Feasibility Study on Use of Near Infrared Spectroscopy for Rapid and Non-Destructive Determination of Gossypol Content in Intact Cottonseeds

Cheng Li
Zhejiang University
Bangsong Su
Zhejiang University
Tianlun Zhao
Zhejiang University
Cong Li
Zhejiang University
Jinhong Chen
Zhejiang University
Shuijin Zhu (shjzhu@zju.edu.cn)
Zhejiang University https://orcid.org/0000-0001-6209-9630

Research

Keywords: intact cottonseed, chemometrics, gossypol, near infrared spectroscopy

Posted Date: August 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-53311/v1

Abstract

Background

Gossypol found in cottonseeds is toxic to human beings and monogastric animals and is a primary parameter for integrated utilization of cottonseed products. It is usually determined by the techniques relied on complex pretreatment procedures and the samples after determination cannot be used in breeding program, so it is of great importance to predict the gossypol content in cottonseeds rapidly and non-destructively to substitute the traditional analytical method.

Results

Gossypol content in cottonseeds was investigated by near-infrared spectroscopy (NIRS) and High-performance liquid chromatography (HPLC). Partial least squares regression, combined with spectral pretreatment methods including Savitzky-Golay smoothing, standard normal variate, multiplicative scatter correction, and first derivate, were tested for optimizing the calibration models. NIRS technique was efficient in predicting gossypol content in intact cottonseeds, as revealed by the root-mean-square error of cross-validation (RMSECV), root-mean-square error of prediction (RMSEP), coefficient for determination of prediction (R_p^2), and residual predictive deviation (RPD) values for all models, being 0.05–0.07, 0.04–0.06, 0.82–0.92, and 2.3–3.4, respectively. The optimized model pretreated by Savitzky-Golay smoothing + standard normal variate + first derivate resulted in good determination of gossypol content in intact cottonseeds.

Conclusions

Near infrared spectroscopy coupled with different spectral pretreatments and PLS regression has exhibited the feasibility in predicting gossypol content in intact cottonseeds, rapidly and non-destructively. It could be used as an alternative method to substitute for traditional one to determine the gossypol content in intact cottonseeds.

1. Introducton

Cotton (*Gossypium*. spp) is one of the important industrial and economic crops.[1] Cottonseed, the main by-product of cotton production, can be used to produce food, animal feed, and other products. Cottonseed contains many kinds of nutrients, including proteins, oils, fatty acids, and amino acids, making it a potential food resource for human beings with the rapid growth of global population[2]. However, the *Gossypium* species are characterized by the presence of gossypol, which is toxic to human beings and monogastric animals,[3] such that the utilization of cottonseed products is limited.

Gossypol, 1, 1’, 6, 6’, 7, 7’-hexahydroxy-5, 5’-diisopropyl-3, 3’-dimethyl-(2, 2’ binaphthalene)-8, 8’-dicarbaldehyde, is a terpenoid compound that help cotton defend against biotic stress.[4–6] Due to the toxicity of gossypol, breeding for both lower gossypol content in cottonseeds and higher gossypol content in cotton plants has been practiced in many cotton-planting countries. The cottonseed breeding work often requires analyzing a large number of cottonseed samples to measure gossypol content. Conventionally, gossypol content is assayed by UV spectrophotometry which not only involves reagents with great toxicity, but also is not accurate and reliable.

High-performance liquid chromatography (HPLC) is generally expensive and time-consuming, although it was high accuracy and sensitivity for gossypol determination. In addition, both classical analytical methods cause undesired destruction of the testing samples which frequently needed to be planted in cotton breeding program. So, a rapid and non-destructive method for gossypol determination is required.

Near infrared (NIR) spectroscopy combined with chemometrics is a rapid, convenient, and environmentally-friendly analytical technique in the quality analysis for crops.[7–20] However, it is a challenge to determine gossypol content in intact cottonseeds by NIR, due to (i) cottonseed being bigger than other crop seeds, so large voids are left between packed samples in sample cells; (ii) some of immature and wizened cottonseeds can be mixed in the samples, which can introduce irrelevant information into the spectra data; and (iii) the tough and thick shell of cottonseed can impact the penetration of NIR light and result in a lower S/N ratio and poor information. Because of these factors, the spectral data of intact cottonseeds are far more complex than that of other crop seeds, which may contain a large amount of useless and uncorrelated information such as noise and background. To overcome these difficulties, sophisticated chemometric methods are applied to extract useful information from NIR spectra and calibrate robust models for gossypol content in intact cottonseeds. Essentially, these include regression methods such as principal component regression (PCR)[21], partial least squares (PLS)[22], support vector machines (SVM)[23], least squares support vector machines (LS-SVM)[24], and artificial neural networks (ANN)[25], coupled with spectral pretreatments such as standard normal variate (SNV)[26], Savitzky-Golay (SG) smoothing[27], multiplicative scatter correction (MSC)[28], and first derivate[29].

Due to undesired destruction of the test sample, previous NIR models which can be used in detection of gossypol in cottonseed meal, can be barely applied in breeding trails.[30] In this present study, spectroscopy was investigated the feasibility of analyzing gossypol in intact cottonseeds based on NIR spectrometer. The main aim of this study was to establish an optimal model which could provide a powerful technical support for cotton breeders and other people who work on cottonseeds.

2. Materials And Methods

2.1 Samples and preparation

A total of 268 samples of cottonseeds were collected from different cultivated areas, including Hangzhou (Zhejiang, China), Xiaoshan (Zhejiang, China), Sanmen (Zhejiang, China), Sanya (Hainan, China), Wuhu (Anhui, China), and Yancheng (Jiangsu, China), in 2012, 2013, and 2014, which kept in cold storage at 4℃. The cottonseed samples were delinted and dried at 30℃ to constant weight. After spectral acquisition by NIR spectroscopy, the intact cottonseed samples were hulled, and then ground to cottonseed kernel powder for HPLC analysis. The preparations were implemented in the same experimental condition in order to reduce the influence of other physical factors.

2.2 Gossypol extraction

0.1 g of cottonseed kernel powder was suspended in 5 mL acetone and sonicated in an ultrasonic bath for 45 minutes. Then, the suspension was filtered through quantitative filter paper followed by a filtration with a 0.45 μm syringe filter (Agela, Newark, USA). The sediment was washed three times by acetone. After this procedure, the extract was adjusted to 25 mL using acetone.

2.3 HPLC analysis

HPLC analysis was performed on an Agilent 1100 HPLC system (Agilent, Santa Clara, USA), equipped with an auto-sampler and UV detection. A C_{18} column (250 mm × 4.6 mm, 5 μm, Dikma, Richmond Hill, USA) was employed as the stationary phase. The mobile phase consisted of methanol/0.2% H_3PO_4 (80/20, v/v). The injection volume was 10 μL and the flow rate was 1.0 mL min^{-1}. The UV detector was set at 238 nm and the column temperature was 25℃. Each sample was measured three times. The limit of detection (LOD) was obtained at a signal-to-noise (S/N) ratio of three and the limit of quantification (LOQ) at an S/N ratio. To detect the stability of gossypol at room temperature, three samples were randomly employed to determine the changes of peak area within 36 hours. HPLC-grade gossypol was purchased from Sigma (Sigma-Aldrich, St. Louis, USA). Methanol (HPLC grade) was procured from Tianjin Chemical Reagent Company (Tianjin, China). Double deionized water was prepared using Milli-Q-water purification system (Millipore, Molsheim, France).

2.4 NIR spectra acquisition

The NIR spectra of intact cottonseed samples were scanned with a Büchi Flex-N500 NIR spectrometer (Büchi, Flawil, Switzerland), equipped with a solid sample module as followings. The NIR spectra were collected across the range 4000–10000 cm^{-1}, and were recorded with a spectral resolution of 4 cm^{-1}. Samples were measured three times on a rotating cylinder device at 25 ± 0.5℃ and 60% relative air humidity. All the spectra were transformed into absorbance (log (1/R)).

2.5 Spectral pretreatment

Before calibration, the spectral data were pretreated for an optimal performance. Eight pretreatment strategies which included one or some combination of Savitzky-Golay smoothing, SNV, MSC, and first derivate (Norris gap) were compared with the raw spectra.

2.6 Sampling design

Samples were assigned to calibration and prediction sets using Kennard-Stone (KS) selection.[31] The calibration models were established with the calibration set, and the prediction set was used to validate the predictive capabilities and analytical features of the calibration models.

2.7 PLS regression

PLS regression has been widely used as a calibration method to investigate the relationship between the spectral and the corresponding reference data. Before calibration of the PLS models, the data sets (spectral and reference data) were analyzed using 4-fold cross-validation to develop a full-spectra calibration model. The aim of the cross-validation was to find the optimum number of latent variables (LV) for PLS. The root-mean-square error of cross-validation (RMSECV) served as a measure to adjust the parameters, and the number of LV which provide the lowest RMSECV was selected as the best.

2.8 Model evaluation

The estimate of the calibration models was based on following quality parameters:

$$R^2 = 1 - \sum_{i=1}^{n}\left(Y_{nirs} - Y_{ref}\right)^2 / \sum_{i=1}^{n}\left(Y_{ref} - \overline{Y_{ref}}\right)^2 \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\left(Y_{nirs} - Y_{ref}\right)^2 / N} \quad (2)$$

$$RPD = SD_{Y_{ref}} / RMSEP \quad (3)$$

where N is the total number of samples, Y_{nirs} is the predicted value by calibrationmodels, Y_{ref} is the reference value by HPLC, and SD is the standard deviation.

The coefficient for determination of prediction (R_p^2), the root mean square error of prediction (RMSEP), the coefficient for determination of calibration (R_c^2), the root mean square error of cross-validation (RMSECV), and the residual predictive deviation (RPD) were used as criterion to evaluate model performance. An acceptable model should have high R_c^2 and R_p^2 values and low RMSECV and RMSEP values. Meanwhile, the model is considered as robust if the RPD is higher than 2.5.

2.9 Software

NIR spectroscopic data (268 samples × 1501variables) were exported in text format, organized in Excel spreadsheets, and then transferred into MATLAB R2011a (Math Works, Natick, USA) for chemometric analysis. All the algorithms in spectral pretreatments, sampling design, and regressions were implemented with MATLAB R2012a.

3. Results

3.1 HPLC analysis

The regression equation, correlation coefficient (r^2), limits of detection (LOD), limits of quantification (LOQ), and average recovery of gossypol were illustrated in Table 1. The retention times of gossypol standard and gossypol extraction were 9.91 and 9.60 minutes, respectively (Fig. 1). Table 2 shows the stability for peak area of gossypol determined by HPLC for 24 hours. All the results indicated that the improved HPLC method could be used to detect gossypol content, and the cottonseed extract should be analyzed within 24 hours.

Table 1
HPLC-VU results.

Regression equation	r^2	LOD (μg/mL)	LOQ (μg/mL)	Average recovery (%)
Y = 102.42 X − 85.055	0.999	37.5	125.0	95.16-101.72

r^2, correlation coefficient; LOD, limits of detection; LOQ, limits of quantification; Y, peak area; X, concentration (μg/mL).

Table 2
The stability of gossypol determined for HPLC during 24 h.

Sample number	Time (h)									*RSD* (%)
	0	3	6	9	12	15	18	21	24	
1	2378.8	2398.0	2397.9	2435.2	2414.2	2421.1	2428.1	2452.7	2465.3	1.14
2	2871.7	2848.1	2860.2	2885.0	2876.6	2886.2	2932.9	2939.8	3010.5	1.76
3	2854.8	2839.8	2862.2	2892.6	2888.7	2918.6	2965.1	2972.7	3007.1	2.01

RSD, relative standard deviation.

3.2 NIR spectra analysis

Across the spectral range of 4000–10000 cm^{-1}, absorbance values are mainly associated with the combination and overtone bands of the C-H, N-H, O-H, and S-H bonds,[32] which were quite sensitive to the compositional variations in complex samples. Figure 2A shows the raw intact cottonseed spectra in the NIR spectral region. The spectra showed six broad absorption peaks around the 4200, 4700, 5150, 5580, 6900, and 8400 cm^{-1}. The small peak observed at 4200 cm^{-1} fell within the regions associated with the combination bands of C-H. At 5150 and 6900 cm^{-1}, these could be attributed to the combination and the first overtone bands of O-H, respectively, which were identified as water absorption. The gentle peaks at 5580 and 8400 cm^{-1}overlapped with the second and first C-H overtone regions, respectively. It was worth mentioning that the peak at 4700 cm^{-1} was attributed to the first C-H combination bands of alkenes and aromatic hydrocarbons, which could be identified as the absorption of polyphenolic terpenes, including gossypol and its derivatives.

The raw spectra were homogeneous, so the presence of noise could not be directly identified. Consistent baseline offsets and bias were present in the spectra, which are common features in the NIR spectra. Hence, eight pretreatment strategies were performed to optimize the raw spectra before establishment of the calibration models. The pretreatment spectra of several types of representative strategies were shown in Fig. 2B, Fig. 2C, and Fig. 2D. To different degrees, all these pretreatments could reduce the physical change among samples due to scattering and remove both additive and multiplicative effects in the spectra. It was noted that ten variables were lost after SG smoothing. Hence, the 1491 variables were used for calibration among the models using SG smoothing during the spectral pretreatments.

3.3 Kennard-Stone sampling design

The Kennard-Stone algorithm is an effective method for extracting a sample subset in multidimensional space, which includes all the most diverse samples and enables the selection of a subset of representative samples. Therefore, it has been confirmed that a calibration set extracted using KS selection has a better predictive capability than a set randomly built or constructed by other data selection methods such as Kohonen self-organized mapping[33] and D-optimal designs[34]. In this study, the total 268 intact cottonseed samples were divided into calibration and prediction sets based on KS algorithm, wtih the former set consisting of 218 samples and the later one 50 samples. The statistical values of gossypol contents in all cottonseed samples for calibration and prediction set were demonstrated in Table 3, which indicated that the range of variation for gossypol content was broad enough to develop NIR calibration models.

Table 3
Statistical values of gossypol content for calibration and prediction set samples.

Data sets	N	Minimum (g kg^{-1})	Maximum (g kg^{-1})	Mean (g kg^{-1})	SD
Calibration set	218	0.32	1.04	0.63	0.16
Prediction set	50	0.35	0.95	0.65	0.15
All samples	268	0.32	1.04	0.64	0.16

N, number of samples; SD, standard deviation.

3.4 PLS regression

The calibration models of gossypol content in intact cottonseeds based on PLS regression were established in the NIR spectral range of 4000–10000 cm^{-1}, and the results were summarized in Table 4. The number of LV were selected with the aid of cross-validation using the first minimum RMSECV for all models. The RMSECV and RMSEP values for all the calibration models were between 0.05–0.07 and 0.04–0.06 for calibration and prediction sets, respectively. The values of R_p^2 and R_c^2 ranged from 0.82 to 0.93 and from 0.87 to 0.97, respectively. The RPD values ranged from 2.3 to 3.4.

Table 4
Performance comparison results for calibration models using different spectral pretreatment strategies.

Spectral pretreatments	LVs	Calibration set		Prediction set		
		RMSECV	R_c^2	RMSEP	R_p^2	RPD
Raw	9	0.06	0.94	0.06	0.82	2.3
SG	9	0.07	0.87	0.06	0.86	2.5
SNV	10	0.06	0.92	0.05	0.87	2.8
MSC	8	0.06	0.92	0.05	0.88	2.8
1st D	9	0.05	0.95	0.05	0.87	2.7
SNV + 1st D	10	0.06	0.96	0.05	0.89	3.0
MSC + 1st D	9	0.06	0.96	0.05	0.88	2.9
SG + SNV + 1st D	9	0.05	0.97	0.04	0.92	3.4
SG + MSC + 1st D	10	0.05	0.96	0.05	0.89	3.0

LVs, latent variables; RMSECV, root mean square error of cross-validation; R_c^2, coefficient of determination of calibration; RMSEP, root mean square error of prediction; R_p^2, coefficient of determination of prediction; RPD, residual predictive deviation; Raw, raw spectra; SG, Savitzky-Golaysmoothing; SNV, standard normal variate; MSC, multiplicative scatter correction; 1st D, first derivate.

4. Discussion

Since NIR spectra of intact cottonseeds were complex and overlapped, suitable spectral pretreatments should be used to optimize the NIR spectra and extract the effective information. In this work, the raw spectra were transformed using by eight pretreatment strategies, including the singe pretreatment strategies (SG smoothing, SNV, MSC, and first derivate), two pretreatments strategies (SNV + first derivate and MSC + first derivate), and three pretreatments strategies (SG smoothing + SNV + first derivate and SG smoothing + MSC + first derivate). In analyzing the results obtained from singe pretreatment strategies, the PLS model using eight latent variables based on application of MSC produced better results with low values of RMSECV and RMSEP (0.06 and 0.05, respectively), and the RPD value was increased by 20.36% compared with that of the direct regression model based on raw spectra (Fig. 4). Figure 3B shows the correlation of model using MSC, presented by plotting predicted and reference values for gossypol content in intact cottonseeds. The samples near the diagonal line indicated that their predicted values were more closed to reference ones and vice versa. In the aspect of two pretreatments strategies, the calibration model based on SNV + first derivate, presented a better predictive ability than that on MSC + first derivate, with the R_c^2 and R_p^2 values of 0.962 and 0.887, respectively. The RPD value of that model was 3.0, increased by 28.14% compared to the model using raw spectra. From all the results of calibration models established, the best model was pretreated using the strategy of SG + SNV + first derivate, and it had the highest R_c^2 (0.97) andR_p^2 (0.93), and the RPD (3.4) increased by 46.28% compared with that of raw spectral model. Furthermore, RMSECV (0.05) and RMSEP (0.04) were the lowest among all the models. The correlation plots between predicted and reference values were focused on the diagonal line (Fig. 3D). It was indicated that the model using SG + SNV + first derivate and PLS was accurate and robust enough to substitute the conventional gossypol analysis methods to measure gossypol in intact cottonseeds.

The NIR spectra of these intact seeds generally contained a mass of undesirable features, including noise, overlapping peaks, baseline effects and some systematic behaviors, caused by the seed size, shell, and some other physical factors. Hence, a suitable pretreatment strategy was required for widespread application of NIR technology in crop seed analysis. In this work, it was indicated that an advisable pretreatment strategy before regression was important to refine the effective information from spectral data and eliminate spectral deviation to calibrate an accurate and robust NIR model.

The calibration models reported here confirmed the feasibility of the using of NIR technology for rapid and non-destructive determination of gossypol, an important parameter to cottonseed products, in intact cottonseeds for the first time. The high RPD values (3.4) suggested that this technology could be an effective method for the measurement of gossypol in intact cottonseeds. The optimal model could substitute conventional analysis methods for gossypol, including UV spectrophotometry and HPLC. Because of the potential of high sample throughput and low costs, as well as a significant reduction in toxic chemicals, the application of NIR method could be encouraged and popularized to other similar agricultural products.

5. Conclusions

The calibration and validation statistics obtained in the current work showed the potential of NIRS to predict microelement gossypol content in intact cottonseeds. The optimized model was that pretreated by Savitzky-Golay smoothing + standard normal variate + first derivate, with RMSECV, RMSEP, R_p^2, and RPD of 0.05, 0.04, 0.92, and 3.4, respectively, which provided a method to determine gossypol content in intact cottonseeds feasibly.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All co-authors have consent for submission of manuscript.

Availability of data and materials

All relevant data are within this article.

Competing interests

The authors declare that they have no competing interests.

Funding

The research work was funded by The National Key Technology R&D program of China (2016YFD0101404), China Agriculture Research System (CARS-18-25), and Jiangsu Collaborative Innovation Center for Modern Crop Production.

Author's contributions

Li C (Cheng) and Zhu SJ designed the experiments and wrote the manuscript. Li C (Cheng), Zhao TL and Su BS analyzed the data, Li C(Cheng), SU BS, Li C (Cong) participated in experiment. Chen JH assisted in editing the article. Zhu SJ and Chen JH conducted and supervised the experiments.

Acknowledgments

We are grateful to Mrs Yu Liu for her technical assistance.

References

1. Sunilkumar G, Campbell CL, Puckhaber L, et al. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *PNAS* **103**:18054–18059 (2006).
2. Sawan MZ, Hafez AS, Basyony EA, et al. Cottonseed, protein, oil yields and oil properties as affected by nitrogen fertilization and foliar application of potassium and a plant growth retardant. *World J Agri Sci* **1**:56-65 (2006).
3. Lordelo M.M, Davis AJ, Calhoun MC, et al. Relative toxicity of gossypol enantiomers in broilers. *Poultry Sci* **84:**:1376–1382 (2005).
4. Kong GC, Daud KM and SJ Z. Effects of pigment glands and gossypol on growth, development and insecticide-resistance of cotton bollworm (*Heliothis armigera* (Hübner)). *Crop Prot* **29**:813-819 (2010).
5. Lin TS, Schinazi RF, Zhu JL, et al. Anti-Hiv-1 activity and cellular pharmacology of various analogs of gossypol. *Biochem Pharmacol* **46**:251-255 (1993).

6. Blanco A, Aoki A, Montamat E, et al. Effect of gossypol upon motility and ultrastructure of *Trypanosoma cruzi*. **30**:649-651 (1983).
7. Sohn M, Himmelsbach SD, Barton EF, et al. Near-infrared analysis of whole kernel barley: comparison of three spectrometers. *Appl Spectrosc* **62**:427-432 (2008).
8. Huang ZR, Sha S, Rong ZQ, et al. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed. *Ind Crop Prod* **43**:654– 660 (2013).
9. Weinstock A, Janni J, Hagen L, et al. Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl Spectrosc* **60**:9-16 (2006).
10. Rosales A, Galicia L, Oviedo E, et al. Near-infrared reflectance spectroscopy (NIRS) for protein, tryptophan, and lysine evaluation in quality protein maize (QPM) breeding programs. *J Agric Food Chem* **59**:10781–10786 (2011).
11. Bellato S, Frate DV, Redaelli R, et al. Use of near infrared reflectance and transmittance coupled to robust calibration for the evaluation of nutritional value in naked oats. *J Agric Food Chem* **59**:4349–4360 (2011).
12. Bala M and Singh M. Non destructive estimation of total phenol and crude fiber content in intact seeds of rapeseed–mustard using FTNIR. *Ind Crop Prod* **42**:357– 362 (2013).
13. Hacisalihoglu G, Larbi B and Settles A. Near-infrared reflectance spectroscopy predicts protein, starch, and seed weight in intact seeds of common bean (Phaseolus vulgaris L.). *J Agric Food Chem* **58**:702-706 (2010).
14. Mendoza AF, Cichy AK, Sprague C, et al. Prediction of canned black bean texture (*Phaseolus vulgaris* L.) from intact dry seeds using visible/near-infrared spectroscopy and hyperspectral imaging data. *J Sci Food Agric* **98**: 283–290 (2018).
15. Lee H, Kim M, Song Y, et al. Non-destructive evaluation of bacteria-infected watermelon seeds using visible/near-infrared hyperspectral imaging. *J Sci food Agric* **97**:1084–1092 (2017).
16. Tierno R, López A, Riga P, et al. Phytochemicals determination and classification in purple and red fleshed potato tubers by analytical methods and near infrared spectroscopy. *J Sci food Agri* **96**:1888–1899 (2016).
17. Yang N and Ren QX. Application of near-infrared reflectance spectroscopy to the evaluation of rutin and d-chiro-inositol contents in tartary buckwheat. *J Agric Food Chem* **56**:761–764 (2008).
18. Lin C, Chen X, Jian L, et al. Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chem* **162**:10-15 (2013).
19. Kovalenko VI, Rippke RG and Hurburgh RC. Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *J Agric Food Chem* **54**:3485-3491 (2006).
20. Fassio A and Cozzolino D. Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy. *Ind Crop Prod* **20**:321–329 (2004).
21. Xie YL and Kalivas HJ. Local prediction models by principal component regression. *Anal Chim Acta* **348**:29-38 (1997).
22. Haaland MD and Thomas VE. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal Chem* **60**:1193-1202 (1988).

23. Nie Z, Han J, Liu T and Liu X. Hot topic: Application of support vector machine method in prediction of alfalfa protein fractions by near infrared reflectance spectroscopy. *J Dairy Sci* **91**:2361-2369 (2008).
24. Shao YN, Zhao CJ, Bao YD, et al. Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food Bioprocess Technol* **5**:100–107 (2012).
25. Makinoa Y, Ichimura M, Oshita S, et al. Estimation of oxygen uptake rate of tomato (Lycopersicon esculentum Mill.) fruits by artificial neural networks modelled using near-infrared spectral absorbance and fruit mass. *Food Chem* **121**:533–539 (2010).
26. Barnes R, Dhanoa M and Lister S. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc* **43**:772-777 (1989).
27. Savitzky A and Golay M. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* **36**:1627-1639 (1964).
28. KP H. The evolution of chemometrics. *Anal Chim Acta* **500**:365–377 (2003).
29. Rinnan Å, van den Berg F and Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *Trend Anal Chem* **28**:1201-1222 (2009).
30. Li C, Zhao TL, Li C, et al. Determination of gossypol content in cottonseeds by near infrared spectroscopy based on Monte Carlo uninformative variable elimination and nonlinear calibration methods. *Food Chem* **221**:990–996 (2017).
31. Kennard RW and Stone LA, Computer aided design of experiments. *Technometrics* **11**:137-148 (1969).
32. Macho S and Larrechi MS, Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends in Analytical Chemistry* **21**:799-806 (2002).
33. Kohonen T, Analysis of a simple self-organizing process. *Biol Cybern* **44**:135-140 (1982).
34. de Aguiar PF, Bourguignon B, Khots MS, et al. D-optimal designs. *Chemometr Intell Lab Syst* **30**:199-210 (1995).

Figures

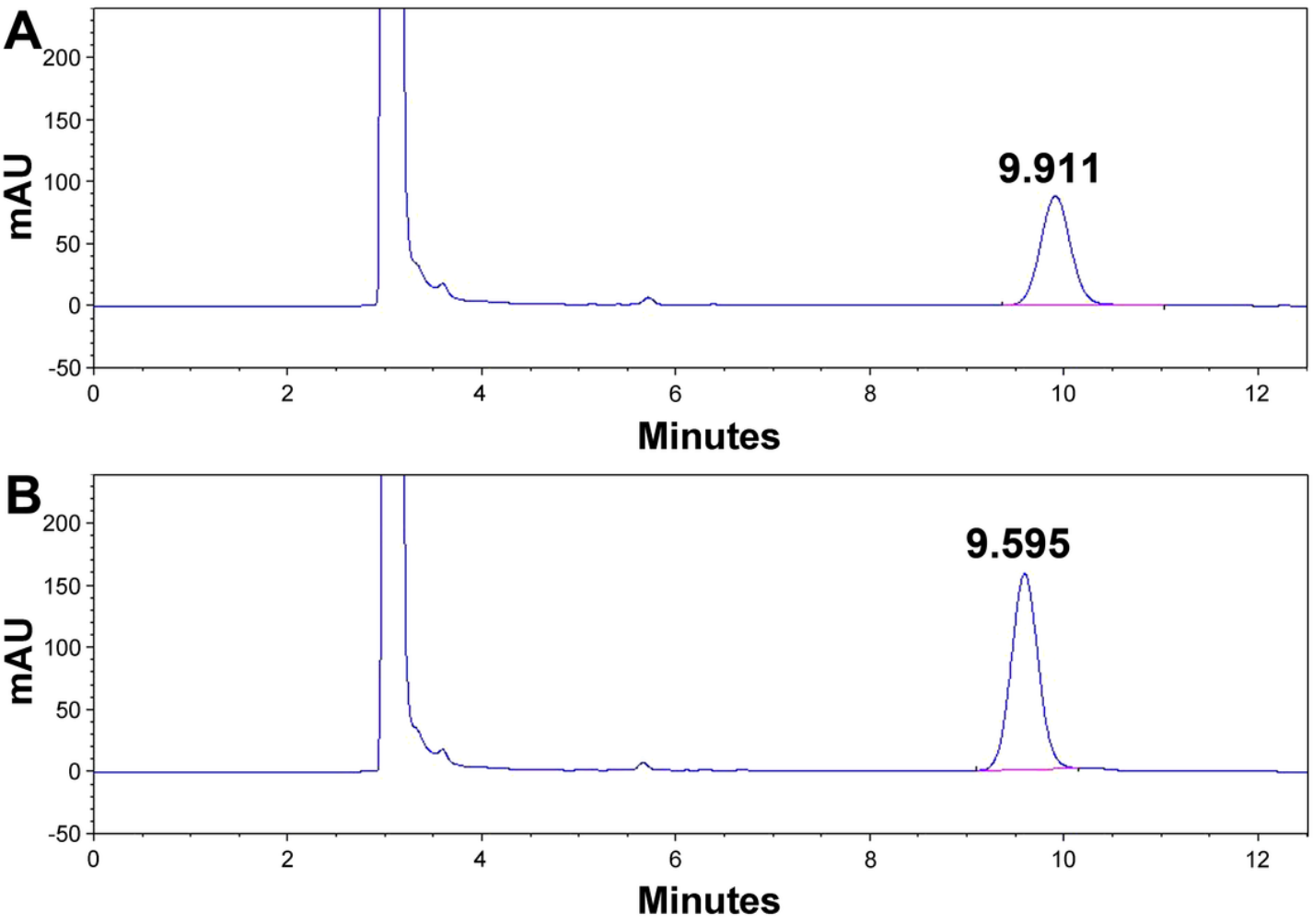

Figure 1

Chromatograms of (A) standard gossypol and (B) gossypol extracted in cottonseeds.

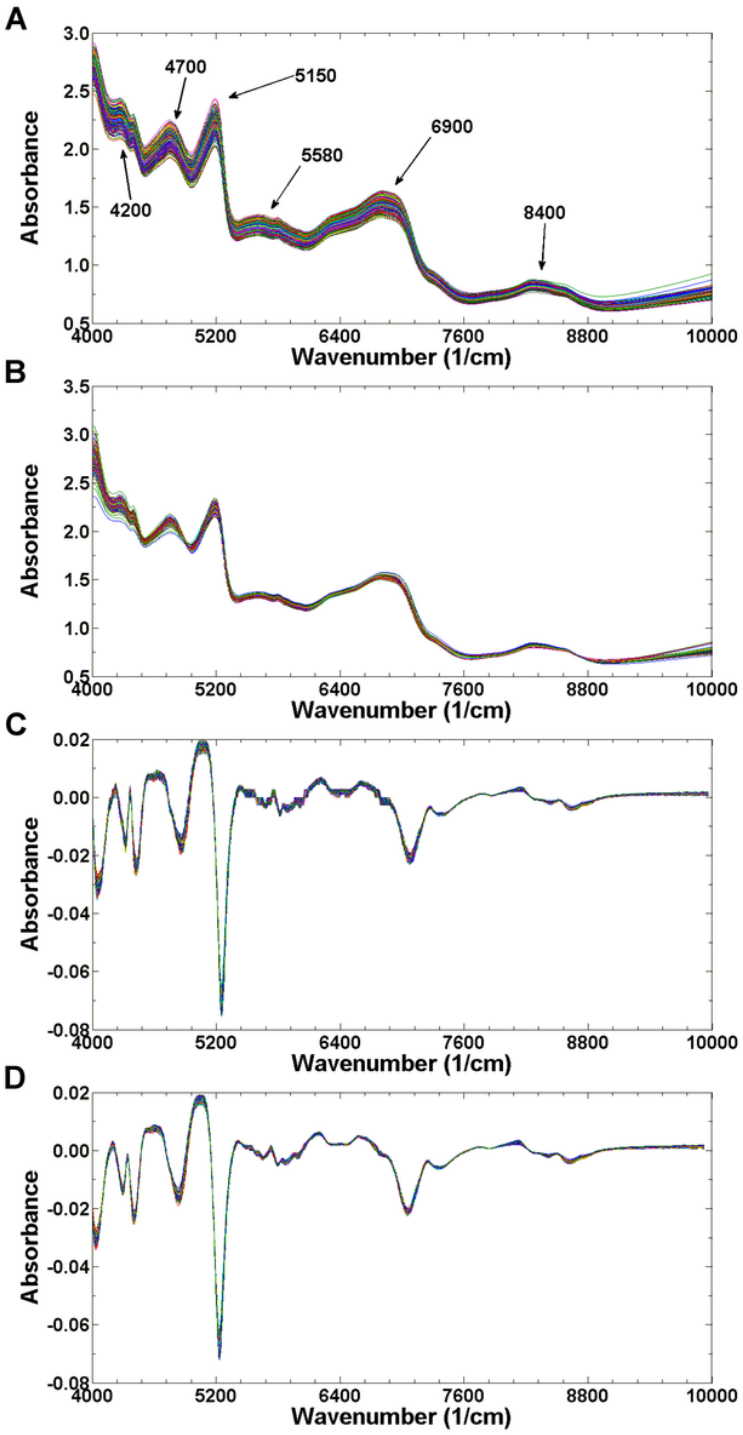

Figure 2

The NIR spectra of intact cottonseeds. (A) The raw spectra, (B) the spectra pretreated by MSC, (C) the spectra pretreated by SNV+ first derivate, and (D) the spectra of pretreated by SG smoothing+ SNV+ first derivate.

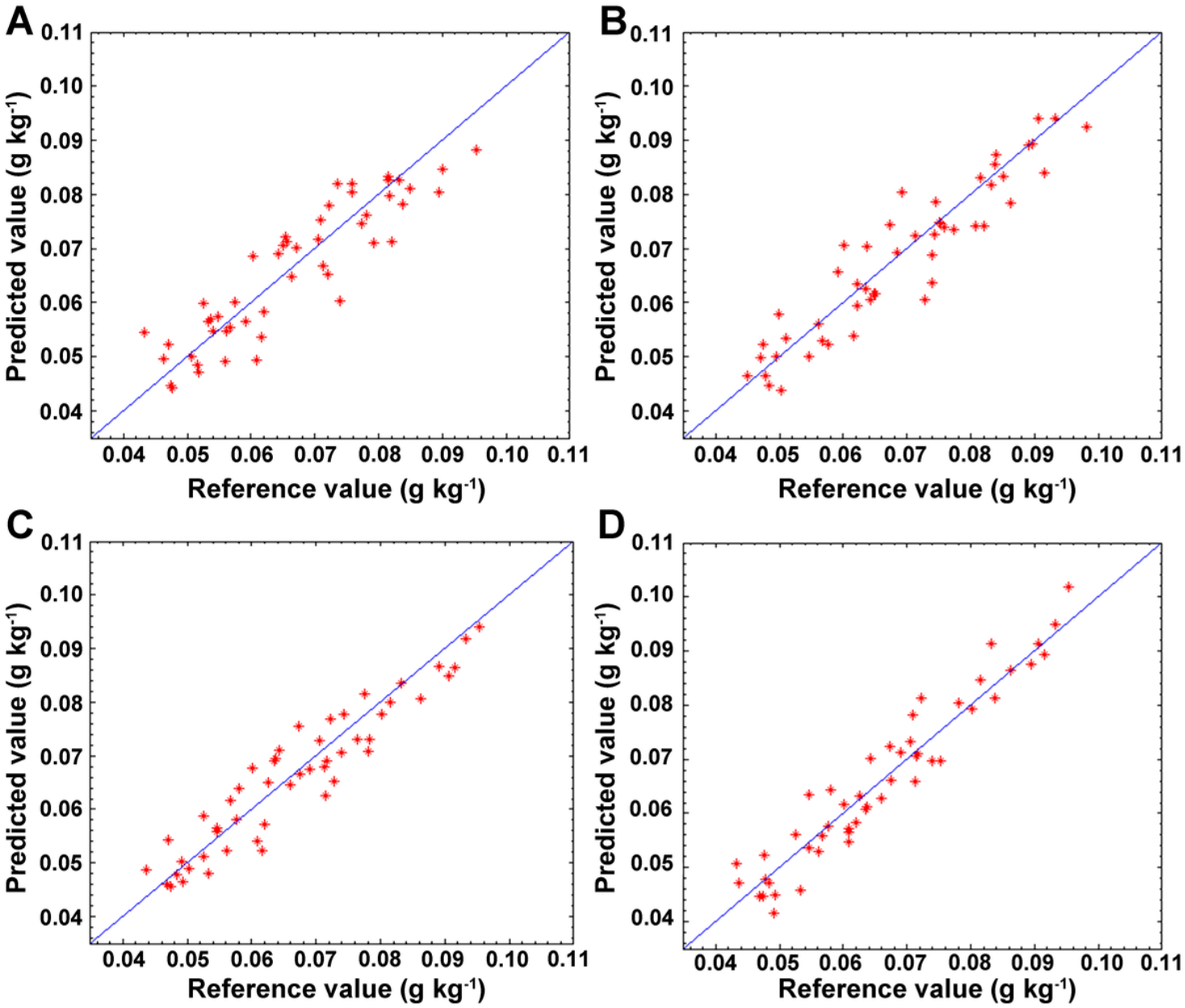

Figure 3

The correlation between predicted and reference values for models of intact cottonseeds. (A) the PLS model based on raw spectra, (B) the PLS model based on the pretreatment of MSC, (C) the PLS model based on the pretreatment of SNV+ first derivate, and (D) the PLS model based on the pretreatment of SG smoothing+ SNV+ first derivate.

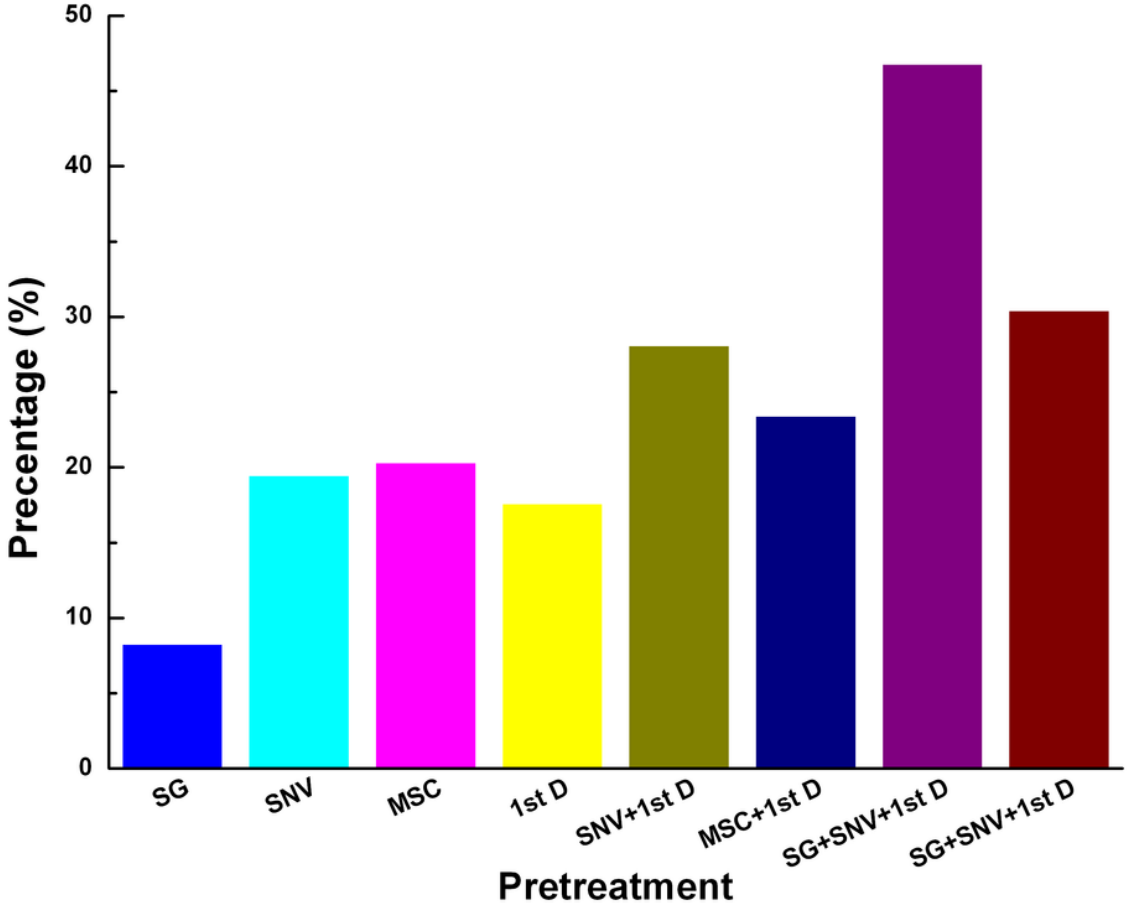

Figure 4

The residual predictive deviation (RPD) for PLS models based on different pretreatment strategies compared with the model using raw spectra.