.); chh0530@foxmail.com (H.C.)

<sup>2</sup> Institute of Soil Science, State Key Laboratory of Soil and Sustainable Agriculture, Chinese Academy of Sciences, Nanjing 210008, China

<sup>3</sup> Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

<sup>4</sup> School of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China; yulei@mail.ccnu.edu.cn

<sup>5</sup> Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China

<sup>6</sup> Key Laboratory of Geographic Information System of the Ministry of Education, Wuhan University, Wuhan 430079, China

\* Correspondence: chenyy@whu.edu.cn (Y.C.); yfliu610@sina.com (Y.L.); Tel.: +86-27-6877-8381 (Y.C.)

Received: 24 November 2017; Accepted: 21 December 2017; Published: 24 December 2017

**Abstract:** Soil organic matter (SOM) is an important parameter of soil fertility, and visible and near-infrared (VIS–NIR) spectroscopy combined with multivariate modeling techniques have provided new possibilities to estimate SOM. However, the spectral signal is strongly influenced by soil moisture (SM) in the field. Interest in using spectral classification to predict soils in the moist conditions to minimize the influence of SM is growing. The objective of this study was to investigate the transferability of two approaches, SM-based cluster method with known SM (classifying the VIS–NIR spectra into different SM clusters to develop models separately), the normalized soil moisture index (NSMI)-based cluster method with unknown SM (utilizing NSMI to indicate the SM and establish models separately), to predict SOM directly in moist soil spectra. One hundred and twenty one soil samples were collected from Central China, and eight SM levels were obtained for each sample through rewetting experiments. Their reflectance spectra and SOM concentrations were measured in the laboratory. Partial least square-support vector machine (PLS-SVM) was employed to construct SOM prediction models. Specifically, prediction models were developed for NSMI-based clusters with unknown SM data. The models were assessed through three statistics in the processes of calibration and validation: the coefficient of determination ( $R^2$ ), root mean square error (RMSE) and the ratio of the performance to deviation (RPD). Results showed that the variable SM led to reduced VIS–NIR reflectance nonlinearly across the entire spectral range. NSMI was an effective spectral index to indicate the SM. Classifying the VIS–NIR spectra into different SM clusters in known SM states could improve the performance of PLS-SVM models to acceptable prediction accuracies ( $R^2_{cv} = 0.69–0.77$ , RPD = 1.79–2.08). The estimation of SOM, when using the NSMI-based cluster method with unknown SM (RPD = 1.95–2.04), was similar to the use of the SM-based cluster method with known SM (RPD = 1.79–2.08). The predictive results (RPD = 1.87–2.06) demonstrated that the NSMI-based cluster method has potential for application outside the laboratory for SOM prediction without knowing the SM explicitly, and this method is also easy to carry out and only requires spectral information.

**Keywords:** visible and near-infrared spectroscopy; soil moisture; soil organic matter; fuzzy k-means clustering; normalized soil moisture index; partial least square-support vector machine

## 1. Introduction

It is well recognized that soil organic matter (SOM) can greatly influence the physical and chemical properties in soil, and plays a positive role in crop growth [1]. Fast, accurate and cost-effective determination of the SOM content across large areas is crucial for local agricultural development [2,3]. The traditional chemical methods of SOM analysis conducted in laboratory environments are relatively accurate, but these methods are tedious and labor-intensive, and cannot rapidly monitor the SOM on broad scales.

In recent years, visible (VIS, 400–700 nm) and near-infrared (NIR, 700–2500 nm) spectroscopy combined with multivariate modeling techniques, provided an alternative tool to characterize SOM [4,5]. Besides, this VIS–NIR technique can be applied either in the laboratory or in the field, and when a calibration model between spectral data and their corresponding soil property reference value is developed, this model can be used to predict other soil samples in specific areas only containing their VIS–NIR spectra [6]. Compared with laboratory-based VIS–NIR spectroscopic measurements, field-based VIS–NIR spectroscopic measurements can greatly improve the scanning efficiency, due to avoiding the collection and preparation of soil samples (e.g., transporting, air drying, grinding, sieving, etc.), a lot of time and labor would be saved [7]. However, the field spectral data is more susceptible to interference of external environmental factors than laboratory deployment, such as variable soil moisture, temperature, natural aggregation, the condition of soil surface and so on, which may lead to the prediction accuracy in the field being less accurate than that of laboratory-based stable environments [8]. Among these factors, the presence of soil moisture (SM) does have a substantial, complex and nonlinear effect on reflectance spectra [9–11]. Generally, with increasing SM, the reflectance spectra across the entire spectral range (350–2500 nm) decrease, and in the field, SM may vary significantly [10]. However, direct analysis of one-dimensional spectral data is not efficient enough to explore the effect of SM on the reflectance spectra because of the nature of varying SM. A better approach is to use two-dimensional correlation spectroscopy that can provide more detailed spectral information [12].

Most recently, many researchers have investigated the effect of SM on reflectance spectra, and some methods for removing or minimizing the SM and improving the prediction accuracy of SOM have been also put forward and explored, such as external parameter orthogonalization (EPO) [9,13–15], direct standardization (DS) and piecewise direct standardization (PDS) [11,16–18], “spiking” method [19,20], first derivative [21], slope bias correction (SB) [22], orthogonal signal correction (OSC) and generalized least squares weighting (GLSW) [23,24], spectral classification [25,26] and so on. The above-mentioned EPO, DS and PDS strategies usually require dry soil spectral libraries (SSLs) at a specific scales (global, continental, national or regional) and then use a projection matrix (or transfer matrix) to correct the moist spectra. These SSLs contain both laboratory-based VIS–NIR spectra and corresponding soil property data, which can be applied to develop spectroscopic calibration models relating to moist spectra. The aforementioned “spiking” strategy is based on the idea of strengthening the leverage of the validation samples (moist samples) by increasing the diversity of the calibration samples, and thus can improve the generalization capacity of the model for the validation samples. Likewise, this method is usually associated with SSLs. Spectral preprocessing of first derivative conducted by Wu et al. [21] serve as another method to alleviate the effect of SM on reflectance spectra, in which they found that the reflectance spectra processed by first derivative within some specific wavelength ranges were insensitive to SM, and these specific regions could be used to determine soil parameters under field conditions. However, the effect influenced by SM on reflectance spectra are nonlinear and very complex, thus this method might neglect some important variables. The hypothesis of the SB method is that the prediction systematic error caused by SM can be corrected by a linear slope and bias correction. Similarly, this method also cannot solve the complex and nonlinear effect of SM because of only using simple linear corrections for target variables [22]. Jiang et al. [23] utilized OSC and GLSW algorithms to remove relative SM effects and verified the transferability of OSC-partial least square (OSC-PLS) and GLSW-PLS models between different

SM levels. Successful practices were reported in their study. One concern with their research is that different SM levels are difficult to determine if directly applied in the field.

Mouazen et al. [25] successfully employed the factorial discriminant analysis (FDA) method to classify the soil VIS–NIR spectra into different SM groups to minimize the effect of SM, and pointed out that the spectral classification would be useful to improve the prediction accuracy for other soil properties (e.g., C and N). Nocita et al. [26] introduced the idea of spectral classification to determine soil organic carbon (SOC) content for moist samples, and computed the normalized soil moisture index (NSMI) as a proxy of SM to spectrally classify soil VIS–NIR spectra into different clusters. Then separate PLS models relating VIS–NIR spectra to SOC content were established for both known SM clusters and unknown SM clusters determined by the NSMI. Their results showed that the predictive accuracies of SOC after NSMI classification (with unknown SM) were similar to those of known SM. Thus, efforts to improve the prediction accuracy of SOM with a wide range of SM may benefit from dividing soil VIS–NIR spectra into smaller sub-variations of SM. The main advantage of the spectral classification method using NSMI is that it can be used especially when the SM is unknown and greatly varies across large areas [26]. Moreover, it does not require sample transportation and preparation if the VIS–NIR spectra of moist soil can directly predict soil properties in the field [27]. However, as described in the work of Nocita et al. [26], the definition of the spectral classification of what it meant to be clear or mixed mainly depended on visual observation, so their method of spectral classification lacked objective criteria. The fuzzy k-mean (FKM) clustering can be applied to identify the optimal number of clusters that do not need set artificial thresholds, and also can deal with the continuous and complex relationships existing in the spectral data [3,28,29]. Although the NSMI calculated from the reflectance values at 1800 and 2119 nm is straightforward to use for reducing the impact of SM, its extension to other datasets finds difficulty, since a NSMI index derived from one specific dataset may not be suitable for another one, and it is difficult to utilize a general NSMI index to indicate SM. Therefore, a more practical NSMI index needs to be taken into account when constructing specific predictive models.

Soil VIS–NIR spectra are mostly non-specific, containing many weak, broad and overlapping bands and thus multivariate statistics would be needed to relate spectra with soil parameters to calibrate prediction models. PLS is a commonly used technique achieving this goal, but it is only a linear calibration method. The attention paid to applying non-linear modeling techniques is continuously increasing, because there are rarely linear relationships between reflectance spectra and soil parameters, especially variation in SM including non-linear nature [9]. In particular, support vector machine (SVM) based on kernel-based learning methods has attracted extensive attention in soil VIS–NIR spectra [5]. Thus, the modeling technique of the combination of PLS and SVM is expected to be more superior to PLS alone.

In this study, we aimed to: (1) investigate the influence of SM on the reflectance spectra; (2) explore the feasibility of classifying the global model into different clusters for known SM (SM-based cluster); (3) verify if the NSMI could be used as an indicator of SM, and specifically, assess the model transferability between SM-based cluster method and NSMI-based cluster method in the corresponding cluster. We hope this research would provide a theoretical guidance to monitor the SOM for moist samples with unknown SM.

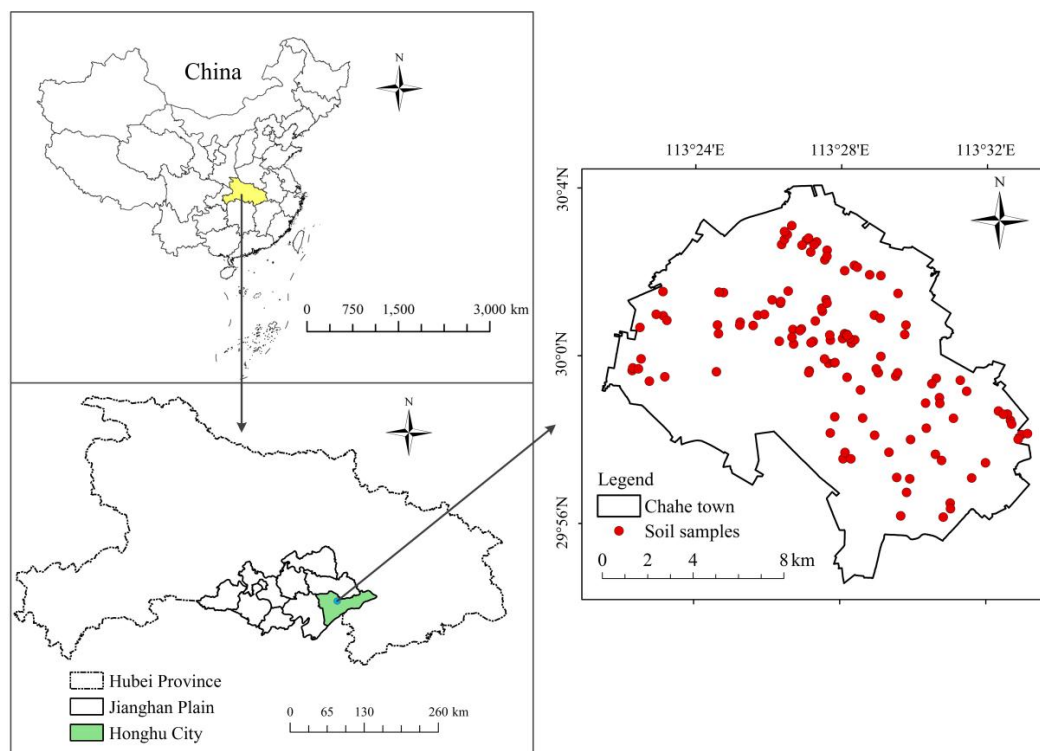
## 2. Materials and Methods

### 2.1. Study Area and Field Sampling

The study area (Chahe town) is situated in the east of Jiangnan plain (Hubei Province, China) (Figure 1), which is characterized by a typical subtropical humid monsoon climate with abundant sunlight, rainfall and distinctive seasons, and is famous for its important agricultural region. The elevation of this study region is around 2–35 m, covering a geographical area of 153 km<sup>2</sup>.

In December 2011, July 2012, November 2012 and April 2013, 121 soil samples were collected (Figure 1). At each site, approximately 1 kg of surface soils was obtained (0–20 cm), and we also used a handheld global position system (GPS) to record the corresponding geographical coordinates (positional

error < 10 m). These fresh soil samples were packed in sealed plastic bags, labeled and taken to the laboratory. According to the Chinese Soil Taxonomic Classification, the soil types of these 121 samples belong to paddy soils and fluvo-aquic soils, and the major land use types are irrigable lands.



**Figure 1.** Study area and the location of each sampling site.

## 2.2. Laboratory Analyses and Rewetting Experiment

The collected soil samples were further air-dried, and crushed to pass through a 2 mm sieve. Stones, roots and the vegetation litter were avoided from the soil. Each soil sample was divided into two portions, with one part for SOM analyses and the other part for spectral measurement, respectively. The SOM concentrations of all the 121 samples were chemically determined using the potassium dichromatic oxidation titration method [30].

Prior to rewetting experiment, all soil samples were oven-dried at 105 °C for 24 h to eliminate soil moisture. Sample rewetting was conducted in only one batch ( $n = 121$ ). Approximately 100 g of oven-dried soil for each sample was weighed using a scale (accuracy = 0.01 g) in the laboratory, then placed in a petri dish and labeled. These samples were wetted with 40 g of deionized water each and weighed. The 40 g of water were added slowly to each sample with a spray flask, and the dishes were immediately covered with a lid for 24 h to avoid evaporation to obtain uniform moisture distribution within samples. These dishes were weighed to determine the varying SM weight and then scanned the first set of moist spectra. In the next few days, the samples were uncovered to air-dry at room temperature, and weighed every day, and their spectra were also recorded simultaneously. As a result, a total of eight different SM levels were collected. The average SM for 8 SM levels were 32.66%, 29.10%, 25.50%, 21.62%, 16.95%, 11.85%, 6.87%, 2.55% (gravimetric, dry basis).

## 2.3. Spectral Measurement and Pre-Processing

Reflectance spectra of each soil sample was acquired in a dark room using an ASD FieldSpec®3 Portable Spectrometer (Analytical Spectral Devices, Boulder, CO, USA), with sampling interval of 1.4 nm (350–1000 nm) and 2 nm (1000–2500 nm). The main geometric parameters of the spectrometer

set-up were illustrated as follows: a 50 W halogen lamp with a 45° incident angle was used as unique light source; the lamp away from petri dish was set as 30 cm; the probe was mounted vertically about 15 cm above the dish, and the field of view of the probe was much smaller than the diameter of the dish. Each sample was scanned in four directions (each rotating the dishes by 90°), and five scans were collected at every direction (a total of 20 scans), and then averaged to one spectrum for each sample. Every ten samples, we would optimize the spectrometer using a standardized white Spectralon® panel as a white reference. The re-sampling interval of the ASD spectrometer was 1 nm.

Before the original spectral data exported, splice corrections were proceeded using viewSpec™ software (version 6.2.0, ASD Inc.: Longmont, CO, USA) to solve breakpoint phenomena around 1000 and 1800 nm. The reflectance of each spectrum was narrowed to 400–2400 nm, and then Savitzky-Golay smoothing with 11 filter widths and a second-order polynomial was applied to filter the reflectance curves [31]. Every spectrum was then resampled by averaging ten successive wavelengths to simplify the dimensionality of spectral matrix, and the final wavelength number was 201 for each spectral curve.

#### 2.4. Spectral Angle and Two-Dimensional Correlation Spectroscopy

Spectral angle (SA) is a tool that can measure the spectral similarity between a test spectrum  $t$  and a reference spectrum  $r$ , by calculating the “angle” [32]. The SA ( $\theta$ ) is defined by:

$$\theta = \arccos \frac{\sum_{i=1}^n t_i r_i}{\sqrt{\sum_{i=1}^n t_i^2} \sqrt{\sum_{i=1}^n r_i^2}}, \dots \theta \in [0, \frac{\pi}{2}] \quad (1)$$

where  $t_i$  and  $r_i$  are the spectral reflectance at specific wavelength,  $i$ ;  $n$  represent the total number of wavelengths. Here,  $n = 201$ .

Two-dimensional correlation spectroscopy developed by Noda is a powerful spectral analysis method to analyze complex spectral intensity variation obtained successively under certain form of perturbation, such as temperature, pressure, or even concentration, and so on [12,33]. In two-dimensional correlation spectroscopy, a group of spectral data is transformed into a correlation intensity map defined by two independent spectral axes. Such two-dimensional correlation spreading spectra into 2-D space would provide more spectral features than conventional one-dimensional spectra, because some spectral features may not be readily observed from one-dimensional spectra. Moreover, some overlapped peaks can also be easily differentiated in real data. Two-dimensional correlation spectra mainly contain three basic properties: synchronous spectrum, asynchronous spectrum and disrelation spectrum. In our case, we used the synchronous correlation spectra to investigate the influence of SM on VIS–NIR spectra in the calibration dataset. Readers can be referred to Noda [12] for additional details on two-dimensional correlation spectroscopy.

#### 2.5. Principal Component Analysis and Fuzzy K-Mean Clustering

Principal component analysis (PCA) is a mathematical method for data compression or reduction, and it is commonly applied to extract the informative features from high-dimensional datasets. Through an orthogonal transformation process, a set of original spectral matrix with possibly correlated variables is converted to a group of new uncorrelated variables that are linear combinations of the original variables, namely principal components (PCs). PCA uses a modest number of PCs to characterize as much of the variation in the original data as possible, so the first few PCs might help us to interpret the original dataset. We employed nonlinear iterative partial least squares algorithm to implement the PCA to compute PCs and scores [34]. This algorithm avoids calculating the covariance matrix of the spectral matrix, which can effectively reduce the computational time. In general, the spectral cluster analysis is firstly dependent on the PC1, namely the reflectance spectra intensity, and then dependent on the spectral shape (PC2) [28].



Fuzzy k-mean (FKM) clustering is a commonly employed approach in unsupervised clustering. Unlike the k-means clustering and discriminant analysis techniques, FKM clustering algorithm, without setting the threshold manually, can provide an objective criterion to determine the optimal number of clusters, and this is a competitive advantage over other methods [28]. The basic idea of the FKM clustering is to divide a set of datasets (in our case known as the PCA scores) into  $k$  classes to seek out the iterative minimization of the objective function. Three evaluation parameters are obtained from the FKM clustering algorithm (i.e., fuzziness performance index (FPI), modified partition entropy (MPE) and clustering separation index (S)). FPI is a measure of the continuity between the classifications, and a value close to 0 indicates there is little shared membership and the partition of classifications is obvious. MPE is a comprehensive index to measure the fuzzy degree among various classifications. S represents the relative distinction between the classifications. The optimal number of classifications can be determined when these three values approach 0 simultaneously. For a more comprehensive description of the FKM clustering algorithm, readers are directed to Shi et al. [28]. FuzME 3.0 software [35] was used to perform the FKM clustering analyses. The maximum number of iterations, the convergence threshold, and the fuzzy weighted index in the FKM clustering algorithm was set to 300, 0.001 and 1.5, respectively [28].

## 2.6. Normalized Soil Moisture Index

Normalized soil moisture index (NSMI) is a non-dimensional measure of reflectance spectra, calculated from normalized difference of two wavelengths using mathematical operations [36]. For the calculation of the NSMI, all possible wavelength combinations in the 400–2400 nm region were explored for their correlation with SM to select the optimal spectral indices to indicate the SM condition, and its mathematical expression is characterized by Equation (2):

$$NSMI = \frac{R_i - R_j}{R_i + R_j} \quad (2)$$

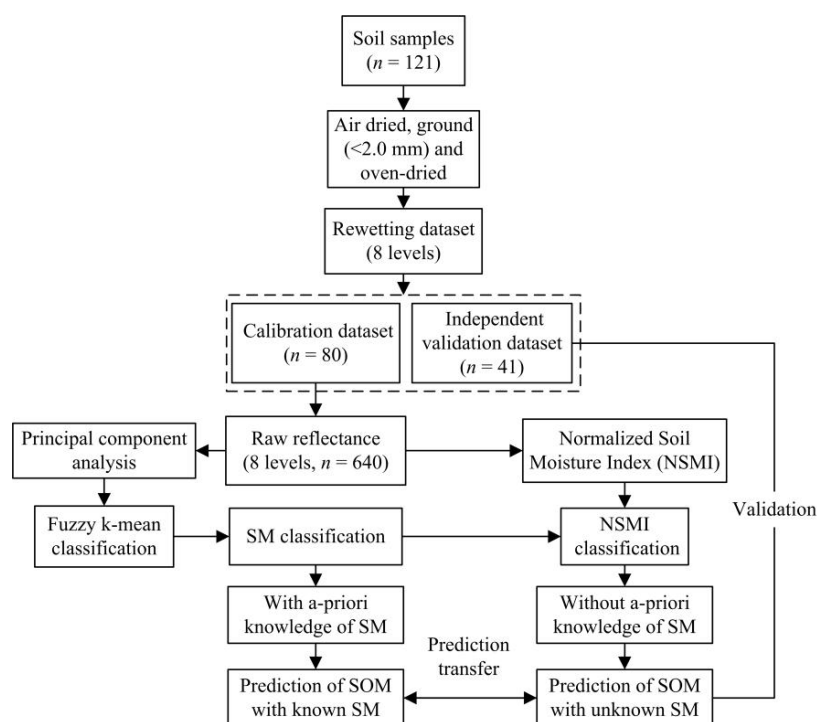
where  $R_i$  and  $R_j$  represent the reflectance values at  $i$ th and  $j$ th nm, respectively. Two-dimensional correlation map was used to show their possible combination, and developed using a program in Matlab R2014a. The NSMI was easy to use and had good interpretability [26,36].

## 2.7. Calibration and Validation

The whole dataset ( $n = 121$ ) was sorted in ascending order according to SOM content, and we used stratified sampling approach to separate 121 samples into 41 strata with two or three intervals, and one sample was selected from each strata as independent validation dataset for model validation (a total of 41 samples). The remaining samples were selected as a calibration dataset for model calibration (a total of 80 samples). In each dataset, 8 SM levels were presented for each sample.

Leave-one-out cross-validation was applied to identify the optimal PLS latent factors that obtained the first minimum value of the root mean squared error of cross-validation. The component factors derived from the PLS were then input to SVM for establishing PLS-SVM models. We selected the e-SVM algorithm and radial basis function for modeling, and a grid search technique with 5-fold cross-validation were chosen for model optimization [37,38]. The coefficient of determination ( $R^2$ ), root mean squared error (RMSE) and ratio of the performance to deviation (RPD) between the predicted and measured SOM in the processes of calibration, cross-validation and validation were selected to evaluate the model performance. In terms of RPD,  $RPD < 1.4$  indicated unacceptable models/predictions;  $1.4 \leq RPD < 1.8$  indicated fair models/predictions;  $1.8 \leq RPD < 2.0$  indicated good models/predictions;  $2.0 \leq RPD < 2.5$  indicated very good models/predictions;  $RPD \geq 2.5$  indicated excellent models/predictions [39,40]. Generally, the larger  $R^2$ , RPD and the smaller RMSE were indicators of a superior model. All data analyses were carried out in Matlab R2014a (The MathWorks Inc.: Natick, MA, USA).

We labeled samples by different SM levels rather than individual samples in order to avoid pseudo-replication of soil samples. The procedures of two classification methods for model calibration and validation are shown in Figure 2 and compared as follows: (1) the calibration dataset with 8 SM levels ( $n = 640$ ) was used to perform PCA, and the first few PCs were then clustered with FKM clustering (referred to as SM-based cluster). Then, PLS-SVM was used to develop a separate calibration model with known SM in each cluster. (2) The calibration dataset with 8 SM levels ( $n = 640$ ) was also applied to calculate NSMI. According to the number of the SM classification, the NSMI index was also divided into the same number of clusters (referred to as NSMI-based cluster). Likewise, a separate PLS-SVM model was established for each NSMI-based cluster, and the results of these models were compared to those of the corresponding SM-based cluster. (3) Models calibrated from NSMI-based cluster were further tested on independent validation dataset (8 SM levels,  $n = 328$ ) with unknown SM to evaluate the classification effect of NSMI method. Mouazen et al. [25] employed correct classification (CC) method to assess the performance of classification, which was calculated by dividing the number of correctly grouped samples by the total number of samples in that cluster, and we also adopted this method to evaluate the performance of NSMI classification.

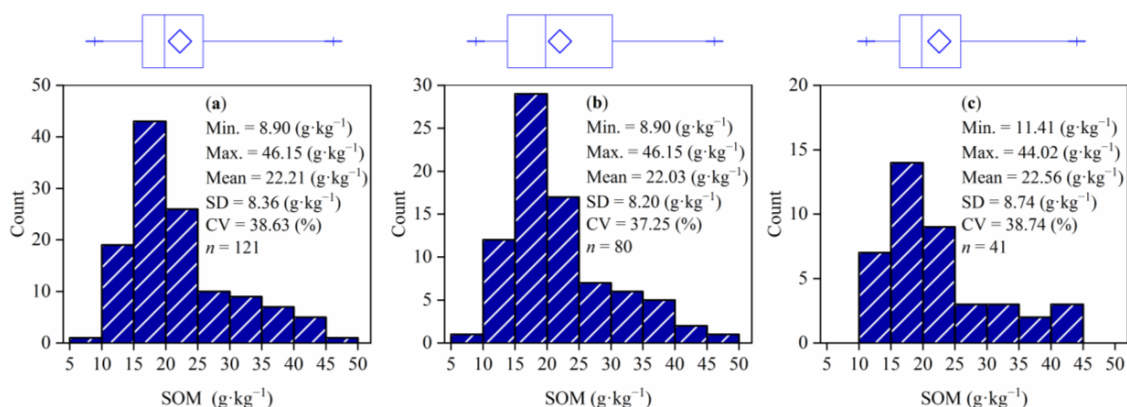


**Figure 2.** The flow chart of two classification methods (SM-based cluster method and NSMI-based cluster method) for SOM estimation at different SM levels in this study.  $n$ : the number of soil samples, same as below.

### 3. Results

#### 3.1. Descriptive Statistics of SOM

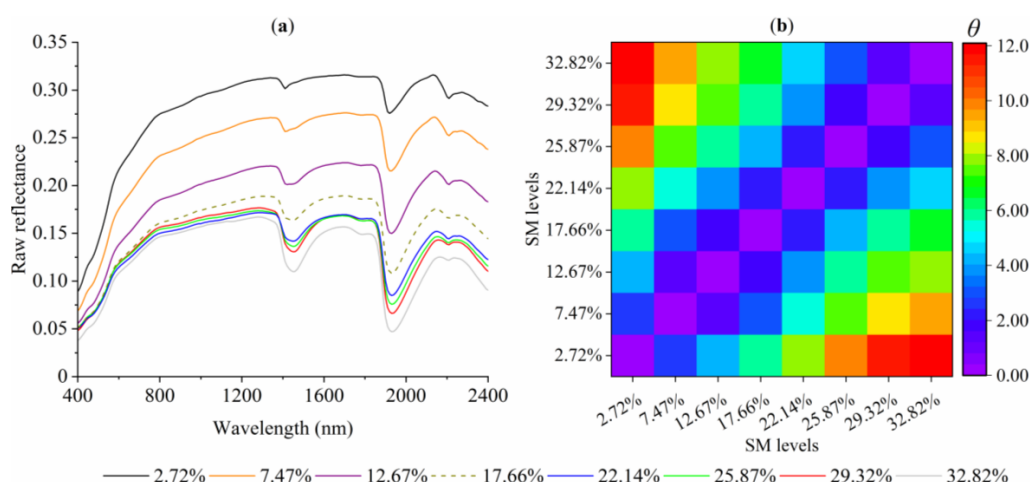
The summary statistics of SOM contents measured by traditional chemical methods for the whole, calibration and independent validation datasets are provided in Figure 3. The calibration dataset varied from 8.90 to 46.15  $\text{g}\cdot\text{kg}^{-1}$  with a mean value of 22.03  $\text{g}\cdot\text{kg}^{-1}$ , and the range in the independent validation dataset was from 11.41 to 44.02  $\text{g}\cdot\text{kg}^{-1}$  with an average of 22.56  $\text{g}\cdot\text{kg}^{-1}$ . Overall, the characteristic statistics of both the calibration and independent validation dataset were similar to the whole dataset, indicating that they were well divided to represent the whole dataset.



**Figure 3.** Box-plots, histograms and descriptive statistics of SOM: (a) the whole dataset; (b) calibration dataset; and (c) independent validation dataset. Min.: minimum, Max.: maximum, SD: standard deviation, CV: coefficient of variation,  $n$ : the number of soil samples.

### 3.2. Influence of SM on VIS–NIR Spectra

To analyze the influence of SM on reflectance spectra, spectral reflectance at different SM levels from the calibration dataset ( $n = 80$ ) were investigated (Figure 4). Spectral curves at different SM levels showed similar shapes but with different intensities (Figure 4a). Three obvious absorption peaks around 1420, 1940, 2200 nm were exhibited in all SM levels. The reflectance spectra across the entire spectral range tended to decrease as SM increased, but the shifts were not homogeneous along the wavelengths. SM had an evident effect on reflectance spectra: for low SM levels ( $SM \leq 17.66\%$ ), the decrease in reflectance spectra was more evident, while when SM was higher than 17.66%, the sensitivity of reflectance spectra to variable SM was less noticeable. To better explain the impact of SM, we calculated the spectral angle ( $\theta$ ) between mean spectral curves at different SM levels (Figure 4b). Overall, the SA ( $\theta$ ) varied greatly ( $0\text{--}12.06^\circ$ ). Taking reflectance spectra at SM level of 2.72% as an example, its SA with SM levels of 7.47%, 12.67%, 17.66%, 22.14%, 25.87%, 29.32% and 32.82% ranged from 0 to  $12.06^\circ$ : with the SM increasing, the differences of SA became more and more obvious, further proving that SM affected the reflectance spectra very significantly.

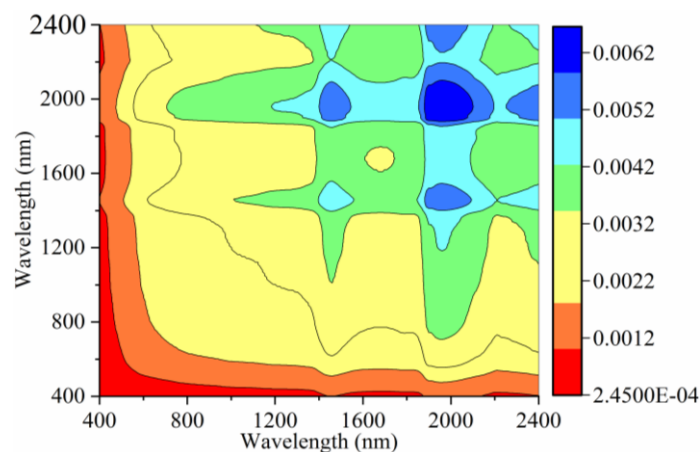


**Figure 4.** (a) Mean reflectance at different SM levels in the calibration dataset and (b) two-dimensional sample-sample spectral angle (SA, by angle) between different SM levels.  $\theta$ : spectral angle.

Two-dimensional synchronous correlation spectra on the averaged reflectance at different SM levels in the calibration dataset are also performed (Figure 5). According to the color bar illustrated in the



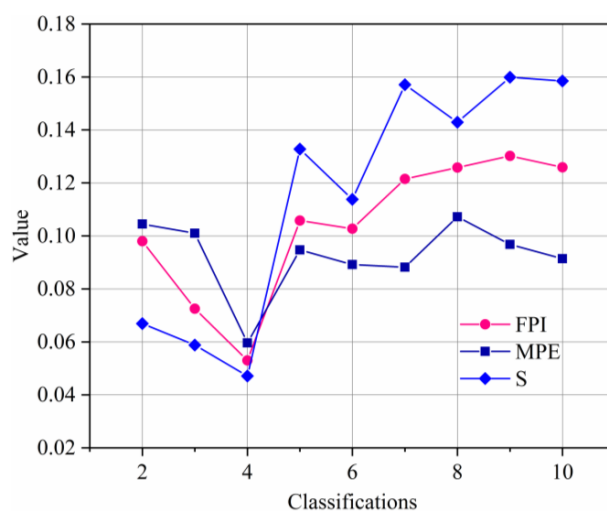
figure, the influence of SM on reflectance spectra in NIR range (1000–2400 nm) was clearly stronger than that in visible range. Besides, two autocorrelation peaks at diagonal position near 1450 nm and 1940 nm could be easily observed. Compared to the autocorrelation peak around 1450 nm, the autocorrelation peak around 1940 nm was more obvious, which indicated the wavebands around 1940 nm were more sensitive to the influence of SM, while the wavebands around 1450 nm were relatively insensitive.



**Figure 5.** Two-dimensional synchronous correlation spectra of the mean reflectance at different SM levels (in the calibration dataset).

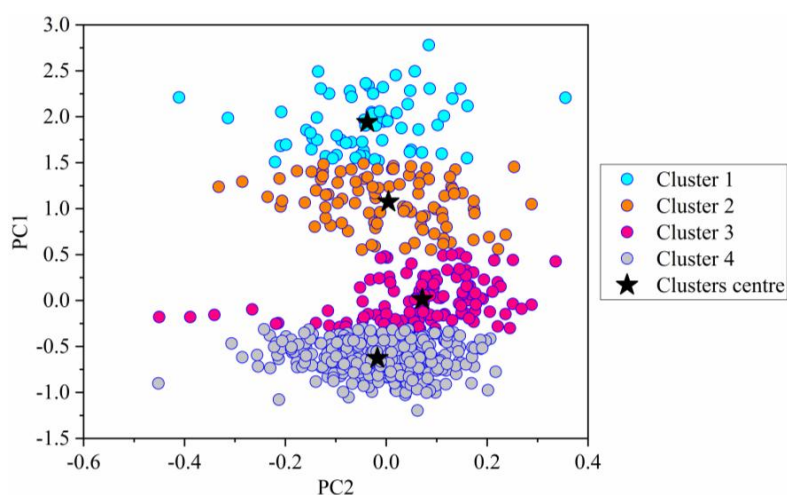
### 3.3. SM Classification

We first mean centered the reflectance spectra of the calibration dataset ( $n = 640$ ), and then performed PCA on the pretreated spectral dataset to reduce the dimensionality. The first two PCs together accounted for more than 95% of the total spectral variations (i.e., 97.53% and 1.70% for PC1 and PC2, respectively). The FKM clustering was then utilized to divide the scores of the first two PCs into spectrally similar clusters. A series of numbers of classes (2–10) were examined to identify the optimal number of classifications. The values of FPI, MPE and S of different classes are calculated and compared in Figure 6, from which we can determine that the best number of classifications is equal to 4, where the FPI, MPE and S obtained the minimum values simultaneously.



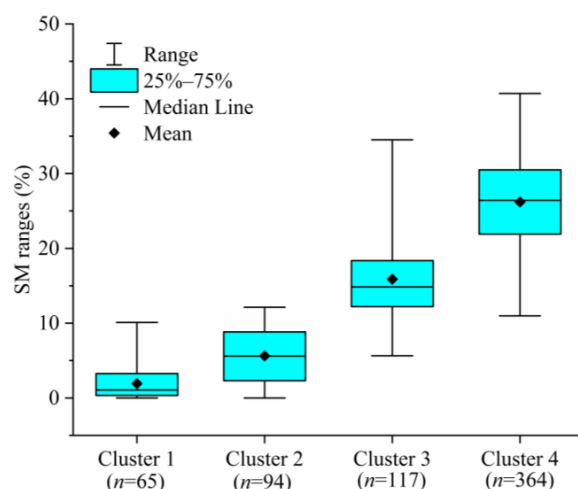
**Figure 6.** FPI, MPE and S values versus different numbers of classifications in the calibration dataset.

Thus, the scores of the first two PCs at different SM levels of the calibration dataset ( $n = 640$ ) were divided into four clusters, and its overview map is shown in Figure 7. As SM increased, the PC space distribution varied from cluster 1 to cluster 4 (Figure 7). The PC1 values of cluster 3 and cluster 4 were relatively concentrated, whereas cluster 1 and cluster 2 demonstrated a wide distribution. These phenomena were in agreement with Figure 4a, which shows for the higher SM levels, the reflectance spectra were not sensitive to variable SM. Besides, the ranges of PC2 values in four clusters were not similar to each other, manifesting that there was some difference in the spectral shape, and this result was also in accordance with Figure 4a.



**Figure 7.** Scatter plot of the first two principal components (PC1, PC2) of four spectral clusters in the calibration dataset ( $n = 640$ ).

Likewise, the 640 soil samples were regrouped to four clusters on the basis of the aforementioned results, and their corresponding descriptive statistics of SOM contents are listed in Table 1. Cluster 2 had the largest variability of SOM contents with a CV (the coefficient of variation) of 38.90%, including 94 soil samples; Cluster 1 had the smallest variability (with a CV of 32.51%) and consisted of 65 soil samples varying from 8.90 to 36.54  $\text{g} \cdot \text{kg}^{-1}$ . Cluster 3 comprised 117 soil samples with a CV of 37.84%. In particular, among these four clusters, cluster 4 had the largest number of soil samples, and was about 6 times larger than that of cluster 1, which indicated that when SM increased to higher levels (Figure 8), soil samples would locate in nearly similar spectral spaces.



**Figure 8.** Box-plots of SM in four clusters using SM-based cluster method (in the calibration dataset).

**Table 1.** Statistical characteristics of SOM using SM-based cluster method in the calibration dataset.

SM Classification	No. of Samples	Min. (g·kg <sup>−1</sup> ) <sup>a</sup>	Max. (g·kg <sup>−1</sup> ) <sup>b</sup>	Mean (g·kg <sup>−1</sup> )	SD (g·kg <sup>−1</sup> ) <sup>c</sup>	CV (%) <sup>d</sup>
Cluster 1	65	8.90	36.54	19.45	6.32	32.51
Cluster 2	94	8.90	46.15	22.60	8.79	38.90
Cluster 3	117	8.90	46.15	20.74	7.85	37.84
Cluster 4	364	8.90	46.15	22.75	8.26	36.33

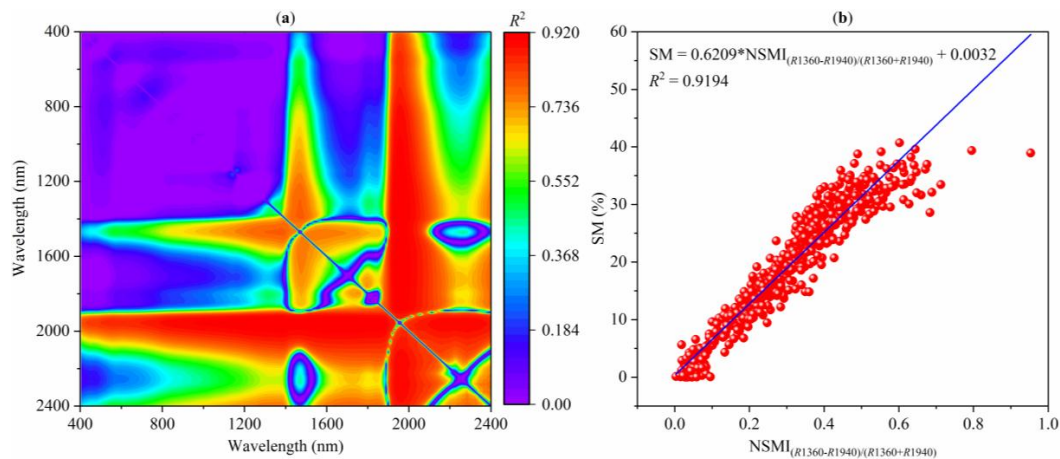
<sup>a</sup> Minimum; <sup>b</sup> Maximum; <sup>c</sup> Standard deviation; <sup>d</sup> Coefficient of variation.

### 3.4. NSMI Classification

The NSMI indices were computed wavelength-by-wavelength in the range of 400–2400 nm and then the coefficient of determination ( $R^2$ ) between SM and NSMI indices were calculated (Figure 9a). Results showed that there was a strong relationship between SM and NSMI, and the wavelength combinations with good correlation mainly located within 1200–2400 nm (red regions). The highest coefficient of determination of 0.9194 was obtained at 1360 nm on the x-axis and 1940 nm on the y-axis (referred to as the  $\text{NSMI}_{(R1360-R1940)/(R1360+R1940)}$ ).

For a deeper investigation of the best NSMI index, the NSMI values of each soil sample calculated from the corresponding reflectance spectra at 1360 nm and 1940 nm were obtained, and the overall relationship between the SM and  $\text{NSMI}_{(R1360-R1940)/(R1360+R1940)}$  could be fitted using a linear regression equation (Figure 9b):

$$\text{SM} = 0.6209 \times \text{NSMI}_{(R1360-R1940)/(R1360+R1940)} + 0.0032 \quad (3)$$



**Figure 9.** (a) 2-D correlogram of the coefficient of determination ( $R^2$ ) between SM and NSMI indices ( $n = 640$ ) and (b) the correlation between SM and the optimal NSMI index  $\{(R_{1360} - R_{1940}) / (R_{1360} + R_{1940})\}$  at different SM levels ( $n = 640$ ). The blue line illustrated in the (b) is the regression line.

The coefficient of determination ( $R^2$ ) between the  $\text{NSMI}_{(R1360-R1940)/(R1360+R1940)}$  and SM was 0.9194, and it was obvious that the  $\text{NSMI}_{(R1360-R1940)/(R1360+R1940)}$  was strongly correlated with SM. We employed Equation (3) to predict the independent validation dataset ( $n = 328$ ) at different SM levels, and the model gave a validation  $R^2$  of 0.8824, indicating the NSMI could be applied as a proxy of soil moisture.

To establish the transferability between SM-based cluster method and NSMI-based cluster method, the 640 soil samples were also partitioned into four clusters according to SM-based cluster, and the respective threshold criteria in the NSMI values were divided: (1) cluster 1,  $0.5045 \leq \text{NSMI} < 0.9534$ ; (2) cluster 2,  $0.4123 \leq \text{NSMI} < 0.5045$ ; (3) cluster 3,  $0.3419 \leq \text{NSMI} < 0.4123$ ; (4) cluster 4,  $0.0034 \leq \text{NSMI} < 0.3419$ . The descriptive statistics of SOM for each cluster are summarized in Table 2. SOM contents in cluster 1 displayed a narrow range of 8.90–41.61 g·kg<sup>−1</sup>, with a CV of 35.26%, whereas cluster 2, cluster 3 and cluster 4 were characterized by slightly larger ranges (Min.–Max.) and CVs of

SOM contents, compared with cluster 1. Although some differences existed between the respective cluster obtained from SM-based cluster method and NSMI-based cluster method (Tables 1 and 2), by and large, some comparable results of SOM contents for each corresponding cluster could be observed. For instance, the statistical characteristics of cluster 4 processed by the SM-based cluster method were similar to those of cluster 4 treated by the NSMI-based cluster method (a same range of SOM), only minor differences existed (mean = 22.75 g·kg<sup>-1</sup>, CV = 36.33% for SM-based cluster method, while mean = 21.86 g·kg<sup>-1</sup>, CV = 37.33% for NSMI-based cluster method).

**Table 2.** Statistical characteristics of SOM contents using NSMI classification method in the calibration dataset.

NSMI Classification	No. of Samples	Min. (g·kg <sup>-1</sup> ) <sup>a</sup>	Max. (g·kg <sup>-1</sup> ) <sup>b</sup>	Mean (g·kg <sup>-1</sup> )	SD (g·kg <sup>-1</sup> ) <sup>c</sup>	CV (%) <sup>d</sup>	CC <sup>e</sup>
Cluster 1	65	8.90	41.61	21.27	7.50	35.26	75.38%
Cluster 2	94	8.90	46.15	21.99	7.97	36.26	82.98%
Cluster 3	117	8.90	46.15	22.99	8.65	37.64	71.79%
Cluster 4	364	8.90	46.15	21.86	8.16	37.33	90.93%

<sup>a</sup> Minimum; <sup>b</sup> Maximum; <sup>c</sup> Standard deviation; <sup>d</sup> Coefficient of variation; <sup>e</sup> Correct classification.

### 3.5. Estimation of SOM with PLS-SVM Model

Separate PLS-SVM models for SOM estimation were built for each cluster generated from the SM-based cluster method with known SM and NSMI-based cluster method with unknown SM, and the cross-validation results are shown in Table 3. Overall, in SM-based cluster method, the best model was obtained for cluster 4 ( $R^2_{cv} = 0.77$  and RPD = 2.08). According to the five-level interpretations of RPD (Section 2.7), a very good model could be observed for cluster 1 (RPD = 2.05); a fair model and good model were obtained for cluster 2 (RPD = 1.79) and cluster 3 (RPD = 1.90), respectively.

**Table 3.** PLS-SVM cross-validation results for SOM estimation based on SM-based cluster method, NSMI-based cluster method, and the entire calibration dataset ( $n = 640$ ).

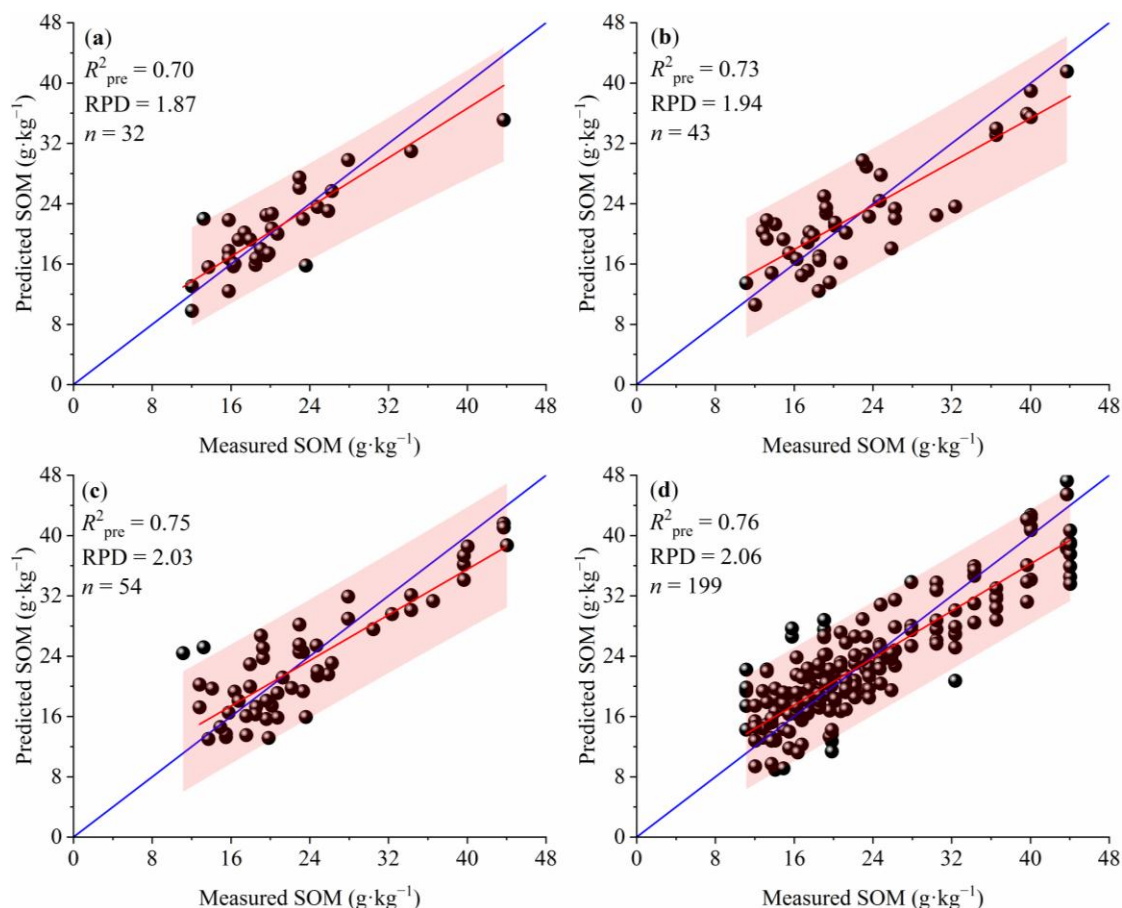
Classification Methods	Classifications	No. of Samples	Latent Variables	$R^2_{cv}$	RMSE <sub>cv</sub> (g·kg <sup>-1</sup> )	RPD
SM-based cluster	Cluster 1	65	9	0.76	3.09	2.05
	Cluster 2	94	8	0.69	4.90	1.79
	Cluster 3	117	9	0.72	4.14	1.90
	Cluster 4	364	12	0.77	3.98	2.08
NSMI-based cluster	Cluster 1	65	11	0.73	3.84	1.95
	Cluster 2	94	7	0.74	4.09	1.95
	Cluster 3	117	8	0.75	4.30	2.01
	Cluster 4	364	10	0.76	4.00	2.04
-	-	640	12	0.59	5.23	1.56

In NSMI-based cluster method (Table 3), the PLS-SVM models of cluster 2 and cluster 3 performed slightly better than the corresponding cluster from the SM-based cluster method, with RPD = 1.95 and 2.01, respectively, while the cross-validation accuracies for cluster 1 and cluster 4 were lower (in terms of  $R^2_{cv}$  and RPD) compared with SM-based cluster method. Thus, it could indicate that, in comparison with the SM-based cluster method, the PLS-SVM cross-validation models using the NSMI-based cluster method would obtain similar accuracies.

To explore the feasibility of improving estimation of SOM at different SM levels by splitting the calibration dataset into smaller sub-clusters, PLS-SVM model was also performed on the whole calibration dataset ( $n = 640$ , global calibration) to compare the performance of the sub-models with global model (Table 3). We observed better results for sub-models than global model: the range of RPD in SM-based cluster method was 1.79 to 2.08, and in NSMI-based cluster method was 1.95 to 2.04, while in global calibration the RPD was only equal to 1.56.

The calibration models of different clusters developed by the NSMI-based cluster method were further applied to test the independent validation dataset (8 SM levels,  $n = 328$ ), and these 328 soil samples were assigned to four clusters on the basis of the calculated threshold criteria of NSMI values

(Section 3.4), and the validation results between laboratory-measured SOM and VIS–NIR predicted SOM are plotted in Figure 10. Overall, with increasing SM, the prediction accuracies of four clusters showed a progressive increase of RPD values from 1.87 to 2.06. The prediction accuracy of cluster 4 was superior to the other three clusters, yielding the largest predictive  $R^2$  of 0.76 and RPD of 2.06. For other three clusters (cluster 1, cluster 2, and cluster 3), the results exhibited good and very good prediction with RPD values of 1.87, 1.94, and 2.03, respectively.



**Figure 10.** Validation scatter plots between laboratory-measured SOM and VIS–NIR predicted SOM obtained from PLS–SVM using NSMI-based cluster method: (a) cluster 1; (b) cluster 2; (c) cluster 3; and (d) cluster 4. The regression statistics ( $R^2_{pre}$ , RPD) for independent validation dataset are illustrated in the upper left corner of the four subplots. The blue lines are the 1:1 line, and the red lines are the regression line. The colored regions in the four subplots represent 95% prediction confidence intervals for covering the future data points.  $n$ : the number of soil samples.

## 4. Discussion

### 4.1. The Influence of SM on Reflectance Spectra

Many studies have demonstrated that VIS–NIR spectra can provide a viable alternative for the rapid determination of SOM content, mainly owing to the various chemical bonds, such as C–H, O–H, N–H, and so on [6]. However, the widespread application of VIS–NIR spectra in the field is limited due to the presence of some environment factors, and the variation in SM is a major obstacle. In our case, we obtained eight levels of SM (ranging from 2.55% to 32.66%, mean value) in the laboratory to simulate the influence of SM. SM had an evident effect on reflectance spectra, and our result was consistent with Minasny et al. [9] and Lobell and Asner [10]. It proved that for lower SM levels ( $SM \leq 17.66\%$ ), the decrease in reflectance spectra across all the wavebands was relatively pronounced, while for the



higher SM levels ( $SM > 17.66\%$ ), the decrease was not obvious (Figure 4a). These may be ascribed to the results reported by Lobell and Asner [10], who demonstrated that, as SM increased, once most of soil surfaces absorbed enough SM, the remaining SM that sequentially filled into the micro and macro pores would have little effect on the reflectance spectra. Mouazen et al. [25] established six SM levels by adding water (ranging from 0% to 27.5%, by weight) for the single-field samples, and reported that when SM was higher than 15%, the sensitivity of reflectance spectra to variable SM decreased. Their results were in line with the findings reported by Nocita et al. [26]. Similar results were also reported in our research. Thus, we believe our experimental design can serve as a reference for future research when SM is lower than the field holding capacity.

Figure 5 displays that the influence of SM was more evident for the longer wavelengths (1000–2400 nm), particularly the strong SM absorption peak around 1940 nm, which even masked the absorption peak signals around 2200 nm associated with organic functional groups. Published studies, however, indicated that some special wavebands related to SOM (or SOC) were located within these areas influenced by the SM [5,28]. For example, Knadel et al. [41] summarized that key components in organic matter had a peak around 1930 nm. Vasques et al. [42] reported that wavebands around 1400 and from 1800 to 2400 nm were especially important for SOC estimation. Therefore, a method for minimizing the effects of SM on reflectance spectra in the estimation of SOM is indispensable.

#### 4.2. Clustering the Modeling Dataset into Different SM Levels

Although the reflectance spectra are highly influenced by SM, some studies confirm that if the calibration dataset and validation dataset come from specific SM conditions (approximately similar SM levels), the effects of SM on SOM/SOC estimation would be reduced, and the moist VIS–NIR spectra can also be applied to predict soil properties using modeling strategies [2]. For instance, Rodionov et al. [43] found that it was practicable to predict SOC for each corresponding SM level (5% to 25%, 5% interval) with RPD ranging from 2.25 to 3.07. The study from Wang et al. [44] also pointed out that when SM was smaller than 22%, SOM could be reliably predicted if the range of SM at each SM level was well-defined (with a similar SM level).

In the current work, we assumed that the more similar two soil samples were in terms of their VIS–NIR spectra, the more similar they could be in terms of SM. That is to say, in a given set of soil samples at different SM levels, the variation of SM can be explained to a certain degree by the variation of spectral similarity/dissimilarity. We introduced the FKM clustering to divide the calibration dataset at nine SM levels ( $n = 640$ ) into smaller clusters, which could divide the ranges of SM into different specific SM conditions, and would reduce the non-linear effects of SM on SOM estimation. We observed an improvement in accuracies of cross-validation for clustered models than for global model (RPD = 1.56). Castaldi et al. [45] confirmed that with a priori knowledge of SM, the predictive models from four SM classes could improve the estimation accuracy of clay. Our results support their findings.

#### 4.3. NSMI

According to Figure 9b, it could be observed that NSMI was highly correlated with SM across all SM levels, and a strong linear relationship was obtained. Moreover, the wavelengths used to compute the NSMI are located at 1360 nm and 1940 nm, and only two spectral wavelengths are required, without a-priori knowledge of SM [26,35]. Thus, the NSMI derived from the VIS–NIR spectral feature space is a useful and potential index for monitoring SM, and the proposed methodology is also simple and easy to implement.

The results of correct classification (CC) obtained by the NSMI-based cluster method are provided in Table 2. For cluster 4, the CC of 90.93% was obtained (33 soil samples were misclassified), and the classification result was the best in all clusters. The order of the classification accuracies for the other three clusters was cluster 2 (CC = 82.98%, 16 soil samples were misclassified) > cluster 1 (CC = 75.38%, 16 soil samples were misclassified) > cluster 3 (CC = 71.79%, 33 soil samples were misclassified).

This suggested NSMI-based cluster method gave slightly worse performance than SM-based cluster method, but the results of CC in four clusters were still striking.

It was notable that the accuracy of cross-validation from the corresponding cluster in the NSMI-based cluster method was comparable to that obtained from the corresponding cluster in the SM-based cluster method (Table 3). The independent validation scatterplots processed by NSMI-based cluster method yielded prediction categories ranging from good to very good (RPD = 1.87 to 2.06) (Figure 10). The prediction accuracy of cluster 4 was superior to the other three clusters. One reason might be attributed to the result summarized by Stenberg [2], who experienced a similar result that rewetting samples showed positive effect for estimating SOC content.

With the application of a kernel function, the PLS-SVM model can be flexible to solve the complicated and non-linear regression problems. Other researchers also concluded that SVM was a suitable multivariate method when using VIS–NIR spectral calibration on field moist samples (e.g., Li et al. [46]; Xu et al. [47]). Perhaps in future studies, we can further refine our technique by combining the NSMI with OSC and GLSW algorithms to investigate the potential in the removal of SM.

## 5. Conclusions

The results derived from our study clearly demonstrated the need to interpret the influence of variable soil moisture (SM) on the prediction of soil organic matter (SOM) via VIS–NIR spectroscopy. Variable SM led to reduced VIS–NIR reflectance nonlinearly across the entire spectral range. When fuzzy k-mean clustering was applied to partition the calibration dataset into four spectrally similar clusters (SM-based cluster method), the model accuracies in all clusters improved compared with the whole calibration dataset ( $n = 640$ , global calibration). This indicated that the non-linear effect of SM was reduced through clustering. The normalized soil moisture index (NSMI) had a strong correlation with SM across all SM levels (validation  $R^2$  was 0.8824). The SOM estimation based on the NSMI-based cluster method with unknown SM presented comparable modeling accuracies compared to the ones estimated by SM-based cluster method with known SM. Good or very good model predictions for SOM (RPD = 1.87–2.06) were obtained using the NSMI-based cluster method. Moreover, the NSMI-based cluster method is easy to carry out since it only considers the spectral information, which might facilitate the prediction of SOM in the field, without explicit knowledge of SM. Because in the field, different well-defined SM levels are difficult to obtain, the present study was conducted in a controlled laboratory environment. In further studies, the effects of soil surface roughness, vegetation cover and the composition of soil (clay, sand) on reflectance spectra need to be taken into account. In addition, future studies should be encouraged to explore the potential of the NSMI-based cluster method associated with more advanced modeling technologies in other study areas.

**Acknowledgments:** This research was financially supported by the National Natural Science Foundation of China (41771440 and 41501444).

**Author Contributions:** Yongsheng Hong, Lei Yu and Yiyun Chen conceived and designed the research. Yongsheng Hong performed all the modelling. Yi Liu and Hang Cheng performed the experiments. Yaolin Liu and Yanfang Liu participated in the data analyses. Yongsheng Hong and Yiyun Chen were involved in drafting and revising the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Powlson, D.S.; Brookes, P.C.; Whitmore, A.P.; Goulding, K.W.T.; Hopkins, D.W. Soil organic matters. *Eur. J. Soil Sci.* **2011**, *62*, 1–4. [[CrossRef](#)]
2. Stenberg, B. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on VIS–NIR predictions of clay and soil organic carbon. *Geoderma* **2010**, *158*, 15–22. [[CrossRef](#)]
3. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Dematte, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. A global spectral library to characterize the world's soil. *Earth-Sci. Rev.* **2016**, *155*, 198–230. [[CrossRef](#)]

4. Bao, N.; Wu, L.; Ye, B.; Yang, K.; Zhou, W. Assessing soil organic matter of reclaimed soil from a large surface coal mine using a field spectroradiometer in laboratory. *Geoderma* **2017**, *288*, 47–55. [[CrossRef](#)]
5. Viscarra Rossel, R.A.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [[CrossRef](#)]
6. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Chapter five-visible and near infrared spectroscopy in soil science. *Adv. Agron.* **2010**, *107*, 163–215.
7. Cambou, A.; Cardinael, R.; Kouakoua, E.; Villeneuve, M.; Durand, C.; Barthès, B.G. Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VIS–NIR) in the field. *Geoderma* **2016**, *261*, 151–159. [[CrossRef](#)]
8. Wang, C.; Pan, X. Improving the prediction of soil organic matter using visible and near infrared spectroscopy on moist samples. *J. Near Infrared Spectrosc.* **2016**, *24*, 231–241. [[CrossRef](#)]
9. Minasny, B.; McBratney, A.B.; Bellon-Maurel, V.; Roger, J.-M.; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* **2011**, *167–168*, 118–124. [[CrossRef](#)]
10. Lobell, D.B.; Asner, G.P. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722–727. [[CrossRef](#)]
11. Ji, W.; Viscarra Rossel, R.A.; Shi, Z. Accounting for the effects of water and the environment on proximally sensed VIS–NIR soil spectra and their calibrations. *Eur. J. Soil Sci.* **2015**, *66*, 555–565. [[CrossRef](#)]
12. Noda, I. Generalized two-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy. *Appl. Spectrosc.* **1993**, *47*, 1329–1336. [[CrossRef](#)]
13. Ge, Y.F.; Morgan, C.L.S.; Ackerson, J.P. VIS–NIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma* **2014**, *221*, 61–69. [[CrossRef](#)]
14. Ackerson, J.P.; Dematte, J.A.M.; Morgan, C.L.S. Predicting clay content on field-moist intact tropical soils using a dried, ground VIS–NIR library with external parameter orthogonalization. *Geoderma* **2015**, *259*, 196–204. [[CrossRef](#)]
15. Wijewardane, N.K.; Ge, Y.F.; Morgan, C.L.S. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma* **2016**, *267*, 92–101. [[CrossRef](#)]
16. Roudier, P.; Hedley, C.B.; Lobsey, C.R.; Rossel, R.A.V.; Leroux, C. Evaluation of two methods to eliminate the effect of water from soil VIS–NIR spectra for predictions of organic carbon. *Geoderma* **2017**, *296*, 98–107. [[CrossRef](#)]
17. Ji, W.; Viscarra Rossel, R.A.; Shi, Z. Improved estimates of organic carbon using proximally sensed VIS–NIR spectra corrected by piecewise direct standardization. *Eur. J. Soil Sci.* **2015**, *66*, 670–678. [[CrossRef](#)]
18. Chen, Y.Y.; Qi, K.; Liu, Y.L.; He, J.H.; Jiang, Q.H. Transferability of hyperspectral model for estimating soil organic matter concerned with soil moisture. *Spectrosc. Spect. Anal.* **2015**, *35*, 1705–1708. (In Chinese)
19. Guerrero, C.; Zornoza, R.; Gomez, I.; Mataix-Beneyto, J. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. *Geoderma* **2010**, *158*, 66–77. [[CrossRef](#)]
20. Viscarra Rossel, R.A.; Cattle, S.R.; Ortega, A.; Fouad, Y. In situ measurements of soil color, mineral composition and clay content by VIS–NIR spectroscopy. *Geoderma* **2009**, *150*, 253–266. [[CrossRef](#)]
21. Wu, C.Y.; Jacobson, A.R.; Laba, M.; Baveye, P.C. Alleviating moisture content effects on the visible near-infrared diffuse-reflectance sensing of soils. *Soil Sci.* **2009**, *174*, 456–465. [[CrossRef](#)]
22. Wijewardane, N.K.; Ge, Y.; Morgan, C.L.S. Prediction of soil organic and inorganic carbon at different moisture contents with dry ground VNIR: A comparative study of different approaches. *Eur. J. Soil Sci.* **2016**, *67*, 605–615. [[CrossRef](#)]
23. Jiang, Q.H.; Chen, Y.Y.; Guo, L.; Fei, T.; Qi, K. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS–NIR spectroscopy. *Remote Sens.* **2016**, *8*, 755. [[CrossRef](#)]
24. Liu, Y.L.; Jiang, Q.H.; Shi, T.Z.; Fei, T.; Wang, J.J.; Liu, G.L.; Chen, Y.Y. Prediction of total nitrogen in cropland soil at different levels of soil moisture with VIS/NIR spectroscopy. *Acta Agric. Scand. Sect. B-Soil Plant Sci.* **2014**, *64*, 267–281. [[CrossRef](#)]
25. Mouazen, A.M.; Karoui, R.; De Baerdemaeker, J.; Ramon, H. Characterization of soil water content using measured visible and near infrared spectra. *Soil Sci. Soc. Am. J.* **2006**, *70*, 1295–1302. [[CrossRef](#)]
26. Nocita, M.; Stevens, A.; Noon, C.; van Wesemael, B. Prediction of soil organic carbon for different levels of soil moisture using VIS–NIR spectroscopy. *Geoderma* **2013**, *199*, 37–42. [[CrossRef](#)]
27. Wang, D.C.; Zhang, G.L.; Rossiter, D.G.; Zhang, J.H. The prediction of soil texture from visible-near-infrared spectra under varying moisture conditions. *Soil Sci. Soc. Am. J.* **2016**, *80*, 420–427. [[CrossRef](#)]

28. Shi, Z.; Wang, Q.L.; Peng, J.; Ji, W.J.; Liu, H.J.; Li, X.; Viscarra Rossel, R.A. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Sci. China-Earth Sci.* **2014**, *57*, 1671–1680. [CrossRef]
29. Fajardo, M.; McBratney, A.; Whelan, B. Fuzzy clustering of VIS–NIR spectra for the objective recognition of soil morphological horizons in soil profiles. *Geoderma* **2016**, *263*, 244–253. [CrossRef]
30. Walkley, A.; Black, I.A. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [CrossRef]
31. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
32. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]
33. He, A.Q.; Zeng, X.Z.; Xu, Y.Z.; Noda, I.; Ozaki, Y.; Wu, J.G. Investigation on the behavior of noise in asynchronous spectra in generalized two-dimensional (2D) correlation spectroscopy and application of butterworth filter in the improvement of signal-to-noise ratio of 2D asynchronous spectra. *J. Phys. Chem. A* **2017**, *121*, 7524–7533. [CrossRef] [PubMed]
34. Martens, H.; Næs, T. *Multivariate Calibration*; John Wiley & Sons: Chichester, UK, 1989; p. 39.
35. Minasny, B.; McBratney, A. *Fuzme Version 3.0*; Australian Centre for Precision Agriculture, The University of Sydney: Camperdown, Australia, 2002. Available online: <http://www.usyd.edu.au/su/agric/acpa> (accessed on 28 August 2017).
36. Haubrock, S.N.; Chabrillat, S.; Lemmertz, C.; Kaufmann, H. Surface soil moisture quantification models from reflectance data under field conditions. *Int. J. Remote Sens.* **2008**, *29*, 3–29. [CrossRef]
37. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*. [CrossRef]
38. Peng, X.; Shi, T.; Song, A.; Chen, Y.; Gao, W. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* **2014**, *6*, 2699–2717. [CrossRef]
39. Terhoeven-Urselmans, T.; Vagen, T.-G.; Spaargaren, O.; Shepherd, K.D. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Sci. Soc. Am. J.* **2010**, *74*, 1792–1799. [CrossRef]
40. Chang, C.W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [CrossRef]
41. Knadel, M.; Thomsen, A.; Schelde, K.; Greve, M.H. Soil organic carbon and particle sizes mapping using VIS–NIR, EC and temperature mobile sensor platform. *Comput. Electron. Agric.* **2015**, *114*, 134–144. [CrossRef]
42. Vasques, G.M.; Grunwald, S.; Harris, W.G. Spectroscopic models of soil organic carbon in florida, USA. *J. Environ. Qual.* **2010**, *39*, 923–934. [CrossRef] [PubMed]
43. Rodionov, A.; Pätzold, S.; Welp, G.; Damerow, L.; Amelung, W. Sensing of soil organic carbon using visible and near-infrared spectroscopy at variable moisture and surface roughness. *Soil Sci. Soc. Am. J.* **2014**, *78*, 949–957. [CrossRef]
44. Wang, C.K.; Pan, X.Z.; Wang, M.; Liu, Y.; Li, Y.L.; Xie, X.L.; Zhou, R.; Shi, R.J. Prediction of soil organic matter content under moist conditions using VIS–NIR diffuse reflectance spectroscopy. *Soil Sci.* **2013**, *178*, 189–193. [CrossRef]
45. Castaldi, F.; Palombo, A.; Pascucci, S.; Pignatti, S.; Santini, F.; Casa, R. Reducing the influence of soil moisture on the estimation of clay from hyperspectral data: A case study using simulated PRISMA data. *Remote Sens.* **2015**, *7*, 15561–15582. [CrossRef]
46. Li, S.; Shi, Z.; Chen, S.; Ji, W.; Zhou, L.; Yu, W.; Webster, R. In situ measurements of organic carbon in soil profiles using VIS–NIR spectroscopy on the Qinghai–Tibet plateau. *Environ. Sci. Technol.* **2015**, *49*, 4980–4987. [CrossRef] [PubMed]
47. Xu, S.; Zhao, Y.; Wang, M.; Shi, X. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by VIS–NIR spectroscopy. *Geoderma* **2018**, *310*, 29–43. [CrossRef]

