Research Article

Prediction Model of the Key Components for Lodging Resistance in Rapeseed Stalk Using Near-Infrared Reflectance Spectroscopy (NIRS)

Jie Kuai,[1] Shengyong Xu [ID],[2] Cheng Guo,[1] Kun Lu,[2] Yaoze Feng,[2] and Guangsheng Zhou [ID][1]

[1]College of Plant Science & Technology, Huazhong Agricultural University, Wuhan 430070, China
[2]College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

Correspondence should be addressed to Guangsheng Zhou; zhougs@mail.hzau.edu.cn

The chemical composition of rape stalk is the physiological basis for its lodging resistance. By taking the advantage of NIRS, we developed a rapid method to determine the content of six key composition without crushing the stalk. Rapeseed stalks in the mature stage of growth were collected from three cultivation modes over the course of 2 years. First, we used the near-infrared spectroscope to scan seven positions on the stalk samples and took their average to form the spectral data. The stalks were then crushed and sieved; then the ratio of carbon and nitrogen, ratio of acid-insoluble lignin and lignin, and the content of soluble sugar and cellulose were determined using the combustion method, weighing method, and colorimetric method, respectively. The partial least squares regression (PLSR) method was used to establish a prediction model between the spectral data and the chemical measurements, and all models were evaluated by an internal interaction verification and an external independent test set sample. To improve the accuracy of the model and reduce the computing time, some optimization methods have been applied. Some outliers were removed, and then the data were preprocessed to determine the best spectral information band and the optimal principal component number. The results showed that elimination of outliers effectively improved the precision of the prediction model and that no spectral pretreatment method exhibited the highest prediction accuracy. In summary, the NIRS-based prediction model could facilitate the rapid nondestructive detection in the key components of rapeseed stalk.

1. Introduction

Rapeseed is one of the most widely cultivated oilseeds crops worldwide, and it is the main source of vegetable oil and protein. In the field, stalk lodging is the key limiting factor for the further improvement of rapeseed yield. Similar to other crops, the content of lignin and cellulose in the rapeseed stalk is closely related to its lodging resistance [1]. In addition, rapeseed stalk is an important renewable resource; the rapid quantification of its major components effectively improves the utilization rate of the stalk [2]. Determination of the main components of stalk lodging resistance in regard to massive rapeseed breeding practice can accelerate the breeding of rapeseed with lodging resistance. However, the commonly used chemical testing methods are time- and labor-consuming with low efficiency. On this basis, a rapid method to evaluate the key components of lodging resistance is desired. Such an advance could provide technical support for the study of lodging resistance in rapeseed.

The near-infrared reflectance spectroscopy (NIRS) technology is rapid and highly efficient and has been applied in the agriculture and many other crops, such as in rice [3], citrus [4], and rapeseed [5]. With the NIRS and multiple linear regression methods, Ding et al. established a quantitative analytical model of glucose and fructose to evaluate their content in honey [6]. Bagherpour et al. researched an optical method based on near-infrared spectroscopy (900–1600 nm) to determine the content of soluble solid and sucrose in sugar beet [7]. Rungpichayapichet et al. developed

calibration models for firmness, total soluble solids (TSS), titratable acidity (TA), and ripening index (RPI) using PLSR analysis. The results indicated that NIRS can be used as a reliable nondestructive technique for mango quality assessment [8]. Mathison et al. adopted NIRS for the prediction of the nutritive value of straw with 195 samples of barley straw. They showed that NIRS was a useful method for predicting chemical composition of straw and estimating its ruminal degradability characteristics [9]. Kaur et al. developed a calibration equation for oil content using NIRS in *Brassica juncea* and *Brassica napus*. The reference values of oil content were generated by nuclear magnetic resonance (NMR). The predicted model was validated in the case of *B. juncea* with an r^2 value 0.85. The results indicated that NIRS could be used satisfactorily for rapid determination of oil content in *B. juncea* [10]. Teye et al. assessed the feasibility of measuring total fat content in cocoa beans by using Fourier transform near-infrared (FT-NIR) spectroscopy based on a systematic study of spectral variable selection via multivariate regression. Experimental results showed that the model based on the novel Si-SVMR algorithm was superior to the others [11]. Shetty and Gislum presented a method of NIRS combined with chemometrics which was used to quantify fructan concentration in samples from seven grass species. The PLSR approach was used on the full spectra to model NIR spectroscopy data [12]. NIRS predictive equations were used to provide accurate high-throughput phenotyping of seed content, opening new perspectives in gene identification following QTL mapping and genome-wide association studies [13].

Tremendous efforts have been made in the study on the straw composition using multispectral technology. Niu et al. used hyperspectral imaging technology to rapidly determine the content of five elements: nitrogen, carbon, hydrogen, sulfur, and oxygen, in the straw of maize, rice, wheat, and rapeseed. Their analysis was achieved through applying the optimal method combined with a competitive adaptive reweighted sampling-partial least squares algorithm (CARS-PLS). Their results indicated that the nitrogen and oxygen contents in the straw could be effectively assessed via their methodology [14]. Li et al. used NIRS combined with PLSR to establish the optimal quantitative model for the biocomponents (nitrogen, carbon, hydrogen, sulfur, and oxygen) of straw [15]. XUE investigated an online analysis method of proximate component (moisture, ash, volatile matter, and fixed carbon) and lignocellulose components (cellulose, semicellulose, and lignin) of coarse crushed corn stover using NIRS. After optimized pretreatment, all the NIRS models were effectively developed using the PLSR method [16]. Sheng et al. provided models of three types of agroforestry biomass (moisture, ash, volatile, and fixed carbon) and caloric value of pine, China fir, and cotton stalk via spectral techniques, indicates that the traditional industrial analysis methods could be completely replaced by visible near-infrared spectroscopy in order to achieve the rapid determination of components and calorific value of the agroforestry biomass [17]. Fu et al. collected and prepared samples of various varieties of rice straw from different locations and created a quantitative analysis model based on

the stepwise multiple linear regression (SMLR) of near-infrared spectroscopy, PLSR, and principal component regression (PCR). Data suggested that their model was capable of rapidly testing the content of soluble sugar in rice straw [18]. Lohr developed a partial least squares calibration models for glucose, fructose, sucrose, and starch in leaves of chrysanthemum and pelargonium cuttings by using a stepwise enzymatic-photometric method [19].

Overall, many studies have been carried out to evaluate the content of stem and leaf components in crops using multispectral analysis, and they can effectively overcome the primary drawbacks of traditional chemical testing methods (time consumption and the tedium of the procedures). However, in most of the cases, samples were smashed before the evaluation in these studies. It is difficult to control the homogeneity of material when the sample is smashed, which may increase the error of spectral scanning. In addition, the overall process is usually complicated, and the smashed material is not able to be used after the study is complete. Currently, a method to determine the key components of lodging resistance in rapeseed by spectral analysis without smashing and sieving the rapeseed samples remains unavailable. In the present study, we propose a fast, accurate, and nondestructive NIRS-based method to predict six components including carbon, nitrogen, lignin, acid quality lignin, soluble sugar, and cellulose in noncrushed rapeseed stalk. Our method provides important and fundamental data for the selection of rapeseed with strong lodging resistance and allows the stalk to be used after analysis for other purposes.

2. Materials and Methods

2.1. Experimental Materials. The plant materials were obtained from the experimental station of Huazhong Agricultural University in both 2017 and 2018. Rapeseeds were bunch planted in September and thinned at the third to fifth leaf stages. The field was flat with medium fertility, and the previous crop was rice harvested during early September. The plots were $10 \, \text{m} \times 2 \, \text{m}$. *Brassica napus* hybrids HYZ62 and FY520, common Swede-type rapeseeds HH901 and HS5, and the DH population consisting of 150 lines (TN) were used as plant materials. Samples were collected during the maturity stage. Plants representing the average yield and lodging status were selected. After removing the roots of samples, the plants were dried and bagged separately. The sample will be dried twice at 105°C and at 80°C until achieving constant weight separately. The cultivation methods in the present study were as follows:

(1) Three-factor split plot experiment with three replicates: the four rapeseed varieties served as the main plot, fertilizer type served as the primary split plot, and the level of N fertilizer (0 kg/hm², 180 kg/hm², and 360 kg/hm² of pure nitrogen), P (0 kg/hm², 120 kg/hm², and 240 kg/hm² of P_2O_5), and K (0 kg/hm², 150 kg/hm², and 300 kg/hm² of K_2O) served as the secondary split plot, so a total of 108 treatments with different NPK ratios were carried out.

(2) Three-factor split plot experiment with three replicates: the four rapeseed varieties served as the main plot, the levels of N fertilizer served as the primary split plot (120 kg/hm^2, 240 kg/hm^2, and 360 kg/hm^2 and urea with 46.7% nitrogen content was used as N source), and three planting densities (15×104 plants/hm^2, 30×104 plants/hm^2, and 45×104 plants/hm^2) served as the secondary split plot. Phosphorus (calcium superphosphate with 12% P content), potassium (KCl with 60% K), and borate fertilizer were applied once as the base fertilizer with 150 kg/hm^2 for P and K and 7.5 kg/hm^2 of borax.

(3) Complete random block design with three replicates: each variety was planted one row with a length of 250 cm, a row spacing of 30 cm, and a plant spacing of 21 cm. Compound fertilizer ($15 : 15 : 15$) of 750 kg/hm^2 with 7.5 kg/hm^2 of borax was used as base fertilizer, and 75 kg/hm^2 of urea was additionally applied during the seedling stage. The filed management was carried out as regular practice.

A Fourier near-infrared spectroscope BRUKERFT-NIR (VECTOR33N, Bruker Inc., Germany) was equipped with PbS detector, quartz rotating sample cup, gold-plated integrating sphere, and OPUS analysis software with the wave number of 12000 cm^{-1}–4000 cm^{-1} and the resolution of 8 cm^{-1}. The near-infrared spectroscope was preheated for 20 min before scanning at room temperature. The sample was put into a sample cell at the same position each time and then covered by the gold-plated integrating sphere to avoid light leak. To ensure the accuracy and integrity of the scanning procedure, seven positions of each sample would be scanned for seven spectra, and then the average of the seven spectra data was taken to build the model. Using a blade, the sample was evenly cut into three parts, and then the middle piece of each part was cut off. As shown in Figure 1(a), samples demarcated as 1 and 2 were the end faces of the stalk. 3 and 4 were the fresh cut surface to avoid variation errors due to long time exposures in the air, and 5–7 were cross sections of the stalk. Figure 1(b) shows the spectra of the samples. All spectral preprocessing was performed in the OPUS software (OPUS 7.0, BRUKER OPTICS, Germany). The SPASS software (IBM SPSS Statistics 24.0, IBM, USA) was adopted for data analysis and mapping.

The stalks were dried after scanning and then ground using a universal high-speed smashing machine (FW100, Tianjin TAISITE Instrument Co., LTD, China) to measure their chemical components. Carbon and nitrogen were determined using an elemental analyzer (Elementar vario max cn, Germany). Oxygen was added to the samples of 250 mg for 90 s, and various forms of nitrogen and carbon in the sample were converted to stable nitrogen and carbon dioxide by the combustion method (the combustion tube of $900°C$ and the reduction tube of $830°C$), and then the contents were tested by an infrared detection module. The ratio of acid-insoluble lignin and lignin was determined using the sulfuric acid method [20]. The cellulose ratio was evaluated by colorimetry [21]. The ratio of soluble sugar was detected by the anthrone colorimetry method [22]. The results of the chemical analysis are shown in Table 1.

2.2. Constructive Process of the Model. The accuracy of model creation is an important step to ensure the predicted results are close to the true chemical values in the plant matter [23]. The mean spectra were obtained by averaging the seven spectra scanned using OPUS software, and the model was further established using the PLSR method and the obtained chemical data. We conducted the following procedures.

2.2.1. Partitioning of Correction Set and Validation Set. The experimental data were arranged randomly. The division ratio of this paper is approximately $3 : 1$, which is very close to the traditional division ratio of $7 : 3$. Figure 2 shows the partition of correction set and validation set of the stalk components in rapeseed. Of these, CA and VA represented the calibration set and validation set, respectively. The data in Table 2 indicated that the samples set exhibited the desired uniformity.

2.2.2. Elimination of Outliers. Changes in environmental factors are also important, such as temperature, humidity, and air flow rate, which affect the accuracy of the spectrum [24]. Therefore, eliminating outliers from the sample set is crucial. In the present study, the outliers were removed based on the principle of predicting concentration residuals [25]. Leave-one-out cross validation was performed each time a sample was deleted, and the samples that increased the correlation between the true value and predicated value (R) while decreasing the root-mean-square error of cross validation (RMSECV) were considered outliers. About 1/6 to 1/5 of the samples in the prediction set were classified as outliers and eliminated from the sample set.

2.2.3. Optimization of Spectral Preprocessing. A total of 11 methods were employed to preprocess the spectra after eliminating the outliers, including elimination constant offset, straight line subtraction, vector normalization, min-max normalization, multiplication scatter correction, the first-order derivative correction, the second-order derivative correction, the first-order derivative and straight line subtraction, first-order derivative and vector normalization, first-order derivative and multiplication scattering correction (MSC), and no preprocessing [26]. Using the sample after removing the outlier, the optimal processing method was selected by comparing R and RMSECV.

2.2.4. Spectral Band Optimization. Choosing the right band can improve the accuracy of the model and reduce the running time. After determining the optimal spectral preprocessing, the spectrum was divided into 45 bands, and the correlation coefficient and the mean square deviation of internal verification at different bands were observed. The
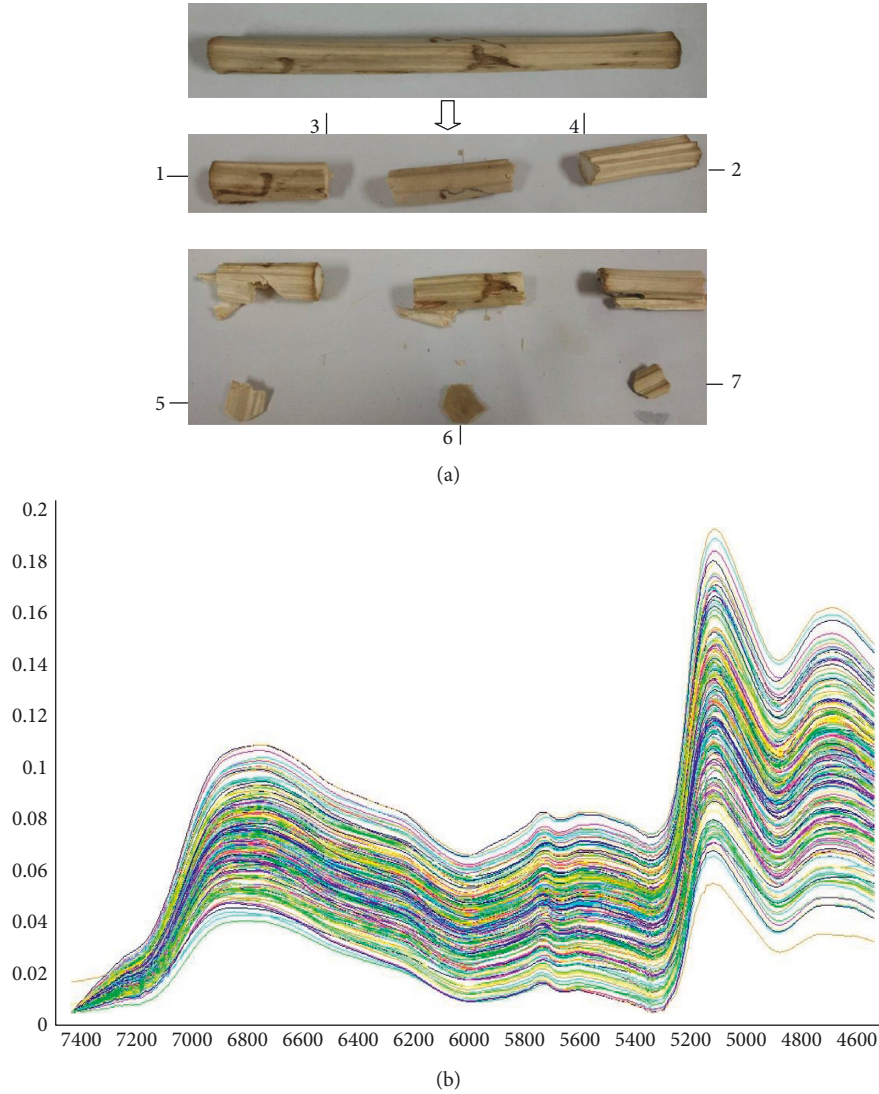
(a)

(b)

FIGURE 1: (a) Schematic of the scanning position on rapeseed stalks. (b) Spectra of the samples.

TABLE 1: Statistical results of six rapeseed stem components.

Content	Number	Maximum (%)	Minimum (%)	Mean (%)	Standard deviation	Coefficient of variation
Carbon	340	46.37	41.13	44.18	1.05	0.024
Nitrogen	340	0.68	0.12	0.34	0.09	0.260
Lignin	240	34.96	18.64	25.27	3.41	0.135
Acid quality lignin	240	38.51	13.69	23.92	4.58	0.192
Soluble sugar	200	5.68	1.52	3.18	0.94	0.295
Cellulose	200	38.27	9.69	23.47	6.23	0.265

optimal wave number was chosen as the one with the smallest RMSECV.

2.2.5. Determination of the Optimal Principal Component Number. Determining the optimal principal component can reduce the dimension. It not only lowers the number of input data needed to create a model but also speeds up the operation of the model with improved prediction accuracy. During the model creation, 1–10 were tested,

respectively, as the principal component number, and the value showing the maximum R and minimum RMSECV of the PLSR model was determined as the optimal principal component number.

2.2.6. Validation of the Calibration Model. To verify whether the accuracy of the model meets the requirement of component prediction, the spectra of the samples from the validation set were introduced into the established model to obtain
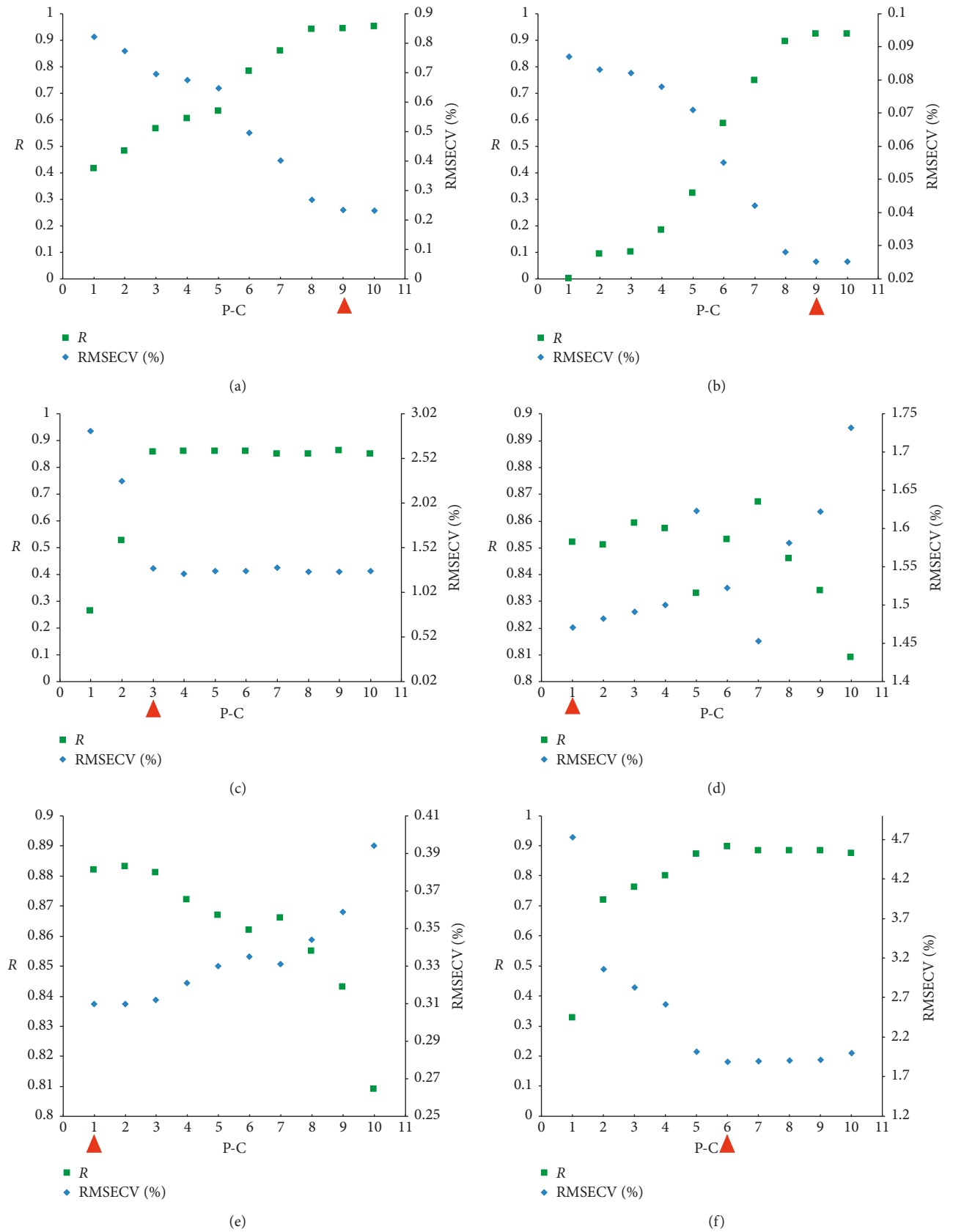
Figure 2: Relationship between principal component number and R/RMSECV of the prediction model.

TABLE 2: Division of the calibration set and validation set.

Index	Set type	Quantity of samples	Maximum value (%)	Minimum value (%)	Mean value (%)	Standard deviation
Carbon	CA	250	46.37	41.13	44.17	1.03
	VA	90	46.37	41.23	44.21	1.12
Nitrogen	CA	250	0.68	0.12	0.34	0.08
	VA	90	0.68	0.13	0.33	0.10
Lignin	CA	180	34.96	18.64	25.27	3.16
	VA	60	34.95	19.64	25.25	4.08
Acid quality lignin	CA	180	34.80	14.83	23.65	4.12
	VA	60	38.51	13.70	24.72	5.71
Soluble sugar	CA	150	5.68	1.52	3.17	0.88
	VA	50	5.55	1.53	3.21	1.12
Cellulose	CA	150	38.27	9.69	23.42	5.80
	VA	50	38.27	11.36	23.63	7.43

predication values (predicted by the model) and true values (from the chemical test). These values were used to calculate the R and RMSECV between prediction values and standard values.

3. Results and Analysis

3.1. Effect of Eliminating Outliers. In the modeling based on NIRS, there were two types of outliers in the calibration set. One type was the sample with significant difference between the chemical determination value and the predicted value, which may be caused by the large error of chemical measurement, the spectral measurement error, or the error during data entry. These abnormal samples must be eliminated before modeling. The other type was high-leverage samples. Comparing with other samples, this type sample contains extreme components and far away from the average of the overall sample. These abnormal samples were obviously not useful for global modeling because of destroying the uniformity of the sample distribution. But those were useful for enriching the calibration set and improving the accuracy of subsequent samples. The presented method in this paper adopted the eliminating outlier method as shown in the literature [24]. In the literature, the proportion of eliminating outliers reached 36.7%. Considering the balance performance between global modeling and local modeling, the eliminating rate was set at approximately 20%. The eliminating rate of carbon, nitrogen, soluble sugar, and cellulose was 20%. The eliminating rate of lignin and acid quality Lignin was 16.7%. The remaining samples were then used to create the model using the PLSR method. Compared to the results obtained without eliminating outliers, using the principal component number of nine for the carbon calibration set, R between standard value and predicted value was increased from 0.960 to 0.973, and RMSECV was decreased from 0.286% to 0.240% with a reduction of predictive residual to 0.600%. The principal component number of nitrogen calibration set remained at nine, R increased from 0.929 to 0.96, RMSECV was reduced from 0.030% to 0.024%, and the predictive residual lowered to 0.6%. The principal component number of lignin calibration set

remained as three, R elevated from 0.860 to 0.923, RMSECV reduced from 1.610% to 1.250%, and predictive residual lowered to 0.6%. The principal component number of acid-insoluble lignin calibration set remained as one, R was increased from 0.893 to 0.928, and RMSECV was reduced from 1.850% to 1.480%, and in this satuation the predictive residual was 4.000%. The principal component number of soluble sugar calibration set changed from one to nine, exhibiting an increase of R from 0.910 to 0.939, the reduction of RMSECV from 0.900% to 0.310%, and the decrease of predictive residual to 3.000%. The principal component number of celluloses in the calibration set was unchanged to 6 with the increase of R from 0.917 to 0.945, the decrease of RMSECV from 2.310% to 1.880%, and the predictive residual lowered to 5.000%. In conclusion, the accuracy of the prediction model was improved slightly after eliminating outliers in calibration set. The increase rate of R is between 1.3% and 6.3%, and the decrease rate of RMSECV is between 0.026% and 0.59%. The prediction model may also have the capability to modeling accuracy without eliminating outliers.

3.2. Optimization of the Spectral Preprocessing. When using the full band spectrum to build the model, the spectral information content is large, resulting in enormous computation and processing time. Occasionally, there are other factors such as collinearity in spectral information, which leads to the incapability to extract relevant information from the models with full optical spectrum. Such concerns can be avoided after optimizing the spectral band. The comparison of 11 spectral pretreatments for the prediction models of carbon content is shown in Table 3. The R and RMSECV indicated that the best approach was without spectral preprocessing, and their values were 0.973 and 0.240%, respectively. The suboptimal choice was a first-order derivative, and R and RMSECV were 0.942 and 0.356%, respectively. R and RMSECV of the first-order derivative were decreased by 0.031 and 0.116%, respectively. The least desirable processing was vector normalization, in which R was the minimum and RMSECV was the maximum. Overall, the best way to build the model for carbon content was with

Table 3: The result of carbon spectrum using eleven spectral preprocessing methods.

Number	Pretreatment method	Wave number (cm^{-1})	Principal component number	R	RMSECV (%)
1	No spectral pretreatment	7501.7–4246.5	9	0.973	0.240
2	First-order derivative	7501.7–4246.5	10	0.942	0.356
3	Straight line subtraction	7501.7–4597.5	10	0.928	0.394
4	Constant offset elimination	7501.7–4597.5	10	0.923	0.407
5	First-order derivative + straight line subtraction	7501.7–4246.5	9	0.922	0.411
6	First-order derivative + MSC	7501.7–4246.5	8	0.916	0.424
7	First-order derivative + vector normalization	7501.7–4246.5	8	0.915	0.426
8	Second-order derivative	7501.7–4246.5	9	0.911	0.436
9	Multiplication scattering correction	7501.7–6097.8 5453.7–4246.5	6	0.908	0.443
10	Min-max normalization	6799.8–6097.8 5453.7–4246.5	8	0.897	0.469
11	Vector normalization	7501.7–4246.5	9	0.896	0.469

no spectral preprocessing and/or first-order derivative processing. The same approach was applied to another five elements, and no preprocessing was optimal for all in each case.

Generally, samples must be ground and placed in a sample cell for the determination of components in straw. For example, when evaluating the nutrients in corn stalk, the best spectral preprocessing of soluble sugar and acid detergent fiber is the first-order derivative and MSC [2]. However, in the present study, no spectral pretreatment was the optimal preprocessing for all spectra, and only the optimal number of principal component number and spectral range differed, indicating that the original spectrogram obtained by scanning was the best way to build a model. This may be due to (1) in the present study, the scanning object is an intact rapeseed stalk without treatments, such as crushing and sieving, which reduced operation processes and possible errors, and (2) the gold-plated integrating ball on top of the sample cell fills the gap between rapeseed stalk and the cell, allowing stalks with various diameters to hold good positioning in the machine, which improved the homogeneity of samples and prevents the interference of natural light.

3.3. Optimization of Spectral Band. Using the approach of no spectral preprocessing and PLSR to build the model, the full spectrum was divided into 45 bands. The relationship of wave number and RMSECV in the prediction model of N from our measurements is shown in Table 4. In these models for nitrogen, the wave number of 7501.7 cm^{-1}–5449.8 cm^{-1} and 4601.3 cm^{-1}–4246.5 cm^{-1} exhibited the minimal RMSECV of 0.024%. The worst/highest RMSECV value of 0.081% was obtained at 11998.9 cm^{-1}–7497.9 cm^{-1}. Our data suggested that selecting the appropriate waveband can effectively improve the accuracy of the model. The same method was applied for the other 5 elements, and the optimal wave number was 7501.7 cm^{-1}–4246.5 cm^{-1}, 7501.7 cm^{-1}–6097.8 cm^{-1}, and 4601.3 cm^{-1}–4423.9 cm^{-1}; 11998.9 cm^{-1}–7497.9 cm^{-1} and 5453.7 cm^{-1}–4246.5 cm^{-1}; 11998.9 cm^{-1}–5449.8 cm^{-1}; 7501.7 cm^{-1}–5449.8 cm^{-1}; and

4601.3 cm^{-1}–4246.5 cm^{-1} for carbon, lignin, acid-insoluble lignin, soluble sugar, and cellulose, respectively.

3.4. Determination of the Principal Component Number. Without spectral preprocessing, at the best wave number, the principal compoment numer will be determined with the maximum R and minimum RMSECV. The prediction model of carbon content at the waveband of 7501.7 cm^{-1}–4246.5 cm^{-1} with no preprocessing is shown in Figure 2(a), where the x-axis P-C represents principal component number. Within the number of principal components of 1–8, R was elevated as the increase in the number, and RMSECV was gradually decreased, indicating that the extracted spectral information was gradually completed during the establishment of the model. When the number of principal component number reached 10, R was lower and RMSECV was higher than those with the number of 9, suggesting the existence of over fitting when excessive information was extracted. The optimal number of principal component number was 9 with the highest R between true value and predicted value of 0.973 and the RMSECV of 0.240%. Therefore, the best number of principal components was determined as 9. The same approach was applied to the other five contents as shown in Figures 2(b)~2(f), and the optimal numbers of principal component number were 9, 9, 3, 1, 1, and 6, respectively. In generally, the R/RMSECV is decreased or increased gradually and continuously. The relationship between the number and R/RMSECV changes regularly, except for the model of acid quality lignin. However, Figure 2(d) shows that when the principal component number was 5 or 7, R/RMSECV is fluctuating drastically and deviated from the regular law. Therefore, 7 was regarded as an outlier and then 1 was considered as the best principal component number.

3.5. Validation of the Corrected Model. The predication results of all six components are shown in Table 5. The models were evaluated based on the previous method [27] in which the correlation coefficient of the calibration set R served as the

TABLE 4: The relationship between spectral wave number and RMSECV for nitrogen.

Number	Wave number (cm^{-1})	RMSECV (%)
1	7501.7–5449.8 4601.3–4246.5	0.024
2	7501.7–4246.5	0.025
3	7501.7–5449.8 4601.3–4423.9	0.030
4	6101.7–4246.5	0.030
5	6475.8–5449.8 4601.3–4246.5	0.031
6	7501.7–6097.8 4601.3–4246.5	0.033
7	6475.8–5449.8 4601.3–4423.9	0.033
8	7501.7–6475.8 4601.3–4246.5	0.036
9	7501.7–6097.8 5453.7–4246.5	0.036
10	7501.7–5449.8 4423.9–4246.5	0.037
11	7501.7–6475.8 4423.9–4246.5	0.038
12	6101.7–4597.5	0.040
13	6101.7–5449.8 4601.3–4246.5	0.041
14	7501.7–4597.5	0.041
15	5453.7–4246.5	0.042
16	7501.7–5449.8	0.042
17	7501.7–5449.8	0.042
18	6475.8–5449.8 4423.9–4246.5	0.044
19	7501.7–6097.8 5453.7–4597.5	0.050
20	6475.8–5449.8	0.050
21	7501.7–6475.8 4601.3–4423.9	0.051
22	4601.3–4246.5	0.052
23	4601.3–4246.5	0.052
24	7501.7–6097.8	0.056
25	6101.7–5449.8	0.058
26	7501.7–6475.8	0.059
27	11998.9–7497.9 6101.7–4246.5	0.060
28	11998.9–4246.5	0.061
29	5453.7–4597.5	0.061
30	4601.3–4423.9	0.063
31	11998.9–5449.8 4601.3–4246.5	0.062
32	11998.9–6097.8 5453.7–4246.5	0.063
33	4423.9–4246.5	0.064
34	11998.9–6097.8 4601.3–4246.5	0.065

TABLE 4: Continued.

Number	Wave number (cm^{-1})	RMSECV (%)
35	11998.9–7497.9 6101.7–5449.8 4601.3–4246.5	0.065
36	11998.9–4597.5	0.066
37	11998.9–7497.9 5453.7–4246.5	0.066
38	11998.9–7497.9 6101.7–4597.5	0.0657
39	11998.9–7497.9 4601.3–4246.5	0.066
40	11998.9–6097.8 5453.7–4597.5	0.068
41	11998.9–7497.9 5453.7–4597.5	0.069
42	11998.9–5449.8	0.071
43	11998.9–7497.9 6101.7–5449.8	0.072
44	11998.9–6097.8	0.072
45	11998.9–7497.9	0.081

criteria. If $0.900 < R < 0.950$, the correction model was considered successful, and if $R > 0.950$, the correction model was considered extremely successful. In the present study, the correction model of carbon and nitrogen content was extremely successful, while the models for lignin, acid-insoluble lignin, soluble sugar, and cellulose were successful. RMSECV and root-mean-square error of prediction (RMSEP) are also factors used to evaluate the model, both of which are generally less than one when creating a model. Combining all three values, the models of carbon, nitrogen, and soluble sugar were extremely successful, while others were successful, and all models met the requirement of prediction.

4. Conclusion

In the present study, we used 340 samples to build a model for carbon and nitrogen, 240 samples for lignin and acid-insoluble lignin, and 200 samples for cellulose and soluble sugar. As compared to the previous similar studies, the sample size in the present study was larger, the treatments were more various, and the applicability was wider. The content of carbon, nitrogen, acid-insoluble lignin, soluble lignin, and cellulose obtained using chemical methods was comparable to those from [28, 29], indicating these chemical methods were reliable, so the data could serve as a good base line value.

Overall, the tested materials in this experiment were highly representative, and the above models met the needs of component determination. Compared to sample crushing, this approach ensures the integrity of rapeseed samples by avoiding damage to the rapeseed stalk during the grinding process, so the tested samples can be further used for other purposes. If the content of samples from various locations or different years is desired, other components can be added to modify the model to meet wider applicability requirements.

TABLE 5: The test result of the model.

Index	Pretreatment method	Wave number (cm^{-1})	Principal component number	Calibration set		Validation set	
				R	RMSECV (%)	R	RMSEP (%)
Carbon	No spectral pretreatment	7501.7–4246.5	9	0.973	0.240	0.887	0.348
Nitrogen	No spectral pretreatment	7501.7–5449.8 4601.3–4246.5	9	0.959	0.024	0.875	0.033
Lignin	No spectral pretreatment	7501.7–6097.8 5453.7–4246.5	3	0.923	1.250	0.760	1.691
Acid-insoluble lignin	No spectral pretreatment	11998.9–7497.9 5453.7–4246.5	1	0.928	1.480	0.766	2.448
Soluble sugar	No spectral pretreatment	11998.9–6097.8	1	0.939	0.310	0.833	0.428
Cellulose	No spectral pretreatment	7501.7–5449.8	6	0.945	1.880	0.863	2.600

In addition, the problems and solutions encountered in creating the model herein can provide reference for other similar studies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

Kuai Jie and Xu Shengyong are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Zhang, B. Tian, G. Shi, L. Huang, and L. Yang, "Effects of the key enzymes activity on the stem lodging resistance in *Brassicanapus* L.," *Chinese Agricultural Science Bulletin*, vol. 27, no. 18, pp. 135–138, 2011.

[2] J. Xue, L. Han, Z. Yang, G. Shen, and Y. Liu, "On-line measurement of nutrient content of corn stover using NIRS," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 47, no. 7, pp. 216–223, 2016.

[3] S. Oh, M. C. Lee, M. C. Yu, S. Lee, and M. Rauf, "Protein contents in collected rice landraces germplasm using near-infrared reflectance spectroscopy (NIRS)," *Korean Journal of Plant Resources*, vol. 30, no. 4, pp. 450–465, 2017.

[4] I. Torres, D. Pérez-Marín, M.-J. De la Haba, and M.-T. Sánchez, "Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors," *Biosystems Engineering*, vol. 153, pp. 140–148, 2017.

[5] R. Font, M. d. Río-Celestino, E. Cartea, and A. de Haro-Bailón, "Quantification of glucosinolates in leaves of leaf rape (*Brassica napus* ssp. *pabularia*) by near-infrared spectroscopy," *Phytochemistry*, vol. 66, no. 2, pp. 175–185, 2005.

[6] J. Ding, Q. Zhang, S. Li, and H. Liu, "Fast detection of glucose and fructose content in honey using NIR spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 36, no. 10, pp. 197-198, 2016.

[7] H. Bagherpour, S. Minaei, N. M. Abdollahian, and F. M. E. Khorasani, "Non-destructive determination of sugar content in root beet by near infrared spectroscopy (NIRS)," *Iranian Journal of Food Science and Technology*, vol. 12, no. 46, pp. 219–228, 2015.

[8] P. Rungpichayapichet, B. Mahayothee, M. Nagle, P. Khuwijitjaru, and J. Muller, "Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango," *Postharvest Biology & Technology*, vol. 111, pp. 31–40, 2016.

[9] G. W. Mathison, H. Hsu, R. Soofi-Siawash et al., "Prediction of composition and ruminal degradability characteristics of barley straw by near infrared reflectance spectroscopy," *Canadian Journal of Animal Science*, vol. 79, no. 4, pp. 519–523, 1999.

[10] B. Kaur, M. K. Sangha, and G. Kaur, "Development of near-infrared reflectance spectroscopy (NIRS) calibration model for estimation of oil content in *Brassica juncea* and *Brassica napus*," *Food Analytical Methods*, vol. 10, no. 1, pp. 1–7, 2016.

[11] E. Teye and X. Huang, "Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression," *Food Analytical Methods*, vol. 8, no. 4, pp. 945–953, 2015.

[12] N. Shetty and R. Gislum, "Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR," *Field Crops Research*, vol. 120, no. 1, pp. 31–37, 2011.

[13] S. Jasinski, A. Lecureuil, M. Durandet, P. Bernardmoulin, and P. Guerche, "ArabidopsisSeed content QTL mapping using high-throughput phenotyping: the assets of near infrared spectroscopy," *Frontiers in Plant Science*, vol. 7, no. 3, pp. 1682–1693, 2016.

[14] Z. Niu, X. Li, and H. Gao, "Rapid detection of element content in straw biomass using hyperspectral imaging technology," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 30, no. 22, pp. 181–187, 2014.

[15] X. Liu, K. Zhu, Z. Niu, and X. Cheng, "NIRS rapid detection of elemental analysis for straw biomass based on PLS

algorithm," *Journal of Huazhong Agricultural University*, vol. 34, no. 2, pp. 131–135, 2015.

[16] J. Xue, Z. Yang, L. Han, Y. Liu, Y. Liu, and C. Zhou, "On-line measurement of proximates and lignocellulose components of corn stover using NIRS," *Applied Energy*, vol. 137, pp. 18–25, 2015.

[17] K. Sheng, Y. Shen, H. Yang, W. Wang, and W. Luo, "Rapid determination of componential contents and calorific value of selected agricultural biomass feedstocks using spectroscopic technology," *Spectroscopy and Spectral Analysis*, vol. 32, no. 10, pp. 2805–2809, 2012.

[18] M. Fu, M. Liu, Z. Niu, Y. Fang, and Y. Xiang, "Rapidly detecting the content of soluble sugar in rice straw with near infrared reflectance spectroscopy," *Journal of Huazhong Agricultural University*, vol. 35, no. 2, pp. 115–121, 2016.

[19] D. Lohr, P. Tillmann, U. Druege, S. Zerche, T. Rath, and E. Meinken, "Non-destructive determination of carbohydrate reserves in leaves of ornamental cuttings by near-infrared spectroscopy (NIRS) as a key indicator for quality assessments," *Biosystems Engineering*, vol. 158, pp. 51–63, 2017.

[20] G. Fan, L. Zhou, Q. Lu et al., "Determination of lignin content in four common crop straws by sulfuric acid method," *Jiangsu Agricultural Sciences*, vol. 5, no. 5, pp. 395–398, 2010.

[21] Y. Pei, Y. Li, Y. Zhang et al., "G-lignin and hemicellulosic monosaccharides distinctively affect biomass digestibility in rapeseed," *Bioresource Technology*, vol. 203, pp. 325–333, 2016.

[22] B. Xie, Z. Rao, and J. Xiang, "Evaluation of uncertainty of the determination of soluble sugar in tobacco by anthrone colorimetry," *China Measurement & Test*, vol. 41, no. 1, pp. 10–13, 2015.

[23] C. Jia, S. Zhu, and G. Zhao, "Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR," *Food Chemistry*, vol. 221, pp. 1939–1946, 2017.

[24] J. Moros, I. Llorca, M. L. Cervera, A. Pastor, S. Garrigues, and M. de la Guardia, "Chemometric determination of arsenic and lead in untreated powdered red paprika by diffuse reflectance near-infrared spectroscopy," *Analytica Chimica Acta*, vol. 613, no. 2, pp. 196–206, 2008.

[25] S. de la Guardia, Y. Wang, and J. Wu, "Outlier sample eliminating criterions and methods for building calibration model of near infrared spectroscopy analysis," *Transactions of The Chinese Society of Agricultural Machinery*, vol. 35, no. 4, pp. 115–119, 2004.

[26] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.

[27] D. F. Malley, C. McClure, P. D. Martin, K. Buckley, and W. P. McCaughey, "Compositional analysis of cattle manure during composting using a field-portable near-infrared spectrometer," *Communications in Soil Science & Plant Analysis*, vol. 36, no. 4-6, pp. 455–475, 2005.

[28] T. Liu and C. Guan, "Grey relational analysis between lodging iidex and biochemistry components of stem agronomic characteristics in rapeseed (*Brassica napus* L)," *Chinese Journal of Oil Crop Sciences*, vol. 30, no. 2, pp. 152–156, 2008.

[29] J. Kuai, X.-Z. Du, M. Hu et al., "Effect of symbiotic periods and plant densities on growth and yield of rape-seed intercropping cotton," *Acta Agronomica Sinica*, vol. 42, no. 4, pp. 591–599, 2016.