

Spectral inversion model of the crushing rate of soybean under mechanized harvesting

Man CHEN¹ , Youliang NI¹, Chengqian JIN^{1*}, Zheng LIU¹, Jinshan XU¹

Abstract

Rapid and timely acquisition of the crushing rate can help in assessing the performance of combine harvesters, which is very important for agricultural production. The spectral reflectance of soybean provides an alternative method to the classical physical and chemical analysis of the crushing rate of soybean in laboratory. Therefore, hyperspectral imaging can be used to rapidly obtain the crushing rate of soybean. In this study, the hyperspectral method was employed, and the application of inter-correlation analysis was explored in the optimization and quantitative analysis of hyperspectral bands. The crushing rate of 130 soybean samples collected from a combine harvester was investigated through physical analysis in the laboratory. Subsequently, the raw hyperspectral reflectance of soybean samples was measured using a spectroradiometer equipped with a high intensity contact probe under darkroom conditions. Next, the raw spectral reflectance (REF) and the logarithmic reciprocal pretreatment spectrum data (LR) were analyzed and compared. The effective wavelengths were selected according to the results of the inter-correlation analysis. Regression models of the crushing rate with different indices were established using a least squares support vector machine (LS-SVM). The inversion results of the model were validated and compared with each other. The experimental results show that sensitive bands from REF are 1061, 1068, 1074, 1090, 2085, 2092, 2095, and 2103 nm. Sensitive bands from LR are 677, 1039, 1078, 1093, 1101, 1956, 2088, and 2107 nm. The results showed that REF was the optimal spectral index in the LS-SVM regression model (R_c^2 was 0.939, and R_p^2 was 0.915). The inter-correlation analysis method could not only support efficient selection of hyperspectral bands, but also retain the original sample information. The REF hyperspectral inversion model based on LS-SVM can realize rapid on-line monitoring of the performance (crushing rate) of grain combine harvesters in the future.

Keywords: soybean; crushing rate; mechanized operations; least squares support vector machine; hyperspectral remote sensing.

Practical Application: This manuscript is relevant for the soybean production industry, especially for the online testing of soybean harvest quality, because this study provides unedited online testing data on mechanically harvested soybean quality. In this study, a spectral inversion model is used to realize an online, rapid, and accurate detection of soybean quality. This method meets the requirements for online soybean quality detection of the agricultural machinery and agronomy.

1 Introduction

The crushing rate is an important index for assessing the performance of combined harvesting machinery for soybean. According to the National Standard, Operation Quality of Soybean Combine Harvester (NY/T 738-2003), the crushing rate of soybean combined harvesting machinery should be less than 5%. At present, the efficiency of combined harvesting machinery is extremely low because the calculation of the crushing rate still relies on manual methods, and there is no mature online monitoring system for determining the crushing rate in real time (Ni et al., 2019; Xu et al., 2021; Chen et al., 2018a).

In recent years, with the development of spectral detection technologies, hyperspectral imaging has become widely used in modern agricultural production because it offers rapid measurements and high accuracy (Wu et al., 2021; Xiao et al., 2021; Lu et al., 2021; Wei, 2021; Zhu & Wu, 2020; Yikmiş, 2020; Rambo et al., 2020; Cui et al., 2020). Zhang et al. used hyperspectral remote sensing to predict the nutrient content of black soil. The spectrum of black soil was resampled, 5th logarithmic reciprocal and first order differential were derived, and envelope removal and multivariate scattering correction

were performed. Furthermore, a quantitative extraction model of organic matter, nitrogen, phosphorus, and potassium was established. The applicability of five spectral transformation methods was studied by comparing errors between predicted and measured values (Zhang et al., 2018a). Steinberg et al. analyzed the spectral curve absorption characteristics of soils with different organic matter contents. The experimental results showed that the partial least squares regression model based on logarithmic processing of reciprocal reflectivity provided the best prediction effect (Steinberg et al., 2016). Liu et al. used the ASD Field Spec3 Field portable hyperspectral spectrometer to collect canopy hyperspectral data of winter wheat in Laoling city. They used 10 spectral pretreatment methods and combined 3 models (partial least squares regression, BP neural network and random forest algorithm) to establish the hyperspectral prediction model of the nitrogen nutrition index of winter wheat (Liu et al., 2018). Chen et al. discussed the feasibility of remote sensing technology to invert photosynthesis parameters of crop canopy. Unmanned aerial vehicles, equipped with a 6-band multi-spectral camera, were used as the remote sensing

Received 22 Nov., 2021

Accepted 04 Jan., 2022

¹Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing, China

*Corresponding author: 412114402@qq.com

platform. The canopy multi-spectral remote sensing images were collected for different budding stages of cotton to extract canopy spectral reflectance information. Through correlation analysis of four photosynthetic parameters and six bands of spectral reflectance, multiple regression was performed to establish an inversion model of different photosynthetic parameters at different times (Chen et al., 2018b). In order to accurately monitor the early stage of powdery mildew infection in wheat, Huang et al. collected wheat leaves in the early stage of the infection as the test material. They used hyperspectral image data to divide the leaf area and disease spot area according to image features and quantitatively calculated the severity of the disease (Huang et al., 2018). Zhang et al. used near-infrared spectroscopy combined with chemometrics to realize rapid nondestructive detection of soluble protein content in soybean leaves. They applied 7 spectral pretreatment methods to establish a partial least squares prediction model of soluble protein content in soybean leaves (Zhang et al., 2018b). In order to understand the difference of lignin in different crop straws, Yang et al. selected representative cotton stalk, corn stalk, and wheat stalk as experimental materials to extract lignin. Three kinds of straw lignin were characterized through Fourier transform infrared spectroscopy (Yang et al., 2018). Considering the successful application of hyperspectral technology in the agricultural field, this study attempts to apply hyperspectral imaging to the inversion and prediction of the crushing rate of soybean.

Soybean samples were collected from a combine harvester. Physical and chemical analyses and spectroscopic determination of soybean samples were performed in a laboratory. Differences between the original spectrum (REF) and the logarithmic reciprocal pretreatment spectrum data (LR) with logarithmic reciprocal were compared. Furthermore, combined with band autocorrelation analysis, characteristic bands that can reflect the impurity rate of soybean samples were extracted. Then, the least squares support vector machine (LS-SVM) regression was applied to establish an inversion model of soybean crushing rate.

2 Materials and methods

2.1 Soybean test sample collection

The experiment was carried out in the farms subordinate to the professional cooperative of grain and cotton planting in Liang Shan county. The soybean variety planted is zheng dou 1307. On October 11 and 12, 2021, the 4YZL-5S combine harvester manufactured by Shandong Yafeng Agricultural Machinery and Equipment Company was used to carry out mechanized soybean harvesting operation over an experimental area of 6.67 ha. The quality of mechanized operation of the combine harvester exhibited significant differences under different working conditions. In this experiment, 9 factors and 3 levels of mechanical harvesting were set up. The nine factors included operating speed F_1 , drum speed F_2 , threshing section concave plate screen F_3 , separating section concave plate screen clearance F_4 , top guide plate angle F_5 , fan speed F_6 , guide plate angle F_7 , fish scale screen front opening F_8 , and fish scale screen back opening F_9 , as shown in Table 1. A total of 130 sets of field trials were carried out. Basic data of soybean crushing rate under different harvesting conditions

were obtained. During the test, 130 samples were collected at the grain outlet of the combine harvester. Then, the samples were placed in a sealed bag, numbered, and brought back to the laboratory for sample breakage rate determination and spectral analysis.

2.2 Spectrum data collection

The hyperspectral data of soybean samples were acquired using the ASD FieldSpec 4 wide-res Field Spectroradiometer (Malvern Panalytical). The spectral range was 350–2500 nm, the sampling intervals were 1.4 nm (350–1000 nm) and 1.1 nm (1001–2500 nm), and the wavelength repeatability was 0.1 nm.

As shown in Figure 1, spectral measurements were taken in a dark room. The light source was an Illuminator Reflectance Lamp, specifically designed by ASD for indoor Reflectance spectral measurement. A view angle of 25° was applied for optical fiber probe spectrum measurement. The distance from the light source to the surface of the soybean L , light incident Angle A , and fiber

Table 1. Mechanized harvest test.

Factor	Level		
F_1 (km·h ⁻¹)	6	7	8
F_2 (r·min ⁻¹)	4 50	500	550
F_3 (mm)	15	20	25
F_4 (mm)	20	25	30
F_5 (°)	17	21.5	26
F_6 (r·min ⁻¹)	1000	1150	1300
F_7 (°)	9	24	39
F_8 (mm)	15	17	19
F_9 (mm)	11	12	15

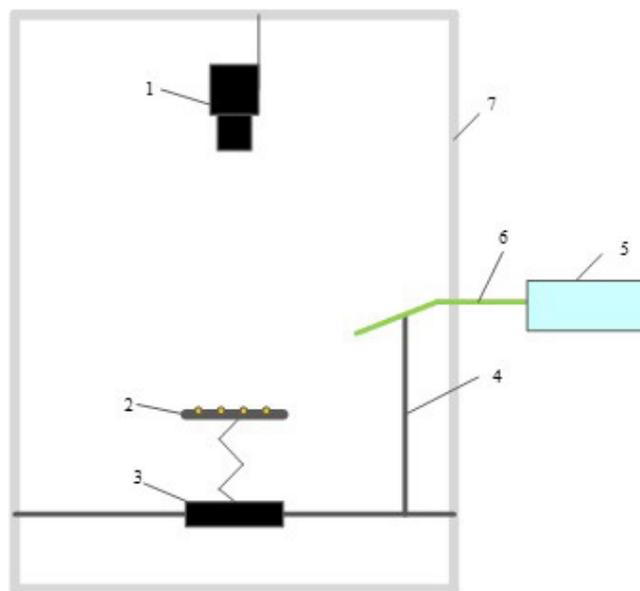


Figure 1. Schematic of the hyperspectral system. 1 Illuminator Reflectance Lamp; 2 Sample stage; 3 Motorized precision rotary stages; 4 Fiber fixing bracket 5 FieldSpec 4 Wide-Res; 6 Fiber optics; 7 Darkroom.

optic probe height from soybean surface H , were adopted from studies of Zhang and Hong et al. (Zhang et al., 2018c; Hong et al., 2016), in which ideal parameters of the indoor geometry test were: $L = 50$ cm, $A = 30^\circ$, and $H = 15$ cm. The soybean measurement samples were tiled across the entire sampling table. Measurements were taken in four directions (3 rotations of 90° each) for each sample. A total of 12 curves was obtained by generating 3 spectral curves in each direction. After algorithm averaging with ViewSpec Pro software (Malvern Panalytical Ltd.), the actual reflectance spectrum data of samples were obtained.

2.3 Determination of soybean crushing rate

The crushing rate of soybean samples was determined according to the National Standard for Operation Quality of Soybean Combine Harvester (NY/T 738-2003). The soybean samples were cleaned and treated. Samples with impurities were removed, the remaining were divided into four parts to obtain about 200 g of samples, and the sample quality was then measured. Broken grains in the sample were selected and weighed. The calculation formula of crushing rate is (Equation 1):

$$P_p = \frac{W_{pz} - W_{pq}}{W_{pz}} \times 100\% \quad (1)$$

where P_p represents the crushing rate, %; W_{pz} represents the sample mass, g; W_{pq} represents the mass of samples after the removal of broken grains, g.

The crushing rate of each group of test samples was measured 5 times by this method, and the mean value was calculated as the quality index of the test samples.

2.4 Least Squares Support Vector Machine (LS-SVM)

LS-SVM is an improvement and extension of traditional SVM. It uses the least squares linear system as the loss function and the equation as the constraint to transform the quadratic transformation problem into the solving problem of linear equations (Xu & Li, 2018). The principle of regression prediction of LS-SVM is as follows:

For a given training sample set, $T = \{(x_i, y_i) | i = 1, 2, \dots, N\}$, where $x_i \in R^N$ is the input and $y_i \in R$ is the output. According to the regularization theory, Suykens et al. changed the constraint condition and risk function of standard SVM, that is, the insensitive loss function was replaced by the quadratic square term of error, and the inequality constraint condition was transformed into the equality constraint condition. (Suykens & Vandewalle, 1999) Therefore, LS-SVM transforms solving quadratic programming problems into solving the following linear equations:

$$\begin{cases} \min[\frac{1}{2}\|w\|^2 + \frac{1}{2}C \sum_{i=1}^N e_i^2]_{i=1,2,\dots,N} \\ s.t. [y_i - (w \cdot \varphi(x_i) + b) = e_i] \end{cases} \quad (2)$$

where w is the weight vector, b is the bias term, C is the penalty parameter, $e_i \in R$ is the error, and $e \in R^N$ is the error vector. To

solve the optimization problem of Equation 2, the following Lagrange function is constructed:

$$L(w, b, e, \alpha) = \frac{1}{2}\|w\|^2 + \frac{1}{2}C \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{w \cdot \varphi(x_i) + b + e_i - y_i\} \quad (3)$$

Take the partial derivative of Equation 3, and according to the KKT (Karush-Kuhn-Tucher) condition in the optimization theory, the following equation and constraint conditions can be obtained:

$$\begin{cases} w = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \sum_{i=1}^N \alpha_i = 0 \quad i = 1, 2, \dots, N \\ \alpha_i = C e_i \\ w \cdot \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (4)$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$, $Q = (1, 1, \dots, 1)^T$, $Y = (y_1, y_2, \dots, y_N)^T$, and I is the identity matrix. Solving the equations simultaneously and eliminating w and e_p , Equation 4 can be modified as Equation 5:

$$\begin{pmatrix} 0 & Q^T \\ QK + C^{-1} & I \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ Y \end{pmatrix} \quad (5)$$

Finally, the regression model of LS-SVM is obtained (Equation 6):

$$f(x) = \sum_{i=1}^N \alpha_i k(x_i, x_j) + b \quad (6)$$

where K is a kernel function matrix, and $k(x_p, x_j) = \varphi(x_p) \cdot \varphi(x_j)$.

For a given training set $T = (x_1, x_2, \dots, x_N)$ and kernel function $k(x, x')$, the matrix of the elements in the kernel ($k(x_u, x_v)$, $u, v = 1, 2, \dots, N$) is defined as the kernel matrix or Gram matrix. For all training sets T , if its kernel matrix is symmetric and semi-positive definite, then the function K is the effective kernel.

The kernel of RBF is (Equation 7):

$$K(x, x') = \exp\left(-\frac{(x-x')^2}{\sigma^2}\right) \quad (7)$$

where σ is the kernel width.

Hence, its kernel matrix is (Equation 8):

$$K(x_u, x_v) = \exp\left(-\frac{\|x_u - x_v\|^2}{\sigma^2}\right) \rightarrow \begin{cases} 0 & x_u = x_v \\ \exp\left(-\frac{\|x_u - x_v\|^2}{\sigma^2}\right) & x_u \neq x_v \end{cases} \quad (8)$$

The kernel matrix $K(x_u, x_v)$ of RBF kernel function has the following properties (Equation 9):

$$K(x_u, x_v) = \begin{cases} 0 & x_u = x_v \\ k(x_u, x_v) = k(x_v, x_u) & x_u \neq x_v \end{cases} \quad (9)$$

The B-spline kernel is (Equation 10):

$$K(x, x'; t_1, t_2, \dots, t_m) = \sum_{i=1}^m (x - t_i)_+^p (x' - t_i)_+^p \quad (10)$$

Hence, its kernel matrix is (Equation 11):

$$K(x_u, x_v; t_1, t_2, \dots, t_m) = \sum_{i=1}^m (x_u - t_i)_+^{2p} \quad (11)$$

The kernel matrix $K(x_u, x_v)$ of the B-spline kernel function has the following properties (Equation 12):

$$K(x_u, x_v; t_1, t_2, \dots, t_m) = \begin{cases} \sum_{i=1}^m (x_u - t_i)_+^{2p} & x_u = x_v \\ k(x_u, x_v) = k(x_v, x_u) & x_u \neq x_v \end{cases} \quad (12)$$

Hence, the kernel matrix of the compound kernel function is (Equation 13):

$$K(x_u, x_v; t_1, t_2, \dots, t_m) = \alpha \cdot \exp\left(-\frac{(x_u - x_v)^2}{\sigma^2}\right) + (1 - \alpha) \cdot \sum_{i=1}^m (x_u - t_i)_+^p (x_v - t_i)_+^p \quad (13)$$

where α is the weighting coefficient and σ is the kernel width.

The kernel matrix $K(x_u, x_v)$ of the compound kernel function has the following properties (Equation 14):

$$K(x_u, x_v) = \begin{cases} (1 - \alpha) \cdot \sum_{i=1}^m (x_u - t_i)_+^{2p} & x_u = x_v \\ k(x_u, x_v) = k(x_v, x_u) & x_u \neq x_v \end{cases} \quad (14)$$

2.5 Model establishment and evaluation

The soybean samples were sorted from large to small according to the determined crushing rate P_p . One sample was taken from every two samples as a verification set sample, and 43 (33.1%) samples were obtained. The remaining 87 (66.9%) samples were taken as modeling set samples. Then, the least squares support vector machine was applied in Matlab (version number 8.0) software, to establish the hyperspectral data inversion model for the crushing rate of soybean samples. The effect of the model was evaluated by modeling the determination coefficient (R_c^2), and verifying the determination coefficient (R_p^2), root-mean-square error (R_{rmse}), and relative analysis error (R_{rpd}).

The coefficient of determination is an important indicator for measuring the fitting effect. The prediction is good at $0.66 \leq R^2 \leq 0.80$, better at $0.81 \leq R^2 \leq 0.90$, and the best at $R^2 \geq 0.90$ (Zhang et al., 2017).

The calculation formulas of R_{rmse} and R_{rpd} are as follows (Equation 15 and 16):

$$R_{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2} \quad (15)$$

$$R_{rpd} = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (p_i - p)^2}}{R_{rmse}} \quad (16)$$

where p_i represents the predicted crushing rate of the soybean samples, %; r_i represents the measured crushing rate of the soybean samples, %; p represents the average predicted crushing rate of the soybean samples, %; n is the number of samples.

A relative analysis error of above 2.5 indicates that the model has excellent prediction ability. Between 2.0 and 2.5, the value indicates good quantitative prediction ability of the model, and between 1.8 and 2.0, it indicates quantitative prediction ability. When the relative analysis error is between 1.4 and 1.8, it shows that the model has general quantitative prediction ability, and between 1.0 and 1.4, it indicates that the model has the ability to distinguish high values from low values. A relative analysis error less than 1.0 indicates that the model does not have the prediction ability (Yin et al., 2014).

3 Results and discussion

3.1 Soybean crushing rate analysis

Figure 2 shows the crushing rate of soybean samples collected under different working conditions of the combine harvester. In all 130 pairs of trials, the crushing rate of the mechanized soybean harvesting was the maximum at 5.98% when the operating speed was 7 km/h, the drum speed was 500 r/min, threshing section concave plate screen was 15 mm, separating section concave plate screen clearance was 30 mm, top guide plate angle was 21.5°, fan speed was 1150 r/min, guide plate Angle was 24°, fish scale screen front opening was 15 mm, and fish scale screen back opening was 13 mm. The crushing rate of mechanized soybean harvesting was the minimum at 0.98% when the operating speed was 8 km/h, the drum speed was 450 r/min, threshing section concave plate screen was 15 mm, separating section concave plate screen clearance was 25 mm, top guide plate angle was 21.5°, fan speed was 1150 r/min, guide plate angle was 24°, fish scale screen front opening was 17 mm, and fish scale screen back opening was 13 mm. Under different working conditions of the combine harvester, the mean crushing rate, standard deviation, and variation coefficient of the harvester were 2.42%, 0.85, and 35.2%, respectively. Under different working conditions of the main components, the crushing rate exhibited obvious differences. Therefore, it is necessary to establish a reliable and

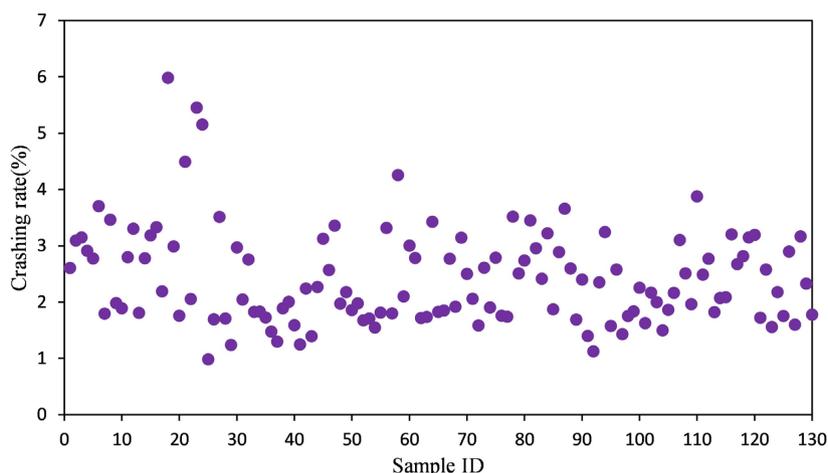


Figure 2. Crushing rate of soybean.

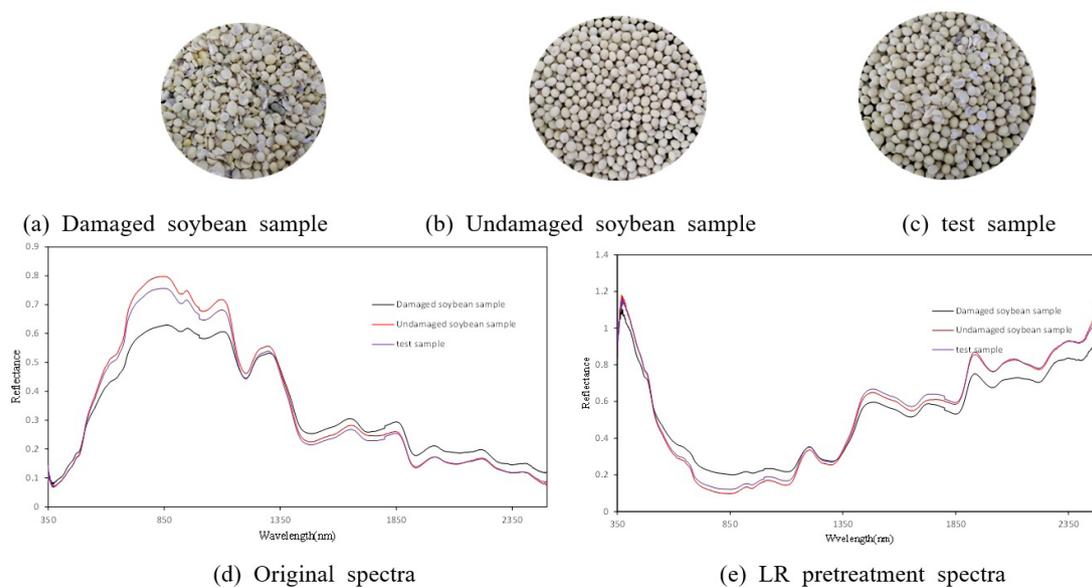


Figure 3. Types and spectra of soybean samples.

stable rate of the inversion model to realize the rapid detection of soybean crushing rate, timely adjustment of the working state of the combined harvest machine, and improve the quality of the soybean harvest mechanization operation.

3.2 Spectral pretreatment analysis

In the process of mechanized soybean harvesting, the working state of the fan speed, drum speed, and threshing clearance of the combine harvester directly affects the quality of mechanized soybean harvesting. Among them, threshing clearance is one of the main factors affecting the crushing rate of the harvester. If the clearance is less than the set rating value, the crushing rate will be relatively large (Wang & Li, 2017). Figure 3 shows graphs of the soybean sample morphology, corresponding to the spectrum and logarithmic reciprocal pretreatment in the mechanized operation. Figure 3a

shows broken grains in the soybean sample. Figure 3b shows the soybean sample after removing the broken seeds. Figure 3c shows the random test samples. As shown in the figure, the morphology of the broken soybean seeds is significantly different from that of intact seeds. Chen et al. showed that the spectral characteristics of different forms of substances were different and could be identified qualitatively and quantitatively based on the difference (Chen et al., 2015), which laid a foundation for identifying the crushing rate of soybean using spectral methods. Figure 3d shows the original spectral curve of broken seeds, non-broken seeds, and random samples. The spectral curves of the three samples are significantly different. Within the range of 600–1350 nm, the spectral reflectance of samples with broken grains removed was the highest, while that of samples with broken grains was the lowest, and three kinds of samples could be identified according to the reflectance. In the range of 1350–1900 nm, the reflectance of the broken grain

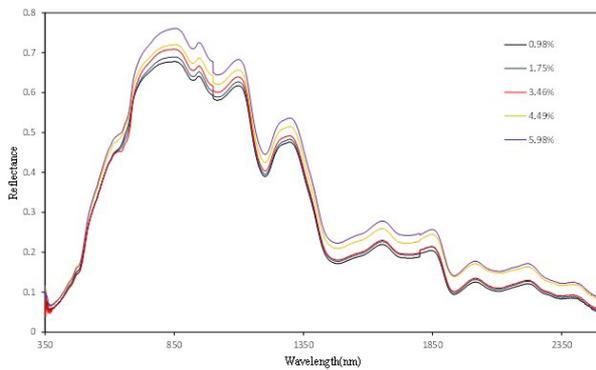
samples is the largest, and the spectral reflectance of the samples excluding the broken grain is relatively small. Figure 3e shows the logarithmic reciprocal spectral curve of the pretreatment sample, soybeans without impurities, and random samples. The spectral curve waveforms of the three samples are basically similar, but the reflectance of each band is obviously different. In the range of 600–1350 nm, the spectral reflectance of the samples without broken grains was lower than that of samples with broken grains. Within the range of 1350–1900 nm, the spectral reflectance of the samples without broken grains was higher than that of samples with broken grains. The spectral difference of samples facilitates the detection of sample impurity.

Figure 4 shows the original spectra and the logarithmic reciprocal pretreatment curve of 5 typical soybean samples collected by the combine harvester under different operating conditions. Figure 4a shows the original spectral curve. The waveforms of the 5 spectral curves are basically similar, especially in the range of 600–1900nm, and the spectral reflectance of soybean samples increases with the increase of the crushing rate. Figure 4b shows the logarithmic reciprocal pretreatment spectral curve. The figure

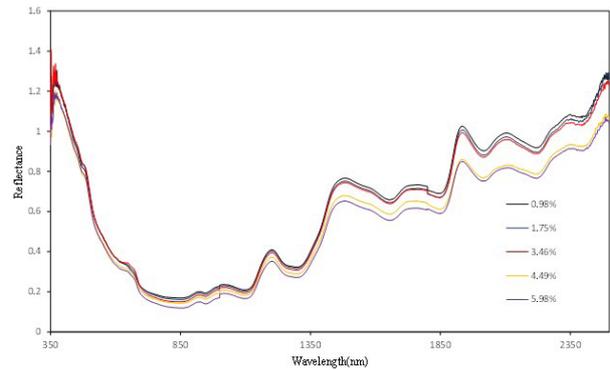
reveals significant differences in spectral reflectance among the 5 selected test samples.

3.3 Spectral autocorrelation analysis

As the spectral data of soybean samples revealed low SNR in the bands of 350–400 nm and 2301–2500 nm, the spectral data exhibit high volatility and poor stability. Therefore, data of these two bands were discarded. After removing the bands with low SNR, 1900 bands still remained in the spectral data of soybean samples. A large amount of redundant and collinearity information was observed between adjacent bands with high correlation data, which will increase the computational amount and complexity of the model. This problem of data redundancy can be effectively solved through autocorrelation analysis (Ji et al., 2014; Deng et al., 2014). Within the range of 401–2300 nm, the absolute value of the correlation coefficient $abs(R)$ was calculated by pairwise combination of each band, and the $abs(R)$ matrix of 1900×1900 was constructed, as shown in Figure 5. Figure 5a and Figure 5b present the autocorrelation matrices of the original spectrum and the LR pretreatment spectrum, respectively.

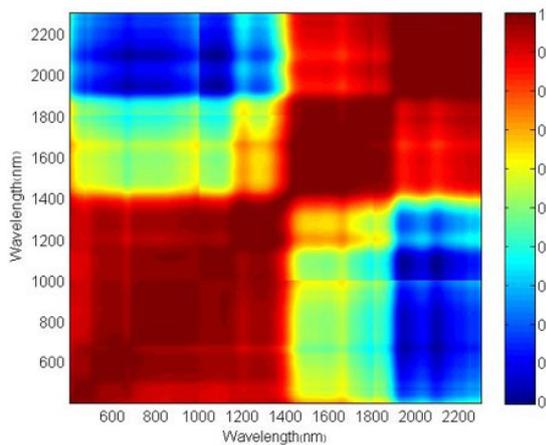


(a) Original spectra

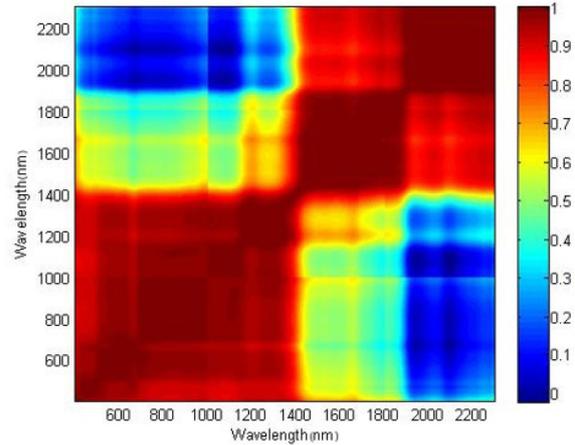


(b) Log 1/R spectra

Figure 4. Spectra of five soybean samples.



(a) Original spectra



(b) Log 1/R spectra

Figure 5. Inter-Correlation matrix.

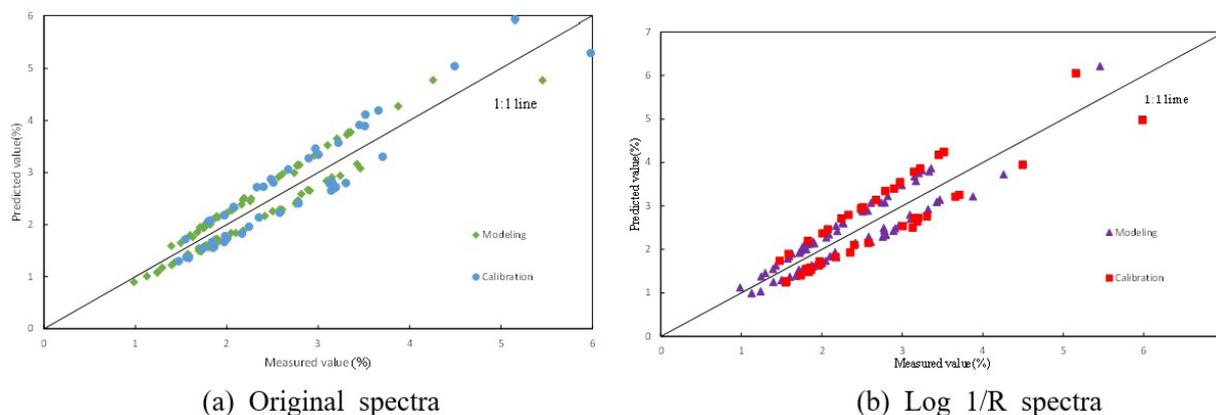


Figure 6. Inversion results of impurity content of soybean samples.

Table 2. SVM model of the impurity content of soybean samples.

Spectral indices	R_c^2 (the determination coefficient)	R_p^2 (the verifying the determination coefficient)	R_{rmse} (the root-mean-square error)	R_{rpd} (the relative analysis error)
REF (the raw spectral reflectance data)	0.939	0.915	0.317	2.94
LR (the logarithmic reciprocal pretreatment spectrum data)	0.912	0.885	0.399	2.36

The lighter the color, the lower the correlation between bands, the smaller the data redundancy, and the richer the effective information contained. According to the first 100 points with the smallest correlation coefficient in the autocorrelation matrix, the frequency of each band was counted to select the characteristic band with the most abundant information (Ding et al., 2017). The characteristic bands of the original spectrum were selected as 1061, 1068, 1074, 1090, 2085, 2092, 2095, and 2103 nm. The characteristic bands of logarithmic reciprocal pretreatment spectra with reciprocal values were screened out as 677, 1039, 1078, 1093, 1101, 1956, 2088, and 2107 nm.

3.4 Modeling analysis based on LS-SVM

Using the raw spectrum of the soybean samples and the spectral data preprocessed using the logarithmic reciprocal pretreatment of the soybean samples, a scatter diagram of the result of the inversion model established based on LS-SVM is shown in Figure 6. The relative error of the original spectral data index modeling set was 16.21% at the maximum, 7.63% at the minimum, and 11.45% on average. The absolute error value of the verification set was 18.77% at the maximum, 9.37% at the minimum, and 13.50% on average. The relative error of the spectral data modeling set preprocessed by logarithmic reciprocal pretreatment was 20.43% at the maximum, 9.54% at the minimum, and 14.31% on average. The relative error of the verification set was 24.97% at the maximum, 13.82% at the minimum, and 17.73% on average. Therefore, the inversion model established using the original spectral data has a better effect.

The comprehensive evaluation results of the model are shown in Table 2. The inversion model established using the

original spectral data of soybean samples appears to be more suitable than that established with the logarithmic reciprocal pretreatment of spectral data. The modeling determination coefficient of the original spectral data reached 0.939, whereas that of the verification data reached 0.915. Overall, the fitting effect of the model established using the original spectral data is highly applicable. Furthermore, the relative analysis error of the original spectral data modeling reached 2.94, indicating that the model based on LS-SVM has a good quantitative prediction ability. After the logarithmic reciprocal pretreatment of the original spectral data, the error of the model increased slightly, but the model still showed good fitting and prediction effect.

4 Conclusions

In this study, the relationship between the spectrum of soybean samples harvested by mechanized equipment and the crushing rate was analyzed. Based on the original spectral data of soybean samples and spectral data preprocessed by logarithmic reciprocal pretreatment, optimal characteristic bands were extracted by using the autocorrelation analysis between bands, and a regression inversion model of crushing rate was established based on the least squares support vector machines regression analysis. The following conclusions can be drawn:

- (1) There were significant differences in spectral responses between broken and intact soybean samples. The original spectra and the logarithmic reciprocal pretreatment curve waveforms of samples with different crushing rates were basically similar, and the spectral reflectance of soybean

samples with the original spectrum within the range of 600–1900 nm increased with increased crushing rates.

- (2) According to the autocorrelation analysis of spectral bands, sensitive bands of soybean crushing rate under various pretreatment conditions are as follows: 1061, 1068, 1074, 1090, 2085, 2092, 2095, and 2103 nm for the original spectrum, and 677, 1039, 1078, 1093, 1101, 1956, 2088, and 2107 nm for the logarithmic reciprocal pretreatment spectra.
- (3) The least squares support vector machine regression analysis and prediction model of soybean crushing rate were established for the spectral data under different pretreatments. The prediction accuracy of the original spectra data inversion model was higher than that of the pre-processed spectra inversion model with logarithmic reciprocal pretreatment. The modeling determination coefficient of the original spectra data inversion model reached 0.939, whereas the verification determination coefficient reached 0.915, and the relative analysis error reached 2.94. Therefore, the model based on LS-SVM has good prediction ability.

Acknowledgements

The authors express appreciation for the financial support provided by Jiangsu Agriculture Science and Technology Innovation Fund (No. CX(20)1007), and Jiangsu Province Modern Agricultural Machinery Equipment and Technology Demonstration Promotion Project (Grant No. NJ2020-60, and NJ2020-25) and Special fund project for the construction of modern agricultural industrial technology system (No. CARS-04-PS26), and Special Funds for Fundamental Scientific Research Operation of Central-level Public Welfare Scientific Research Institutes (No. S202007, and No. S202015). The authors also thank the editors and anonymous reviewers for providing helpful suggestions for improving the quality of this manuscript.

References

- Chen, B., Wang, Q., Xiao, C. H., Lin, H. R., & Zou, N. (2015). Exploration and recognition of spectral characteristic of cotton leaf suffered nine stresses. *Xibei Nongye Xuebao*, 24(10), 64-73. <http://dx.doi.org/10.7606/j.issn.1004-1389.2015.10.010>.
- Chen, J. Y., Chen, S. B., Zhang, Z. T., Fu, Q. P., Bian, J., & Cui, T. (2018b). Investigation on photosynthetic parameters of cotton during budding period by multi-spectral remote sensing of unmanned aerial vehicle. *Zhongguo Nong-Jihua Xuebao*, 49(10), 230-239. <http://dx.doi.org/10.6041/j.issn.1000-1298.2018.10.026>.
- Chen, J., Gu, Y., Lian, Y., & Han, M. N. (2018a). Online recognition method of impurities and broken paddy grains based on machine vision. *Nongye Gongcheng Xuebao*, 34(13), 187-194. <http://dx.doi.org/10.11975/j.issn.1002-6819.2018.13.022>.
- Cui, L. L., Xu, J. Y., Feng, Z. W., Yan, M. X., Piao, X. M., Yu, Y., Hou, W., Jin, Y. P., & Wang, P. P. (2020). Simultaneous determination and difference evaluation of volatile components generated from ginseng fruit by HS-SPME Coupled with GC-MS according to fruit color. *Food Science and Technology*, 40(3), 532-536.
- Deng, X. L., Li, M. Z., Zheng, L. H., Zhang, Y., & Sun, H. (2014). Estimating chlorophyll content of apple leaves based on preprocessing of reflectance spectra. *Nongye Gongcheng Xuebao*, 30(14), 140-147. <http://dx.doi.org/10.3969/j.issn.1002-6819.2014.14.018>.
- Ding, Y. J., Zhang, J. J., Sun, H., & Li, X. H. (2017). Sensitive bands extraction and prediction model of tomato chlorophyll II in glass greenhouse. *Guangpuxue Yu Guangpu Fenxi*, 37(1), 194-199. <https://doi.org/10.1590/fst.26718>.
- Hong, Y. S., Yu, L., Geng, L., Zhang, W., Nie, Y., & Zhou, Y. (2016). Using direct standardization algorithm to eliminate the effect of laboratory geometric parameters on soil hyperspectral data fluctuate characteristic. *Journal of Central China Normal University. Nature and Science*, 50(2), 303-308. <http://dx.doi.org/10.19603/j.cnki.1000-1190.2016.02.023>.
- Huang, L. S., Zhang, Q., Zhang, D. Y., Lin, F. F., Xu, C., & Zhao, J. L. (2018). Early diagnosis of wheat powdery mildew based on Relief-F band screening. *Hongwai Yu Jiguang Gongcheng*, 47(5), 8. <http://dx.doi.org/10.3788/IRLA201847.0523001>.
- Ji, R. H., Zheng, L. H., Deng, X. L., Zhang, Y., & Li, M. Z. (2014). Forecasting chlorophyll content and moisture of apple leaves in different tree growth period based on spectral reflectance. *Zhongguo Nong-Jihua Xuebao*, 45(8), 269-275. <http://dx.doi.org/10.6041/j.issn.1000-1298.2014.08.043>.
- Liu, C. H., Fang, Z., Chen, Z. C., Zhou, L., Yue, X. Z., Wang, Z., Wang, C. Y., & Miao, Y. X. (2018). Nitrogen nutrition diagnosis of winter wheat based on ASD Field Spec3. *Nongye Gongcheng Xuebao*, 34(19), 162-169. <http://dx.doi.org/10.11975/j.issn.1002-6819.2018.19.021>.
- Lu, X. G., Li, X. H., Zhang, S. M., Zhang, M. Q., & Jiang, H. T. (2021). A method for detecting sucrose in living sugarcane with visible-NIR transmittance spectroscopy. *Guangpuxue Yu Guangpu Fenxi*, 41(12), 3747-3752. [http://dx.doi.org/10.3964/j.issn.1000-0593\(2021\)12-3747-06](http://dx.doi.org/10.3964/j.issn.1000-0593(2021)12-3747-06).
- Ni, Y. L., Jin, C. Q., Chen, M., Wang, T. E., Li, Z. F., & Yuan, W. S. (2019). Research progress on mechanized production technology and equipment of soybean in China. *Zhongguo Nong-Jihua Xuebao*, 40(12), 17-25. <http://dx.doi.org/10.13733/j.jcam.issn.2095-5553.2019.12.04>.
- Rambo, M. K. D., Ferreira, M. M. C., Melo, P. M., Santana, C. C. Jr, & Bertuol, D. A. (2020). Prediction of quality parameters of food residues using NIR spectroscopy and PLS models based on proximate analysis. *Food Science and Technology*, 40(2), 444-450. <http://dx.doi.org/10.1590/fst.02119>.
- Yikmiş, S. (2020). Effect of ultrasound on different quality parameters of functional sirkencubin syrup. *Food Science and Technology*, 40(1), 258-265. <http://dx.doi.org/10.1590/fst.40218>.
- Steinberg, A., Chabrilat, S., Stevens, A., Segl, K., & Foerster, S. (2016). Prediction of common surface soil properties based on vis-nir airborne and simulated enmap imaging spectroscopy data: prediction accuracy and influence of spatial resolution. *Remote Sensing*, 8(7), 613-627. <http://dx.doi.org/10.3390/rs8070613>.
- Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293-300. <http://dx.doi.org/10.1023/A:1018628609742>.
- Wang, K. R., & Li, S. K. (2017). Progresses in research on grain broken rate by mechanical grain harvesting. *Zhongguo Nong Ye Ke Xue*, 50(11), 2018-2026. <http://dx.doi.org/10.3864/j.issn.0578-1752.2017.11.007>.
- Wei, J. (2021). Research on soybean seed appearance quality testing technique based on BP neural network. *Heilongjiang Science*, 12(16), 12-16. <http://dx.doi.org/10.3969/j.issn.1674-8646.2021.16.004>.
- Wu, Y. L., Chen, Y. Y., Lian, X. Q., Liao, Y., Gao, C., Guan, H. N., & Yu, C. C. (2021). Study on the identification method of citrus leaves based on hyperspectral imaging technique. *Guangpuxue Yu Guangpu Fenxi*, 41(12), 3837-3843. [http://dx.doi.org/10.3964/j.issn.1000-0593\(2021\)12-3837-07](http://dx.doi.org/10.3964/j.issn.1000-0593(2021)12-3837-07).

- Xiao, H. B., Guo, P. Y., & Wang, Y. (2021). Nutritional health risk grading of bacon hyper-spectrum based on deep learning. *Journal of Chinese Institute of Food Science and Technology*, 22(1), 275-281.
- Xu, Y. Q., & Li, C. X. (2018). Predicting fluctuating wind velocity using optimized combined kernel functions based LSSVM. *Journal of Shanghai University*, 24(4), 627-633. <http://dx.doi.org/10.12066/j.issn.1007-2861.1819>.
- Xu, Z. J., Qi, Y. J., Xing, X. H., Tong, F., & Wang, X. (2021). Analysis and evaluation of agronomic and quality traits in soybean germplasms from Huang-Huai-Hai Region. *Zhiwu Yichuan Ziyuan Xuebao*. <http://dx.doi.org/10.13430/j.cnki.jpgr.20210915001>.
- Yang, Z. L., Mei, J. Q., Cao, C., Ji, G. Y., & Han, L. J. (2018). Traits of milled wood lignin isolated from different crop straw based on FT-IR. *Nongye Gongcheng Xuebao*, 34(19), 219-224. <http://dx.doi.org/10.11975/j.issn.1002-6819.2018.19.028>.
- Yin, Z., Lei, T. W., Chen, Z. P., Yan, Q. H., & Dong, Y. Q. (2014). Adaptability of near-infrared sensor for moisture measurement of different soils. *Zhongguo Nong-Jihua Xuebao*, 45(3), 148-151, 190. <http://dx.doi.org/10.6041/j.issn.1000-1298.2014.03.025>.
- Zhang, D. H., Zhao, Y. J., Qin, K., Zhao, N. B., & Yang, Y. C. (2018a). Influence of spectral transformation methods on nutrient content inversion accuracy by hyperspectral remote sensing in black soil. *Nongye Gongcheng Xuebao*, 34(20), 141-147. <http://dx.doi.org/10.11975/j.issn.1002-6819.2018.20.018>.
- Zhang, Q. X., Zhang, H. B., Zhang, H. J., Wang, X. S., & Liu, W. K. (2017). Hybrid inversion model of heavy metals with hyperspectral reflectance in cultivated soils of main grain producing areas. *Zhongguo Nong-Jihua Xuebao*, 48(3), 148-155. <http://dx.doi.org/10.6041/j.issn.1000-1298.2017.03.019>.
- Zhang, Y. K., Luo, B., Song, P., Pan, D. Y., Lu, W. C., Zhou, Y. N., Wang, C., & Zhao, C. J. (2018b). Rapid determination of soluble protein content for soybean leaves based on near infrared spectroscopy. *Nongye Gongcheng Xuebao*, 34(8), 187-193. <http://dx.doi.org/10.11975/j.issn.1002-6819.2018.08.023>.
- Zhang, Z. T., Wang, H. F., Arnon, K., Chen, J. Y., & Han, W. T. (2018c). Inversion of soil moisture content from hyperspectral based on ridge regression. *Zhongguo Nong-Jihua Xuebao*, 49(5), 240-248. <http://dx.doi.org/10.6041/j.issn.1000-1298.2018.05.028>.
- Zhu, L. T., & Wu, W. B. (2020). Artificial neural network for determining the hedonic score of texture of and distinguishing different grades of ham sausages. *Food Science and Technology*, 40(1), 46-54. <http://dx.doi.org/10.1590/fst.31018>.