

Article

# Hyperspectral-Based Estimation of Leaf Nitrogen Content in Corn Using Optimal Selection of Multiple Spectral Variables

Lingling Fan <sup>1,2</sup>, Jinling Zhao <sup>1,\*</sup> , Xingang Xu <sup>2,\*</sup>, Dong Liang <sup>1</sup>, Guijun Yang <sup>2</sup>, Haikuan Feng <sup>2</sup>, Hao Yang <sup>2</sup>, Yulong Wang <sup>1,2</sup>, Guo Chen <sup>2</sup> and Pengfei Wei <sup>2</sup>

<sup>1</sup> National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230601, China

<sup>2</sup> Beijing Engineering Research Center of Agricultural Internet of Things, Beijing 100097, China

\* Correspondence: zhaojl@ahu.edu.cn (J.Z.); xxg2007@aliyun.com (X.X.);

Tel.: +(0551)-6295-0280 (J.Z.); +86-010-5150-3676 (X.X.)

Received: 19 May 2019; Accepted: 19 June 2019; Published: 30 June 2019



**Abstract:** Accurate and dynamic monitoring of crop nitrogen status is the basis of scientific decisions regarding fertilization. In this study, we compared and analyzed three types of spectral variables: Sensitive spectral bands, the position of spectral features, and typical hyperspectral vegetation indices. First, the Savitzky-Golay technique was used to smooth the original spectrum, following which three types of spectral parameters describing crop spectral characteristics were extracted. Next, the successive projections algorithm (SPA) was adopted to screen out the sensitive variable set from each type of parameters. Finally, partial least squares (PLS) regression and random forest (RF) algorithms were used to comprehensively compare and analyze the performance of different types of spectral variables for estimating corn leaf nitrogen content (LNC). The results show that the integrated variable set composed of the optimal ones screened by SPA from three types of variables had the best performance for LNC estimation by the validation data set, with the values of  $R^2$ , root means square error (RMSE), and normalized root mean square error (NRMSE) of 0.77, 0.31, and 17.1%, and 0.55, 0.43, and 23.9% from PLS and RF, respectively. It indicates that the PLS model with optimally multitype spectral variables can provide better fits and be a more effective tool for evaluating corn LNC.

**Keywords:** hyperspectral; leaf nitrogen content (LNC); successive projections algorithm (SPA); partial least squares (PLS) model; random forest (RF) model

## 1. Introduction

Hyperspectral data with high spectral resolution could reveal small changes in the biochemical components of plant leaves, and its acquisition was rapid and non-destructive. In fact, rapid, non-destructive monitoring of plant leaf biochemical components by hyperspectral means has now become an important part of the evaluation of vegetative growth status. However, although hyperspectral data with hundreds or even thousands of bands could provide more detailed and richer spectral information than multispectral data, it suffered from significant data redundancy, high correlation between adjacent bands, etc. [1]. Therefore, more research was required to determine how best to extract from hyperspectral data the characteristic spectral variables to effectively monitor the biochemical components of crop targets.

At present, spectral variables based on extracted hyperspectral data might be divided into three categories generally:

(1) Characteristic reflected bands. Hyperspectral data offer more accurate bands, which can better reflect the characteristics of vegetation. Bai et al. [2] applied the successive projections to extract eight

sensitive bands correlated with total nitrogen content of winter wheat leaves (1985, 2474, 1751, 1916, 2507, 1955, 2465, and 344 nm). These bands had an extremely negative correlation, great significance and established a highly accurate and stable successive projections algorithm- partial least squares (SPA-PLS) model to estimate leaf content in the wheat jointing stage. He et al. [3] used the variable importance for projection (VIP) and the grey relational analysis (GRA) to extract five optimal first derivative spectra in the range of 350–2500 nm in winter wheat. On the whole, those bands had a pretty strong relationship and established better results with leaf nitrogen content (LNC) in validation.

(2) Position of reflected features. Reflectance and absorption features that characterize hyperspectral data are also related to specific physical and chemical crop characteristics [4]. Wei et al. [5] extracted the red edge position by using six different methods, and analyzed the relationship between the red edge position extracted from canopy spectra and associated LNC of the vegetation above ground. Cho et al. [6] extracted the red edge position (REP) from rye canopy, and corn leaf and mixed grass/herb leaf stack hyperspectral data via a new technique and REPs extracted using this new technique (linear extrapolation method) showed high correlations with a wide range of foliar nitrogen concentrations (NC).

(3) Vegetation index characteristics. Vegetation indices are combinations of linear and nonlinear characteristics in the visible-near-infrared band and quantitatively reflect vegetation growth under certain conditions. Chen et al. [7] compared the double-peak canopy nitrogen index (DCNI), with some existing vegetation indices such as modified chlorophyll absorption ratio index (MCARI), canopy chlorophyll index (CCI), Medium Resolution Imaging Spectrometer (MERIS) terrestrial chlorophyll index (MTCI), etc. and determined that the DCNI of wheat and corn provided the best spectral index for evaluating the efficiency of crop nitrogen treatment. Finally, through the correlation analysis among the normalized differential red edge index (NDRE), water sensitivity index (WI), and crop leaf area index (LAI). Shu et al. [8] constructed a new red-edge resistance water vegetable index (RRWVI) to improve the accuracy of hyperspectral inversion of the crop leaf area index (LAI). Tan et al. [9] comprehensively analyzed the correlation and predictability of ratio vegetation index (RVI), normalized difference vegetation index (NDVI), difference vegetation index (DVI), etc. and summer corn LAIs. The results suggested that the correlation between those commonly used spectral vegetation indices and LAI reached the significant level of 0.05.

On the one hand, in this paper, the characteristics of corn canopy spectra were systematically summarized and studied from three perspectives. Those studies demonstrate that hyperspectral variable features have been utilized in research. Although these features could be used to monitor and evaluate crop growth parameters, most studies only use a single type of spectral variable. The various types of spectral variables provide useful information from different viewpoints about crop growth parameters. Therefore, if making use of different spectral variables to improve the accuracy offered a possible way with which crop-target parameters were monitored, little information was currently available on how to comprehensively exploit multiple hyperspectral variables to better exploit the rich information of hyperspectral spectrum and thereby we might improve the accuracy of crop nutrient status. On the other hand, hyperspectral data were characterized by multiple collinearity. Multiple linear regression analysis (MLRA) model, PROPERTIES SPECTRA (Prospect) model, Decision Support System for Agrotechnology Transfer (DSSAT) model, and so on were commonly used methods for LNC inversion. The PLS regression method was an extension of the multiple linear-regression model, which was widely used because it could reduce the problem of collinearity between data variables [10–12]. Furthermore, although the random forest (RF) model, which was used mostly in biology and had high predictive and learning ability, resolved the problem of singular values between response variables and explanatory variables [13], few reports had used it to monitor the nitrogen content in corn. Thus, the present study applied these three categories of spectral variables based on extracted hyperspectral information and combines them with sensitive variables selected by using the successive projections algorithm (SPA) to estimate the LNC of corn. Furthermore, we compared the PLS and RF modeling

methods for monitoring corn LNC to obtain new ideas and methods for evaluating the nitrogen nutrition spectrum of crops.

## 2. Materials and Methods

### 2.1. Study Area and Preprocessing

#### 2.1.1. Study Area

The experiment was done at the National Precision Agriculture Research and Demonstration Base in 2012. The base is located northeast of Xiaotangshan Town, Changping District, Beijing ( $40^{\circ}00'–40^{\circ}21' N$ ,  $116^{\circ}34'–117^{\circ}00' E$ , 36 m). The climate was a temperate continental monsoon climate. The soil in the experimental area was a silt-clay loam, which the PH value reached 8.0. The average soil nutrients of the site were as follows: Organic matter 1.14%; alkaline nitrogen  $49.9 \text{ mg}\cdot\text{kg}^{-1}$ ; available phosphorus  $17.0 \text{ mg}\cdot\text{kg}^{-1}$ ; and available potassium  $145 \text{ mg}\cdot\text{kg}^{-1}$ . Three nitrogen levels were used: No nitrogen content of  $0 \text{ kg N}\cdot\text{ha}^{-1}$ , normal nitrogen content of  $337 \text{ kg N}\cdot\text{ha}^{-1}$ , and excess nitrogen content of  $765 \text{ kg N}\cdot\text{ha}^{-1}$ . The experiments were done in replicates of three. We used 18 study plots. 1#–9# plots were planted 'Nongda 108' (Mid-drape type) and 10#–18# plots were planted 'Jinghua 8' (compact type). The total study area was  $924 \text{ m}^2$  and each area was  $7 \times 7 \text{ m}^2$ . The blank line was  $42 \text{ m}^2$  in the middle. Figure 1 shows the study plots. Corn samples were extracted at four growth stages of V6, V14, R1, and R2 in 18 plots, respectively. Sowing was done on 21 June 2012 and harvesting on 15 October 2012 of corn. All plots followed the local standard practices (weed control, pest management, and fertilizer application).

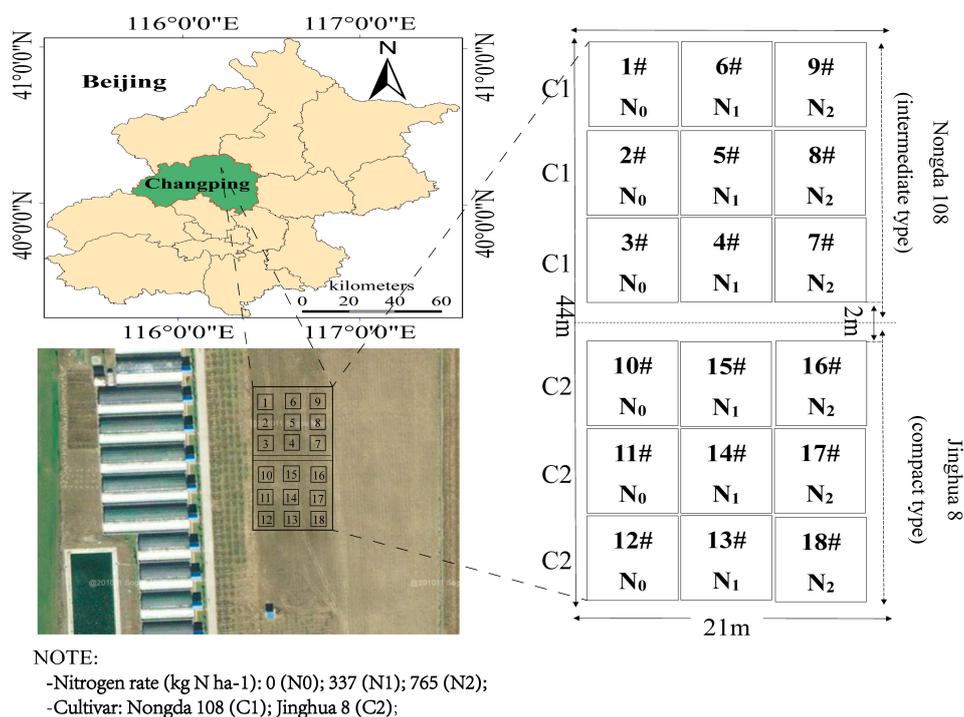


Figure 1. Study plots.

#### 2.1.2. Spectrum Acquisition

The experiment was done during the different growth stages of corn: July 30 (V6—6 leaf), August 15 (V14—14 leaf), August 26 (R1—silking), and September 14 (R2—blister stage). Corn canopy hyperspectral data were collected from each experimental plot.

The spectral reflectance was acquired by using an ASD FieldSpec FR2500 Spectrometer (Analytical Spectral Device, Boulder, CO, USA) with a spectral range of 350–2500 nm. The resolution was 1.4 nm

from 350 to 1000 nm, and 1 nm from 1000 to 2500 nm. Generally, measurements were done at 10:00 a.m. and 14:00 p.m. Beijing time during clear, windless, cloudless conditions. The probe was oriented vertically downward when viewed. The height was 1.3 m from the ground and the field of view angle was 25°. Each measurement was corrected before and after by using the reference plate.

### 2.1.3. Plant Sample and LNC Acquisition

After making spectral measurements of each experimental plot, the stems and leaves were separated, and the leaves were placed in a paper bag. The leaves were then placed in an oven at 105 °C for 30 min, and then baked at 80 °C for 48 h or more until weighed. The dried-leaf samples were weighed, and then the leaves were pulverized and their nitrogen content was determined by using a Kjeldahl analyzer (Buchi B-339, FOSS, Sweden). The total statistics were 72 of four growth stage in 18 plots; 48 for calibration, and 24 for validation. Table 1 showed the total statistics of green LNC measured. The LNC range for the calibration dataset in 2012 was from 0.92 to 2.83, with an average of 1.91 and a standard deviation of 0.59. Similarly, the statistical parameters for the test dataset in 2012 was 0.82–2.68, 1.81, and 0.65, respectively.

**Table 1.** Descriptive statistics of leaf nitrogen content (LNC).

| Dataset             | Year | Samples | Max  | Min  | Mean | SD   | Coefficient of Variation |
|---------------------|------|---------|------|------|------|------|--------------------------|
| Calibration dataset | 2012 | 48      | 2.83 | 0.92 | 1.91 | 0.59 | 0.31                     |
| Validation dataset  | 2012 | 24      | 2.68 | 0.82 | 1.81 | 0.65 | 0.36                     |

## 2.2. Principles and Methods

### 2.2.1. Preprocessing of Hyperspectral Data

To eliminate part of the noise in the spectrum, we applied a Savitzky-Golay (SG) convolution smoothing method [14,15]. Based on the results of preliminary experiments, maximum denoising was achieved with a moving window width of 17 and a polynomial frequency of two. We calculated various spectral variables, such as first derivative (FD), position features, and vegetation indices, based on the spectral reflectance after SG denoising. The FD formula was:

$$FD_{\lambda(i)} = \frac{R_{\lambda(j+1)} - R_{\lambda(j)}}{\lambda(j+1) - \lambda(j)} \quad (1)$$

where FD is the first derivative of reflectance at wavelength midpoint  $i$  between wavebands  $j$  and  $j + 1$ ,  $R_{\lambda(j)}$  is the reflectance at waveband  $j$ ,  $R_{\lambda(j+1)}$  is the reflectance at waveband  $j + 1$ , and  $\lambda(j+1) - \lambda(j)$  is the difference in wavelength between wavebands  $j$  and  $j + 1$ .

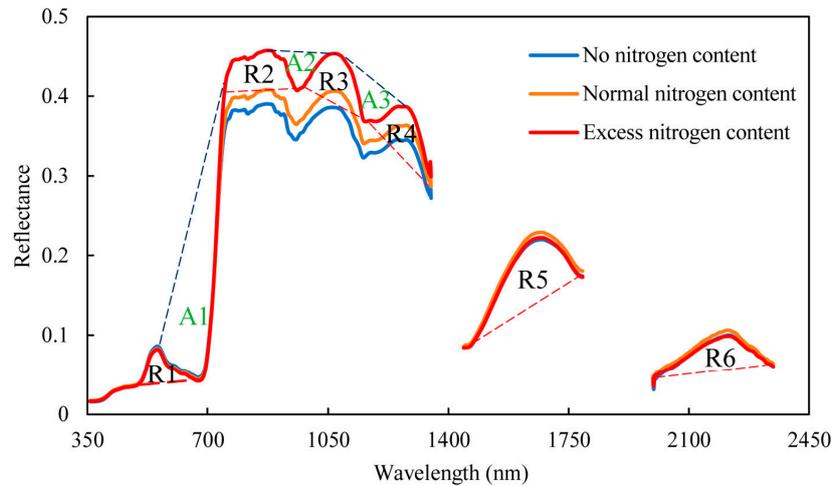
### 2.2.2. Spectral Position Features

Figure 2 showed the characteristic absorption and reflections of the summer corn canopy for three nitrogen treatments. The figure showed the three absorptions (560–760, 920–1080, and 1120–1280 nm) and six reflections (500–670, 780–970, 980–1200, 1200–1350, 1480–1720, and 2000–2300 nm) that were used to study the characteristic absorption and reflection positions [16]. In the present study, we explored only three parameters: Depth, area, and normalized depth [17,18].

The absorption depth ( $A\_Depth_i$ ) was calculated as follows:

$$A\_Depth_i = 1 - R'_i(\lambda_{min}) = 1 - R_i(\lambda_{min})/R_{ci}(\lambda_{min}), \quad (2)$$

where  $R'_i(\lambda_{min})$  is the continuum-removal reflectance and is defined as the ratio of  $R_i(\lambda_{min})$  that is the reflectance at corresponding wavelength  $\lambda$  in the absorption region to the continuum line  $R_{ci}(\lambda_{min})$  in the corresponding band. The index  $i$  identifies the number of absorption positions ( $i = 1, 2, 3$ ).



**Figure 2.** Characteristic (A) absorption and (R) reflection positions of the summer corn for the three nitrogen treatments. The red dotted line was the inner continuous line  $R_{ci}(\lambda_{\max})$  in the reflection position, the blue dotted line was outer continuum line  $R_{ci}(\lambda_{\min})$  in the absorption region.

The absorption area ( $A\_Area_i$ ) was calculated as follows:

$$A\_Area_i = \int_{\lambda_j}^{\lambda_k} (R_{ci}(\lambda) - R_i(\lambda)) d\lambda. \quad (3)$$

The absorption area ( $A\_Area_i$ ) is the integral of the difference between the reflectance of continuum line  $R_{ci}(\lambda)$  and the reflectance  $R_i(\lambda)$  at the corresponding wavelength  $\lambda$  in the absorption region. The wavelengths  $\lambda_j$  and  $\lambda_k$  are the initial and final wavelengths in each absorption region.

The normalized absorption depth ( $A\_ND_i$ ) was:

$$A\_ND_i = A\_Depth_i / A\_Area_i. \quad (4)$$

$A\_ND_i$  is the ratio of the absorption depth to the integrated absorption wavelength.

Each reflection depth ( $R\_Depth_i$ ) was defined as:

$$R\_Depth_i = 1 - R'_i(\lambda_{\max}) = 1 - R_{ci}(\lambda_{\max}) / R_i(\lambda_{\max}). \quad (5)$$

The reflection depth is the difference between unity and the continuum-removed reflectance  $R'_i(\lambda_{\max})$ . The continuum-removed reflectance  $R'_i(\lambda_{\max})$  is the ratio of the inner continuous line  $R_{ci}(\lambda_{\max})$  in the reflection position and the maximum reflectance value  $R_i(\lambda_{\max})$  at the corresponding band. The index  $i$  indicates the number of the corresponding band ( $i = 1, 2, 3$ ).

The reflection area ( $R\_Area_i$ ) was defined as:

$$R\_Area_i = \int_{\lambda_j}^{\lambda_k} (R_i(\lambda) - R_{ci}(\lambda)) d\lambda, \quad (6)$$

which is the definite integral of the difference between the reflectance  $R_i(\lambda)$  in the corresponding band  $\lambda$  at the reflectance region and the inner continuum line  $R_{ci}(\lambda)$ . The wavelengths  $\lambda_j$  and  $\lambda_k$  are the initial and final wavelengths in each reflectance region, respectively. The index  $i$  is the number of the band ( $i = 1, 2, 3$ ).

The normalized reflection depth ( $R\_ND_i$ ) was the ratio of the reflectance depth ( $R\_Depth_i$ ) to the reflectance area ( $R\_Area_i$ ):

$$R\_ND_i = R\_Depth_i / R\_Area_i. \quad (7)$$

For more information, please see the relevant literature [19–21]. Table 2 showed the positional bands used in this study.

**Table 2.** Partial characteristic variables of the hyperspectral data.

| Variables   | Definition and Description  |
|-------------|---|
| Db          | Maximum value of the 1st derivative with a blue edge (490–530 nm)   |
| $\lambda_b$ | Wavelength at Db  |
| Dy          | Maximum value of the 1st derivative with a yellow edge (560–640 nm) |
| $\lambda_y$ | Wavelength at Dy  |
| Dr          | Maximum value of the 1st derivative with a red edge (680–760 nm)    |
| $\lambda_r$ | Wavelength at Dr  |
| Rg          | Maximum reflectance with a green peak (510–560 nm)                  |
| $\lambda_g$ | Wavelength at Rg  |
| Ro          | Lowest reflectance with a red well (650–690 nm)                     |
| $\lambda_o$ | Wavelength at Ro  |
| SDb         | Sum of the 1st derivative values within the blue edge               |
| SDy         | Sum of the 1st derivative values within the yellow edge             |
| SDr         | Sum of the 1st derivative values within the red well                |

### 2.2.3. Vegetation Indices

Many different optical indices have been reported in the literature and have proven to be well correlated with vegetation parameters. Consequently, we selected 34 vegetation indices (VIs) for estimating the corn canopy LNC (Table 3). Six of these indices were nitrogen-sensitive hyperspectral VIs [7], such as the optimal vegetation index ( $V_{i_{opt}}$ ), the normalized difference vegetation index green-blue<sup>#</sup> (NDVI<sub>g-b</sub><sup>#</sup>), the ratio vegetation index I<sup>#</sup> (RVI I<sup>#</sup>), RVI II<sup>#</sup>, the combined index (MCARI/MTVI2), the double-peak canopy nitrogen index<sup>#</sup> (DCNI<sup>#</sup>), NDVI I<sup>#</sup>, RVI III<sup>#</sup>, DVI I<sup>#</sup>, SAVI I<sup>#</sup>, normalized difference red-edge (NDRE), etc. Another 23 indices were used: The anti-atmospheric vegetation index (ARVI), the difference vegetation index (DVI II<sup>#</sup>), the enhanced vegetation index (EVI), the green normalized vegetation index (GNDVI), the modified nonlinear vegetation index (MNLI), the second modified SAVI (MSAVI2), the modified simple ratio (MSR), NDVI II<sup>#</sup>, the nonlinear vegetation index (NLI), the optimized SAVI (OSAVI), RDVI, the ratio vegetation index IV (RVI IV), SAVI II<sup>#</sup>, TVI and the modified triangle vegetation index (MTVI2), the red-edge-related index NDVI<sub>Red-edge</sub>, CI<sub>Red-edge</sub>, MTCI, the water-related index (WI, NDWI), the normalized difference infrared index (NDII), the water stress index (DSWI), the standardized LAI-determining index (sLAIDI\*), etc. The VIs that related to the wide-band information were obtained from hyperspectral calculation using the spectral response functions of the corresponding sensors [22].

## 2.3. Screening and Modeling Methods

### 2.3.1. Successive Projections Algorithm

In recent years, the successive projections algorithm (SPA) [23,24] has been ever more widely used for screening and extracting sensitive variables and is a forward-variable-selection algorithm that effectively eliminates the collinearity problem in spectral information. By reducing the redundancy between variables and selecting representative feature parameters for modeling, the efficiency of modeling analysis could be greatly improved.

Table 3. Summary of partial vegetation indices.

| Index               | Name  | Formula   | Reference |
|---------------------|---|---|-----------|
| $V_{opt}$           | Optimal vegetation index                            | $(1 + 0.45) ((R_{800})^2 + 1)/(R_{670} + 0.45)$   | [25]      |
| $NDVI_{g-b}^{\#}$   | Normalized difference vegetation index <sup>#</sup> | $(R_{573} - R_{440})/(R_{573} + R_{440})$   | [26]      |
| RVI I <sup>#</sup>  | Ratio vegetation index I <sup>#</sup>               | $R_{810}/R_{660}$   | [27]      |
| RVI II <sup>#</sup> | Ratio vegetation index II <sup>#</sup>              | $R_{810}/R_{560}$   | [28]      |
| MCARI/MTVI2         | Combined index                                      | MCARI/MTVI2<br>MCARI: $(R_{700} - R_{670} - 0.2(R_{700} - R_{550})) (R_{700}/R_{670})$<br>MTVI2: $1.5(1.2(R_{800} - R_{550})) / \sqrt{(2R_{800} + 1)^2 - (6R_{800} - 5\sqrt{R_{670}}) - 0.5}$ | [29]      |
| DCNI <sup>#</sup>   | Double-peak canopy nitrogen index I <sup>#</sup>    | $(R_{720} - R_{700})/(R_{700} - R_{670})/(R_{720} - R_{670} + 0.03)$  | [7]       |
| NDVI I              | Normalized difference vegetation index I            | $(R_{800} - R_{670})/(R_{800} + R_{670})$   | [30]      |
| RVI III             | Ratio vegetation index III                          | $R_{800}/R_{670}$   | [31]      |
| DVI I               | Difference vegetation index I                       | $R_{800} - R_{670}$   | [32]      |
| SAVI I              | Soil-adjusted vegetation index I                    | $1.5(R_{800} - R_{670})/(R_{800} + R_{670} + 0.5)$  | [33]      |
| NDRE                | Normalized difference red edge                      | $(R_{790} - R_{720})/(R_{790} + R_{720})$   | [34]      |
| ARVI                | Atmospherically-resistant vegetation index          | ARVI = $(R_{NIR} - RB)/(R_{NIR} + RB)$<br>RB = $R - \gamma(B - R)$ , $\gamma = 1$   | [35]      |
| DVI II              | Difference vegetation index II                      | $DVI = R_{NIR} - R_R$   | [36]      |
| EVI                 | Enhanced vegetation index                           | $EVI = 2.5(R_{NIR} - R_R)/(R_{NIR} + 6R_R - 7.5R_B + 1)$  | [37]      |
| GNDVI               | Green normalized difference vegetation index        | $GNDVI = (R_{NIR} - R_R)/(R_{NIR} + R_R)$   | [38]      |
| MNLI                | Modified nonlinear vegetation index                 | $MNLI = 1.5(R_{NIR}^2 - R_R)/(R_{NIR}^2 + R_R + 0.5)$   | [39]      |
| MSAVI2              | The second modified SAVI                            | $MSAVI2 = (2R_{NIR} + 1 - \sqrt{(2R_{NIR} + 1)^2 - 8(R_{NIR} - R_R)})/2$  | [40]      |
| MSR                 | Modified simple ratio                               | $MSR = (R_{NIR}/R_R - 1) / (R_{NIR}/R_R + 1)$   | [41]      |
| NDVI II             | Normalized difference vegetation index II           | $NDVI = (R_{NIR} - R_R) / (R_{NIR} + R_R)$  | [42]      |
| NLI                 | Nonlinear vegetation index                          | $NLI = (R_{NIR}^2 - R_R)/(R_{NIR}^2 + R_R)$   | [43]      |
| OSAVI               | Optimization of soil-adjusted vegetation index      | $OSAVI = (1 + 0.16) (R_{NIR} - R_R)/(R_{NIR} + R_R + 0.16)$   | [44]      |
| RDVI                | Renormalization difference vegetation index         | $RDVI = (R_{NIR} - R_R) / (\sqrt{R_{NIR} + R_R})$   | [45]      |
| RVI IV              | Ratio vegetation index                              | $RVI = R_{NIR}/R_R$   | [33]      |
| SAVI II             | Soil-adjusted vegetation index II                   | $SAVI = 1.5(R_{NIR} - R_R) / (R_{NIR} + R_R + 0.5)$   | [46]      |
| TVI                 | Triangular vegetation index                         | $TVI = 60(R_{NIR} - R_C) - 100(R_R - R_C)$  | [47]      |
| MTVI2               | Modified triangular vegetation index                | $MTVI2 = 1.5(1.2(R_{NIR} - R_C) - 2.5(R_R - R_C)) / (\sqrt{(2R_{NIR} + 1)^2 - (6R_{NIR} - 5\sqrt{R_R}) - 0.5})$   | [48]      |
| $NDVI_{Red-edge}$   | Red-edge NDVI                                       | $NDVI_{Red-edge} = (R_{NIR} - R_{Red-edge}) / (R_{NIR} - R_{Red-edge})$   | [49]      |
| $CI_{Red-edge}$     | Red-edge Chlorophyll Index                          | $CI_{Red-edge} = (R_{NIR}/R_{Red-edge}) - 1$  | [50]      |
| MTCI                | MERIS Terrestrial Chlorophyll Index                 | $MTCI = (R_{NIR} - R_{Red-edge}) / (R_{Red-edge} - R_{NIR})$  | [51]      |
| WI                  | Water Index   | $WI = R_{900}/R_{970}$  | [52]      |
| NDWI                | Normalized difference water index                   | $NDWI = (R_{860} - R_{1240}) / (R_{860} + R_{1240})$  | [53]      |
| NDII                | Normalized difference infrared index                | $NDII = (R_{819} - R_{1600}) / (R_{819} + R_{1600})$  | [54]      |
| DSWI                | Disease water stress index                          | $DSWI = (R_{803} - R_{549}) / (R_{1659} + R_{681})$   | [55]      |
| sLAIDI *            | Standardized LAI-determining index                  | $sLAIDI^* = s(R_{1050} - R_{1250}) / (R_{1050} + R_{1250})R_{1555}$ , $s = 1$   | [56]      |

\* I, II, III, IV, V were just only for the same planting indices that distinguished different bands.

In order to solve the collinearity problems, a minimally redundant subset of wavebands is selected in SPA and it belongs to the class of forward selection methods [57]. SPA starts with one wavelength and selects a new one at each iteration by using projection operators in a vector space until reaching the predefined number of wavelengths. Root means square error (RMSE) was used as the evaluation index. The final number of variables selected by the SPA is defined based on the lower RMSE value obtained.

### 2.3.2. Partial Least Squares Regression

Partial least squares regression [58] is a statistical method that included principal component analysis, canonical correlation analysis, and multiple linear regression methods [59]. PLS regression is a modeling technique for studying multi-dependent variables or single-dependent variables and multi-independent variables. It can screen out low-collinearity components in the case of small sample size.

Consider  $m$  dependent variables  $y_1, y_2, \dots, y_m$  and  $n$  arguments  $x_1, x_2, \dots, x_n$ . The quantities  $E_0, E_1, \dots, E_r, F_0, \dots, F_r$  are standardized observation data arrays of two sets of variables from which we extract the components  $t_1, \dots, t_r (r \leq m)$ ,  $t_h$  is a linear combination from the independent-variable set  $X = (x_1, \dots, x_m)^T$  and carries the maximum information possible from  $X$ . At the same time,  $t_h$  has the greatest explanatory power for the dependent-variable system  $F_0$ . If we extract  $r$  components  $t_1, \dots, t_r$  from the independent-variable set, the PLS regression will establish the regression equation for  $y_1, y_2, \dots, y_m$  and  $t_1, \dots, t_r$ , and then express  $y_1, y_2, \dots, y_m$  and the regression equation of the original independent variables; that is, the PLS regression equation:

$$y_j = a_{j1}x_1 + \dots + a_{jm}x_m, (j = 1, 2, \dots, p). \quad (8)$$

### 2.3.3. Random Forest

Random forest (RF) [60,61] used bootstrapping to randomly draw samples that were resampled and put back. The extracted samples are used to construct a classification decision tree, and the non-extracted samples constitute the out-of-bag (OOB) data set.

Given  $n$  features, RF arbitrarily extracts less than  $m$  ( $m < n$ ) features at each node of each tree, selects the classification of the decision tree with the largest amount of information among the  $m$  features, and does not prune the classification decision tree.

A plurality of regression decision trees is constructed by using the extracted samples to form a RF, and then the data are classified, and the result is decided by voting.

Each time a RF forms, the OOB data set is used to evaluate the classification results, following which we evaluate the combined classifier. The variable that generates the decision tree is randomly selected each time from the training set, so the random forest has a stable error rate, and each OOB generated can be used to evaluate the classifier performance.

### 2.3.4. Statistical Analysis Method

Between the sensitive bands, the location characteristics, VIs, and corn canopy LNC were analyzed by using Rstudio 3.5.3. The validation samples were one-third of the samples (i.e., 24 samples) and did not participate in the validation. The operation of partial least squares and random forest algorithm was done in MATLAB R2014a. The determination coefficient  $R^2$ , the RMSE, and the normalized root mean square error (NRMSE) serve as indicators to explain and quantify the relationship with nitrogen in canopy leaves. They were calculated as follows:

$$R^2 = (\sum_{i=1}^n y_i - \bar{y})^2 / (\sum_{i=1}^n x_i - \bar{y})^2, \quad (9)$$

$$\text{RMSE} = \sqrt{\sum_{i=1, j=1}^n (x_i - y_j)^2 / n}, \quad (10)$$

$$\text{NRMSE} = \sqrt{\sum_{i=1, j=1}^n (x_i - y_j)^2 / n / \bar{y}}, \quad (11)$$

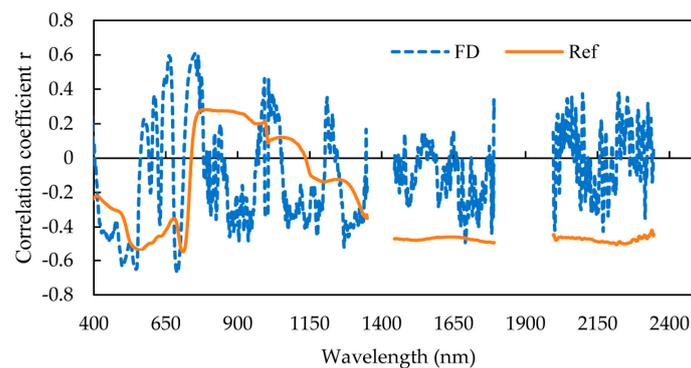
where  $x_i$  is the measured nitrogen content in corn canopy leaves,  $y_i$  is the predicted nitrogen content in corn canopy leaves,  $\bar{y}$  is the mean nitrogen content in corn canopy leaves, and  $n$  is the number of samples.

### 3. Results

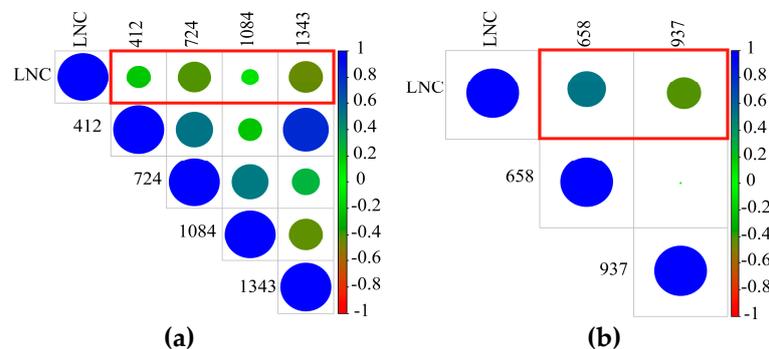
#### 3.1. Optimal Spectral Characteristics

##### 3.1.1. Sensitive Reflectance Feature Data Set

Figure 3 shows a correlation between the corn canopy reflectance spectra and the FD spectra and LNC. The reflectance spectrum in Figure 3 suggests a negative correlation between the blue (630 nm), the red (711 nm), and the short-wave near-infrared (1996–2346 nm) reflectance spectra, and a positive correlation in the near-infrared (739–1135 nm). The FD spectrum shows a positive correlation at 661 and 751 nm, and a negative correlation at 691 nm. Spectral first derivative (FD) operation reduces the background noise and raises the efficiency of the target. The correlation coefficient of the FD spectrum is greater than the reflection spectrum in the visible light and NIR range (400–1400 nm). Using the SPA algorithm, four reflectance wavelengths and two FD wavelengths were selected to form a sensitive spectral dataset (Figure 4). The 724, 1343 nm (Ref), 658, and 937 nm (FD) wavelengths were well correlated with LNC, and 724 nm, 658 nm, and 937 nm fell in the visible light and NIR bands, indicating that the corn canopy reflectance spectrum and the FD spectrum were strongly correlated with LNC in the visible range (400–700 nm) and NIR range (700–800 nm).



**Figure 3.** Correlation between model set reflectance spectra (Ref) and first derivative spectra (FD) and LNC of the training set.



**Figure 4.** Correlation between selected sensitive spectral bands and LNC (in red boxes,  $n = 48$ ). (a) 412, 724, 1084, and 1343 nm are reflectance spectra; and (b) 658 and 937 nm are first derivative spectra.

The original spectrum is easily affected by illumination, soil background, atmosphere, and other factors. However, derivative transformation can reduce or eliminate the influence of background and atmospheric scattering and improve the contrast of different absorption characteristics. Therefore

the reflection spectral bands and the first derivative bands selected were used separately in PLS and RF models.

The SPA screened out six sensitive spectra (including four reflectance spectral wavelengths: 412, 724, 1084, and 1343 nm and two first derivative spectral wavelengths: 658 and 937 nm) with the least linearity of leaf nitrogen content. A model was established between the four reflectance spectra and the LNC based on the PLS and RF regression. Similarly, a model was established between the two first derivative spectra and LNC based on the PLS and RF regression. The results are given in Table 4. The PLS model was used to estimate the nitrogen content of leaves. The coefficient  $R^2$ , the RMSE, and the NRMSE of the reflection spectral bands were 0.59, 38.2%, and 0.20, respectively. For the FD bands, these values are 0.54, 39.7%, and 0.21, respectively. The RF model was used to estimate the nitrogen content of leaves. The coefficient  $R^2$ , the RMSE, and the NRMSE of the bands of the reflection spectrum were 0.61, 42.1%, and 0.22, respectively; and for the FD bands these values were 0.59, 37.9%, and 0.20, respectively. These values represented good results for the modeling.

In the validation set, when PLS was used to estimate the LNC,  $R^2$  for the reflectance spectrum was 0.22 greater than for the RF model, the RMSE was 0.2 less, and the NRMSE was 11.2% less. The coefficient  $R^2$  of the RF model for the FD value was 0.02 less than the Ref, the RMSE was 0.05 less, and the NRMSE was 2.6% less. These results showed that the inversion of the corn LNC by the PLS model was better than the RF model for the reflectance spectra, which suggested that the PLS model should provide more accurate predictions of the LNC. However, the RF model was more stable than the PLS model.

**Table 4.** Accuracy between the sensitive band and FD spectrum and LNC.

| Algorithm                   | Feature Types | Calibration Set ( $n = 48$ ) |      |       | Validation Set ( $n = 24$ ) |      |       |
|-----------------------------|---------------|------------------------------|------|-------|-----------------------------|------|-------|
|                             |               | $R^2$                        | RMSE | NRMSE | $R^2$                       | RMSE | NRMSE |
| Partial Least Squares (PLS) | Ref           | 0.59                         | 0.38 | 19.8% | 0.82                        | 0.28 | 15.2% |
|                             | FD            | 0.54                         | 0.40 | 20.8% | 0.64                        | 0.38 | 21.0% |
| Random Forest (RF)          | Ref           | 0.61                         | 0.42 | 22.1% | 0.60                        | 0.48 | 26.4% |
|                             | FD            | 0.59                         | 0.38 | 19.9% | 0.58                        | 0.43 | 23.6% |

### 3.1.2. Position Feature Data Set

We selected the position characteristics of 40 hyperspectral reflectance wavelengths and calculated the positional correlation of each calibration set (75%; Figure 5). The results showed that the LNC was strongly correlated with Db, Dr,  $\lambda_b$ , Rg,  $\lambda_g$ , and SDb, whereas the LNC was weakly correlated with the other parameters. Two positional parameters SDb and Dr with smaller collinearity were selected by the SPA algorithm and were modeled using the LNC (Table 5). The results of estimating the LNC in the optical layer for  $R^2$ , RMSE, and NRMSE were 0.50, 41.2%, 0.22 and 0.57, 39.9%, 0.21 for the PLS and RF models, respectively. The coefficient  $R^2$  of the RF model was 0.07 greater, the RMSE was 0.01 less, and the NRMSE was 0.7% less than for the PLS model, which indicated that the RF model was more stable and the PLS model had better results than the RF model. The coefficient  $R^2$  of the PLS model in the validation set was 0.1 greater, the RMSE was 0.07 lower, and the NRMSE was 3.2% lower than the RF model.

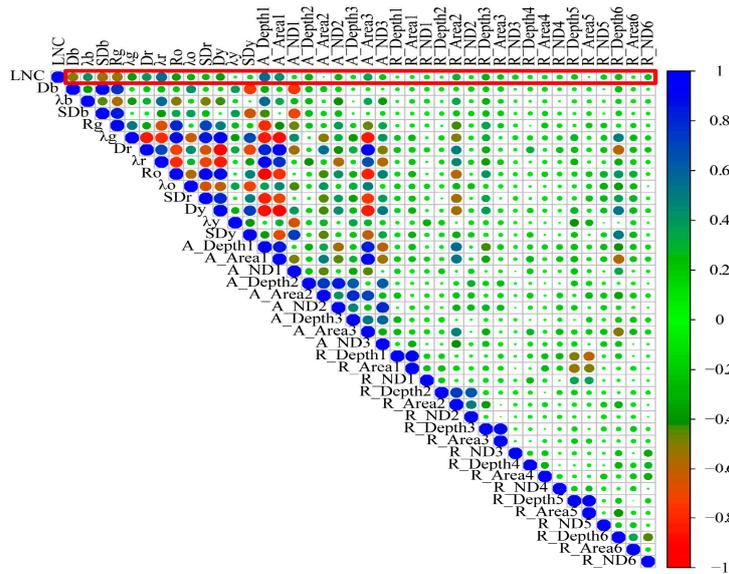


Figure 5. Correlations between LNC and special positions (in red boxes,  $n = 48$ ).

Table 5. Modeling results between position features and LNC.

| Algorithm                   | Feature Types | Calibration set ( $n = 48$ ) |      |       | Validation set ( $n = 24$ ) |      |       |
|-----------------------------|---------------|------------------------------|------|-------|-----------------------------|------|-------|
|                             |               | R <sup>2</sup>               | RMSE | NRMSE | R <sup>2</sup>              | RMSE | NRMSE |
| Partial Least Squares (PLS) | Positions     | 0.50                         | 0.41 | 21.6% | 0.62                        | 0.41 | 22.9% |
| Random Forest (RF)          | Positions     | 0.57                         | 0.40 | 20.9% | 0.52                        | 0.47 | 26.1% |

### 3.1.3. Vegetation Indices Data Set

For this study, 34 vegetation indices (VIs; Figure 6) were selected to study the correlation between them and LNC. The results show that DVI I, DVI II, and TVI had a weaker relevance with LNC than others. Such results are expressed in Figure 6, where these VIs are represented by smaller circles and lighter color when compared to the other VIs.

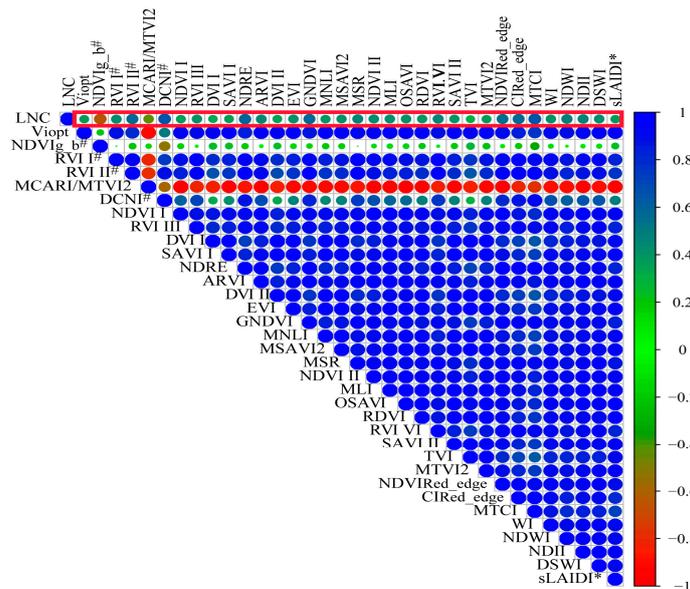


Figure 6. Correlation and significance of vegetation indices (VIs) and LNC (in red boxes,  $n = 48$ ).

The SPA algorithm selected the eigenvalues  $NDVI_{g-b}^{\#}$  and DVI II, which had small collinearity between VIs. The two parameters were modeled with the LNC (Table 6), and the results show that  $R^2$ , RMSE, and NRMSE were 0.68, 33.3%, 0.17 and 0.64, 35.6%, 0.19, when the canopy LNC was estimated by the PLS model and the RF model, respectively. For the estimation, the coefficient  $R^2$  was 0.04 greater, the RMSE was 0.03 less, and the NRMSE was 0.8% less for the PLS model than for the RF model. For the validation model, the coefficient  $R^2$  was 0.2 greater, the RMSE was 11.2% less, and the NRMSE was 0.06 less for the PLS model than for the RF model. These results indicated that the RF model was more stable and that the PLS model might provide more accurate results to estimate the LNC.

Table 6. Modeling results between VIs and LNC.

| Algorithm                   | Feature Types | Calibration Set ( $n = 48$ ) |      |       | Validation Set ( $n = 24$ ) |      |       |
|-----------------------------|---------------|------------------------------|------|-------|-----------------------------|------|-------|
|                             |               | $R^2$                        | RMSE | NRMSE | $R^2$                       | RMSE | NRMSE |
| Partial Least Squares (PLS) | VIs           | 0.68                         | 0.33 | 17.4% | 0.80                        | 0.31 | 16.9% |
| Random Forest (RF)          | VIs           | 0.64                         | 0.36 | 18.6% | 0.60                        | 0.42 | 23.1% |

### 3.2. Composite Spectral Features

To further improve the accuracy of the spectral estimates of the corn canopy LNC, six sensitive spectral features (four reflective spectral features, two FD features), two positional features, and two VIs were obtained from the three spectral variables. The SPA algorithm was then used to further screen the sensitive characteristic parameters to model, which were combined into a new set of spectral variables. The analysis shows that the spectral reflection bands at 724 and 1343 nm, FD band at 658 nm, and  $NDVI_{g-b}^{\#}$  became the new sensitive spectral variables. Table 7 gives the results of using these new characteristic parameters to estimate the corn canopy LNC. For estimating the corn canopy LNC,  $R^2$  was 0.14 greater, RMSE was 7.5% lower, and the NRMSE was 0.03 lower for the PLS model than for the RF model. Figure 7a,b show the results of the validation model. The fit was better between the measured value and the predicted value,  $R^2$  was 0.22 greater, RMSE was 0.12 less, and NRMSE was 6.8% less for the PLS model than for the RF model. The results of the RF model did not differ significantly, and the results were relatively stable. However, the result of the PLS model was better.

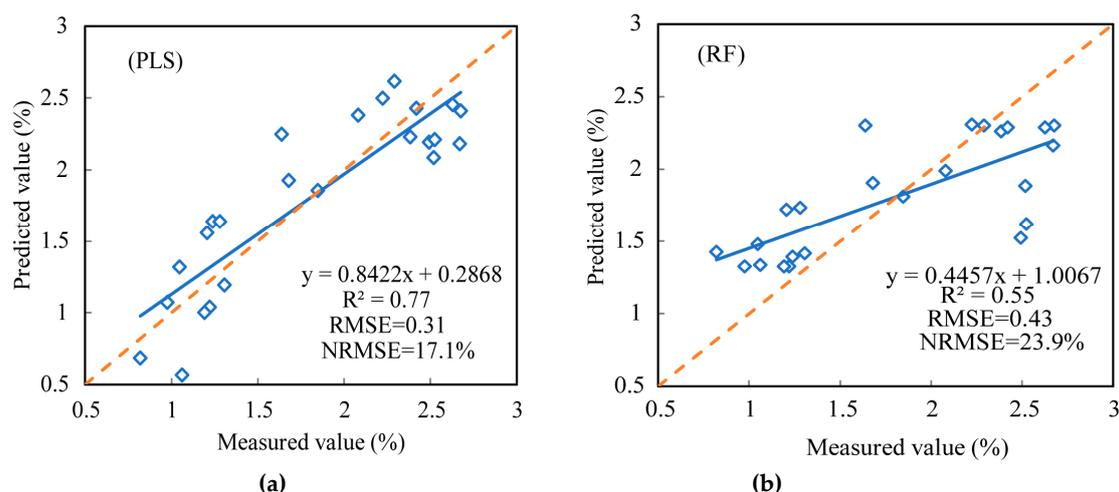


Figure 7. Accuracy of the measured and predicted values of the validation set: (a) Partial least squares (PLS) model, (b) random forest (RF) model.

**Table 7.** Modeling results between comprehensive parameters and LNC.

| Algorithm                   | Feature Types   | Calibration Set ( $n = 48$ ) |      |       | Validation Set ( $n = 24$ ) |      |       |
|-----------------------------|-----------------|------------------------------|------|-------|-----------------------------|------|-------|
|                             |                 | R <sup>2</sup>               | RMSE | NRMSE | R <sup>2</sup>              | RMSE | NRMSE |
| Partial Least Squares (PLS) | Integrated data | 0.71                         | 0.32 | 16.7% | 0.77                        | 0.31 | 17.1% |
| Random Forest (RF)          | Integrated data | 0.57                         | 0.39 | 20.4% | 0.55                        | 0.43 | 23.9% |

#### 4. Discussion

In order to reduce the influence of water vapor and other factors of hyperspectral data [62,63], we chose 400–1353 nm, 1437–1799 nm, and 1992–2354 nm to study the spectra. We selected reflectance spectra of 412, 724, 1084, and 1343 nm and first derivative spectra of 658 and 937 nm. Kokaly and Clark [18] got the spectral characteristics of absorption and reflection positions via using the continuum-removal method. We selected two positions characteristic using the same method: SDb and Dr. The vegetation indices were a linear and non-linear combination of different bands, and the functional relationship of vegetation characteristic parameters was more stable and reliable than a single band [64]. We selected NDVI<sub>g-b</sub><sup>#</sup>, DVI II of the two optimal VIs.

Serious multi-collinearity problems arose in sensitive bands, positions and VIs. The optimal sensitive band features, position features, and VIs of hyperspectral data were selected by using SPA and they had a good correlation with LNC (Figures 4–6), but the correlation between the position features and LNC was low. Bands that the optimal parameters used were mainly focused on visible- and near-infrared-band. This result was consistent with the results of a previous study [65]. In this paper, the  $R^2$  between the optimal reflectance spectra, VIs and LNC achieved 0.82 and 0.80. The parameters were mainly concentrated on blue-light, red-edge, and NIR bands, it might be the influence of internal factors such as chlorophyll and cell of plants, and the position parameters were red-shifted due to the difference of leaves nitrogen content. The integrated spectral features (reflectance at 724 and 1343 nm, FD at 658 nm, and NDVI<sub>g-b</sub><sup>#</sup>) determined  $R^2$ , the RMSE, and the NRMSE for the calibration set (validation set) of the PLS model and the RF model to be 0.71, 31.8%, 0.17 (0.77, 31.0%, 0.17) and 0.57, 39.3%, and 0.20 (0.55, 43.3%, and 0.24), respectively. Chen et al. [7] suggested that the  $R^2$  values were 0.72 for corn. Ours results were increased by 0.05 than theirs in the PLS model and there were no significant differences in values. The composite spectral features integrated characteristics of three variable sets and the results of PLS model were more stable, when comparing the results for calibration and validation datasets, than any other three variable datasets used independently. The reflectance bands and position features were easily affected by external light, water and nitrogen content, and so on. VIs had the ability to eliminate effects of soil background factors, especially NDVI<sub>g-b</sub><sup>#</sup>.

Most previous studies focused on a single variable of the spectrum [2,11,66] to study corn leaves nitrogen content, whereas few studies discussed the comprehensive processing of data or compared models with similar variables. The present study used two models for the analysis: The PLS and RF models. The LNC model established by the PLS algorithm was the best in the two models. The method of linear model was obviously better than machine learning. The results of the PLS model could decompose and filter the data by leveraging the number of input samples. A high precision model was established for the comprehensive variable with the strongest explanatory power of the dependent variable [67,68]. The RF model is a machine learning algorithm with simple implementation, good precision, and strong over-fitting ability [69]. This is indicative of a strong learning ability, which is consistent with the results of Feng et al. In the process of modeling with the four variable datasets as the independent variable and LNC as dependent variables, the modeling result of the RF model was not very different, but the result of the PLS model was better, which could better predict LNC. Based on sensitive variables obtained from screened multi-variety and multi-growth data over one year, the next step is to lengthen the study (multiple years) and use more regional data for an in-depth analysis.

## 5. Conclusions

In this paper, the results showed that the spectral bands, absorption and reflection positions, and VIs were usually good predictors. LNC had better correlation with the optimal sensitive bands and VIs (Figures 4 and 5), but the optimal positions have a bad correlation with LNC (Figure 6). These hyperspectral features were mostly concentrated on the 300–1400 nm region, and the features in visible light and NIR regions was able to better realize the monitoring of corn LNC [70].

After screening out the original spectral bands for sensitive reflect feature dataset by SPA, for the validation set, the  $R^2$ , RMSE, and NRMSE of the PLS model and RF model were 0.82, 27.5%, and 0.15 and 0.64, 37.9%, and 0.21, respectively. The  $R^2$ , RMSE, and NRMSE for the first derivative bands of the PLS model and RF model were 0.60, 47.7%, and 0.26 and 0.58, 42.8%, and 0.24, respectively. The  $R^2$ , RMSE, and NRMSE for the position feature dataset of the PLS model and RF model were 0.62, 41.5%, and 0.23 and 0.52, 47.2%, and 0.26, respectively. The  $R^2$ , RMSE, and NRMSE for the vegetation indices feature dataset of the PLS model and RF model were 0.80, 0.31, and 16.9% and 0.60, 0.42, and 23.1%, respectively. The  $R^2$ , RMSE, and NRMSE for the reflect feature integration dataset of the PLS model and RF model were 0.77, 0.31, and 17.1% and 0.55, 0.43, and 23.9%, respectively. For estimating the corn LNC, the RF model had a good learning ability and stable results. However, the results of  $R^2$ , RMSE, and NRMSE were poor in the validation set with a small sample size, while the results of PLS were good, especially in the integration dataset, which could better estimate the LNC.

**Author Contributions:** X.X. provided toward creating the general idea of this paper. L.F. analyzed the data and wrote the draft of the manuscript. X.X. and J.Z. helped edit the draft and provided critical comments to improve the paper. G.Y., H.F. and H.Y. collected the data. D.L., Y.W., G.C., and P.W. supervised the process of data handling.

**Funding:** The project was supported by the National Natural Science Foundation of China (41571416, 61672032), the National Key Research and Development Program (2017YFD0201501), and Natural Science Research Project of Anhui Provincial Education Department (KJ2018A0009).

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

## References

1. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; Van der Meer, B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* **2004**, *91*, 354–376. [[CrossRef](#)]
2. Bai, L.; Li, F.; Chang, Q.; Zeng, F.; Cao, J.; Lu, G. Increasing accuracy of hyper-spectral remote sensing for total nitrogen of winter wheat canopy by use of SPA and PLS methods. *J. Plant. Nutr. Fertil.* **2018**, *24*, 52–58.
3. Peng, H.; Xingang, X.; Baolei, Z.; Zhenhai, L.; Haikuan, F.; Guijun, Y.; Yongfeng, Z. Estimation of leaf chlorophyll content in winter wheat using variable importance for projection (VIP) with hyperspectral data. *Remote Sens. Agric. Ecosyst. Hydrol. XVII* **2015**, *12*. [[CrossRef](#)]
4. Strachan, I.; Pattey, E.; Boisvert, J. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sens. Environ.* **2002**, *80*, 213–224. [[CrossRef](#)]
5. Dandan, W.; Xiaobing, L.; Hong, W.; Han, W.; Wanyu, W. Comparative study on estimation of nitrogen content in the heterogenous typical steppe using various red edge position extraction techniques. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012. [[CrossRef](#)]
6. Cho, M.; Skidmore, A. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sens. Environ.* **2006**, *101*, 181–193. [[CrossRef](#)]
7. Chen, P.F.; Haboudane, D.N.T.; Wang, J.H.; Vigneault, P.; Li, B.G. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote Sens. Environ.* **2010**, *114*, 1987–1997. [[CrossRef](#)]
8. Shu, M.; Gu, X.; Sun, L.; Zhu, J.; Yang, G.; Wang, Y.; Zhang, L. High spectral inversion of winter wheat LAI based on new vegetation index. *Sci. Agric. Sin.* **2018**, *51*, 57–67.
9. Tan, C.; Huang, Y.; Huang, W.; Wang, J.; Zhao, C.; Liu, L. Study on colony leaf area index of summer maize by remote sensing vegetation indexes method. *J. Anh. Agric. Univ.* **2004**, *31*. [[CrossRef](#)]

10. Li, S.; Ren, Z. Quantitative analysis of near infrared spectroscopy based on wavelet coefficients and partial least-squares regression. *J. Changchun Univ.* **2018**, *28*, 5.
11. Zhang, L.; Chen, S. Measurements and analysis of marginal effect of water use efficiency by scientific and technological innovation based on PLS-PATH method. *Adv. Sci. Technol. Water Res.* **2018**, *38*, 8.
12. Qi, B.; Zhang, N.; Zhao, T.; Xing, G.; Zhao, J.; Gai, J. Using canopy hyperspectral reflectance to predict growth traits and seed yield of soybeans from middle and lower yangtze valleys through partial least squares regression. *Soybean Sci.* **2015**, *34*, 7.
13. Zhang, L.; Wang, L.; Zhang, X.; Liu, S.; Sun, P.; Wang, T. The basic principle of random forest and its applications in ecology: A case study of *Pinus yunnanensis*. *Acta Ecol. Sin.* **2014**, *34*, 3. [[CrossRef](#)]
14. Lu, N.; Han, P.; Wang, J. Prediction on firmness of strawberry based on hyperspectral imaging. *Softw. Guide* **2018**, *17*, 3.
15. 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.
16. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.* **1984**, *89*, 6329. [[CrossRef](#)]
17. Mutanga, O.; KSkidmore, A. Hyperspectral band depth analysis for a better estimation of grass biomass (*Cenchrus ciliaris*) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. [[CrossRef](#)]
18. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* **1999**, *67*, 267–287. [[CrossRef](#)]
19. Wang, L. Study on Nutrition Diagnosis of Nitrogen Content in Maize Leaves of Cold Region Based on Hyperspectral Imaging. Master's Thesis, Northeast Agricultural University, Harbin, China, June 2017. (In Chinese)
20. Xie, W. Estimation of chlorophyll content in maize leaves based on hyperspectral under the action of microorganism. *West. Dev. Land Dev. Eng. Res.* **2018**, *3*, 31–37.
21. He, T. Hyperspectral remote sensing estimation models for nitrogen nutrition monitoring of maize. Master's Thesis, Shenyang Agricultural University, Shenyang, China, June 2016. (In Chinese).
22. Li, X.C.; Zhang, Y.J.; Bao, Y.S.; Luo, J.H.; Jin, X.L.; Xu, X.G.; Song, X.Y.; Yang, G.J. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sens.* **2014**, *6*, 6221–6241. [[CrossRef](#)]
23. Yuan, Y.; Wang, W.; Chu, X.; Xi, M.J. Selection of characteristic wavelengths using SPA and qualitative discrimination of mildew degree of corn kernels based on SVM. *Spectrosc. Spect. Anal.* **2016**, *36*, 226–230. [[CrossRef](#)]
24. Wu, D.; Ning, J.; Liu, X.; Liang, M.; Yang, S.; Zhang, Z. Determination of anthocyanin content in grape skins using hyperspectral imaging technique and successive projections algorithm. *Food Sci.* **2014**, *35*, 5.
25. Reyniers, M.; Walvoort, D.; De Baardemaaker, J. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. *Int. J. Remote Sens.* **2006**, *27*, 21. [[CrossRef](#)]
26. Hansen, P.M.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* **2003**, *86*, 542–553. [[CrossRef](#)]
27. Zhu, Y.; Yao, X.; Tian, Y.C.; Liu, X.J.; Cao, W.X. Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 1–10. [[CrossRef](#)]
28. Xue, L.H.; Cao, W.X.; Luo, W.H.; Dai, T.B.; Zhu, Y. Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agron. J.* **2004**, *96*, 135–142. [[CrossRef](#)]
29. Eitel, J.U.H.; Long, D.S.; Gessler, P.E.; Smith, A.M.S. Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status. *Int. J. Remote Sens.* **2007**, *28*, 4183–4190. [[CrossRef](#)]
30. Adams, M.L.; Philpot, W.D.; Norvell, W.A. Yellowness index: An application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *Int. J. Remote Sens.* **1999**, *20*, 3663–3675. [[CrossRef](#)]
31. Serrano, L.; Penuelas, J.; Ustin, S.L. Remote sensing of nitrogen and lignin in mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens. Environ.* **2002**, *81*, 355–364. [[CrossRef](#)]

32. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [[CrossRef](#)]
33. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
34. Fitzgerald, G.J.; Rodriguez, D.; Christensen, L.K.; Belford, R.; Sadras, V.O.; Clarke, T.R. Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precis Agric.* **2006**, *7*, 233–248. [[CrossRef](#)]
35. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 261–270. [[CrossRef](#)]
36. Richardson, A.J.; Wiegand, C.L. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.
37. Liu, H.Q.; Huete, A. Feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Geosci. Remote Sens. Soc.* **1995**, *33*, 457–465. [[CrossRef](#)]
38. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
39. Gong, P.; Pu, R.; Biging, G.S.; Larrieu, M.R. Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1355–1362. [[CrossRef](#)]
40. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
41. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [[CrossRef](#)]
42. Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. *Nasa Spec. Publ.* **1974**, *351*, 309.
43. Goel, N.S.; Qin, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer simulation. *Remote Sens. Rev.* **1994**, *10*, 309–347. [[CrossRef](#)]
44. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
45. Roujean, J.L.; Breon, F.M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [[CrossRef](#)]
46. Tang, S.; Zhu, Q.; Wang, J.; Zhou, Y.; Zhao, F. Theoretical bases and application of three gradient difference vegetation index. *Sci. China Ser. D* **2003**, *33*, 9. (In Chinese)
47. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, *76*, 156–172. [[CrossRef](#)]
48. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [[CrossRef](#)]
49. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. spectral features and relation to chlorophyll estimation. *J. Plant Phys.* **1994**, *143*, 286–292. [[CrossRef](#)]
50. Gitelson, A.A.; Viña, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett. Banner* **2003**, *30*, 1248. [[CrossRef](#)]
51. Dash, J.; Curran, P.J. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
52. Penuelas, J.; Filella, I.; Serrano, L.; Savé, R. Cell wall elasticity and water index (R970 nm/R900 nm) in wheat under different nitrogen availabilities. *Int. J. Remote Sens.* **1996**, *17*, 373–382. [[CrossRef](#)]
53. Gao, B.C.; Goetz, A.F. Retrieval of equivalent water thickness and information related to biochemical components of vegetation canopies from AVIRIS data. *Remote Sens. Environ.* **1995**, *52*, 155–162. [[CrossRef](#)]
54. Hardisky, M.; Klemas, V.; Smart, R.M. The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of spartina alterniflora canopies. *Photogramm. Eng. Remote Sens.* **1983**, *48*, 77–84.
55. Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting sugarcane ‘orange rust’ disease using EO-1 Hyperion hyperspectral imagery. *Int. J. Remote Sens.* **2004**, *25*, 489–498. [[CrossRef](#)]

56. Delalieux, S.; Somers, B.; Hereijgers, S.; Verstraeten, W.W.; Keulemans, W.; Coppin, P. A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. *Remote Sens. Environ.* **2008**, *112*, 3672–3772. [[CrossRef](#)]
57. Xu, L.; Zhang, W.J. Comparison of different methods for variable selection. *Anal. Chim. Acta* **2001**, *446*, 477–483. [[CrossRef](#)]
58. Wold, S.; Ruhe, H.; Wold, H.; Dunn, W.J. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *Siam J. Sci. Stat. Comput.* **1984**, *5*, 735–743. [[CrossRef](#)]
59. Kamruzzaman, M.; Elmasry, G.; Sun, D.W.; Allen, P. Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Anal. Chim. Acta* **2012**, *714*, 57–67. [[CrossRef](#)]
60. Lü, J.; Hao, N.; Cui, X. Inversion model for copper content in farmland of tailing area based on visible-near infrared reflectance spectroscopy. *Trans. Chin. Soc. Agric. Eng.* **2015**, *6*, 265–270.
61. Han, Z.; Zhu, X.; Fang, X.; Wang, Z.; Wang, L.; Zhao, G.; Jiang, Y. Hyperspectral estimation of apple tree canopy LAI based on SVM and RF regression. *Spectrosc. Spect. Anal.* **2016**, *36*, 6.
62. Darvishzadeh, R.; Atzberger, C.; Skidmore, A.K.; Abkar, A.A. Leaf area index derivation from hyperspectral vegetation indices and the red edge position. *Int. J. Remote Sens.* **2009**, *30*, 6199–6218. [[CrossRef](#)]
63. Yue, J.B.; Feng, H.K.; Yang, G.J.; Li, Z.H. A Comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sens.* **2018**, *10*, 66. [[CrossRef](#)]
64. Xu, X. *Remote Sensing Physics*; Peking University Press: Beijing, China, 2006.
65. Wang, K.R.; Pan, W.C.; Li, S.K.; Chen, B.; Xiao, H.; Wang, F.Y.; Chen, J.L. Monitoring models of the plant nitrogen content based on cotton canopy hyperspectral reflectance. *Spectrosc. Spect. Anal.* **2011**, *31*, 1868–1872. [[CrossRef](#)]
66. Liang, J.; Hu, K.; Tian, M.; Wei, D.; Li, H.; Bai, Y.; Jun-zhen, Z. Diagnosis of nitrogen content in upper and lower corn leaves based on hyperspectral data. *Spectrosc. Spect. Anal.* **2013**, *33*, 1032–1037.
67. Luo, P.; Guo, J.; Li, Q. Modeling construction based on partial least-squares regression. *J. Tianjin Univ.* **2002**, *35*, 783–786.
68. Wang, H.B.; Zhao, Z.Q.; Lin, Y.; Feng, R.; Li, L.G.; Zhao, X.L.; Wen, R.H.; Wei, N.; Yao, X.; Zhang, Y.S. Leaf area index estimation of spring maize with canopy hyperspectral data based on linear regression algorithm. *Spectrosc. Spect. Anal.* **2017**, *37*, 1489–1496.
69. Breiman, L. Random forests. *Mach. Learn* **2001**, *45*, 5–32. [[CrossRef](#)]
70. Xu, X.; Zhao, C.; Wang, J.; Li, C.; Yang, X. Associating new spectral features from visible and near infrared regions with optimal combination principle to monitor leaf nitrogen concentration in barley. *Int. J. Infrared Millimeter Waves* **2013**, *32*, 9. [[CrossRef](#)]

