electronics

MDPI

Article

Estimation of Total Nitrogen Content in Rubber Plantation Soil Based on Hyperspectral and Fractional Order Derivative

Rongnian Tang, Xiaowei Li, Chuang Li, Kaixuan Jiang, Wenfeng Hu and Jingjin Wu *

College of Mechanical and Electrical Engineering, Hainan University, Haikou 570228, China; rn.tang@hainanu.edu.cn (R.T.); lxwwsh@126.com (X.L.); lc@hainanu.edu.cn (C.L.); jiangkaixuanl@sina.cn (K.J.); 990531@hainanu.edu.cn (W.H.)

* Correspondence: jingjin.wu@hainanu.edu.cn

Abstract: Soil total nitrogen (TN) is a vital nutrient element that affects the growth and rubber production of rubber trees. Especially in the coastal environment, soil nutrients will show significant differences. Using hyperspectral technology to detect soil nitrogen ion content in the offshore environment can provide technical support for nutrient management. Preprocessing hyperspectral data is a crucial step in accurate spectral model estimation. At the same time, it is considered that the traditional first-order and second-order derivatives are easily unbalanced between the signal-to-noise ratio, resulting in the loss of adequate information. Therefore, this work focuses on the feasibility of fractional order derivative (FOD) combined with partial least squares regression (PLSR) to estimate its TN content. By collecting soil samples from rubber plantations, the TN content of the soil samples was determined, and the spectral reflectance was measured. The FOD of the original spectrum was preprocessed with an interval of 0.2, and 11 spectral curves were obtained. Then, successive projections algorithm (SPA) was used to extract spectral features, and partial least squares regression (PLSR) models of soil TN content were established. The research results show that compared with the traditional integer derivative, FOD has a tremendous advantage in balancing spectral information and noise and can provide more abundant characteristic variables, which helps establish a more robust estimation model. In the range of orders 0–2, the model established by the 1.8-order is the best. Under that circumstance, the determination coefficients of validation (R^2v) is 0.649, and the ratio of the performance to deviation (RPD) is 1.72. Combined with FOD, it is feasible and practical to establish an accurate and rapid estimation model of soil TN content, which can provide an important reference for large-scale detection of soil TN content in rubber plantations.

Keywords: hyperspectral; nitrogen detection; machine learning; fractional order derivative; spectral sensors

Citation: Tang, R.; Li, X.; Li, C.; Jiang, K.; Hu, W.; Wu, J. Estimation of Total Nitrogen Content in Rubber Plantation Soil Based on Hyperspectral and Fractional Order Derivative. *Electronics* **2022**, *11*, 1956. https://doi.org/10.3390/electronics11131956

Academic Editor: Kenji Suzuki

Received: 12 May 2022
Accepted: 19 June 2022
Published: 22 June 2022

1. Introduction

Soil nutrient content significantly affects the growth of rubber trees [1–3]. Compared with other rubber-producing areas, Hainan Province is surrounded by the sea. The erosion of the marine environment and the high temperature and rainy climate have caused soil weathering, resulting in the decline in soil fertility in Hainan rubber plantations, especially the severe loss of nitrogen nutrients. At the same time, with the growth of rubber trees year by year, the TN content in the soil will show a downward trend [4], which requires timely supplementary nutrients. In order to better manage the nutrients in rubber plantations and ensure the increase in rubber production and quality, it is necessary to regularly detect the TN content of the soil. The Kjeldahl method is used to determine the TN content in the soil nutrient detection of rubber plantations [5,6]. The detection method, based on the Kjeldahl method of determining TN content, involves measuring several points in the area and taking the average value to represent the nutrient content of the entire area. Due to the considerable variation in soil nutrients in the offshore environment, this

method cannot accurately represent the TN content of the soil. Moreover, this method is time-consuming, labor-intensive, costly, and unsuitable for batch use. Therefore, it is crucial to find an efficient and low-cost detection method for the scientific decision making of soil nutrient management.

Hyperspectral diagnosis technology is a rapid non-destructive detection technology that combines traditional nutrient diagnosis methods and artificial intelligence scientific decision making [7,8]. In recent years, hyperspectral has also made remarkable achievements in soil element inversion [9–11]. Lin L et al. improved the random forest method by mixing-based weight learning, which significantly improved the estimation of soil TN [12]. Tan K et al. estimated the distribution trend of soil heavy metals in mining areas based on hyperspectral images based on ensemble learning [13]. Through intensive sampling, a higher-precision estimation model can be established. Thereby, hyperspectral detection technology can measure soil nutrient content quickly and in large quantities, which further improves fertilization decision making.

Establishing a hyperspectral rubber plantation soil TN content model can accurately, quickly, and cost-effectively predict the soil TN content. However, when acquiring spectral data, the final quantitative analysis is often affected by the complexity of the sample itself, the untreated sample, or the interference of the instrument itself [14]. The research results show that spectral data preprocessing can effectively reduce or remove these interferences, thereby improving the quantitative analysis and prediction ability of the model [15,16]. Specifically, integer-order derivatives are often used in hyperspectral data processing. However, some researchers have shown that fractional order derivatives (FOD) have more potential than integer derivatives [17,18]. By exploring the potential of FOD for the prediction of soil copper content, Cui S et al. proved that the FOD is better than the integer-order, as it can effectively eliminate noise and highlight the spectral sensitivity of copper [19]. Lao C et al. used FOD with an interval of 0.05 and a range of 0–2 to pretreat soil spectra to predict soil salinity and soluble ion content. The results showed that the optimal FOD orders of different ions were different [20]. In order to estimate soil organic matter content, Hong Y et al. processed the VIS-NIR spectral data FOD, and the results showed that the PLS-SVM with 1.25-order was the best model [21]. However, even now, the application of hyperspectral-based techniques for assessing TN content in rubber plantation soils in offshore environments is still unknown. At the same time, there are few studies on the soil TN content model using FOD. Therefore, it is exceptionally urgent to study the TN content of rubber plantation soils in the offshore environment using a hyperspectral prediction model combined with FOD.

In this study, we tried to study Hainan rubber plantation soil with substantial soil variability in the offshore environment, evaluate the applicability of hyperspectral in the diagnosis of nutrient components in rubber plantations, and provide a scientific basis for fertilization decision making. On this basis, the FOD was introduced to study the optimal estimation model of hyperspectral soil TN based on FOD, which further improved the estimation ability of the model.

2. Materials and Methods

2.1. Study Area, Soil Sampling and Preparation

The soil sample collection sites, rubber plantations, are located in Danzhou City, a coastal city in the northwest of Hainan Province. Danzhou has a tropical monsoon climate and is close to the sea. High temperature: the average annual temperature is 23.1 °C, the average temperature of the hottest month (July) is 28.0 °C, and the coldest month (January) is 17.8 °C. Rainy: Danzhou has an annual rainfall of 900–2200 mm, with an annual average of 1815 mm. The dry and wet seasons are apparent, and the annual average sunshine hours are 1967 h. The soil of the sample collection area in this study is latosol soil, and its parent materials are mainly sand shale, basalt, and granite. The latosol soil presents acidic, which is a suitable condition for rubber trees to grow.

Figure 1 shows the location of the study area and the spatial pattern of soil sampling in Hainan province, Southern China. As shown in the sampling spatial distribution map, the red part is the latosol soil area, and we collected samples around this area. From March to April 2021, we collected 105 soil samples within the study area. Each sample was collected under the following collection principles. First, we avoided the fertilization ditch, the compost area and the ridge, which were not suitable for collecting samples. When sampling, the weeds and other coverings from the surface were removed. Then, we used the earth borrower to borrow soil, with a borrowing depth of 0–20 cm. We then collected five sub-samples in a 100 m^2 area, mixed the five sub-samples to form a regional mixed soil sample (about 2.5 kg), and put them in a white transparent sealed bag. At the same time, the longitude and latitude coordinates of the sampling site were recorded synchronously by hand-held GPS and were pasted on the sample bag. Finally, according to the longitude and latitude information of the collection point, the site locations were drawn on the topographic map of Danzhou City using ArcGIS software to form a sampling spatial distribution map. The collected samples were shipped back to the laboratory on the same day and processed. Sample processing was performed in the following order. First, the soil samples were air-dried at room temperature and some debris was removed. After the samples were air-dried, 105 latosol soil samples were passed through a sieve with a size of 2 mm by a vibrating screen machine. Finally, a small amount of soil was taken using the quartering method. The TN content was determined by the Kjeldahl method. The rest of the soil was used as the sample to be tested for spectral by the hyperspectral method [12].

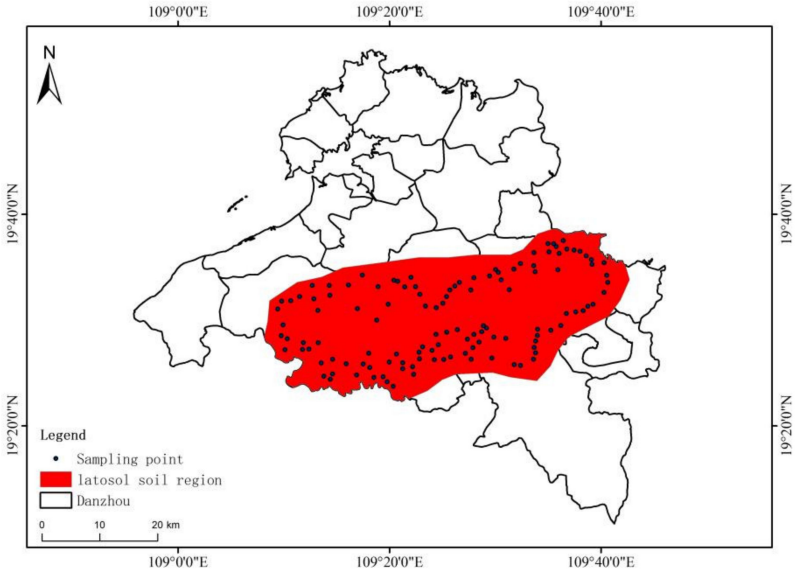

Figure 1. Location of the study area and the spatial pattern of soil sampling in Hainan province, Southern China.

2.2. Laboratory Spectral Measurements and Preprocessing

The soil samples were collected by a hyperspectral imaging system with a spectral reflectance of 866–1701 nm. The hyperspectral imaging system consists of an imaging spectrograph (ImSpector N17E, Specim, Finland) with a spectral resolution of 4 nm, an InGaAs camera with a resolution of 320 × 256, 256 wavelength channels and a camera lens (HSIA-OLE22, Finland), and was provided by Sichuan Shuangli Hepu Technology Co. Ltd., China. After the tested sample was placed in the sample stage, four halogen lamps irradiated the sample. The four 200 W halogen lamps were placed 0.8 m from the soil sample and irradiated the sample at a 45° angle of incidence. Before acquiring hyperspectral data, it was required to perform an initial calibration of the instrument, adjust the height of the lens, and correct the position of the sample so that was able to take the whole soil sample and take a clear spectral image. After the sample was irradiated, the

imaging lens captured the reflected light, and a picture (connotative spatial information and spectral information) was obtained. As the conveyor belt drives the sample continuously, continuous pictures and three-dimensional cube data (spatial information and spectral information) can be obtained.

The steps of soil spectrum collection are as follows. First, the sieved soil samples were placed in black Petri dishes (4.5 cm in radius and 1.5 cm in height). Then, the surface of the sample was smoothed by gently scraping the surface of the Petri dish through a piece of cardboard. Next, the Petri dish was placed on black cardboard with approximately zero reflectivity, and the instrument collected soil spectral data through line scanning. The acquired spectral data are hyperspectral images of all objects within a rectangular area (including soil and non-soil areas). Finally, the hyperspectral image was corrected for black and white, and the average spectrum of the soil sample was calculated. Each soil sample to be tested was measured three times according to the above steps, and the average value of the three values was taken.

When collecting spectral data, we found that the collected spectral curves had intense noise at the beginning and end, mainly due to the influence of the external environment and the instrument itself [22,23]. In order to eliminate the noise caused by the error of the instrument itself, the hyperspectral bands with higher noise were removed (866–978 nm, 1671–1701 nm), and the final hyperspectral data band range was 979–1670 nm.

2.3. Data Division

As shown in Table 1, among the 105 soil samples, outliers and duplicates were removed. By removing the outliers, 91 soil samples were retained. This study used the concentration gradient method to divide the dataset. The ratio of the calibration set to the validation set was 2:1. Among the 91 soil samples in the total sample, 61 soil samples were used as the calibration set, and the remaining 30 soil samples were used for the validation calibration model. The maximum value of TN content of the soil samples was 0.8885, the minimum value was 0.3064, the mean value was 0.5631, the standard deviation was 0.1558, and the coefficient of variation was 28.4%. The coefficients of variation in the calibration set and validation set were 27.87% and 27.75%, respectively. From the coefficient of variation, it can be seen that the calibration set and the validation set have almost the same distribution characteristics as the total sample. Such a divided sample set makes the establishment and evaluation of the model more reasonable and ensures the generalization of the estimated model.

Table 1. Descriptive statistics of total content (TN, g/kg) in the entire set.

Sample Set	n	Minimum	Mean	Median	Maximum	SD	CV
Total	91	0.3064	0.5631	0.5487	0.8885	0.1558	28.4%
Calibration	61	0.3064	0.5613	0.5487	0.8885	0.1564	27.87%
Validation	30	0.3191	0.5666	0.5449	0.8852	0.1572	27.75%

Notes: n, sample number; SD, standard deviation; CV, coefficient of variation.

2.4. Fractional Order Derivative (FOD)

The fractional order derivative (FOD) extends the traditional integer-order derivatives (first-order, second-order) to any order, and different orders of derivatives can be obtained through different intervals. Among many FOD definitions, the Grünwald-Letnikov (G-L) definition is widely used in practical applications because it is easier to obtain numerical solutions [24–26]. Therefore, this study is based on the G-L definition of conducting research analyses. Taking the TN spectral data of rubber plantation-soil as the FOD object, we set the step size to 1. After processing and calculation in MATLAB, 11 fractional order derivatives of soil spectral data, ranging from 0 to 2, with 0.2 as an interval, are obtained.

2.5. Spectral Variable Selection Techniques

Spectral data are characterized by high dimensionality and redundancy. When applied to the quantitative analysis of soil nutrient content, due to the problems of high correlation and much useless information, the soil nutrient diagnostic model constructed based on the full spectral data has low predictive ability and poor robustness. In order to find the profound relationship between spectral data and samples, eliminate useless information, and highlight characteristic bands, wavelength extraction has become a vital issue in nutrient diagnostic models.

The successive projections algorithm (SPA) method is a widely used method. The SPA can reduce the model complexity by effectively eliminating the collinearity between multiple wavelength variables and redundant information, and choosing the feature band [27,28]. The variable selection principle of the SPA method is positive selection, the physical meaning of the selected wavelength is clear, and the model has strong explanatory power. The working principle of the SPA algorithm is divided into the following steps: Assuming that the independent variable of the spectral matrix is X, the number of variables to be extracted is N, and a variable Xi is randomly selected from X as the first variable. Calculating the projection variables of Xi to the remaining spectral variables and recording the most significant projection variables, each iteration adds new variables and stops working when the maximum number of variables is reached. Therefore, the common SPA algorithm is used to extract its features in this study. In this study, the maximum number of variables N was set to 45 and the minimum to 5.

2.6. Partial Least Squares Regression (PLSR)

As one of the most widely used regression modeling methods in spectral data analysis, the partial least squares regression (PLSR) can efficiently and reliably process complex spectral data, solve collinearity problems, and improve the explanatory ability of dependent variables. It belongs to the linear regression model [29]. The basic idea of PLSR assumes that the independent variable of spectral data is X, and the dependent variable is Y. By transforming the matrix data of the independent variable X into the latent variable (LV), the latent variable can well represent the independent variable X (including more information of the X spectral data), the calculated LV is independent and between Y has the most remarkable correlation. In this study, by setting the maximum value of LV, the calculation cycle is performed, and then the optimal number of LVs to meet the accuracy is selected.

2.7. Model Evaluation

Among many model evaluation indicators, the coefficient of determination (R^2) (Equation (1)), root mean square error (RMSE) (Equation (2)), and the ratio of performance to deviation (RPD) (Equation (3)) are the most widely used. Therefore, in this study, the above three indicators were used as the evaluation indicators of the soil TN model. The criteria for a model with better model performance are higher RPD and R^2 values and lower RMSE values [30,31]. Data preprocessing, feature extraction, and model construction were performed in MATLAB 2018b (the MathWorks, Inc., Natick, USA).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$RPD = \frac{SD}{RMSE} \quad (3)$$

where n represents the number of soil samples, y_i, $\hat{y}_i$ represents measured and predicted values of TN content, $\bar{y}_i$ is the measured average value, and SD is the standard deviation, respectively.

3. Results

3.1. Soil Spectral Reflectance

Figure 2 is the spectral curve of rubber plantation soil samples in Danzhou City, Hainan. We can see from it that the overall trend of the spectrum of the 91 latosol soil samples is consistent, but the spectral reflectance is different due to the difference in soil composition content. In the range of 979–1300 mm, the spectral reflectance increases with wavelength increase, and a small peak appears around 1305 nm. In the range of 1305–1449 nm, the spectral reflectance gradually decreased, and in the range of 1449–1670 nm, the spectral reflectance gradually increased. The soil spectral curve has an apparent absorption valley at 1449 nm. Studies have shown that this is related to free water hydroxide ions (OH) in soil and the presence of Al-OH lattice structures in clay minerals [32].

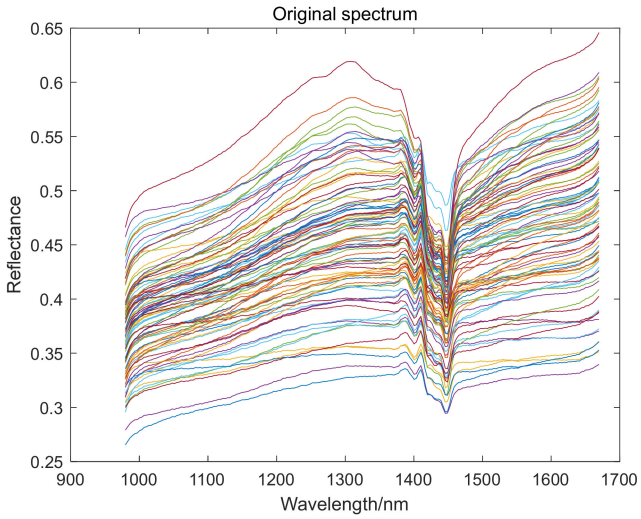

Figure 2. Spectral curve of rubber plantation soil samples in Danzhou City, Hainan.

3.2. FOD Spectral Analysis

To better remove baselines and overlapping peaks, sharpen spectral features, and dig deeper into spectral information, we used FOD to preprocess the spectra. Figure 3 shows the spectral curves of rubber plantation soils under different FOD. With 0.2 as the interval, 11 orders of spectral curves are obtained in 0–2. Figure 3 shows the original spectral curve with 0 order. When the order increases from 0 to 0.2, it can be seen that the reflectivity does not change much, and the peak-valley separation begins to appear around 1449 nm. When the order increases from 0.4 to 1, the peaks and valleys around 1449 nm are further separated. There are four absorption valleys at 1394, 1417, 1430, and 1446 nm, respectively. Four reflection peaks appear at 1407, 1423, 1436, and 1456 nm, respectively. The reflectance value generally tends to be 0. As the order increases from 1 to 2, the spectral curve is smoothed, and the reflectance value decreases to 0. This indicates that the baseline shift effect and overlapping peaks are gradually eliminated as the fractional order increases. Compared to the original spectrum, the FOD spectrum can slowly show changes in spectral detail and improve the resolution of the spectral curve.

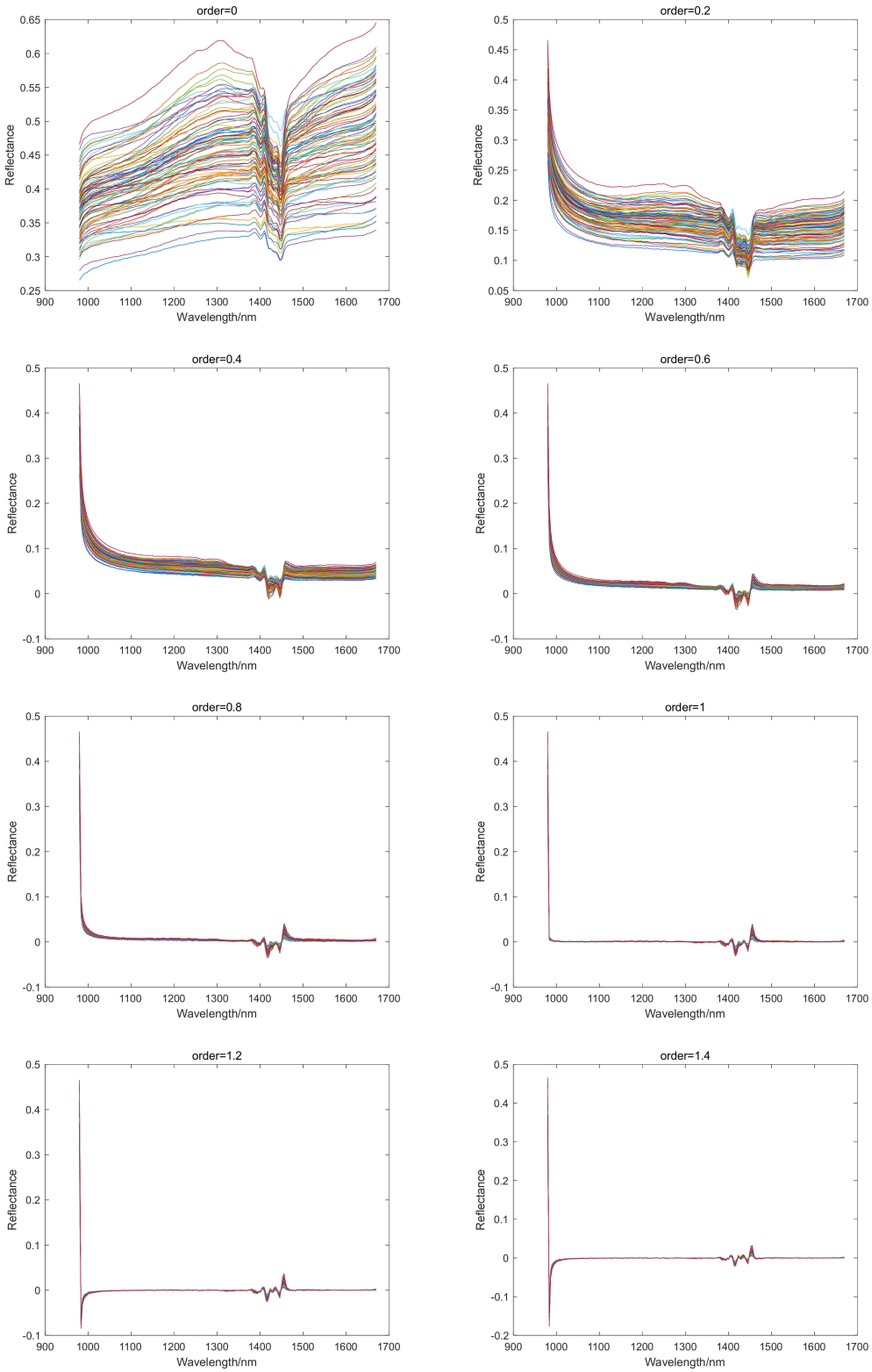

Figure 3. *Cont.*

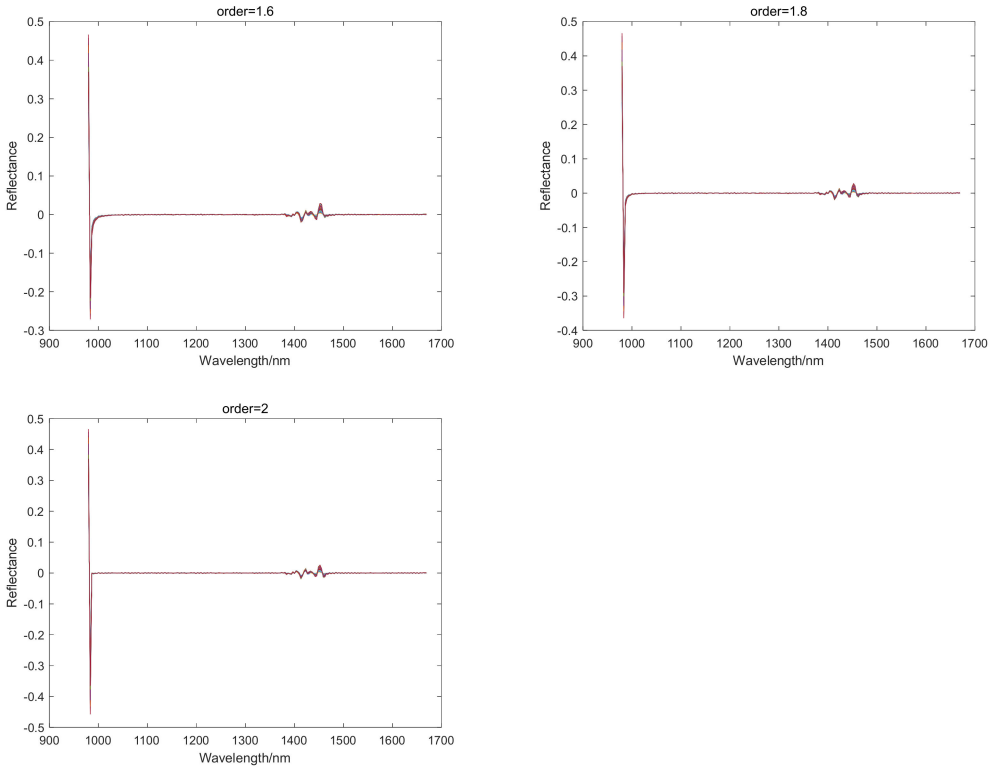

Figure 3. Spectral curves of rubber plantation soils under different FOD.

3.3. Feature Variable Selection and Building Estimation Models

To reduce the redundant information of the spectral data and eliminate the collinearity between the wavelength variables, the SPA algorithm is used to extract features from the 11 fractional-orders of soil spectral data with a range of 0–2 and an interval of 0.2. Figure 4 shows the feature variables selected by SPA under different FOD. It can be seen from Figure 4 that the SPA algorithm extracts the 11 fractional-orders soil spectral data, the number of selected variables is significantly reduced, and the number of bands is reduced from 211 to seven variables, 3% of the total bands. This process reduces the dimension of data and improves work efficiency. Comparing the selected characteristic variables of different orders, it can be seen that the 0.6-order selects the most bands, 30 bands. Additionally, the 1.2-order selects the least, seven bands. The characteristic variables of the 11 fractional orders of soil TN spectrum data extracted by SPA are different. In the 0–1 order range, its characteristic variables are mainly concentrated in the vicinity of 1000 nm and 1400–1500 nm. In contrast, in the 1.2–2 order range, its characteristic variables are mainly concentrated in the vicinity of 1564–1603 nm. The results show that different orders of FOD can separate overlapping peaks and dig out deeper spectral features related to soil TN.

To explore the effect of FOD spectral preprocessing and SPA feature extraction on the prediction model, SPA-PLSR regression prediction models were established. In those models, the spectrum extracted by FOD and SPA is set as the independent variable, and the soil TN content is set as the dependent variable. As shown in Table 2, the modeling results of PLSR and SPA-PLSR are compared. As shown in Table 2, compared with the full spectral data, the SPA-PLSR model has improved the prediction performance of soil TN content. The RPD increased from 1.35 to 1.49 gradually. The results show that the model has developed from almost no estimation capabilities to accurate estimation capabilities.

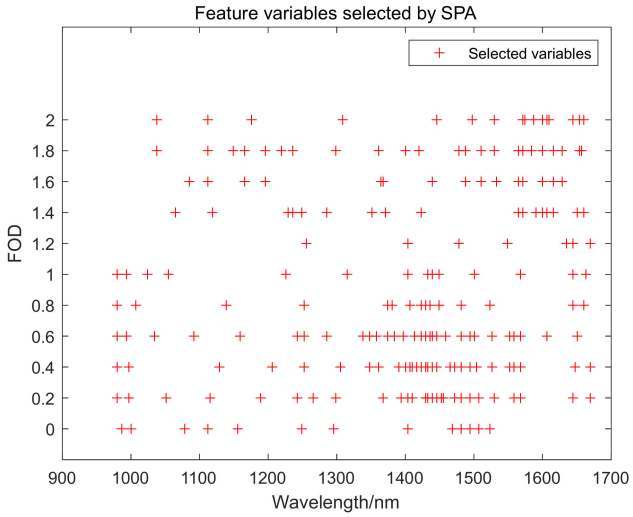

Figure 4. Feature variables selected by SPA under different FOD.

Table 2. Comparison of PLSR and SPA-PLSR modeling results.

Model	Number of Variables	Calibration		Validation		
		R^2c	RMSEC	R^2v	RMSEV	RPD
PLSR	211	0.728	0.081	0.429	0.117	1.35
SPA-PLSR	13	0.387	0.121	0.535	0.105	1.49

Notes: R^2c, determination coefficients of calibration; R^2v, determination coefficients of validation.

Table 3 shows the modeling results of SPA-PLSR under different FOD. It can be seen from Table 3 that the prediction performance of the original, first and second derivative spectral models is not high, with R^2v of 0.535, 0.568, and 0.513; RMSEV of 0.105, 0.102, and 0.108; RPD of 1.49, 1.55 and 1.46, respectively. The RPD of the original and second derivative spectra is less than 1.4, and the model prediction performance is poor. After the FOD processing, the prediction performance of the FOD model is greatly improved, and better results are obtained in the 0.6-order, 1.6-order, and 1.8-order, respectively. The R^2v is 0.633, 0.631 and 0.649, the RMSEV is 0.094, 0.094 and 0.092, and the RPD is 1.68, 1.67 and 1.72, and the best prediction results apparent at 1.8-order. The sensitive bands of soil TN spectrum are 1037.7, 1112.1, 1149.0, 1165.8, 1195.9, 1219.2, 1235.8, 1298.7, 1361.1, 1400.3, 1419.8, 1478.1, 1487.8, 1510.4, 1529.6, 1564.8, 1571.2, 1584.0, 1599.9, 1615.8, 1628.5, 1653.8 and 1656.9 nm. The RPD of the estimated models at 0.2-order, 0.8-order and 1.2-order are all worse than those established by the original spectra. Figure 5 shows the scatter plots of true and predicted TN content for the calibration and validation set of original, first-order, 1.8-order and second-order derivatives. It is demonstrated that the model at 1.8-order can better predict soil TN samples at low and high concentrations compared to the original and integer derivatives. This indicates that the model established after FOD preprocessing can predict samples with a broader range of TN content, and its model is more generalized.

Table 3. SPA-PLSR modeling results under different FOD.

Order	Number of Variables	Calibration		Validation		
		R^2c	RMSEC	R^2v	RMSEV	RPD
0	13	0.387	0.121	0.535	0.105	1.49
0.2	27	0.648	0.092	0.462	0.113	1.39
0.4	29	0.664	0.090	0.584	0.100	1.58
0.6	30	0.731	0.081	0.633	0.094	1.68
0.8	15	0.505	0.109	0.513	0.108	1.46
1.0	14	0.354	0.125	0.568	0.102	1.55
1.2	7	0.346	0.126	0.473	0.112	1.40
1.4	17	0.567	0.102	0.575	0.101	1.56
1.6	15	0.548	0.104	0.631	0.094	1.67
1.8	23	0.612	0.097	0.649	0.092	1.72
2.0	16	0.457	0.114	0.513	0.108	1.46

At the same time, to further compare the preprocessing methods, two classical preprocessing methods, multiplicative scatter correction (MSC) and standard normal variate (SNV), are used to preprocess the spectral data. After SPA feature extraction, a PLSR model is established. The comparison results are shown in Table 4. It can be seen from Tables 3 and 4 that compared with SPA-PLSR (order = 0), the prediction effect of the MSC preprocessed model decreases with $R^2v = 0.5 < 0.535$, and the prediction effect of the SNV preprocessed model improves with $R^2v = 0.569 > 0.535$. All results were worse than that of the FOD preprocessed model. The comparison further proves the superiority of FOD for pretreatment of the rubber plantation soil spectrum.

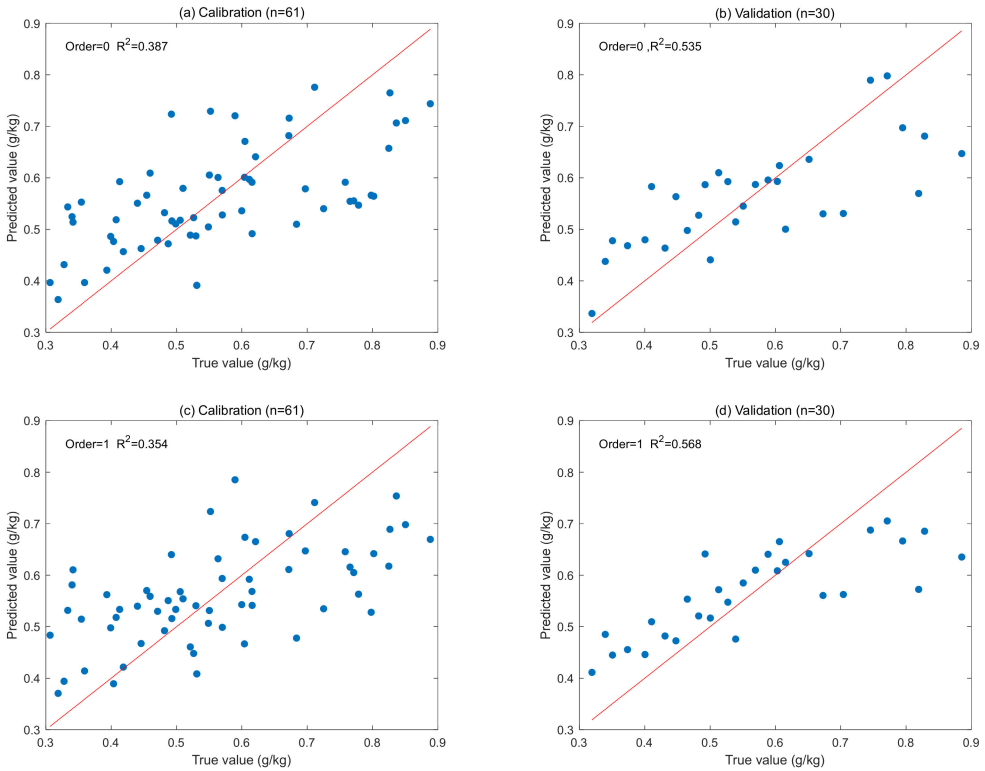

Figure 5. *Cont*.

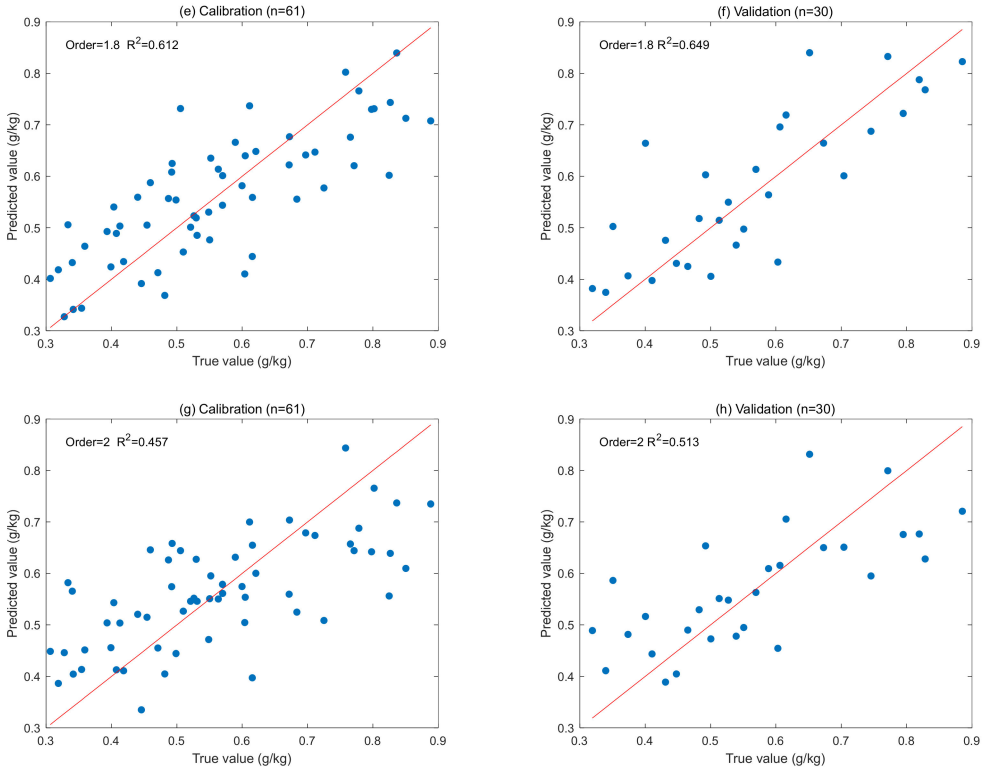

Figure 5. Scatter plots of true and predicted TN content for the calibration and validation set of original, first-order, 1.8-order and second-order derivatives. The red line represents the 1:1 line between the true and predicted TN content. (**a**,**c**,**e**,**g**) are scatter plots of calibration set results; (**b**,**d**), (**f**,**h**) are scatter plots of validation set results.

Table 4. Modeling results under different preprocessing.

Model	Number of Variables	Calibration		Validation		
		R^2c	RMSEC	R^2v	RMSEV	RPD
MSC-SPA-PLSR	17	0.491	0.111	0.500	0.109	1.44
SNV-SPA-PLSR	23	0.631	0.094	0.569	0.102	1.55
1.8-order-SPA-PLSR	23	0.612	0.097	0.649	0.092	1.72

4. Discussion

In this study, hyperspectral detection technology was used to obtain the TN spectrum data of latosol soil in rubber plantations. It is expected that hyperspectral technology can be used to establish a prediction model of TN in latosol soil, providing technical support for guiding the fine fertilization of rubber plantations. The research results show that the prediction effect of the full spectral model is not satisfactory for the diagnosis of TN content in rubber plantations. Therefore, further analysis of the hyperspectral data is required to improve the estimation ability of TN content. To this end, this study uses FOD to preprocess soil spectra to dig deeper into hyperspectral information. SPA is used to extract features from 11-order spectral data to eliminate redundant information, expecting to obtain spectral information that is more sensitive than the original spectrum.

The research results show that compared with the original spectrum-based PLSR model, the SPA-PLSR model established by the sensitive band after SPA feature extraction through FOD preprocessing has a better prediction performance. The modeling effect of the SPA-PLSR model under different FOD preprocessing is different. The difference is due to the changes in spectral information after different FOD processing, resulting in different

characteristic bands selected by SPA in the 11-order soil spectrum. The reason is that SPA selects characteristic wavelengths from the full spectral data to form an optimal subset, which is an optimized process that removes redundant information from spectral data and extracts adequate information. Compared to using full spectral data as the input, SPA reduces the number of bands selected at 11 fractional orders by a minimum of 86% and a maximum of 97%, improving the predictive performance of the model [28,33]. It can be seen from Table 3 that the prediction model of soil TN content established by the SPA method has better performance in reducing redundancy, can better preserve adequate information, and has an excellent response to the inversion parameters. This confirms that the SPA method can effectively extract features from the rubber plantation soil spectral data.

Different orders of FOD impact the performance of soil TN content prediction model. In this study, the interval is set as 0.2, and the 0–2 order range has been divided into 11-order derivatives. The study found that after FOD preprocessing, the soil spectral reflectance decreased and approached 0 gradually. Confidential information, including peaks and valleys, was excavated. This shows that FOD can dig out confidential information and provide richer variables for feature extraction. On the whole, the FOD preprocessed model performs better than the original, first, and second derivative spectra. It was found that the SPA-PLSR model based on the original spectrum (0-order) performed poorly with RPD = 1.49. After the first-order derivative preprocessing, the model slightly improved with RPD = 1.55, which indicates that the first-order derivative improves the modeling accuracy [34]. On the other hand, the prediction performance of the second derivative-based model was lower than that of the original and the first derivative, which may be due to the amplification of spectral data noise or the loss of important spectral information [35]. After the introduction of FOD preprocessing, the prediction accuracy has been dramatically improved. The results in this study were similar to other soil-related content estimation models with FOD preprocessing [19,21]. Chen L et al. confirmed that FOD could better dig out the correlation spectrum of heavy metals in soil than integer derivative, and improve the model estimation ability [36]. The reason may be that although the integer-order (first-order, second-order) spectrum enhances the spectral reflectance, the sensitive information is gradually inclined or lacks sensitivity to the curvature [20], resulting in the loss of spectral information or the amplification of noise. By contrast, FOD can fill up the deficiency. FOD can separate overlapping peaks and baseline effects such as an integer-order. Since it can control the interval, it can affect the change in order and then control the change in slope or curvature. Therefore, the maximum signal-to-noise ratio of the rubber plantation soil spectral data could be found, and the prediction performance of the model could be improved. This study established 11 SPA-PLSR soil TN prediction models, and the 1.8-order-based model showed the best prediction effect. This finding is consistent with the results of Xu X et al. [37] predicting Hg content in soil. It should be noted that some FOD-based models have worse prediction performance than the original and integer-order derivative spectra, including the 0.2-order, 0.8-order, and 1.2-order. The reason may be that the spectral features cannot be amplified during preprocessing due to too large or too small curvature and slope. Even the spectral information related to soil TN content is lost, resulting in poor prediction performance. This also shows that the hyperspectral detection technology combined with FOD to model and analyze the TN spectrum of rubber plantation soil can predict the TN content, which is an effective and practical method. In the future, we will collect more soil samples from different rubber plantation areas by calibrating the model, in the hope of providing a universal model for detecting TN content in rubber plantation soils in different places.

5. Conclusions

This paper discusses the potential of hyperspectral detection technology combined with FOD in detecting TN content in rubber plantation soil. The obtained results show that: (1) FOD can also reduce baseline drift and separate overlapping peaks and is better than the first and second derivatives in finding the maximum signal-to-noise ratio. (2) In

the 0–2 order range, the SPA-PLSR model established with 1.8-order has the best performance in estimating soil TN content, with R^2v of 0.649 and RPD of 1.72. (3) The sensitive bands of the TN spectrum of rubber plantation soil were 1037.7, 1112.1, 1149.0, 1165.8, 1195.9, 1219.2, 1235.8, 1298.7, 1361.1, 1400.3, 1419.8, 1478.1, 1487.8, 1510.4, 1529.6, 1564.8, 1571.2, 1584.0, 1599.9, 1615.8, 1628.5, 1653.8 and 1656.9 nm. In summary, hyperspectral detection technology can predict the TN content of rubber plantations. Combined with the spectral pretreatment of FOD, it can improve the estimation performance of the model and provide an important reference for the large-scale estimation of soil TN content and nutrient management.

Author Contributions: Conceptualization, R.T. and J.W.; methodology, X.L., C.L., K.J. and W.H.; software, X.L., C.L. and K.J.; writing—original draft preparation, R.T., X.L. and J.W.; writing—review and editing, R.T., X.L. and J.W.; supervision, C.L., K.J. and W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Innovative Research Team Project of Hainan Natural Science Found of China] grant number 320CXTD431, [National Natural Science Found of China] grant number 32060413, the scientific research fund of Hainan University (No.kyqd(ZR)1934).

Data Availability Statement: The data are not publicly available due to privacy.

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

References

1. Chen, K.; Li, C.; Tang, R. Estimation of the nitrogen concentration of rubber tree using fractional calculus augmented NIR spectra. *Ind. Crop. Prod.* **2017**, *108*, 831–839. [CrossRef]
2. Tang, R.; Luo, X.; Li, C.; Zhong, S. A study on nitrogen concentration detection model of rubber leaf based on spatial-spectral information with NIR hyperspectral data. *Infrared Phys. Technol.* **2022**, *122*, 104094. [CrossRef]
3. Li, X.; Liu, Z.M.; Ding, H.; Xu, M.; Li, C. Study on estimation models for total nitrogen content of rubber tree leaves based on hyper-spectral data. *Southwest China J. Agric. Sci.* **2015**, *28*, 569–574.
4. Wu, Z.; Xie, G.; Yang, C.; Zhou, Z.; Chen, B. Soil nutrient characteristics of rubber plantations at different age stages in western area of Hainan island. *Nat. Sci. J. Hainan Univ.* **2011**, *29*, 42–48.
5. Nie, P.; Dong, T.; He, Y.; Qu, F. Detection of Soil Nitrogen Using Near Infrared Sensors Based on Soil Pretreatment and Algorithms. *Sensors* **2017**, *17*, 1102. [CrossRef] [PubMed]
6. Zhang, Y.; Li, M.; Zheng, L.; Qin, Q.; Lee, W.S. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma* **2019**, *333*, 23–34. [CrossRef]
7. Qin, J.; Chao, K.; Kim, M.S.; Lu, R.; Burks, T.F. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* **2013**, *118*, 157–171. [CrossRef]
8. Gomez, C.; Rossel, R.A.V.; McBratney, A.B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma* **2008**, *146*, 403–411. [CrossRef]
9. Ou, D.; Tan, K.; Lai, J.; Jia, X.; Wang, X.; Chen, Y.; Li, J. Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction. *Geoderma* **2021**, *385*, 114875. [CrossRef]
10. Shi, Y.; Zhao, J.; Song, X.; Qin, Z.; Wu, L.; Wang, H.; Tang, J. Hyperspectral band selection and modeling of soil organic matter content in a forest using the Ranger algorithm. *PLoS ONE* **2021**, *16*, e0253385. [CrossRef]
11. Yang, C.; Feng, M.; Song, L.; Wang, C.; Yang, W.; Xie, Y.; Jing, B.; Xiao, L.; Zhang, M.; Song, X.; et al. Study on hyperspectral estimation model of soil organic carbon content in the wheat field under different water treatments. *Sci. Rep.* **2021**, *11*, 18582. [CrossRef] [PubMed]
12. Lin, L.; Liu, X. Mixture-based weight learning improves the random forest method for hyperspectral estimation of soil total nitrogen. *Comput. Electron. Agric.* **2022**, *192*, 106634. [CrossRef]
13. Tan, K.; Ma, W.; Chen, L.; Wang, H.; Du, Q.; Du, P.; Yan, B.; Liu, R.; Li, H. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. *J. Hazard. Mater.* **2021**, *401*, 123288. [CrossRef] [PubMed]
14. Duckworth, J. Mathematical data preprocessing. *Near-Infrared Spectrosc. Agric.* **2004**, *44*, 113–132.
15. Diwu, P.; Bian, X.; Wang, Z.; Liu, W. Study on the selection of spectral preprocessing methods. *Spectrosc. Spectr. Anal.* **2019**, *39*, 2800.
16. Dotto, A.C.; Dalmolin, R.S.D.; ten Caten, A.; Grunwald, S. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* **2018**, *314*, 262–274. [CrossRef]
17. Abulaiti, Y.; Sawut, M.; Maimaitiaili, B.; Chunyue, M. A possible fractional order derivative and optimized spectral indices for assessing total nitrogen content in cotton. *Comput. Electron. Agric.* **2020**, *171*, 105275. [CrossRef]

18. Zhang, Z.; Ding, J.; Wang, J.; Ge, X. Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. *CATENA* **2020**, *185*, 104257. [CrossRef]
19. Cui, S.; Zhou, K.; Ding, R.; Cheng, Y.; Jiang, G. Estimation of soil copper content based on fractional-order derivative spectroscopy and spectral characteristic band selection. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *275*, 121190. [CrossRef]
20. Lao, C.; Chen, J.; Zhang, Z.; Chen, Y.; Ma, Y.; Chen, H.; Gu, X.; Ning, J.; Jin, J.; Li, X. Predicting the contents of soil salt and major water-soluble ions with fractional-order derivative spectral indices and variable selection. *Comput. Electron. Agric.* **2021**, *182*, 106031. [CrossRef]
21. Hong, Y.; Chen, Y.; Yu, L.; Liu, Y.; Liu, Y.; Zhang, Y.; Liu, Y.; Cheng, H. Combining Fractional Order Derivative and Spectral Variable Selection for Organic Matter Estimation of Homogeneous Soil Samples by VIS–NIR Spectroscopy. *Remote Sens.* **2018**, *10*, 479. [CrossRef]
22. Yu, H.-D.; Qing, L.-W.; Yan, D.-T.; Xia, G.; Zhang, C.; Yun, Y.-H.; Zhang, W. Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chem.* **2021**, *348*, 129129. [CrossRef]
23. Wang, J.; Wang, T.; Shi, T.; Wu, G.; Skidmore, A.K. A Wavelet-Based Area Parameter for Indirectly Estimating Copper Concentration in Carex Leaves from Canopy Reflectance. *Remote Sens.* **2015**, *7*, 15340–15360. [CrossRef]
24. Scherer, R.; Kalla, S.L.; Tang, Y.; Huang, J. The Grünwald–Letnikov method for fractional differential equations. *Comput. Math. Appl.* **2011**, *62*, 902–917. [CrossRef]
25. Abd Elaziz, M.; Yousri, D.; Al-qaness, M.A.A.; AbdelAty, A.M.; Radwan, A.G.; Ewees, A.A. A Grunwald–Letnikov based Manta ray foraging optimizer for global optimization and image segmentation. *Eng. Appl. Artif. Intell.* **2021**, *98*, 104105. [CrossRef]
26. Varga, D. Full-Reference Image Quality Assessment Based on Grünwald–Letnikov Derivative, Image Gradients, and Visual Saliency. *Electronics* **2022**, *11*, 559. [CrossRef]
27. Khan, I.H.; Liu, H.; Cheng, T.; Tian, Y.; Cao, Q.; Zhu, Y.; Yao, X. Detection of wheat powdery mildew based on hyperspectral reflectance through SPA and PLS-LDA. *Int. J. Precis. Agric. Aviat.* **2018**, *1*, 43–48.
28. Chen, J.; Yu, H.; Jiang, D.; Zhang, Y.; Wang, K. A novel NIRS modelling method with OPLS-SPA and MIX-PLS for timber evaluation. *J. For. Res.* **2021**, *33*, 369–376. [CrossRef]
29. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
30. Xu, S.; Zhao, Y.; Wang, M.; Shi, X. Determination of rice root density from Vis–NIR spectroscopy by support vector machine regression and spectral variable selection techniques. *CATENA* **2017**, *157*, 12–23. [CrossRef]
31. Shi, Z.; Wang, Q.; Peng, J.; Ji, W.; Liu, H.; 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]
32. Xie, S.; Li, Y.; Wang, X.; Liu, Z.; Ma, K.; Ding, L. Research on estimation models of the spectral characteristics of soil organic matter based on the soil particle size. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *260*, 119963. [CrossRef] [PubMed]
33. Sun, J.; Cong, S.; Mao, H.; Wu, X.; Zhang, X.; Wang, P. CARS-ABC-SVR model for predicting leaf moisture of leaf-used lettuce based on hyperspectral. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 178–184.
34. Qiao, X.-X.; Wang, C.; Feng, M.-C.; Yang, W.-D.; Ding, G.-W.; Sun, H.; Liang, Z.-Y.; Shi, C.-C. Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques. *Spectrosc. Lett.* **2017**, *50*, 156–163. [CrossRef]
35. Hong, Y.; Shen, R.; Cheng, H.; Chen, Y.; Zhang, Y.; Liu, Y.; Zhou, M.; Yu, L.; Liu, Y.; Liu, Y. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. *Sci. Total Environ.* **2019**, *651*, 1969–1982. [CrossRef]
36. Chen, L.; Lai, J.; Tan, K.; Wang, X.; Chen, Y.; Ding, J. Development of a soil heavy metal estimation method based on a spectral index: Combining fractional-order derivative pretreatment and the absorption mechanism. *Sci. Total Environ.* **2022**, *813*, 151882. [CrossRef] [PubMed]
37. Xu, X.; Chen, S.; Ren, L.; Han, C.; Lv, D.; Zhang, Y.; Ai, F. Estimation of Heavy Metals in Agricultural Soils Using Vis-NIR Spectroscopy with Fractional-Order Derivative and Generalized Regression Neural Network. *Remote Sens.* **2021**, *13*, 2718. [CrossRef]