The quantitative detection of botanical trashes contained in seed cotton with near infrared spectroscopy method

Wanhuai Zhou¹,², Hao Li¹ and Houjun Liang¹

Abstract
This study is performed to investigate the potential of near infrared (NIR) spectroscopy for the detection of botanical trashes content of seed cotton harvested by cotton-picker (SCHCP). Large quantity of trashes become comingled with cotton fiber in the harvesting process, especially when the cotton is harvested with cotton-picker. In China, trashes content of seed cotton (SC) has to be detected when farmers sell the SC to ginneries because trashes reduce the prices of SC and it should be deducted from the whole weight. The conventional instrumental method used to detect the trashes content of SC, ginning and trashes analysis, is complex and time consuming. In this study, 353 SC samples were collected from three ginneries, the NIR spectra bands from 12,000 to 4000 cm⁻¹ were collected with the FT-NIR spectrometer Nexus. Models between NIR spectra and the trashes contents of these SC samples have been developed with the method of partial least square regression (PLSR), bands of 12,000–4000 cm⁻¹, multiplicative signal correction (MSC) was used to eliminate the negative effects caused by sample shapes, second derivative spectra were used to eliminate the translation and the rotation in the spectral baseline. And the parameters of optimized model: \( R^2 \) is up to 0.985 (calibration set) and 0.973 (prediction set), RMSEC is as low as 0.072 g and RMSEP is 0.158 g. Results of ANOVA also certified the trashes contents calculated with the models are consistent with the actual trashes contents.

Keywords
Trash content, seed cotton, near infrared, PLSR

Date received: 6 September 2021; accepted: 22 January 2022

Introduction
Cotton is an essential natural fiber accounting for approximately 27% of all fibers. China is one of the world’s largest cotton producing countries.¹ In recent years, machined-picked cotton has been rapidly promoted, about 80% of cotton is harvested mechanically up to now. The harvested seed cotton (SC) will be ginned to separate lint from cotton seeds and various types of impurities, such as leaf, bark, stem, seed coat, paper, and plastic bag. The ginning procedure contains a series of sequential steps including several SC cleaning processes, ginning, several lint cotton (LC) cleaning processes, and packaging of LC into bales. If the impurity cleaning efficiency is too low, the high impurity content of LC will adversely affect spinning. According to a survey done by the International Textile Manufacturers Federation (ITMF), 26% of cotton processed by the spinning mills were

¹School of Computer Science and Technology, Anhui University of Finance and Economics, Bengbu, P.R. China
²Cotton Engineering Institute, Anhui University of Finance and Economics, Bengbu, P.R. China

Corresponding author:
Wanhuai Zhou, School of Computer Science and Technology, Anhui University of Finance and Economics, 962 Cao Shan Road, Bengbu 233030, P.R. China.
Email: zhouwanhuai@aliyun.com
found to be moderately or severely contaminated impurities. However, the excessive cleaning will significantly degrade the fiber quality, for example, the length, the length uniformity, the strength, etc. Which not only decreases the monetary value of cotton fiber, but also reduces the overall quality of yarns and cotton textiles. Therefore, the detection of the impurity content is very important in cotton industry.

The commonly used instrumental detection methods for cotton impurity include the gravimetric methods, the geometric methods and the spectroscopy methods. The gravimetric methods are mainly used to separated impurities and fiber. For example, the saw impurity analyzer which is the standard method to measure the impurity content of LC in China separates the botanical impurities and cotton fiber using the differences in densities and volumes of them.

The geometric methods are mainly used to detect the impurity content. For example, Lieberman et al. used neural networks and learn vector quantization methods to detect impurities in LC, the impurities in large cotton lump has also been detected; Li et al. proposed a method which extracted the feature vectors with Gabor Operator and the white foreign fibers were separated in the binary image compositing using the feature vectors. Zhang et al. proposed a method for the identification of impurities in SCHCP with machine vision using the support vector machine classifier optimized with genetic algorithm.

The spectroscopy methods are also mainly used to detect the impurity content. For example, Fortier et al. acquired the FT-NIR spectral characteristics of hull, leaf, seed, stem of cotton, and identified the cotton impurity components with a NIR spectral database, the identification accuracy was as high as 98% when the spectrum of impurity of new sample was compared with the reference spectral library. Rodgers et al. monitored the micronaire of cotton fiber with a portable NIR instrument. Liu et al. classified the cotton samples into different level using the models between cotton samples and their spectra (220–2200 nm), the results indicated that using the model in the bands of 1105–1700 nm could reach an acceptable separation; Liu et al. found there were large differences between the spectra (1200–900 cm⁻¹) of the mature and immature cotton fiber. Allen et al. used a FT-IR spectral database to classify cotton samples into different impurity levels. Gamble and Foulk built the partial least squares (PLS) models of six botanical trash types using fluorescence spectroscopy and the models of leaves and hull were capable of predicting individual trash component with a high degree of confidence. Gaitán-Jurado et al. determined the moisture content and impurity level of SC using NIR spectra. For moisture, the best model was obtained using PLS regression method, the first derivate, drying method, standard normal variate (SNV), and detrending as the pretreatment method.

The geometric methods and the spectroscopy methods are often combined to detect cotton impurities and other features of cotton fiber. For example, Zhang et al. inspected foreign matter the surface of LC using the method of liquid crystal tunable filter hyper-spectral image with spectral ranging from 900 to 1700 nm. Mustafic et al. found fluorescent imaging apparatus with blue and UV light excitation sources could be a promising method for cotton foreign matter detection. Jia and Ding transformed the discriminations of the absorption characteristic of cotton fibers and foreign fibers at the NIR band to image features, then an image segmentation algorithm was selected for extracting foreign fiber objects from cotton background. The high volume instrument (HVI) incorporates both spectroscopy and imaging, can measure multixes of LC, such as the color grade, the length, the strength, and the fineness of cotton fiber, etc. HVI can also measure the impurity content of LC in the way of counting the number of impurities and the percentage of total surface area of impurities.

In summary, it can be found that there are many researches to detect the impurity of LC, while there is almost no research on the detection of the impurities of SC. The impurity content of SC is much larger than the impurity content of LC, especially large quantity of stems, hulls, and leaves of cotton plant become comingled with cotton fiber in SC harvested with cotton-picker (SCHCP); The other reason is SC contains cotton seeds while LC do not. In this research, NIR spectroscopy is investigated to detect the impurity content (mainly refers to stems, hulls, and leaves of cotton plant) of SCHCP. The partial least square (PLS) models between the diffuse reflection NIR spectra and the impurity content of SCHCP.

Materials and methods

The whole experiment mainly contains four steps: sampling and sample preparation, NIR spectra acquisition, separate impurity from cotton, mathematical modeling, as shown in Figure 1.

Sampling and preparation

As we known, different geographical locations have different climatic conditions, so the cotton maturity period will also be different, which further leads to different impurity rate of cotton. In order to ensure the representativeness, three ginneries were selected from the three most representative producing areas in Xinjiang, Chinese main cotton production area. In each ginneries, two cotton modules with the volume of 10 × 2 × 2.5 m³ and the weight of 10 tons were randomly selected. In each module, 60 samples were collected on the module’s both longest sides. On each side, 30 samples were collected along roughly equidistant three lines. Along
each line, 10 samples were collected roughly equidistant under the surface of the module not less than 10 cm. The weight of each sample ranged from 100 to 150 g. In conclusion, a total of 360 samples were collected from three gineries, with 120 samples per ginnery. However, there are seven sample bags were broken during the transport. Therefore, 353 samples were actually used for modeling and analysis.

Before acquiring NIR spectral data, 20 g SC was separated from each sample and kept in the laboratory under the condition at a constant temperature of (20 ± 1)°C and relative humidity of (65 ± 2)%RH for more than 24 h. Then, all of the samples were sealed with PE self sealing bags as shown in Figure 2.

### NIR spectral acquisition

The NIR spectra were acquired with an FT-NIR spectrometer Nexus (Thermo Electron Corp., Madison, Wisc., USA) with a smart diffuse reflectance accessory, an InGaAs detector over a range of 12,000–4000 cm⁻¹ and the light source was a built-in 50 W quartz halogen lamp. The background was the NIR spectra of a Teflon plate. Before spectral acquisition, all samples were kept at the environment of

---

### Figure 1. Experimental procedure: Sampling and sample preparation, NIR spectra acquisition, Separate trash from cotton fibers, and Modeling and optimization.

### Figure 2. 20 g SC used for NIR analysis, Gin, Trashes analysis. The weighing error was less than 0.01 g, the bags were weighted individually one by one.
20 ± 1°C and 65 ± 2%RH last for more than 24h. The spectra were acquired at a resolution of 8 cm⁻¹ and 32 scans over the range of 12,000–4000 cm⁻¹.

The sample pool is a cylinder (Figure 3(a)) made of special metal material with the internal diameter of 10.16 cm and height of 6.35 cm. One end of the cylinder is sealed with low OH quartz glass which hardly absorb near infrared light and the other end is open for loading samples. In this research, the average spectrum of five spectra which were gained from five different positions (1–5 marked in Figure 3(a)) was used to represent the according sample. Because it is difficult to characterize one sample with the spectrum gained from just one position for the spot of the spectrometer is too small while the surface of
the sample is quite large. A gold coated compression tool was used to press the samples to a certain height and density (Figure 3(b)). In this way, it can ensure the consistency of the density and the height of different samples, and the stray light can also be prevented.

**Trash detection**

In this study, the trashes mainly refer to stems, hulls and leaves of cotton and grass (Figure 4) because they account most of the botanical trashes in SCHCP. As shown in Figure 1 (P3), the trashes in SCHCP are mainly separated in the following three steps:

1. In the process of P3.1, most of the large trashes, mainly refers to cotton stems and hulls, were picked out manually.
2. In the process of P3.2, the SCHCP was ginned with a small roller ginning machine SY-20 (roller size 120mm × 205mm, rotational speed of 88 rpm, produced by River machinery plant, Xinxiang, Henan, China), which was specially designed to gin small quantity of SC. In this process, the remaining cotton hulls and a part of cotton plant leaves were separated from cotton fiber.
3. In the process of P3.3, the remaining leaves were separated from the cotton fiber with cotton trash analyzer YG-041 according to the test method for percentage of trash content in raw cotton (GB/T-6499) (roller size 57.15mm × 490mm, rotational speed of the roller 0.9 rpm, produced by Changzhou No.1 Textile Equipment Co., Ltd., Changzhou, Jiangsu, China), which is also specially designed for small quantity of LC.

In fact, although most of trashes could be separated from cotton fiber in the above processes, the LC still contains a small amount of trashes which is difficult to separate. In general, the trashes content of the cotton samples after trashes analysis is deemed as 0, so the content of this part of trashes is ignored in this research.

**Tools and data analysis**

Origin 2017 (OriginLab Corporation, Northampton, MA 01060 USA), Omnic 8.0 and TQ Analyst 8.0 (Thermo Electron Corp., Madison, Wisc., USA) were used for spectral pretreatment and modeling. Origin 2017 and Omnic 8.0 provides commonly used spectral preprocessing and spectral feature extraction methods, such as calculate the average spectra, the first and second derivative spectra, spectral smoothing, peak fitting, and so on. TQ Analyst 8.0 is a specialized software for infrared spectral modeling, which provides commonly used spectral preprocessing (multiplicative signal correction, baseline correction, spectral smoothing, data format transform) and modeling methods (e.g. distance match, discriminant analysis, simple beer’s law, principal component regression, PLS, etc.).

Models were formulated which related the FT-NIR spectra and trashes content of each sample. The prediction ability of the model is given as root mean squares error of calibration (RMSEC) and root mean squares error of prediction (RMSEP).

\[
RMSEC = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - y_{i}^c)^2 },
\]

\[
RMSEP = \sqrt{\frac{1}{m-1} \sum_{j=1}^{m} (y_j - y_j^p)^2 },
\]

\(y_i\) is the tested trash content of the \(i\)th sample of calibration set.

\(y_{i}^c\) is the predicted trash content of the \(i\)th sample of calibration set.

\(y_j\) is the predicted trash content of the \(j\)th sample of prediction set.

\(y_j^p\) is the predicted trash content of the \(j\)th sample of prediction set.

\(n\) number of samples of calibration set.

\(m\) number of samples of prediction set.

The mathematics statistic method Analysis of variance (ANOVA) was used to determine whether there were significant differences between the prediction trash contents and the tested trash contents.

\[
F = \frac{MSA}{MSE} \sim F(k-1,N-k)
\]

\[
MSA = \frac{\sum_{i=1}^{k} N_i (x_i^{avg} - x_i^{avg})^2}{(k-1)}
\]

\[
MSE = \frac{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (x_{ij} - x_i^{avg})^2}{(N-k)}
\]

In this research, number of data sets \(k = 2\),

\(N\) the number of samples in each group,

\(x_i^{avg}\) the mean in group,

\(x_i^{avg}\) the mean of all (in this research, \(x_i^{avg}\) is the mean of the prediction trashes content and the tested trashes content).
Results and discussion

NIR spectra

In section 2.2, the methods of selecting multiple acquisition points for each sample and acquiring multiple spectra (32 scans) for each acquisition point were used to improve the representativeness and the signal-to-noise ratio (SNR). In addition, several other pretreatment methods have been used to improve the quality of the spectra before modeling. The baseline was corrected with the least squares baseline correction algorithm, then the spectra were smoothed with Savitzky-Golay method (the size of the window was 5 points and the polynomial order was 2), at last the spectra were normalized with the method of dividing by the maximum ordinate value (Figure 5).

The average spectra of the samples from different ginnersies over the entire spectral range (12,000–4000 cm\(^{-1}\)) are compared in Figure 6. There are obvious differences in the raw spectra (Figure 6(a)); But the differences are becoming smaller after correcting baseline and smoothing (Figure 6(b) and (c)); At last, the differences almost disappear after normalizing.

It can also be observed that there are two bands representing cotton botanical trashes are observed at about 2050–2100 nm (bands of 4878–4762 cm\(^{-1}\)) with an O-H bend and C-O stretch combination, and bands at about 2200–2270 nm (bands of 4545–4405 cm\(^{-1}\)) with O-H and C-O stretch combination or C-H stretch and CH\(_2\) deformation. Despite this, it was found that there were considerable overlaps among the FT-NIR spectra of cotton fiber and varieties of botanical trashes (stems, leaves, hulls, seed coat, seed meat, etc.). Thus the components of cotton fiber and the trashes could not be uniquely identified with the original FT-NIR spectra.

Trash content

The main statistical characteristics related to the weight of trashes and the content of trashes are shown in Figure 7. The weight of the trashes separated in P3.3 (Figure 7(a)) is a little larger than the weight of the trashes separated in P2.1 and P3.2 (Figure 7(c)). Moreover, the trashes separated in P3.3 distributed more uniformly than the trashes separated in P3.1 and P3.2, because the trashes separated in P3.3 are mainly cotton leaves, while the trashes separated in P3.1 and P3.2 contain not only leaves, but also cotton plant stems and hulls whose sizes are larger than leaves and whose amounts are far less than the amounts of leaves. And the uneven distribution of trashes might lead to poor models.

The detailed statistical parameters of Figure 7 is shown in Table 1. The contents of trashes (shown in Figure 7(b), (d), and (f)) are calculated according to formulas (5)–(7). In which, I.1 represents the trashes separated P3.1 and P3.2, I.2 represents the trashes separated in P3.3.

\[
R.1 = \frac{I.1}{20} \quad (5)
\]
\[
R.2 = \frac{I.2}{20} \quad (6)
\]
\[
R.t = \frac{I.1 + I.2}{20} \quad \text{or} \quad R.1 + R.2 \quad (7)
\]

Models between FT-NIR spectra and trashes content

The quantitative prediction models were established using the software TQ Analyst 8.0. Multiplicative signal correction (MSC) was used to eliminate the negative effects caused by different sample shapes and granularity of fiber and trashes; The PLS algorithm was used to build the relationship between the NIR spectra and the trashes content because it was known that PLS usually was better than...
other methods in cotton analysis for the errors caused by the tightness and thickness could be greatly reduced in the transformation of the data matrix composed both dependent variables and independent variables; To avoid over fitting, each group of SCHCP samples was divided into five internal cross validation sub groups. In this way, one sub group was used as the prediction set and the calibration model was built on the remain four sub groups. This procedure was repeated until each sub group was used as prediction set; The models between the three different spectral types (spectrum, first derivate and second derivate) ranging from 12,000 to 4000 cm$^{-1}$ were established and compared (Table 2).

It is obvious that the second derivate is most suitable for establishing between NIR spectra of SCHCP samples and their trashes contents (before optimizing). The most important reasons are the non-uniform translation in the baseline caused by different sample tightness and the significant non-uniform rotation caused by different sample heights. Although all of the samples were pressed with the same compression tool, the variations in trashes contents and the inconsistent traits of cotton seeds, etc. result in differences in sample tightness; And then the variations in sample tightness resulted in different heights which not only caused the non-uniform translation in the baseline, but also caused non-uniform rotation. It is known that the first derivative and the second derivative can reduce the effect of spectral translation and rotation to a certain extent.24 Therefore, the models between NIR spectra of SCHCP and their trashes contents should be established using the second derivate. Figure 8 shows the relationship between the actual trashes contents and the calculated trashes contents of the SCHCP samples.

Although the results of calibration set are quite good, the results of the prediction set are not as good as the calibration set. It is known that abnormal samples (including physical or chemical properties abnormal, spectral abnormal) might cause significant decrease in model quality. In section 3.2, the samples which are abnormal in trashes content have been distinguished. And here, the methods of mahalanobis distance, chauvenet test, and the leverage were used to distinguish the samples which were abnormal.

Figure 6. The average spectra of the SCHCP samples of different ginneries is the: (a) raw spectra, (b) spectra after correcting the baseline with the least squares method, (c) spectra after smoothing with Savitzky-Golay method, and (d) spectra after normalizing with dividing by the maximum ordinate value method.
Figure 7. The weights of trashes: (a and b) weights of the trashes separated in P3.3 and its ratio, (c and d) weights of the trashes separated in P3.1 and P3.2 and its ratio, and (e and f) total weights and its ratio.

in FT-NIR spectra. In addition, the samples in the prediction set would be deemed as abnormal samples if the trashes contents exceed the trashes content coverage of the calibration set. In total, seven abnormal samples in Aler group, seven abnormal samples in Kuitun group, and four abnormal samples in Shihezi group were excluded from the models. The parameters of optimized models are shown in Table 3.

The relationship between the calculated trashes contents and the actual trashes contents after excluding the abnormal samples is shown in Figure 9. It is obvious that the models have been significantly improved (The average RMSEC reduces to 0.072 g and the average RMSEP reduces to 0.158 g).

The ANOVA method is used to test whether there are significant differences between the actual trashes contents
Table 1. The statistical parameters of the weight of trashes and the content of trashes.

| Source | Amount | Attributes | Statistical parameters |
|--------|--------|------------|------------------------|
|        |        |            | Min | Max | Average | 25% quartile | 75% quartile | Median | SD |
| Shihezi| 118    | I.1        | 0.650 | 1.900 | 1.211 | 1.085 | 1.330 | 1.185 | 0.231 |
|        |        | I.2        | 0.350 | 2.510 | 1.045 | 0.755 | 1.260 | 0.965 | 0.421 |
|        |        | I.t        | 1.140 | 3.700 | 2.256 | 1.885 | 2.565 | 2.220 | 0.506 |
|        |        | R.1        | 0.033 | 0.095 | 0.061 | 0.054 | 0.067 | 0.059 | 0.012 |
|        |        | R.2        | 0.018 | 0.126 | 0.052 | 0.038 | 0.063 | 0.048 | 0.021 |
|        |        | R.t        | 0.057 | 0.185 | 0.113 | 0.094 | 0.128 | 0.111 | 0.025 |
| Kuitun | 115    | I.1        | 0.260 | 0.870 | 0.625 | 0.530 | 0.710 | 0.630 | 0.119 |
|        |        | I.2        | 0.130 | 2.120 | 0.668 | 0.390 | 0.820 | 0.630 | 0.385 |
|        |        | I.t        | 0.590 | 2.650 | 1.292 | 1.020 | 1.490 | 1.270 | 0.410 |
|        |        | R.1        | 0.013 | 0.044 | 0.031 | 0.027 | 0.036 | 0.032 | 0.006 |
|        |        | R.2        | 0.007 | 0.106 | 0.033 | 0.020 | 0.041 | 0.032 | 0.019 |
|        |        | R.t        | 0.030 | 0.133 | 0.065 | 0.051 | 0.075 | 0.064 | 0.021 |
| Aler   | 120    | I.1        | 0.660 | 2.400 | 1.081 | 0.955 | 1.195 | 1.070 | 0.218 |
|        |        | I.2        | 0.320 | 2.800 | 0.985 | 0.655 | 1.265 | 0.895 | 0.442 |
|        |        | I.t        | 1.030 | 4.480 | 2.066 | 1.675 | 2.380 | 2.005 | 0.545 |
|        |        | R.1        | 0.033 | 0.120 | 0.054 | 0.048 | 0.060 | 0.054 | 0.011 |
|        |        | R.2        | 0.016 | 0.140 | 0.049 | 0.033 | 0.063 | 0.045 | 0.022 |
|        |        | R.t        | 0.052 | 0.224 | 0.103 | 0.084 | 0.119 | 0.100 | 0.027 |

Table 2. The parameters of models between trashes contents and NIR spectra of SCHCP samples (before optimizing).

| Spectral type | Group | $R^2$-C | RMSEC | $R^2$-P | RMSEP |
|---------------|-------|---------|-------|---------|-------|
| Spectrum      | Aler  | 0.315   | 0.466 | 0.191   | 0.997 |
|               | Kuitun| 0.363   | 0.421 | 0.243   | 0.472 |
|               | Shihezi| 0.599  | 0.381 | 0.485   | 0.588 |
|               | Average| 0.426  | 0.423 | 0.306   | 0.686 |
| The first derivate | Aler | 0.720   | 0.337 | 0.380   | 0.781 |
|               | Kuitun| 0.887   | 0.206 | 0.282   | 0.514 |
|               | Shihezi| 0.662  | 0.353 | 0.583   | 0.573 |
|               | Average| 0.757  | 0.299 | 0.415   | 0.623 |
| The second derivate | Aler | 0.971   | 0.103 | 0.336   | 0.275 |
|               | Kuitun| 0.864   | 0.227 | 0.709   | 0.355 |
|               | Shihezi| 0.996  | 0.030 | 0.887   | 0.556 |
|               | Average| 0.944  | 0.120 | 0.644   | 0.395 |

and the calculated trashes contents. The results are shown in Table 4 from which, it is obvious that the trashes contents calculated with the models are consistent with the actual trashes contents.

Conclusions

This study is performed to investigate the potential of FT-NIR spectroscopy for the detection of botanical trashes contents of SCHCP samples. PLS models are built with the original FT-NIR spectra, the first derivate and the second derivate in bands of 12,000–4000 cm$^{-1}$, the results indicates that the second derivate is most suitable for establishing the NIR models to predict the botanical trashes contents of SCHCP samples, the correlation coefficient of optimized model is as high as 0.985 (calibration set), 0.973 (prediction set) and the RMSEC is 0.072 g and RMSEP is 0.158 g. The ANOVA analysis results also certify the trash contents calculated with the models are consistent with the actual trashes contents. Future studies will include more calibration samples and a NIR knowledge-based expert system (NIR-KBES) for the detection of trashes content of SC will be developed.
Table 3. The parameters of models between trashes content and NIR spectra of SCHCP samples (after optimizing).

| Spectral type       | Group    | $R^2$-C | RMSEC | $R^2$-P | RMSEP |
|---------------------|----------|---------|-------|---------|-------|
| Original spectra    | Aler     | 0.960   | 0.151 | 0.804   | 0.432 |
|                     | Kuitun   | 0.923   | 0.141 | 0.616   | 0.234 |
|                     | Shihezi  | 0.591   | 0.354 | 0.472   | 0.588 |
|                     | Average  | 0.824   | 0.215 | 0.631   | 0.418 |
| First derivate      | Aler     | 0.874   | 0.203 | 0.485   | 0.456 |
|                     | Kuitun   | 0.887   | 0.199 | 0.539   | 0.221 |
|                     | Shihezi  | 0.634   | 0.311 | 0.792   | 0.576 |
|                     | Average  | 0.798   | 0.238 | 0.605   | 0.418 |
| Second derivate     | Aler     | 0.972   | 0.098 | 0.981   | 0.283 |
|                     | Kuitun   | 0.987   | 0.072 | 0.983   | 0.073 |
|                     | Shihezi  | 0.996   | 0.047 | 0.956   | 0.119 |
|                     | Average  | 0.985   | 0.072 | 0.973   | 0.158 |

Figure 8. The relationship between the calculated trashes contents and actual trashes contents of SCHCP samples of Aler (a, d, and g), Kuitun (b, e, and h), and Shihezi (c, f, and i). The models were built with the original spectra (a–c), first derivate (d–f), second derivate (g–i) before optimizing.
Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful for the financial support provided by the National Natural Science Foundation of China (No. 31601224), the Natural Science Foundation of Anhui Provincial Department of Education No. KJ2019A0650 and KJ2020ZD004, the key research and development plan of Anhui province No. 202104a06020014.

ORCID iD
Wanhuai Zhou https://orcid.org/0000-0002-9134-4831

Table 4. The ANOVA analysis results (in this table, the actual trashes content is the average trashes content tested by ginning and trashes analysis, the calculated trashes content is the average trashes content predicted with the NIR models).

| Group   | Actual trashes content | Calculated trashes content | F      | p      | Significance |
|---------|------------------------|----------------------------|--------|--------|--------------|
| Aler    | 2.053                  | 1.973                      | 1.648  | 0.201  | No           |
| Kuitun  | 1.249                  | 1.241                      | 0.021  | 0.886  | No           |
| Shihezi | 2.236                  | 2.169                      | 1.231  | 0.269  | No           |

We are grateful for the financial support provided by the National Natural Science Foundation of China (No. 31601224), the Natural Science Foundation of Anhui Provincial Department of Education No. KJ2019A0650 and KJ2020ZD004, the key research and development plan of Anhui province No. 202104a06020014.

ORCID iD
Wanhuai Zhou https://orcid.org/0000-0002-9134-4831

Figure 9. The relationship between the calculated trashes contents and actual trashes contents of SCHCP samples of Aler (a, d, and g), Kuitun (b, e, and h), and Shihezi (c, f, and i). The models were built with the original spectra (a–c), first derivate (d–f), second derivate (g–i) after excluding the abnormal samples.

Table 4. The ANOVA analysis results (in this table, the actual trashes content is the average trashes content tested by ginning and trashes analysis, the calculated trashes content is the average trashes content predicted with the NIR models).

| Group   | Actual trashes content | Calculated trashes content | F      | p      | Significance |
|---------|------------------------|----------------------------|--------|--------|--------------|
| Aler    | 2.053                  | 1.973                      | 1.648  | 0.201  | No           |
| Kuitun  | 1.249                  | 1.241                      | 0.021  | 0.886  | No           |
| Shihezi | 2.236                  | 2.169                      | 1.231  | 0.269  | No           |

We are grateful for the financial support provided by the National Natural Science Foundation of China (No. 31601224), the Natural Science Foundation of Anhui Provincial Department of Education No. KJ2019A0650 and KJ2020ZD004, the key research and development plan of Anhui province No. 202104a06020014.

ORCID iD
Wanhuai Zhou https://orcid.org/0000-0002-9134-4831

Figure 9. The relationship between the calculated trashes contents and actual trashes contents of SCHCP samples of Aler (a, d, and g), Kuitun (b, e, and h), and Shihezi (c, f, and i). The models were built with the original spectra (a–c), first derivate (d–f), second derivate (g–i) after excluding the abnormal samples.
References

1. Food and Agriculture Organization of the United Nations. FAOSTAT[EB/OL], https://www.fao.org/faostat/zh (2021, accessed 1 April 2021).
2. Cao JQ, Xv H and Xie ZL. Study on the influence degree of ginning process on impurities and neps in machine-picked cotton. Adv Text Technol 2019; 27(6): 21–25.
3. Xv H, Chao JQ, Xie W, et al. Influence of saw type lint cleaning on performance of machine stripped cotton. Text Res J 2014; 351: 35–39.
4. Zhang H and Li D. Applications of computer vision techniques to cotton foreign matter inspection: a review. Comput Electron Agric 2014; 109: 59–70.
5. Lieberman MA and Patil RB. Clustering and neural networks to categorize cotton trash. Opt Eng 1994; 33(5): 1642–1653.
6. Lieberman MA and Patil RB. Evaluation of learning vector quantization to classify cotton trash. Opt Eng 1997; 36(3): 1642–1653.
7. Lieberman MA, Bragg CK and Brennan SN. Determining gravimetric bark content in cotton with machine vision. Text Res J 1998; 68(2): 94–104.
8. Li GH, Su ZW and Xia XY. Algorithm for inspection of white foreign fibers in cotton by machine vision with irregular imaging function. Trans Chin Soc Agric Mach 2010; 41(5): 164–167.
9. Zhang CL, Li L, Dong QC, et al. Recognition for machine picking seed cotton trashes based on the GA-SVM model. Trans Chin Soc Agric Eng 2016; 32(24): 189–196.
10. Fortier CA, Rodgers JE, Cintrón MS, et al. Identification of cotton and cotton trash components by Fourier transform near-infrared spectroscopy. Text Res J 2011; 81(3): 230–238.
11. Fortier CA, Rodgers JE, Foulk J, et al. Near-infrared classification of cotton lint, botanical and field trashes. J Cotton Sci 2012; 16(1): 72–79.
12. Rodgers J, Fortier C, Montalvo J, et al. Near infrared measurement of cotton fiber micronaire by portable near infrared instrumentation. Text Res J 2010; 80(15): 1503–1515.
13. Liu Y, Gamble G and Thibodeaux D. Uv/visible/near-infrared reflectance models for the rapid and non-destructive prediction and classification of cotton color and physical indices. Trans ASABE 2010; 53(4): 1341–1348.
14. Liu YL, Thibodeaux D and Gamble G. Development of Fourier transform infrared spectroscopy in direct, non-destructive, and rapid determination of cotton fiber maturity. Text Res J 2011; 81(15): 1559–1567.
15. Liu Y and Foulk J. Potential of visible and near infrared spectroscopy in the determination of instrumental leaf grade in lint cottons. Text Res J 2013; 83(9): 928–936.
16. Allen A, Foulk J and Gamble G. Preliminary fourier-transform infrared spectroscopy analysis of cotton trash. J Cotton Sci 2007; 11(1): 68–74.
17. Gamble GR and Foulk JA. Quantitative analysis of cotton (gossypium hirsutum) lint trash by fluorescence spectroscopy. J Agric Food Chem 2007; 55(13): 4940–4943.
18. Gaitán-Jurado AJ, García-Molina M, Peña-Rodríguez F, et al. Near infrared applications in the quality control of seed cotton. J Near Infrared Spec 2008; 16(4): 421–429.
19. Zhang R, Li C, Zhang M, et al. Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification. Comput Electron Agric 2016; 127: 260–270.
20. Mustafic A, Li C and Haidekker M. Blue and UV LED-induced fluorescence in cotton foreign matter. J Biol Eng 2014; 8: 29–40.
21. Mustafic A and Li C. Classification of cotton foreign matter using color features extracted from fluorescent images. Text Res J 2015; 85(12): 1209–1220.
22. Jia DY and Ding T. Detection of foreign fibers in cotton using near-infrared optimal wavelength imaging. Opt Eng 2005; 44(7): 076402.
23. Foulk JA, McAlister D and Kulasekera KB. Detecting the cotton trashes particle size distribution in mill laydown using HVI trashes meter software. J Text App Tech Manage 2006; 52: 1–11.
24. Yan YL, Chen B, Zhu DZ, et al. Near infrared spectroscopy-principles, technology and applications. Beijing: China Light Industry Press Ltd, 2013, pp.45–46.