

RESEARCH

Open Access

Feasibility study on the use of near-infrared spectroscopy for rapid and nondestructive determination of gossypol content in intact cottonseeds



LI Cheng^{1,2}, SU Bangsong³, ZHAO Tianlun³, LI Cong³, CHEN Jinhong^{1,3} and ZHU Shuijin^{1,3*} 

Abstract

Background: Gossypol found in cottonseeds is toxic to human beings and monogastric animals and is a primary parameter for the 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 the breeding program, so it is of great importance to predict the gossypol content in cottonseeds rapidly and nondestructively 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 derivative 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 derivative resulted in a good determination of gossypol content in intact cottonseeds.

Conclusions: Near-infrared spectroscopy coupled with different spectral pretreatments and partial least squares (PLS) regression has exhibited the feasibility in predicting gossypol content in intact cottonseeds, rapidly and nondestructively. It could be used as an alternative method to substitute for traditional one to determine the gossypol content in intact cottonseeds.

Keywords: Intact cottonseed, Chemometrics, Gossypol, Near-infrared spectroscopy

* Correspondence: shjzhu@zju.edu.cn

¹Hainan Institute, Zhejiang University, Sanya 572025, China

³Department of Agronomy, Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou 310058, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Cotton (*Gossypium*. spp) is one of the important industrial and economic crops (Sunilkumar et al. 2006). 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 the global population (Sawan et al. 2006). However, the *Gossypium* species are characterized by the presence of gossypol, which is toxic to human beings and monogastric animals (Lordelo et al. 2005), 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 helps cotton defend biotic stresses (Kong et al. 2010; Blanco et al. 1983). Due to the toxicity of gossypol, breeding for either lower gossypol content in cottonseeds or higher gossypol content in cotton plants has been practiced in many cotton-planting countries. The cottonseed breeding often requires analyzing a large number of cottonseed samples to measure gossypol content. Conventionally, gossypol content is assayed by ultraviolet (UV) spectrophotometry which not only involves reagents with great toxicity but also is inaccurate and unreliable. Despite offering a high level of accuracy and sensitivity, high-performance liquid chromatography (HPLC) is usually costly and time-consuming. In addition, both classical analytical methods cause undesired destruction of the testing samples which frequently needed to be planted in a cotton breeding program. So, a rapid and nondestructive 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 (Sohn et al. 2008; Huang et al. 2013; Weinstock et al. 2006; Rosales et al. 2011; Bellato et al. 2011; Bala and Singh 2013; Hacisalihoglu et al. 2010; Mendoza et al. 2018; Lee et al. 2017; Tierno et al. 2016; Yang and Ren 2008; Lin et al. 2013a, 2013b; Kovalenko et al. 2006; Fassio and Cozzolino 2004). Although the NIR calibration model for determining gossypol content in cotton powder was developed (Li et al. 2017), it could not be used to nondestructively analyze gossypol content in intact cottonseeds, especially in breeding programs where the genetic materials from genetic modification or cross-breeding have limited availability. It is a challenge to determine gossypol content in intact cottonseeds by NIR, because (i) cottonseed being bigger than other crop seeds, so large voids are left between packed samples in sample cells; (ii) some 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) (Xie and Kalivas 1997), partial least squares (PLS) (Haaland and Thomas 1988), support vector machines (SVM) (Nie et al. 2008), least squares support vector machines (LS-SVM) (Shao et al. 2012), and artificial neural networks (ANN) (Makinoa et al. 2010), coupled with spectral pretreatments such as standard normal variate (SNV) (Barnes et al. 1989), Savitzky-Golay (SG) smoothing (Savitzky and Golay 1964), multiplicative scatter correction (MSC) (Hopke 2003), and first derivative (Rinnan et al. 2009).

Due to undesired destruction of the test sample, previous NIR models which can be used in the detection of gossypol in cottonseed meal can be barely applied in breeding trials (Li et al. 2017). In this present study, spectroscopy has 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 powerful technical support for cotton breeders and other people who work on cottonseeds.

Materials and methods

Samples and preparation

A total of 268 samples of cottonseeds were collected from different growing 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. The cottonseed samples were delinted and dried at 30 °C to constant weight. After spectral acquisition by NIR spectroscopy, the intact cottonseed samples were dehulled 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.

Gossypol extraction

A sample of 0.1 g of cottonseed kernel powder was suspended in 5 mL acetone and sonicated in an ultrasonic bath for 45 min. Then, the suspension was filtered

through quantitative filter paper followed by 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.

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 C₁₈ 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% H₃PO₄ (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 °C. 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 the peak area within 36 h. 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).

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 follows. The NIR spectra were collected across the range of 4 000~10 000 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 °C and 60% relative air humidity. All the spectra were transformed into absorbance (lg (1/R)).

Spectral pretreatment

Before calibration, the spectral data were pretreated for 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.

Sampling design

Samples were assigned to calibration and prediction sets using Kennard-Stone (KS) selection (Kennard and Stone

1969). 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.

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 provides the lowest RMSECV was selected as the best.

Model evaluation

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{nirs} - Y_{ref})^2}{\sum_{i=1}^n (Y_{ref} - \bar{Y}_{ref})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{nirs} - Y_{ref})^2}{n}} \tag{2}$$

$$RPD = SD_{Y_{ref}} / RMSEP \tag{3}$$

where *n* is the total number of samples, *Y_{nirs}* is the predicted value by calibration models, *Y_{ref}* is the reference value by HPLC, and SD is the standard deviation.

The coefficient for determination of prediction (*R_p²*), the root mean square error of prediction (RMSEP), the coefficient for determination of calibration (*R_c²*), the root mean square error of cross-validation (RMSECV), and the residual predictive deviation (RPD) were used as criteria to evaluate model performance. An acceptable model should have high *R_c²* and *R_p²* values and low RMSECV and RMSEP values. Meanwhile, the model is considered robust if the RPD is higher than 2.5.

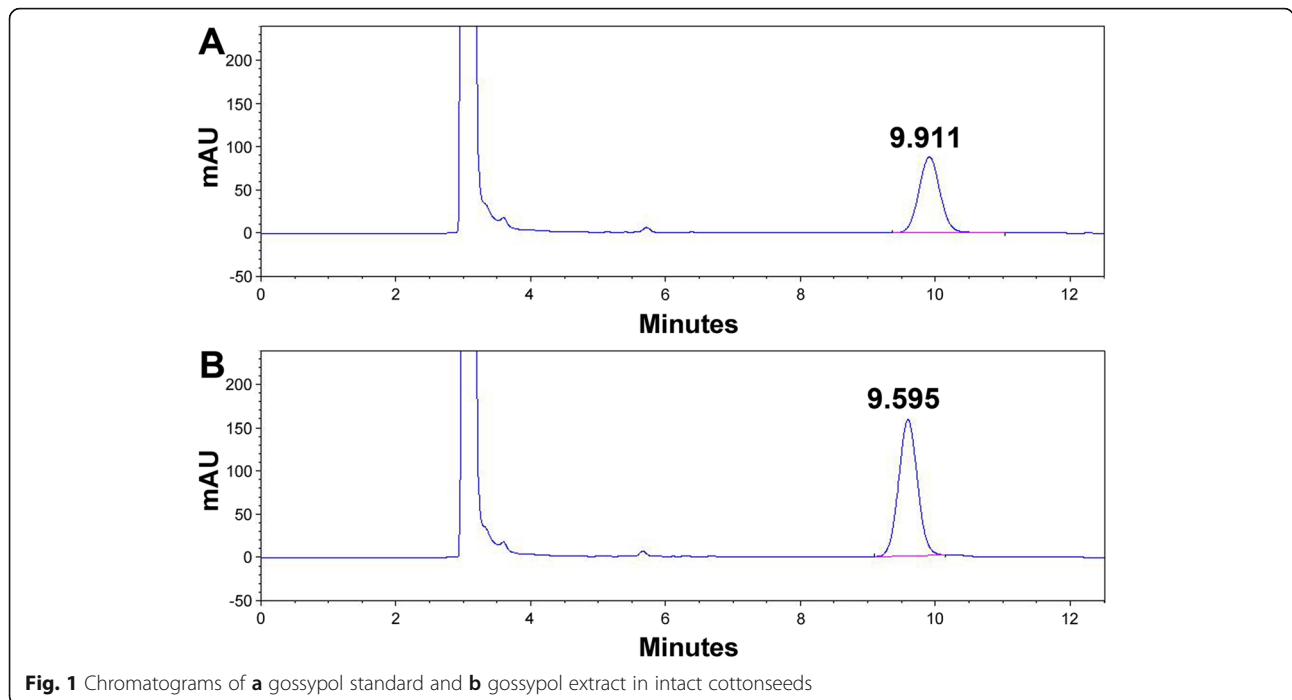
Software

NIR spectroscopic data (268 samples × 1 501 variables) were exported in text format, organized in Microsoft Excel spreadsheets, and then transferred into MATLAB AB R2011a (Math Works, Natick, USA) for chemometric analysis. All the algorithms in spectral pretreatments,

Table 1 HPLC-VU results

Regression equation	r ²	LOD/(μg·mL ⁻¹)	LOQ/(μg·mL ⁻¹)	Average recovery/%
y = 102.42x - 85.055	0.999	37.5	125.0	95.16~101.72

r² correlation coefficient, LOD limits of detection, LOQ limits of quantification, y peak area, x concentration (μg·mL⁻¹)



sampling design and regressions were implemented with MATLAB R2011a.

Results

HPLC analysis

The regression equation, correlation coefficient (r^2), limits of detection (LOD), limits of quantification (LOQ), and an average recovery of gossypol were illustrated in Table 1. The retention time of gossypol standard and gossypol extractions was 9.91 and 9.60 min, respectively (Fig. 1). Table 2 shows the stability for the peak area of gossypol determined by HPLC for 24 h. 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 h.

NIR spectra analysis

Across the spectral range of 4 000~10 000 cm^{-1} , the absorbance values are mainly associated with the combination and overtone bands of the C-H, N-H, O-H, and S-H bonds (Macho and Larrechi 2002), 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 4 200, 4 700, 5 150, 5 580, 6 900, and 8 400 cm^{-1} , respectively. The small peak observed at 4 200 cm^{-1} fell within the regions associated with the combination bands of C-H. At 5 150 and 6 900 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 5 580 and 8 400 cm^{-1} overlapped with the second and first C-H overtone regions, respectively. It was worth mentioning that the peak at 4 700 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 biases 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 the establishment of the calibration

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	2 378.8	2 398.0	2 397.9	2 435.2	2 414.2	2 421.1	2 428.1	2 452.7	2 465.3	1.14
2	2 871.7	2 848.1	2 860.2	2 885.0	2 876.6	2 886.2	2 932.9	2 939.8	3 010.5	1.76
3	2 854.8	2 839.8	2 862.2	2 892.6	2 888.7	2 918.6	2 965.1	2 972.7	3 007.1	2.01

RSD relative standard deviation

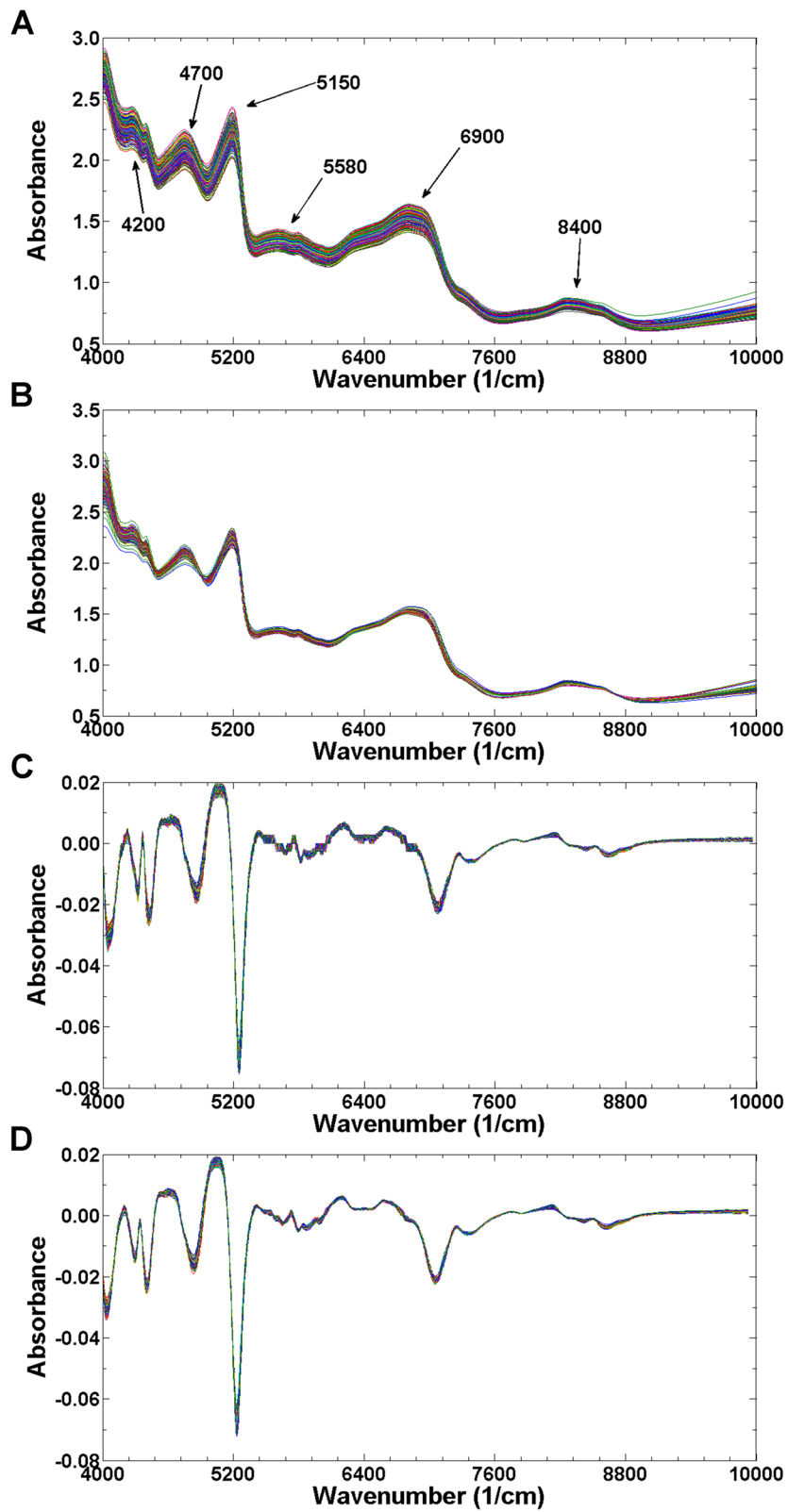


Fig. 2 The NIR spectra of intact cottonseeds. **a** the raw spectra, **b** the spectra pretreated by MSC, **c** the spectra pretreated by SNV+ first derivative, and **d** the spectra of pretreated by SG smoothing+ SNV+ first derivative

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

models. The pretreatment spectra of several types of representative strategies were shown in Fig. 2b, c, and d. 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 1 491 variables were used for calibration among the models using SG smoothing during the spectral pretreatments.

Kennard-Stone sampling design

The Kennard-Stone algorithm is an effective method for extracting a sample subset in the 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 the 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 (Kohonen 1982) or D-optimal designs (de Aguiar et al. 1995). In this study, the total of 268 intact cottonseed samples were divided into calibration and prediction sets based on KS algorithm, with the former set consisting of 218 samples and the latter 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.

PLS regression

The calibration models of gossypol content in intact cottonseeds based on PLS regression were established in the NIR spectral range of 4 000~10 000 cm⁻¹, and the results were summarized in Table 4. The number of LV was 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_p² and R_c² 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.

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 eight pretreatment strategies, including single 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 the analyzing of the results obtained from single pretreatment strategies, the PLS model using eight latent variables based on the 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. 3). Figure 4b showed the correlation of the 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_c² and R_p² values of 0.962 and 0.887, respectively. The RPD value of that model was 3.0, increased by 28.14% compared with the model using raw

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_c² coefficient of determination of calibration, RMSEP root mean square error of prediction, R_p² 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

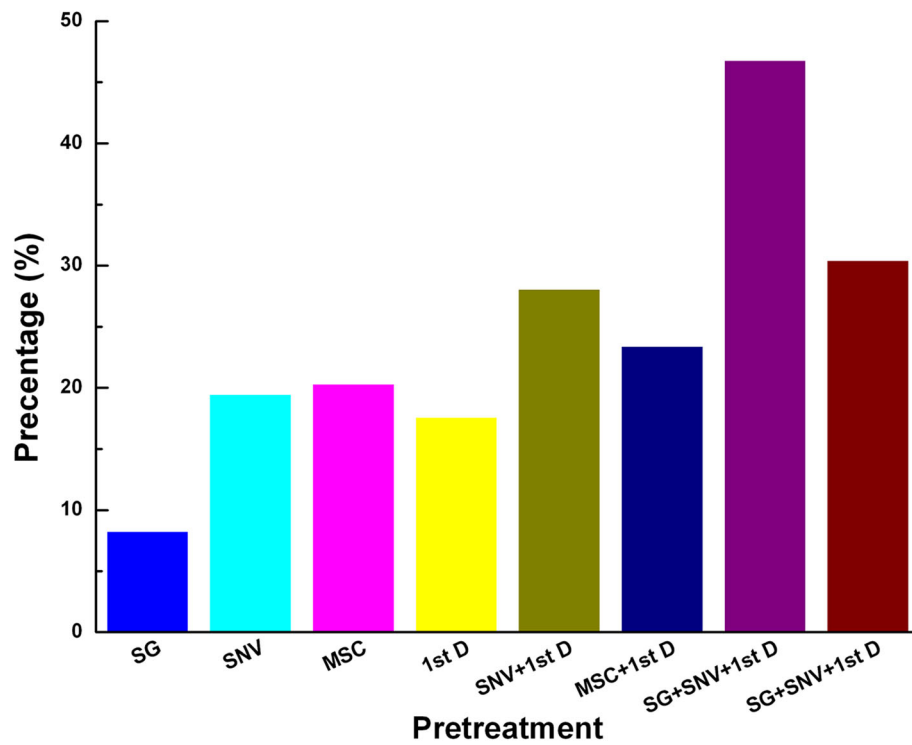


Fig. 3 The residual predictive deviation (RPD) for PLS models based on different pretreatment strategies compared with the model using raw spectra

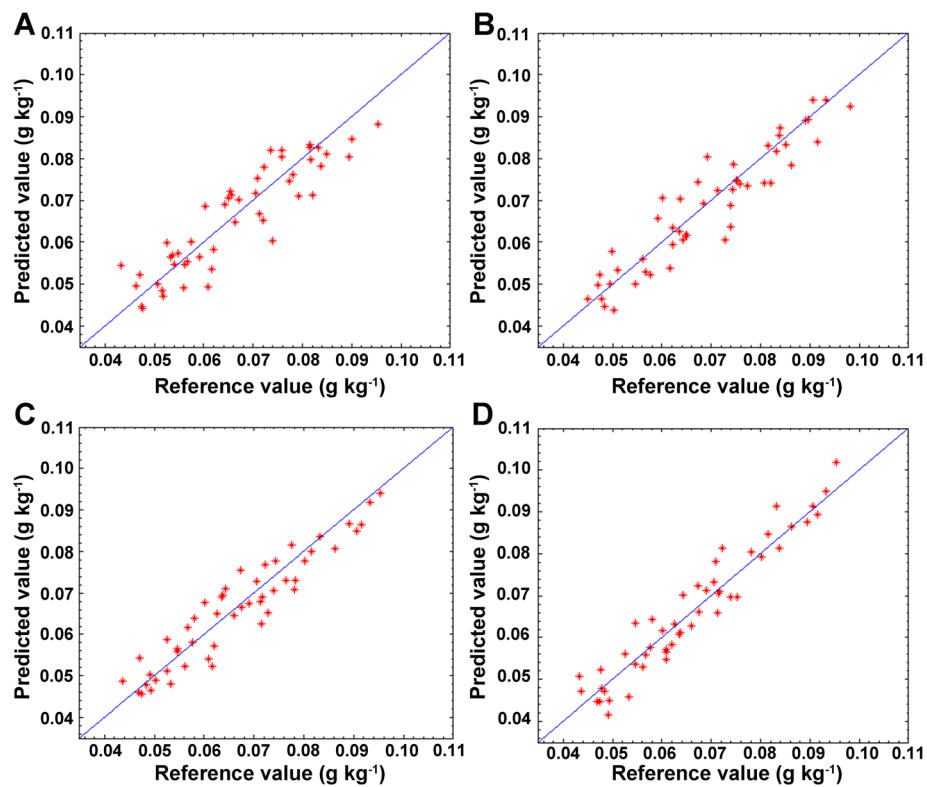


Fig. 4 The correlation between predicted and reference values for models of intact cottonseeds. **a** the PLS model based on raw spectra, **b** the PLS model based on the pretreatment of MSC, **c** the PLS model based on the pretreatment of SNV+ first derivative, and **d** the PLS model based on the pretreatment of SG smoothing+ SNV+ first derivative

spectra. From all the results of calibration models established, the best model was the one that pretreated using the strategy of SG + SNV + first derivative, and it had the highest R_c^2 (0.97) and R_p^2 (0.93), and the RPD (3.4) increased by 46.28% compared with that of the raw spectral model. Furthermore, RMSECV (0.05) and RMSEP (0.04) were the lowest among all the models. The correlation plots between the predicted and reference values were focused on the diagonal line (Fig. 4d). It was indicated that the model using SG + SNV + first derivative and PLS was accurate and robust enough to substitute the conventional gossypol analysis methods (HPLC) 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 the 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 use of NIR technology for rapid and nondestructive 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.

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 derivative, 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.

Acknowledgments

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

Authors' 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 the experiment. Chen JH assisted in editing the article. Zhu SJ and Chen JH conducted and supervised the experiments. The author(s) read and approved the final manuscript.

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.

Availability of data and materials

All relevant data are within this article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All co-authors have consent for submission of manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Hainan Institute, Zhejiang University, Sanya 572025, China. ²Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Science, Shenzhen 518124, Guangdong, China. ³Department of Agronomy, Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou 310058, China.

Received: 4 August 2020 Accepted: 2 April 2021

Published online: 07 May 2021

References

- Bala M, Singh M. Non-destructive estimation of total phenol and crude fiber content in intact seeds of rapeseed–mustard using FTNIR. *Ind Crop Prod.* 2013;42:357–62. <https://doi.org/10.1016/j.indcrop.2012.06.014>.
- Barnes R, Dhanoa M, Lister S. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Appl Spectrosc.* 1989; 43(5):772–7. <https://doi.org/10.1366/0003702894202201>.
- 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.* 2011;59(9):4349–60. <https://doi.org/10.1021/jf200087y>.
- Blanco A, Aoki A, Montamat E, et al. Effect of gossypol upon motility and ultrastructure of *Trypanosoma cruzi*. *J Protozool.* 1983;30(4):649–51. <https://doi.org/10.1111/j.1550-7408.1983.tb05337.x>.
- de Aguiar PF, Bourguignon B, Khots MS, et al. D-optimal designs. *Chemom Intell Lab Syst.* 1995;30(2):199–210. [https://doi.org/10.1016/0169-7439\(94\)00076-X](https://doi.org/10.1016/0169-7439(94)00076-X).
- Fassio A, Cozzolino D. Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy. *Ind Crop Prod.* 2004;20(3): 321–9. <https://doi.org/10.1016/j.indcrop.2003.11.004>.
- Haaland MD, Thomas VE. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal Chem.* 1988;60(11):1193–202. <https://doi.org/10.1021/ac00162a020>.
- Hacisalihoglu G, Larbi B, 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.* 2010;58:702–6. <https://doi.org/10.1021/jf9019294>.
- Hopke H. The evolution of chemometrics. *Anal Chim Acta.* 2003;500(1-2):365–77. [https://doi.org/10.1016/S0003-2670\(03\)00944-9](https://doi.org/10.1016/S0003-2670(03)00944-9).
- 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.* 2013;43:654–60. <https://doi.org/10.1016/j.indcrop.2012.08.015>.

- Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics*. 1969;11(1):137–48. <https://doi.org/10.2307/1266770>.
- Kohonen T. Analysis of a simple self-organizing process. *Biol Cybernetics*. 1982; 44(2):135–40. <https://doi.org/10.1007/BF00317973>.
- Kong GC, Daud KM, Zhu SJ. Effects of pigment glands and gossypol on growth, development and insecticide-resistance of cotton bollworm (*Heliothis armigera* (Hübner)). *Crop Prot*. 2010;29(8):813–9. <https://doi.org/10.1016/j.cropro.2010.03.016>.
- Kovalenko VI, Rippke RG, Hurburgh RC. Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *J Agric Food Chem*. 2006;54(10):3485–91. <https://doi.org/10.1021/jf052570u>.
- 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*. 2017;97(4):1084–92. <https://doi.org/10.1002/jsfa.7832>.
- 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*. 2017;221:990–6. <https://doi.org/10.1016/j.foodchem.2016.11.064>.
- Lin C, Chen X, Jian L, et al. Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chem*. 2013a;162:10–5. <https://doi.org/10.1016/j.foodchem.2014.04.056>.
- Lin TS, Schinazi RF, Zhu JL, et al. Anti-Hiv-1 activity and cellular pharmacology of various analogs of gossypol. *Biochem Pharmacol*. 2013b;46(2):251–5. [https://doi.org/10.1016/0006-2952\(93\)90411-O](https://doi.org/10.1016/0006-2952(93)90411-O).
- Lordelo MM, Davis AJ, Calhoun MC, et al. Relative toxicity of gossypol enantiomers in broilers. *Poult Sci*. 2005;84(9):1376–82. <https://doi.org/10.1093/ps/84.9.1376>.
- Macho S, Larrechi MS. Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends Anal Chem*. 2002;21(12):799–806. [https://doi.org/10.1016/S0165-9936\(02\)01202-5](https://doi.org/10.1016/S0165-9936(02)01202-5).
- 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*. 2010;121:533–9. <https://doi.org/10.1016/j.foodchem.2009.12.043>.
- 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*. 2018;98:283–90. <https://doi.org/10.1002/jsfa.8469>.
- Nie Z, Han J, Liu T, Liu X. Hot topic: application of support vector machine method in prediction of alfalfa protein fractions by near infrared reflectance spectroscopy. *J Dairy Sci*. 2008;91(6):2361–9. <https://doi.org/10.3168/jds.2008-0985>.
- Rinnan Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal Chem*. 2009; 28(10):1201–22. <https://doi.org/10.1016/j.trac.2009.07.007>.
- 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*. 2011;59(20):10781–6. <https://doi.org/10.1021/jf201468x>.
- Savitzky A, Golay M. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36(8):1627–39. <https://doi.org/10.1021/a60214a047>.
- 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 Agric Sci*. 2006;1(6):56–65. <https://doi.org/10.1007/BF02544517>.
- Shao YN, Zhao CJ, Bao YD, He Y. Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food Bioprocess Technol*. 2012;5(1):100–7. <https://doi.org/10.1007/s11947-009-0267-y>.
- Sohn M, Himmelsbach SD, Barton EF, et al. Near-infrared analysis of whole kernel barley: comparison of three spectrometers. *Appl Spectrosc*. 2008;62(4):427–32. <https://doi.org/10.1366/000370208784046768>.
- Sunilkumar G, Campbell CL, Puckhaber L, et al. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc Natl Acad Sci*. 2006;103(48):18054–9. <https://doi.org/10.1073/pnas.0605389103>.
- 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 Agric*. 2016;96(6):1888–99. <https://doi.org/10.1002/jsfa.7294>.
- 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*. 2006;60:9–16. <https://doi.org/10.1366/000370206775382631>.
- Xie YL, Kalivas HJ. Local prediction models by principal component regression. *Anal Chim Acta*. 1997;348(1–3):29–38. [https://doi.org/10.1016/S0003-2670\(97\)00036-6](https://doi.org/10.1016/S0003-2670(97)00036-6).
- Yang N, 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*. 2008;56(3):761–4. <https://doi.org/10.1021/jf072453u>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

