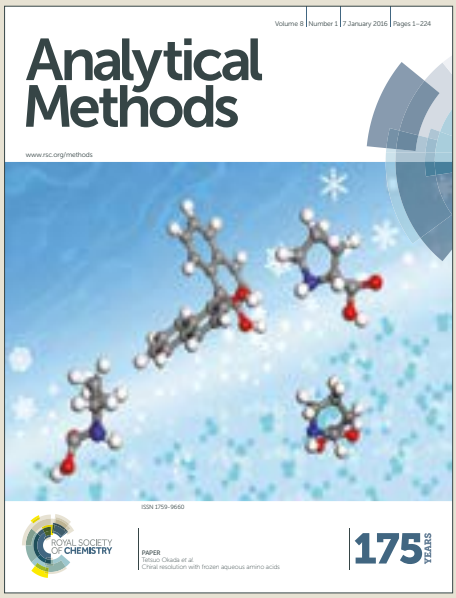
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

rsc.li/methods

View Article Online
DOI: 10.1039/C7AY02115A

NIR hyperspectral imaging with multivariate analysis for measurement of oil and protein content in peanut varieties

Jun-Hu Cheng[a, b], Huali Jin[a*], Zhongyue Xu[c], Fuping Zheng[b**]

[a] School of Food Science and Engineering, South China University of Technology, Guangzhou, 510641, China

[b] Beijing Laboratory for Food Quality and Safety, Beijing Technology and Business University, Beijing 100048, China

[c] School of Business Administration, South China University of Technology, Guangzhou, 510641, China

Abstract: The potential of hyperspectral imaging in the spectral range of 1000-2500 nm with multivariate analysis for prediction of oil and protein concentration in five peanut cultivars was investigated. Quantitative partial least squares regression (PLSR) models were established using the extracted spectral data from hyperspectral images and the reference measured oil and protein concentration. The PLSR models established using the whole spectral data pretreated by multiplicative scatter correction (MSC) method showed good results for predicting oil concentration with the determination coefficient (R^2_P) of 0.945 and root mean square errors by prediction (RMSEP) of 0.196, and for predicting protein concentration with R^2_P of 0.901 and RMSEP of 0.441. In addition, eight and eight optimal wavelengths were selected using the regression coefficients of the PLSR analysis and used for simplifying the obtained models. The simplified PLSR models also presented good performances with R^2_p of 0.933 and 0.912 for predicting oil and protein concentration. The whole results demonstrated that NIR hyperspectral imaging technique coupled with chemometrics analysis is a promising tool for rapid and non-destructive determination of oil and protein concentration in peanut kernels and has the potential to develop a multispectral imaging system for

* Corresponding author. E-mail: hljin2008@126.com (H Jin).
** Corresponding author.

View Article Online
DOI: 10.1039/C7AY02115A

future on-line detection of peanut quality.

Keywords: Hyperspectral imaging, peanut, oil, protein, PLSR, variable selection

1. Introduction

Peanut is one of the important and commercial crops in the world. Peanut has high value nutrients including the oil, plant protein, vitamin, and essential unsaturated fatty acid and some functional and special components (resveratrol and plant sterols).[1] Among them, oil and protein in peanut are important components due to their distinctive biological and nutritional properties. The oil content is a significant standard of monetary assessment in the trade of peanut, and the raw material price also depends on its oil content.[2] In addition, peanut protein is a fundamental source of plant protein for human consumption. Therefore, determination and evaluation of oil and protein content in different peanut kernels is obviously important.

The traditionally used analytical methods for oil content determination are mainly related to the Soxhlet extraction, the American Oil Chemists' Society method, and supercritical fluid extraction (SFE) method. Protein content in peanut is commonly measured by The Kjeldahl and combustion methods. These above-mentioned methods afford an effective and accurate measurement of oil and protein content for quality evaluation and inspection in peanut kernels. However, the process of measurement is tedious, inconsistent, and time-consuming. Moreover, they usually require the use of large amounts of chemical solvents and analytical reagents, which might be hazardous and harmful to analysts and the lab environment. Therefore, non-destructive and rapid measurement of oil and protein content is very necessary.

Near infrared (NIR) spectroscopy as a rapid, non-destructive technique has been widely developed

View Article Online
DOI: 10.1039/C7AY02115A

for measurement and assessment of food compositions.[3] In peanut research field, NIR reflectance spectroscopy with chemometrics analysis has been successfully used for determination of moisture content in peanut kernels,[4] estimation of the peanut essential minerals,[5] and measurement of peanut oil and fatty acid compositions.[6] Although NIR spectroscopy has been recognized as a useful and reliable tool for detection of peanut compositions, it cannot provide the spatial information of the object for indicating the dynamic changes of the measured chemical information. Compared to NIR spectroscopy, hyperspectral imaging (HSI) integrates the traditional NIR spectroscopy and computer vision into one system, making it possible for providing both spectral and spatial information of a certain object simultaneously. It means that one hyperspectral image is a set of images exhibiting the tested object at different wavelengths where each pixel in the image represents a spectrum for this specific point of the object, which provides the corresponding relation between the NIR wavelengths and sample images. In addition, Although HSI device is a little expensive than NIR instrument, the hyperspectral imaging technology can be developed to a multispectral imaging technology based on the variable selection algorithms. A multispectral imaging provides more information and shows more rapid online detection capacity compared with NIR device with similar costs. Thus, the hyperspectral imaging or multispectral imaging technology shows more advantages than NIR technology. The hypercube I (x, y, λ) of the hyperspectral image includes spatial (x and y) and spectral (λ) information, which shows a three-dimensional (3-D) dataset that contains many images of the same object, and each of which is measured at a different wavelength.[7] Thus, HSI technology can provide more information and also can be used for establishing more reliable model for rapid and non-invasive inspection of food and agricultural product compositions.[8-10] As to the food kernel samples, HSI technology with multivariate analysis has been effectively applied for classification of different maize

View Article Online
DOI: 10.1039/C7AY02115A

kernel hardness,[11,12] detection of *Fusarium* in wheat kernel,[13] detection of *Fungal* infection and *Ochratoxin A* contamination in stored wheat,[14] and classification of black beans.[15] Up to now, using hyperspectral imaging technique for non-destructive determination of oil and protein content in peanut kernels has been rarely reported to date.

Therefore, this study was aimed to investigate the possibility of using HSI for predicting oil and protein content in different peanut varieties. The specific objectives of this study were to (1) obtain the hyperspectral images of peanut kernels in the NIR spectral range (1000-2500 nm); (2) extract the spectral data from recognized regions of interests (ROIs) within the acquired hyperspectral images; (3) establish the calibration model between the extracted spectral information and the reference measured oil and protein content; (4) select the optimum wavelengths showing the most relevant information related to oil and protein content prediction; and (5) construct the simplified calibration model based on the selected wavelengths and develop a multispectral imaging system.

2. Material and methods

2.1 Sample preparation and chemical measurement

Five peanut varieties named Huayu (HY), Luohanguo (LHG), Zhonghua (ZH), Dabaisha (DBS), Xiaobaisha (XBS), were purchased from a local seed market in Guangzhou, China. After the shell, a number of 45 kernels with similar size and condition for each variety were selected and a total of 225 peanut samples were obtained. Each peanut kernel was scanned by the hyperspectral imaging system and then used for traditional oil and protein measurement. According to the Kennard-Stone (K-S) method, two thirds of each variety were selected as the calibration set, and the remaining one third samples as the prediction set. Therefore, 150 peanut kernels were used for the calibration model and

View Article Online
DOI: 10.1039/C7AY02115A

75 subsamples were used for the prediction model. The reference oil content was measured using the protocol of the Official Methods and Recommended Practices of the American Oil Chemists' Society (AOCS). The reference protein content was determined by the classical Kjeldahl method. All experiments were conducted in triplicate. All data were expressed as mean ± standard deviation.

2.2 Hyperspectral imaging system

A reflectance hyperspectral imaging system in the spectral range of 1000-2500 nm was used to acquire the hyperspectral images of peanut kernels. The system mainly includes a line-scan imaging spectrograph (Specim V25E, Spectral Imaging Ltd., Oulu, Finland) covering the spectral range of 1000-2500 nm, a high-performance 320 × 256 CCD camera (XC403, Xenics Infrared Solutions, Leuven, Belgium). Fig. 1 shows the schematic diagram of the main components of the hyperspectral imaging system. When the machine works, each peanut kernel was taken on the mobile platform and scanned with the attuned speed and exposure time to acquire the hyperspectral images, each of which has three dimensions (x, y, λ), where x and y are the spatial dimensions and λ is the number of wavebands. Therefore, the original images were created and stored. In order to remove the effects of illumination and detector sensitivity, the raw acquired hyperspectral images (I_0) should be corrected using two reference standards: a white one (W) to set-up the maximum reflectance (~99%) condition, which was obtained for a white calibration tile under the same condition of the raw image; and a black one (B) to define the no reflectance (~0%) condition, which was acquired by completely covering the lens with its black cap. The calibrated image (I) was then calculated by the following equation.

$$I = \frac{I_0 - B}{W - B} \quad (1)$$

All the calibrated peanut kernel images were used to extract the spectral data and the Environment for Visualizing Images software (ENVI v4.8, ITT Visual Information Solutions, Boulder, CO, USA)

View Article Online
DOI: 10.1039/C7AY02115A

was applied to recognize the region of interests (ROIs) with a circle shape, and then the spectral data within ROIs for peanut samples were extracted and averaged at each wavelength to acquire one mean spectrum representing the ROI. All of the extracted spectral data from peanut kernels were then arranged in a matrix where the rows of this matrix represent the number of samples (225 subsamples) and the columns represent the number of variables. According to the inherent features of the used hyperspectral imaging system, a total of 228 wavelengths/variables were obtained. Therefore, the constructed matrix (225 × 228) was used for further multivariate analysis.

2.3 Multivariate data analysis

2.3.1. Spectral preprocessing and PLSR model

In this study, in order to remove the effects of physical phenomena caused by light scattering of particle size and shape, and optical interference during the experiments, multiplicative scatter correction (MSC) as the most widely developed preprocessing technique was applied to compensate for additive and/or multiplicative effects in spectral data.[16] Partial least squares regression (PLSR) is a reliable and effective multivariate chemometrics method, it has prominent advantages that solving multicollinearity problems and allowing variables more than the samples, has been widely used for developing a scientific model.[17] In this study, the PLSR calibration models were established between pre-processed spectral data and the reference measured oil and protein content. PLSR can find a set of independent variables (wavelengths), the X-matrix (225×228), and the dependent variable (oil and protein content values), the Y-matrix (225×1), where the X-matrix represents the average spectral data at the 228 wavelengths for the 225 subsamples, whereas, the Y-matrix shows the values of oil or protein content in the subsamples. The spectral pre-process and multivariate data analysis were carried out using The Unscrambler 9.7(CAMO Software AS, Norway).

View Article Online
DOI: 10.1039/C7AY02115A

2.3.2 Optimal wavelengths selection

The hyperspectral images suffered from the problem of high dimensionality and multi-collinearity among contiguous wavelength bands, which can slow down the calculation speed of the calibration process and influence the model accuracy. Therefore, looking for several optimum wavelengths carrying the most useful information corresponding to the oil and protein content of peanut instead of the whole wavelengths is necessary. In this study, regression coefficients based on the PLSR analysis was applied to select the informative and feature wavelengths from the full spectral range. The weighted regression coefficients method, also called β-coefficients approach, use the regression coefficients resulting from the calibration PLSR model to choose the effective variables.[18] The wavelengths showing the highest absolute values of regression coefficients are selected as the optimum wavelengths and the ones with the lowest absolute values of β-coefficients are completely uninvolved due to little contribution in prediction.

2.3.3 Model evaluation

The PLSR models built with the whole spectra and the optimal spectra should be validated. Full cross validation also called leave-one-out cross-validation was used for validate the calibrated models in this study. The process of this validation method was conducted by removing one sample or a subset of samples from the calibration dataset and a new PLSR model was then constructed based on the remaining calibration samples. In the end, the acquired model was applied to predict the sample left out. The procedure was repeated for every subsample in the dataset, providing a more realistic measurement of the prediction errors of the model. In addition, the optimal number of latent variables (LV) based on the PLSR algorithm for building the calibration model was determined by using the minimum value of predicted residual error sum of squares (PRESS).[19] The robustness and

View Article Online
DOI: 10.1039/C7AY02115A

effectiveness of the PLSR models are usually evaluated by the determination coefficients of calibration (R^2_C), cross-validation (R^2_{CV}) and prediction (R^2_P), and root mean square error of calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP).[20] Generally, a good model should have high R^2_C, R^2_{CV}, and R^2_P, and low RMSEC, RMSECV and RMSEP, as well as a slight difference between them. In details, R^2 shows the proportion of the variance in reference data that can be explained by the variance in the predicted data. The values of RMSEC, RMSECV and RMSEP are measurements of the root mean square errors in the analysis and assessment of the fitting degree of regression during calibration, cross-validation and prediction with lower values implying better predictive capacity. [20] Therefore, it is ideally expected to obtain RMSEs (RMSEC, RMSECV and RMSEP) as close as 0 and R^2 as close as 1. In reality, R^2 in the range of 0.82-0.90 usually shows good performance of a model, while R^2 lower than 0.82 indicates relatively poor performance, and R^2 higher than 0.90 shows excellent performance. Their corresponding calculation equations are as follows.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(y_{cal} - y_{act})^2}{n}} \quad (2)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{act})^2}{n}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{cal} - y_{act})^2}{\sum_{i=1}^{n}(y_{cal} - y_{mean})^2} \quad (4)$$

where n is the number of samples; y_{act} is the actual value; y_{cal} is the calibrated value; y_{pred} is the predicted value; y_{mean} is the mean of the reference measured value. Fig. 2 describes the whole procedure of the experiment by using hyperspectral imaging technique.

3. Results and discussion

View Article Online
DOI: 10.1039/C7AY02115A

3.1 Reference measurement and average spectral analysis

Table 1 shows the relevant statistics of oil and protein content determined by the traditionally chemical methods for peanut samples. It can be found from Table 1, in the calibration set, the variation ranges from 47.72% to 50.61% for oil content and from 23.46% to 28.43% for protein content were obtained. In the prediction set, the oil content changed from 47.75% to 50.34%, and the protein content varied from 23.57% to 28.35%. The acquired differences covered all of the five mentioned peanut kernels chemical information and it is beneficial to construct the reliable calibration and prediction models. Fig. 3a shows the full wavelength scanned from 1000 to 2500 nm for oil content changes in all the peanut kernels. Generally speaking, the most prevailing absorption bands in the NIR region are ascribable to the strong overtone and combination absorptions of hydrogen mainly related to the bonds of O-H in water, C-H in oil, and N-H in protein. In Fig. 3a, nevertheless, the main absorption peaks were found at about 1216, 1738 and 2330 nm, which were apparently attributed to the contribute of 1st and 2nd overtone C-H stretches of methyl or methylene group in oil changes.[21] Fig. 3b shows the average reflectance spectral information extracted from peanut kernels for protein changes. It is clear that the spectral reflectance curves presented a similar trend in the whole wavelength region about the five different peanut kernels, but the amplitude of variation of spectral reflectance were different. Based on the analysis of the value of reflectance, two significant absorption peaks positioned at around 1217 nm and 1500 nm, which were mostly related to the second overtone C-H stretching and the first overtone N-H stretching.[22] Protein absorption peak was considered to be arranged at approximately 2162 nm.[5]

3.2 Oil and protein prediction using the whole wavelengths

View Article Online
DOI: 10.1039/C7AY02115A

Based on all the spectra of peanut samples and their corresponding traditionally measured oil and protein content values, the PLSR models using the whole spectral wavelengths were established. The performances of PLSR models are shown in Table 2 and 3. As to the oil content prediction, the PLSR model showed excellent results with R^2_C, R^2_{CV}, and R^2_P value of 0.924, 0.923, and 0.939 with their corresponding RMSEC, RMSECV and RMSEP value of 0.237, 0.234, and 0.212, which indicated that using NIR hyperspectral imaging with PLSR analysis is suitable for prediction of oil content in different peanut kernels. Also, it can be observed that the used spectral pretreatment of MSC method was helpful to enhance the PLSR calibration capability with an increase of 0.006 and a decrease of 0.016. Similarly, the performance of the MSC-PLSR model for prediction of protein content using the spectra pretreated by MSC method was also improved with R^2_P from 0.885 to 0.901 and RMSEP from 0.465 to 0.441. Based on the statistical data from Table 2 and 3, the established PLSR models with acceptable performances confirmed the suitability of hyperspectral imaging technique with multivariate data analysis for prediction of the oil and protein content in peanut kernels in a rapid, non-destructive and manner.

3.3 Oil and protein prediction using the optimal wavelengths

In order to elude the high dimensionality of the acquired hyperspectral images, diminish unnecessary information among neighboring wavebands and decrease the time of computation, it is necessary to select optimal wavelengths with effective and valuable information for predicting oil and protein content in peanut kernels. In this study, the optimum wavelengths were selected based on the regression coefficients of the established PLSR models. Fig. 4 shows the key wavelengths that were shown with large regression coefficients values (regardless of the sign) selected. As a result, 1127, 1216, 1477, 1738, 1953, 2073, 2143, and 2319 nm as the eight important wavelengths were selected

View Article Online
DOI: 10.1039/C7AY02115A

for oil content prediction and 1153, 1567, 1972, 2143, 2288, 2339, 2389 and 2446 nm as the eight optimal wavelengths were nominated for protein content prediction. As illustrated in Table 2 and 3, the simplified models also exhibited better results for oil and protein prediction with the R^2_P of 0.933 and 0.912, and RMSEP of 0.198 and 0.438. Compared with the original PLSR models, the optimized PLSR models displayed comparable or better performances for prediction of oil and protein content in peanut kernels. It can be concluded that using the selected eight optimal wavelengths to replace the full wavelengths for predicting the oil and protein content in peanuts is promising and acceptable. More importantly, the selected several important wavelengths can be used and developed for a multispectral imaging system, which is beneficial for on-line and real-time detection of peanut quality.

4. Conclusions

The feasibility of NIR hyperspectral imaging with multivariate data analysis for predicting oil and protein content in peanut kernels was explored. The quantitative PLSR models using full spectral data pretreated by means of MSC method were established and exhibited good performances for oil and protein content prediction with high R^2_P of 0.945 and 0.901, and low RMSEP of 0.196 and 0.441. in addition, eight important wavelengths (1127, 1216, 1477, 1738, 1953, 2073, 2143, and 2319 nm) and eight significant wavelengths (1153, 1567, 1972, 2143, 2288, 2339, 2389 and 2446 nm) were selected using the regression coefficients from the PLSR analysis and used for optimizing the models. The simplified PLSR models also showed satisfactory capability and reliability with R^2_P of 0.933 and 0.912 for prediction of oil and protein content in peanuts. On the whole, the above results indicated that NIR hyperspectral imaging technique in tandem with chemometrics has the great potential to determine the oil and protein content in peanut kernels in a rapid and non-destructive way. It is also beneficial to develop a multispectral imaging system for further on-line application with more

View Article Online
DOI: 10.1039/C7AY02115A

varieties of peanuts adopted in future study.

Acknowledgments

The authors gratefully acknowledge the financial support from the Natural Science Foundation of Guangdong Province (2017A030310558), the China Postdoctoral Science Foundation (2017M612672), the Fundamental Research Funds for the Central Universities (2017MS067) and the Guangdong Provincial R & D Centre for the Modern Agricultural Industry on Non-destructive Detection and Intensive Processing of Agricultural Products and the Common Technical Innovation Team of Guangdong Province on Preservation and Logistics of Agricultural Products (2016LM2154).

References

1 Branch, W. D., Brenneman, T. B., & Noe, J. P., 2016. Evidence for a Second RKN Resistance Gene in Peanut. Peanut Science, 43(1), 49-51.

2 Akhtar, S., Khalid, N., Ahmed, I., Shahzad, A., & Suleria, H. A. R., 2014. Physicochemical characteristics, functional properties, and nutritional benefits of peanut oil: a review. Critical Reviews in Food Science and Nutrition, 54(12), 1562-1575.

3 Cheng, J.-H., & Sun, D.-W., 2014. Hyperspectral imaging as an effective tool for quality analysis and control of fish and other seafoods: current research and potential applications. Trends in Food Science & Technology, 37(2), 78-91.

4 Govindarajan, K., Kandala, C., & Subbiah, J., 2009. NIR reflectance spectroscopy for nondestructive moisture content determination in peanut kernels. Transactions of the ASABE, 52(5), 1661-1666.

5 Phan-Thien, K.-Y., Golic, M., Wright, G. C., & Lee, N. A., 2011. Feasibility of estimating peanut essential minerals by near infrared reflectance spectroscopy. Sensing and Instrumentation for Food Quality and Safety, 5(1), 43-49.

6 Sundaram, J., Kandala, C. V., Holser, R. A., Butts, C. L., & Windham, W. R., 2010. Determination of in-shell peanut oil and fatty acid composition using near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society, 87(10), 1103-1114.

7 Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F., 2013. Hyperspectral

View Article Online
DOI: 10.1039/C7AY02115A

imaging applications in agriculture and agro-food product quality and safety control: a review. Applied Spectroscopy Reviews, 48(2), 142-159.

8 Cheng, J.-H., Nicolai, B., & Sun, D.-W., 2017. Hyperspectral imaging with multivariate analysis for technological parameters prediction and classification of muscle foods: A review. Meat Science, 123, 182-191.

9 Ravikanth, L., Jayas, D. S., White, N. D., Fields, P. G., & Sun, D.-W., 2016. Extraction of Spectral Information from Hyperspectral Data and Application of Hyperspectral Imaging for Food and Agricultural Products. Food and Bioprocess Technology, 1-33.

10 Sendin, K., Williams, P. J., & Manley, M., 2016. Near infrared hyperspectral imaging in quality and safety evaluation of cereals. Critical Reviews in Food Science and Nutrition (just-accepted).

11 Williams, P., Geladi, P., Fox, G., & Manley, M., 2009. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. Analytica Chimica Acta, 653(2), 121-130.

12 Williams, P. J., & Kucheryavskiy, S., 2016. Classification of maize kernels using NIR hyperspectral imaging. Food Chemistry, 209, 131-138.

13 Saccon, F. A., Elrewainy, A., Parcey, D., Paliwal, J., & Sherif, S. S., 2016. Detection of Fusarium on wheat using near infrared hyperspectral imaging. In Photonics North (PN), 5, 24-26.

14 Senthilkumar, T., Jayas, D., White, N., Fields, P., & Gräfenhan, T., 2016. Detection of fungal infection and Ochratoxin A contamination in stored wheat using near-infrared hyperspectral imaging. Journal of Stored Products Research, 65, 30-39.

15 Sun, J., Jiang, S., Mao, H., Wu, X., & Li, Q., 2016. Classification of black beans using visible and near infrared hyperspectral imaging. International Journal of Food Properties, 19(8), 1687-1695.

16 Byrne, H. J., Knief, P., Keating, M. E., & Bonnier, F., 2016. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. Chemical Society Reviews, 45(7), 1865-1878.

17 Mahesh, S., Jayas, D., Paliwal, J., & White, N., 2015. Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat. Food and Bioprocess Technology, 8(1), 31-40.

18 Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example: John Wiley & Sons.

19 Wong, T.-T., 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition, 48(9), 2839-2846.

View Article Online
DOI: 10.1039/C7AY02115A

20 Cheng, J.-H., Sun, D.-W., Zeng, X.-A., & Pu, H.-B., 2014. Non-destructive and rapid determination of TVB-N content for freshness evaluation of grass carp (Ctenopharyngodon idella) by hyperspectral imaging. Innovative Food Science & Emerging Technologies, 21, 179-187.

21 Sundaram, J., Kandala, C. V., & Butts, C. L., 2009. Application of near infrared spectroscopy to peanut grading and quality analysis: overview. Sensing and Instrumentation for Food Quality and Safety, 3(3), 156-164.

22 Büning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chemistry, 82(1), 107-115.

View Article Online
DOI: 10.1039/C7AY02115A

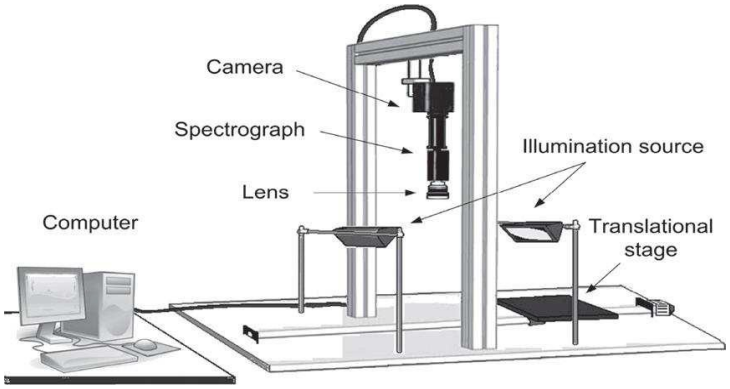


Figure 1. Schematic diagram of the main components of the hyperspectral imaging system

View Article Online
DOI: 10.1039/C7AY02115A

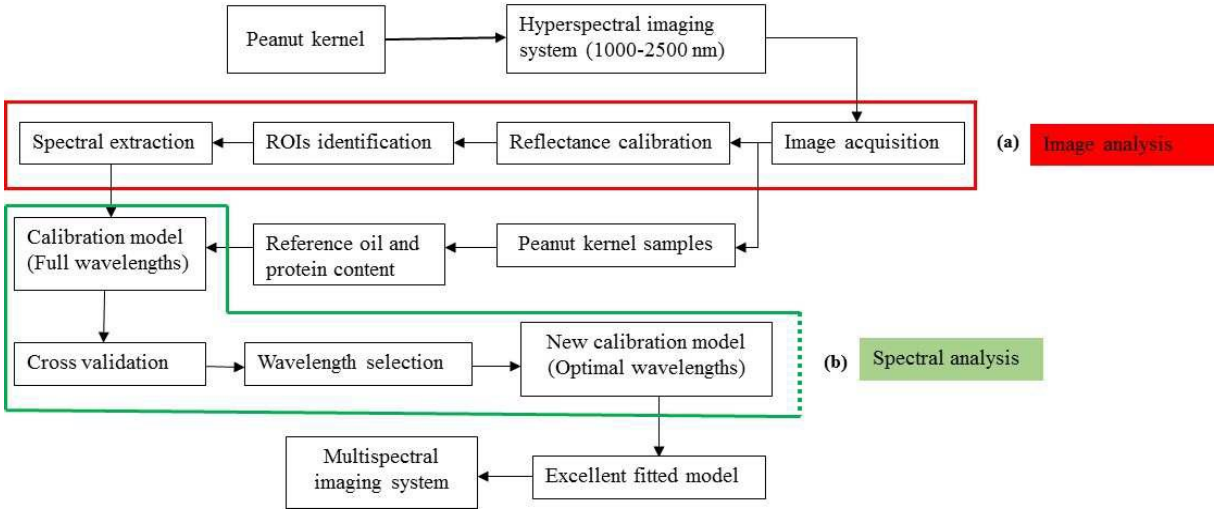
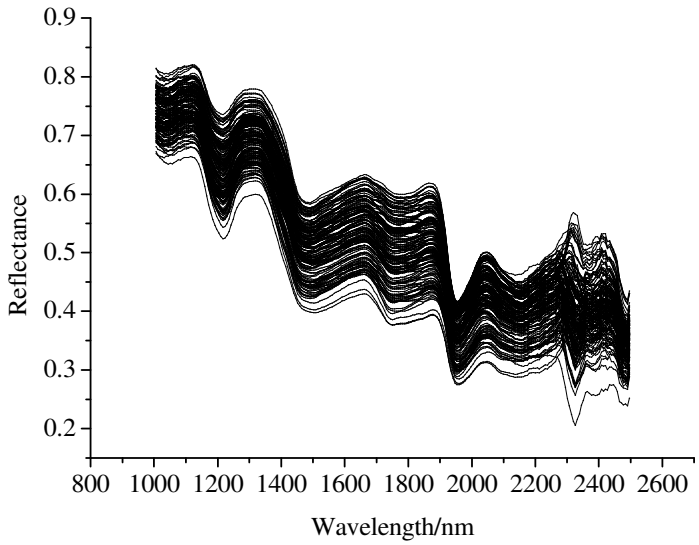
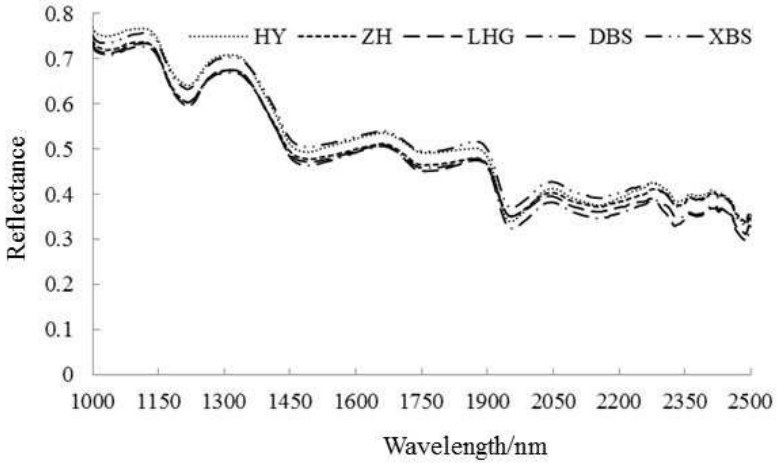


Figure 2. The whole procedure of the experiment by using hyperspectral imaging technique

View Article Online
DOI: 10.1039/C7AY02115A



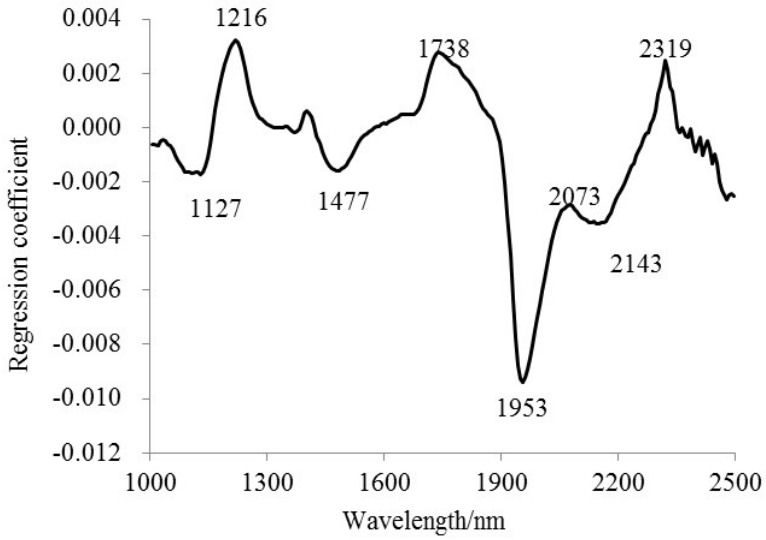
(a) Oil spectra of peanut kernels



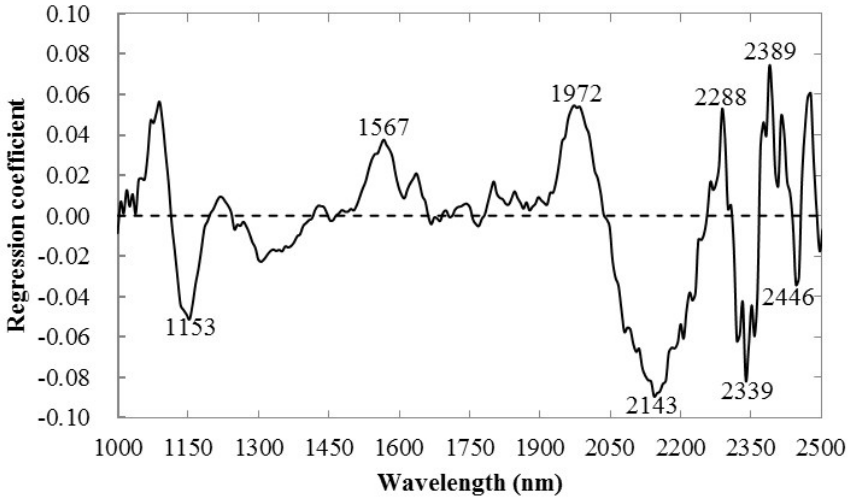
(b) Protein spectra of peanut kernels

Figure 3. Spectral data of the tested peanut samples in NIR spectral range

View Article Online
DOI: 10.1039/C7AY02115A



(a) The selected wavelengths for oil content prediction



(b) The selected wavelengths for protein content prediction

Figure 4. Selection of optimal wavelengths by regression coefficients from PLSR analysis

View Article Online
DOI: 10.1039/C7AY02115A

Table 1 Reference measurement of oil and protein content (%) in peanut kernels using the traditional methods

Index	Variety	Calibration set			Prediction set		
		N	Range	Mean ± SD	N	Range	Mean ± SD
Oil content	HY	30	48.31-49.76	48.86±0.37	15	48.36-49.42	48.77±0.30
	ZH	30	47.83-48.29	48.07±0.15	15	47.83-48.25	48.05±0.13
	LHG	30	47.72-48.36	48.05±0.23	15	47.75-48.26	48.03±0.23
	DBS	30	48.46-49.69	48.98±0.30	15	48.53-49.18	48.96±0.22
	XBS	30	49.38-50.61	49.99±0.34	15	49.62-50.34	49.99±0.24
	Total	150	47.72-50.61	48.79±0.77	75	47.75-50.34	48.76±0.76
Protein content	HY	30	24.71-25.71	25.18±0.30	15	24.81-25.60	25.21±0.27
	ZH	30	25.50-26.67	26.12±0.34	15	25.63-26.61	26.17±0.32
	LHG	30	26.57-27.69	27.19±0.34	15	26.69-27.61	27.24±0.31
	DBS	30	27.25-28.43	27.84±0.36	15	27.36-28.35	27.88±0.35
	XBS	30	23.46-24.75	24.13±0.38	15	23.57-24.65	24.18±0.35
	Total	150	23.46-28.43	26.11±0.35	75	23.57-28.35	26.12±0.32

N: number of samples; HY: Huayu; ZH: Zhonghua; LHG: Luohanguo; DBS: Dabaisha; XBS: Xiaobaisha

View Article Online
DOI: 10.1039/C7AY02115A

Table 2 Results of PLSR models using the full and optimal variables for prediction of oil content in peanut kernels

Model	TV	LV	Calibration		Cross validation		Prediction	
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP
Raw-PLSR	228	5	0.924	0.237	0.923	0.234	0.939	0.212
MSC-PLSR	228	4	0.928	0.228	0.921	0.236	0.945	0.196
RC-PLSR	6	4	0.912	0.227	0.904	0.239	0.933	0.198

TV: total variables; LV: latent variables; MSC: multiplicative scatter correction; RC: regression coefficient; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction

Table 3 Results of PLSR models using the full and optimal variables for prediction of protein content in peanut kernels

Model	TV	LV	Calibration		Cross validation		Prediction	
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP
Raw-PLSR	228	8	0.898	0.441	0.853	0.535	0.885	0.465
MSC-PLSR	228	6	0.906	0.438	0.898	0.487	0.901	0.441
RC-PLSR	8	7	0.905	0.439	0.893	0.489	0.912	0.438

TV: total variables; LV: latent variables; MSC: multiplicative scatter correction; RC: regression coefficient; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction