Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics

Industrial Crops and Products

Yu, Hongwei; Liu, Hongzhi; Erasmus, Sara Wilhelmina; Zhao, Simeng; Wang, Qiang et al

<https://doi.org/10.1016/j.indcrop.2020.112956>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact openscience.library@wur.nl

Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics

Hongwei Yu^{a,b}, Hongzhi Liu^a, Sara W. Erasmus^b, Simeng Zhao^a, Qiang Wang^{a,*}, Saskia M. van Ruth^{b,c}

^a Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, Key Laboratory of Agro-Products Processing, Ministry of Agriculture and Rural Affairs, P.O. Box 5109, Beijing, 100193, China

^b Food Quality and Design, Wageningen University and Research, P.O. Box 17, 6700 AA, Wageningen, the Netherlands

^c Institute for Global Food Security, School of Biological Sciences, Queen's University, 19 Chlorine Gardens, Belfast, BT9 5DL, Northern Ireland, United Kingdom

ARTICLE INFO

Keywords:

Amino acids
Partial least square regression
Peanut
Portable NIRS device
Single kernels
Sucrose

ABSTRACT

The quality traits of peanuts (*Arachis hypogaea* L.) are fundamental to the whole peanut industry. However, many common analyses require the sample to be brought to the laboratory. Therefore, this research explores the feasibility of portable near-infrared spectroscopy combined with a single detection accessory to analyse the composition of peanuts in a single seed level quantitatively. The single detection accessory was specifically designed for spectral data collection considering the internal and external characteristics of single peanuts. Confocal laser scanning microscopy revealed that the oil body and protein body were randomly distributed at cell of single peanuts. The external characteristics of single peanuts were also determined and considered length (11.32–24.25 mm) and width (7.49–12.25 mm). The chemical compositional data (i.e. fat, sucrose, protein, and 16 amino acids) were determined by conventional wet-chemical methods and showed large variation. Principal component analysis on the compositional data showed that peanuts with higher fat contents usually have higher hydrophobic amino acids contents, lower sucrose contents, and lower protein contents. The composition prediction models of single peanuts were estimated using partial least squares regression models that were integrated with different spectral pre-treatments and validated by external sets. The results showed that the prediction models have good performance with a correlation coefficient above 0.88 (calibration) and 0.83 (prediction) and a residual prediction deviation above 1.5 except for a few indicators. Overall, the portable near-infrared spectroscopy offered reliable methods to assess the major components and amino acids quantitatively in a single peanut, which will improve the raw material quality in the peanut industry through the simultaneous and short-term determination of multiple indicators.

1. Introduction

The peanut (*Arachis hypogaea* L.) is a widespread leguminous crop and a dominant commercial agricultural product. Global peanut production was over 48 million tons (Mt) in 2017. China (17.2 Mt), the world's largest peanut-producing country, accounted for nearly 36 % of the total production (Faostat, 2017). Peanuts are nutritious and mainly comprise fat, protein, sucrose, and amino acids. (Wang, 2016, 2018). The fat component accounts for half of peanuts' contents and are the prevailing nutrient constituent in peanuts for producing edible oil (Yu

et al., 2016). Proteins are the second group of constituents in peanuts and consist of eight essential amino acids which satisfy the demands of the Food and Agriculture Organization, except for methionine (Dubinina et al., 2014). The non-essential amino acids not only carry health effects but also have great contributions to the function of proteins such as their soluble and gel characteristics which depend on the molecular structure of amino acids (Wang, 2016; Yu et al., 2017). The sweet and attractive flavour of peanuts and peanut products are derived from its sucrose as well as Maillard reactions with amino acids (Wang et al., 2017). There is no doubt that peanuts consisting of high levels of fat, protein, and

* Corresponding author.

E-mail address: wangqiang06@caas.cn (Q. Wang).

<https://doi.org/10.1016/j.indcrop.2020.112956>

Received 4 July 2020; Received in revised form 14 September 2020; Accepted 16 September 2020

Available online 25 September 2020

0926-6690/© 2020 Published by Elsevier B.V.

sucrose are preferred over others by the peanut industry. Breeding experts, seed retailers, peasants, food and feed companies, and other stakeholders all benefit from high-quality peanut varieties. Hence, it is an ongoing aim of the peanut industry to cultivate new peanut varieties with specific traits.

It is extremely useful for breeders and other stakeholders in the peanut supply chain to master the genetic attributes of peanut qualities and to select desired peanuts by analysing the quality traits in single kernels. The conventional methods for screening peanuts carried out in laboratories are time-consuming, complex operations, and usually result in sample destruction. For breeders and others, sometimes a limited sample size is available and one may not have a lot of materials for evaluation by conventional methods. Therefore, it is essential to explore techniques that are speedy, cost-effective, and non-damaged in nature for the detection of single peanut kernels quality. Consequently, near infrared spectroscopy (NIRS) has great advantages of multiple-trait, non-destructive analysis in a relatively short time (Agelet et al., 2012; Chopra et al., 2019).

Normally, NIRS are bulk sample analysers that require, on average, samples sizes of about 250 g peanut kernels (Agelet and Hurburgh, 2014). The measurements of fat (Sundaram et al., 2010), oleic acids (Shin et al., 2010), protein, and amino acids (Wang et al., 2013) have been conducted by NIRS which proved its potential applications in peanuts. However, research on single peanut kernels only focused on fatty acid determinations (Tillman et al., 2006). By contrast, NIRS of single seeds was also already applied for grains and beans analysis. Specifically, the contents of amino acids and sucrose (Natsuga et al., 2007) of single soybean kernels have been examined. Furthermore, the protein, starch (Jiang et al., 2007), and ergosterol contents (Berardo et al., 2005) of single corn kernels, glucosinolates and indole concentrations of single rapeseed kernels (Hom et al., 2006), as well as caffeine concentrations of single coffee kernels (Fox et al., 2013) have been analysed by NIRS. Meanwhile, with the miniaturization of optical technology, portable NIRS has received more and more attention because of its lower price and higher convenience than the benchtop equivalents, at least for some applications. Although the studies mentioned above adopted benchtop NIRS, some studies have been conducted with portable NIRS and used the technology for quantitative testing, such as for the fatty acid analysis of peanuts (Yu et al., 2020), and for qualitative testing such as organic milk distinction (Liu et al., 2018) and olive oil classification (Yan et al., 2019). Satisfactory conclusions of the above research were made by portable NIRS coupled with chemometrics.

To analyse single peanut kernels, a detection accessory combined with portable NIRS should be designed with consideration of the characteristics of single peanut kernels. However, no reference study about features of single peanut kernels to help design such a detection accessory in order to offer multiple-trait non-destructive analysis was published so far. Therefore, this study aims to investigate the external and internal characteristics of single peanut kernels for designing a single peanut detection accessory for spectral data collection. Furthermore, the results obtained with the portable NIRS device was examined for its prediction capabilities of the fat, protein, sucrose, and amino acids contents of single peanuts. The relationships between spectral data and the different molecular groups of those components were also explored.

2. Materials and methods

2.1. Peanut samples

A set of 110 peanut varieties and strains were randomly selected from 10 main planting provinces in China. Peanuts (approximately 2 kg per variety and strain) were stored at 4 °C in a commercial cold store (Yuangdong Co., Ltd., Tianjing, China). The overview of all peanut varieties and strains collected and analyses are provided as Supplementary data in Table S1 (Appendix A. Supplementary data).

2.2. Compositional analysis

The following number of samples were randomly analysed per conventional compositional analysis: 110 (fat and protein), 100 (amino acids), and 80 (sucrose). More information on the sample/analysis distribution can be found in Table S1 (Appendix A. Supplementary data). Each chemical analysis was performed in duplicate and the results were averaged for data analysis. The reference fat contents were determined based on Soxhlet extraction method (GB5009.6, 2016) by using a Soxtec 2050 instrument (FOSS, Hillerød, Denmark). The protein contents were measured by Kjeldahl method (GB5009.5, 2016) using a 2300 Nitrogen Analyzer (FOSS, Hillerød, Denmark) with a conversion coefficient (5.46). The quantitative analysis of amino acids were conducted through the standard (GB5009.124, 2016). Briefly, about 2 g peanut samples were transferred to hydrolysis tubes with nitrogen after acidic hydrolysis with 15 mL hydrochloric acid solution. The hydrolysis tubes were placed in a hydrolysis furnace (110 °C) for 22 h. After constant volume and drying, the solution was then filtered through a 0.22 µm membrane after which the filtrate was analysed by an L-8900 amino acids automatic analyser (Hitachi, Tokyo, Japan). Sucrose in the peanuts was analysed by high-performance anion-exchange chromatography method (GB5009.8, 2016). Approximately 5 g peanut powder samples were degreased firstly and then extracted in volumetric flasks (100 mL) placed in an ultrasonic bath for 30 min with water, zinc acetate solution, and potassium ferrocyanide solution. The extracts were centrifuged for 30 min at 12,000 rpm. The supernatant was collected and filtered through a 0.45 µm filter after which the filtrate was analysed using a Dionex ICS-3000 (Thermo Fisher Scientific, Waltham, USA). All reagents used in the above analyses were purchased from Sinopharm Chemical Reagent Co., Ltd. (Beijing, China).

2.3. Size analysis of single peanut kernels

The width and length of single peanut kernels were measured using a Vernier Caliper (Shanghai Tool Factory Co., Ltd., Shanghai, China). The width (mm) of peanut kernels is the widest point of the mature plump seeds, whereas the longest length of the mature plump seeds is regarded as the length (mm) of peanut kernels according to a previously reported procedure (Huifang et al., 2006). All varieties and strains ($n = 110$) were analysed in duplicate.

2.4. Confocal laser scanning microscopy analysis

The internal microstructures of the peanut kernels were assessed by confocal laser scanning microscopy (CLSM) using a Zeiss LSM 880 microscope (Carl Zeiss Jena, Munchen, Germany). The microscope was outfitted with 40× objectives and two lasers, including an Argon laser (λ excitation 488 nm) and a diode-pumped-solid-state laser (λ excitation 561 nm). Matured embryos were separated from peanut seeds based on an approach reported previously (Perry and Wang, 2003). In brief, peanut embryos were immediately soaked for 4 h in 2.5 % v/v glutaraldehyde and 1.6 % w/v paraformaldehyde in a 0.1 M phosphate buffer solution (pH 7.4). After flushing, the embryos were soaked in a 2 % w/v osmium tetroxide solution for an additional 4 h followed by dehydration using acetone. The dehydrated embryos were fixed in epoxy resin and semithin sections were attained by a ultramicrotome EM UC6 (Leica, Wetzlar, Germany). All reagents used in the above processing of semithin sections were purchased from Sinopharm Chemical Reagent Co., Ltd. (Beijing, China). 0.1 % w/v Nile red and 0.1 % w/v fluorescein isothiocyanate (Sigma Chemical Co., St Louis, MO, USA) were used for visualization of the fat and protein components. The samples Yuanza9847, Hanghua2, Yueyou7, Jihua13, Yuhua9326, and Huayu25 were selected as representative samples for analysis and were analysed in duplicate (for sample list see Table S1, Appendix A. Supplementary data).

2.5. Portable NIRS and spectral collection

A previously developed box-portable NIRS system was used in this study (Yu et al., 2020). The device comprised a spectrometer-Micro-NIRS 1700 (908–1676 nm, 6 nm interval, VIAVI Solutions, San Jose, USA), a control centre consisting of a tablet computer (Surface, Microsoft Corporation, Redmond, USA), a reference white Teflon board for white reference, a lithium battery for power supply, and a single peanut kernels detection accessory designed in this study. Each peanut sample ($n = 110$) was placed into the single detection accessory to collect the spectral data by measurement in reflection mode. The dark reference was regarded as the collected dark current when the light was off. The measurements (spectral data acquisitions) were conducted after dark and white reference calibration. Twenty peanut kernels per sample were scanned once. The spectral data per sample were averaged for data analysis.

2.6. Data analysis

The first step was to determine outliers. The average spectra and the corresponding components of single peanut kernels were regarded as the dependent variables X and Y , respectively. The X outliers of each component group were determined by principal component analysis (PCA) according to the Hotelling's T^2 (Yu et al., 2020) separately, whilst the Y outliers were analysed by boxplot combined with violin analysis (Dettori et al., 2018; Xu et al., 2019a). The boxplot analysis presented box charts with the median as the middle, the 25th and 75th percentiles as the edges, and the maximum and minimum value as the whiskers (Quintelas et al., 2019), while violin plots showed the distributions of data. The data outside the range of whiskers were considered the Y outliers which were marked with black dots. PCA combined with K-means clustering was used for exploring the relationship between different peanut varieties and strains. After the outliers were eliminated, samples of each component were first sorted by content in descending order. Fifteen samples (ten samples for sucrose) were selected for validation according to the sequence (e.g. select one from every six samples) and the remaining for calibration. The range and standard deviation (SD) of calibration and validation set of each component were calculated.

The partial least square regression analysis (PLSR) prediction models were established to quantify the composition of single peanut kernels (Chen et al., 2017; Kutsanedzie et al., 2018). Before modelling, the spectral data were pre-processed in order to eliminate the overlapping of the original spectra (Leng et al., 2020), such as first derivative (1st der), second derivative (2nd der), standard normalization variable (SNV), detrending (DT), baseline (BL), multiple scattering correction (MSC), and combined pre-treatments including 1st der + DT and 2nd der + SNV, etc. After modelling, the performances of models were validated through cross validation and external prediction. The cross validation (leave-one-out) was used to avoid overfitting of the calibration models. The number of factors was chosen based on the best results of the cross validation. The evaluation index had the correlation coefficient in cross validation (R_{CV}) and standard error in cross validation (SECV). The external prediction, using the validation set, evaluated the robustness of the calibration models which included the parameters such as the correlation coefficient in calibration (R_c) and prediction (R_p), and standard error in calibration (SEC) and prediction (SEP) as given by Eq. (1), as well as bias and residual predictive deviation (RPD). The RPD is equal to the SD of validation set divided by SEP. Different RPD ranges have been reported in different sources (Fearn, 2002; Nicolaï et al., 2007). In this study, according to Lima et al. (2020) and Ye et al. (2016) research, if RPD values ≥ 2 , models are considered excellent and RPD values < 1.5 are considered non-reliable, while the models with RPD value in between are considered good. R 3.5.3 (R Foundation for Statistical Computing, Austria) and Unscrambler 10.3 (CAMO Software AS, Norway) were used for data analysis.

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_{pi} - \hat{y}_{pi})^2}{n-1}} \quad (1)$$

$\hat{y}_{pi}$ and y_{pi} are prediction and reference values of the i th sample in the validation group; n is the number of samples in the validation group. If all variables come from the calibration group, the equation indicates SEC.

3. Results and discussion

3.1. Compositional results of different peanut varieties

The distribution map composed of a boxplot (black box in the violin) and violin plot shows the distributions of major components and amino acids (Fig. 1). The fat content ranged from 43.5–57.6 g/100 g (with only one outlier identified), while the protein content ranged from 18.4–28.6 g/100 g (two outliers identified). Based on the violin plot, most fat and protein contents of peanut varieties ranged from 49–54 g/100 g and 22–27 g/100 g, respectively. This is similar compared with previous studies for fat (48–55 g/100 g) (Li et al., 2018) and protein

Fig. 1. The distribution map of the different components of peanut kernels after deleting spectral outliers. (Outliers marked with black spots) (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

(20–26 g/100 g) (Gong et al., 2018) in peanuts. Sucrose varied from 1.6 to 6.7 g/100 g. Lower values (2.6–6.5 g/100 g) were observed by Bishi (Bishi et al., 2015). The peanuts appeared rich in essential amino acids. Especially, phenylalanine (Phe, 0.98–1.83 g/100 g, four outliers), leucine (Leu, 1.18–2.27 g/100 g, two outliers), and valine (Val, 0.81–1.31 g/100 g, two outliers) were main essential amino acids in the peanuts, while the other essential amino acids such as lysine (Lys, 0.71–1.23 g/100 g, one outlier), isoleucine (Ile, 0.68–1.09 g/100 g, an outlier), methionine (Met, 0.20–0.37 g/100 g, one outlier), and threonine (Thr, 0.50–0.88 g/100 g, an outlier) were relatively low. The non-essential amino acids presenting the highest concentrations were glutamic acid (Glu, 3.27–7.86 g/100 g), aspartic acid (Asp, 2.05–4.33 g/100 g, two outliers), arginine (Arg, 2.21–4.29 g/100 g, three outliers), and glycine (Gly, 1.11–2.20 g/100 g), whereas alanine (Ala, 0.65–1.37 g/100 g, two outliers), tyrosine (Tyr, 0.50–1.59 g/100 g), proline (Pro, 0.87–1.34 g/100 g, three outliers), serine (Ser, 0.89–1.88 g/100 g), and histidine (His, 0.46–0.75 g/100 g, one outlier) displayed the lower concentrations. The amino acids contents were similar to those reported by Radhakrishnan et al. (2014) and Klevorn et al. (2019).

The PCA method was used to explore the relationship between different peanut samples ($n = 70$) containing with all components analysed. The PCA biplot (PCA score plot combined with loading plot) is shown in the Fig. 2. The total numbers of PC1 and PC2 is 70.3 %. K-means as an unsupervised method was used to cluster the peanuts based on all components analysed. The blue points present that samples with a higher fat content, while the yellow ones have higher sucrose contents. Obviously, a negative correlation between fat and sucrose contents exist, which is consistent with previous research (Bishi et al., 2015). It is also found that fat and protein contents are negatively correlated because they are located on opposite sides of the plot, which another study confirmed as well (Sarvamangala et al., 2011). The amino acids have the same direction, which means they are positively related with the fat content. Among them, hydrophobic amino acids such as Leu, Met, Val, and Phe have similar orientation to fat. It appears that peanut varieties and strains with higher fat contents usually have higher hydrophobic amino acids contents, lower sucrose contents, and lower protein contents.

3.2. The internal structure of single peanuts and component distribution

The distribution of the oil body (OB) and protein body (PB) which are the source of NIRS information in the peanut cell was investigated by CLSM. Fig. 3 shows that OB (red) and PB (green) were randomly and evenly (no agglomeration) distributed at cell level, which is consistent with previous results (Zaaboul et al., 2018). The amino acids, especially the essential ones, are known to be located in the OB and PB. Therefore, the collected spectral information would likely contain fat, protein and amino acids information at the same time. This is one of the reasons why spectral pre-processing are performed before modelling. There is a lot of differences between different varieties and strains. For example, the differences between Yuanza9807 (Fig. 3a) and Jihua13 (Fig. 3b) are considerable since most PB in Yuanza9807 are larger but the number is relatively lower because of the cell size and PB size. OB are very densely distributed in cells of two varieties and it could be observed that OB in Jihua13 are larger. The superposition of differences at cell level eventually leads to differences in chemical composition between varieties and strains.

3.3. Development of the single detection accessory

The range of distribution of the length and width of all peanuts ($n = 110$) was 11.32–24.25 mm and 7.49–12.25 mm, respectively. To meet different sizes, two different apertures were designed considering the peanut size range. One was sized 12.5 mm $\times$ 24.5 mm and the other one 9.7 mm $\times$ 17.5 mm (Fig. 4a). The bottom part of the accessory was made of 1 mm quartz glass which has good permeability for light in the NIRS region. The material used for the rest of the accessory was aluminium. According to previous research (Fraser, 2001), wavelengths from 1400 to 1600 nm could penetrate up to 1 mm. Although the thickness of the penetration would be increased as the wavelength goes down, it is still very difficult to penetrate peanuts ($\text{width}_{\min} = 7.49$ mm). Therefore, the reflectance measurements for peanut kernel detection were chosen. A schematic diagram of the portable device with a single peanut accessory and the spectral acquisition route are shown in the Fig. 4b.

Fig. 2. The principal component analysis (PCA) biplot of different peanut samples ($n = 70$, different coloured ovals indicate the different sample clusters and arrows indicate the contributions of compositions). (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

Fig. 3. The confocal images of single peanut kernels, (a) Yuanza9847, (b) Jihua13.

Fig. 4. The (a) single kernel near-infrared spectroscopy detection accessory (the numerals represent the aperture size, unit: mm), and (b) schematic of the portable near-infrared spectroscopy measurement system.

3.4. Spectral data and various data pre-treatment procedures

The original averaged spectral data of the single kernels of all peanut varieties ($n = 110$) are presented in the Fig. 5a. Compared with previous research (Wang et al., 2013; Yu et al., 2016), the curve of the spectral line was the same. Specifically, the peak and valley of the original spectra appeared in the same position. The peaks of the original spectra are 1205 nm and 1460 nm, while 1106 nm and 1298 nm are the valleys, which are caused by the frequency absorption harmony of hydrogen-containing groups including C–H (generally from fat and sucrose), O–H (generally from water), and N–H (generally from protein and amino acids) (Hourant et al., 2000). Hence, the peaks and valleys are characteristic of the main chemical constituents of the peanuts.

Different pre-treatments of the raw spectra could help to attain more useful information for the generation of models. For instance, derivatives could eliminate the baseline and the background effect and improve the resolution and sensitivity of spectral information (Chu, 2011). Detrending is a method to erase the baseline drift of the spectral data (Pérez-Rodríguez et al., 2018). The SNV or MSC pre-processing can remove the unnecessary spectral information induced by seed size (Agelet et al., 2012). After pre-treatment of the spectra, more signals of molecular groups that may be related to fat, sucrose, protein, and amino acids were exposed. Taking the 2nd derivate combined with detrending as an example (Fig. 5b), 927 nm is 3rd C–H stretch overtone from methylene (Workman and Weyer, 2012), while 952 nm is mainly attributed to the second O–H stretch overtone (Williams and Norris, 1987). There are many signals between 1110 nm and 1250 nm linked

with the C–H stretch overtone. More specifically, It is found that 1125 nm and 1181 nm come from the second C–H stretch overtone (Fernandez-Novales et al., 2019) and the second C–H stretching of HC=CH respectively. It is reported that 1212 nm is derived from the second overtone C–H stretching of CH₂–CH₂ and 1243 nm is from the band assignments of 3×C–H stretch (Workman and Weyer, 2012). The first overtone region exhibits the major bonded O–H peak (1280 nm) whereas the C–H combination of C–H₃ is located at 1385 nm (Workman and Weyer, 2012). The wavelength 1428 nm is usually identified as the first stretching of N–H and the signal at 1658 nm is the C–H methyl of carbonyl adjacent (Workman and Weyer, 2012).

The PCA results of the original spectra of peanut varieties ($n = 110$) are indicated in Fig. 5c. The total variation explained by PC1 and PC2 was 99 %. It could be found that some dots are outside the line calculated the Hotelling's T^2 with 0.5 % significance level. These dots are regarded as the outliers and there were five outliers in total. These five outliers also exist in the amino acids data set. After removing all chemical and spectral outliers, descriptive statistics including the range values and the SD of the chemical composition of the reference samples were calculated (Table 1). The wide-range data and the SD of different compositions of peanut kernels show the sample diversity in this study.

3.5. Prediction models for compositional characteristics

Various models were calculated to predict the compositional characteristics of the peanuts from the spectral data. The calibration, cross-validation, and external prediction statistics arguments (PLSR factor, R,

Fig. 5. The (a) raw spectra and (b) pre-processed spectra (2nd derivate combined with detrending) of single peanut kernels ($n = 110$), and the (c) principal component analysis (PCA) scores plot ($n = 110$).

SEC/CV/P, RPD, and bias) of the optimization models for the assessment of the major compounds and each amino acid combined with the optimal pre-treatment are shown in Table 2. Generally, the number of PLSR factors is a critical parameter to evaluate the capability of the calibration models. If the number of PLSR factors used is too small, useful information would be lost. In contrast, the use of an excessive large number of factors would result in overfitting of the model. The best number of factors to use is determined by the results of cross validation. The factors of all components in single peanuts ranged from 7 to 11.

Compared with bulk sample analysis, the spectral data of single kernels are easily affected by kernel size even if pre-treatment has been considered to reduce that noise. Therefore, the more useful information based on higher factors should be used. A similar number of factors were needed as the previous studies, such as for single rice kernels (factor = 9) (Xu et al., 2019b), single coffee beans (factor = 8–10) (Fox et al., 2013), and single cocoa beans (factor = 8–15) (Caporaso et al., 2018) to ensure robustness of the models.

The optimal calibration model for fat prediction was established with

Table 1

Compositional characteristics of sample sets used for the calibration and external validation of the spectral models.

Component (g/100 g)	Calibration set			Validation set		
	n	Range	SD	n	Range	SD
Fat	89	43.5–57.6	3.1	15	45.9–56.7	4.2
Protein	88	18.4–28.6	2.2	15	18.4–26.8	3.4
Sucrose	70	1.6–6.7	1.0	10	2.4–6.0	1.2
Arg	77	2.21–4.29	0.45	15	2.27–4.40	0.49
Asp	78	2.05–4.33	0.47	15	2.58–4.25	0.47
Phe	76	0.98–1.83	0.17	15	1.00–1.72	0.19
Ala	78	0.65–1.37	0.15	15	0.74–1.27	0.15
Gly	80	1.11–2.20	0.24	15	1.36–2.05	0.20
Lys	79	0.71–1.23	0.10	15	0.90–1.08	0.05
Tyr	80	0.50–1.59	0.24	15	0.81–1.40	0.17
Leu	78	1.18–2.27	0.25	15	1.47–2.09	0.18
Pro	77	0.87–1.34	0.11	15	0.88–1.29	0.10
Ile	79	0.68–1.09	0.09	15	0.68–0.88	0.06
Val	78	0.81–1.31	0.11	15	0.92–1.21	0.10
Met	80	0.20–0.37	0.04	15	0.20–0.35	0.05
Glu	80	3.27–7.86	1.03	15	3.82–6.81	0.82
Ser	80	0.89–1.88	0.22	15	1.06–1.76	0.19
Thr	79	0.50–0.88	0.08	15	0.59–0.76	0.06
His	79	0.46–0.75	0.06	15	0.46–0.77	0.07

(Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (N) Number of samples; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (SD) Standard deviation between samples; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

an RPD value of 2.39 ($R_p = 0.91$). This value is higher than the RPD value of 2.13 obtained by Tallada et al. (2009) for single maize kernel samples. The model accuracy to predict the content of sucrose was also high considering the RPD value of 2.02 ($R_p = 0.92$). Very few studies have focused on the prediction of sucrose content before. It was only reported that the model for the prediction of sucrose in green soybeans, using a transmission model, had a modest result with a R_{cv} value of 0.86 (Natsuga et al., 2007). Nevertheless, sucrose is the main contributor to the sweetness of peanuts (Bishi et al., 2013). Therefore, a good sucrose model as presented in this study is certainly relevant for peanuts. The protein content of the single peanut kernels could be predicted best by the model based on the baseline combined with MSC with an RPD value

of 2.36 ($R_p = 0.91$). The result was slightly better than the previously reported models for protein measurements in single soybeans ($RPD_{max} = 2.33$) which were built by three different instruments except for the Light Tube ($RPD = 3.28$) with high resolution (3 nm) (Agelet et al., 2012). It also had good agreement with the RPD value of 2.9 obtained by Fox (Fox et al., 2011) for single barley analysis.

In terms of amino acids, good performance calibration models were established for Arg, Asp, Phe, Ile, Glu, and Ser. Although Arg is not the predominant component among all amino acids of peanuts, it is a characteristic amino acid of peanuts because its content is higher in peanuts compared with those of other crops and grains. The optimal model was built for the prediction of Arg with the highest calibration and validation correlation coefficients ($R_c = 0.90$; $R_{cv} = 0.87$; $R_p = 0.90$) by using 1st der combined with Baseline pre-treatment. The RPD value of 1.98 was slightly lower than for bulk peanut samples ($RPD = 2.88$) (Wang et al., 2013). Though the matrices were different, these correlation coefficients were higher than the ones studied by Carbas et al. (2020) for common beans. There is no doubt that Asp is another characteristic amino acid of peanuts that can adjust the metabolism of the human brain and nervous system. Good quality prediction models were built with the current dataset, showing high correlation coefficients for calibration ($R_c = 0.90$) and validation ($R_p = 0.89$). The RPD value of 2.15 was slightly lower than bulk peanut samples ($RPD = 2.56$) (Wang et al., 2013) and the same as Juan reported ($RPD = 2.22$) (Fernandez--Novales et al., 2019). It is known that Phe is the essential amino acid for most people for the synthesis of important neurotransmitters and hormones. However, the content should be controlled for all phenylketonuria because they cannot metabolise Phe. The model was first established for the prediction of Phe in peanut kernels and had good robustness ($R_c = 0.89$; $R_p = 0.93$; $RPD = 2.65$). These results were better than the Phe prediction model for rice wine (Shen et al., 2010) and common beans (Carbas et al., 2020). The model of Ile, which is one of the essential amino acids, was also built for the first time. Table 2 showed that the model had the best performance ($R_c = 0.84$; $R_p = 0.88$; $RPD = 2.00$) with the pre-treatment, 1st der combined with MSC. Glu constitutes the highest content of amino acids in peanuts and Ser can promote the metabolism of fats and fatty acids. More importantly, these two amino acids had great contributions to the function of peanut protein (Wang et al., 2013). The models generated for detecting Glu and Ser in single peanut kernels had great results ($RPD_{Glu} = 1.99$; $RPD_{Ser} = 2.08$)

Table 2

The calibration, cross validation and external validation results of peanut kernel major components and amino acids.

Components	Pre-treatment	Factor	Calibration		Cross validation		External validation				
			R_c	SEC	R_{cv}	SECV	R_v	SEP	Bias/Mean	Slope	RPD
Fat	2 nd der + MSC	11	0.90	1.38	0.78	2.09	0.91	1.76	0.011	0.917	2.39
Sucrose	2 nd der + SNV	10	0.94	0.39	0.83	0.62	0.92	0.57	0.011	1.164	2.02
Protein	BL + MSC	11	0.90	1.25	0.80	1.76	0.91	1.42	0.016	0.831	2.36
Arg	1 st der + BL	9	0.92	0.19	0.87	0.24	0.90	0.25	−0.009	1.097	1.98
Asp	DT + SNV	10	0.90	0.22	0.82	0.29	0.89	0.22	−0.002	0.697	2.15
Phe	2 nd der + DT	8	0.89	0.09	0.81	0.11	0.93	0.07	0.006	0.797	2.65
Ala	2 nd der + MSC	10	0.90	0.07	0.77	0.11	0.83	0.09	−0.033	0.885	1.67
Gly	2 nd der + BL	9	0.88	0.12	0.77	0.16	0.85	0.12	−0.019	0.990	1.63
Lys	1 st der + BL	9	0.89	0.05	0.73	0.07	0.82	0.03	−0.001	0.973	1.52
Tyr	2 nd der + MSC	7	0.87	0.12	0.76	0.16	0.87	0.10	−0.093	1.053	1.72
Leu	2 nd der + SNV	10	0.91	0.11	0.78	0.17	0.83	0.12	0.050	0.910	1.51
Pro	2 nd der + DT	10	0.88	0.06	0.72	0.09	0.80	0.06	0.019	0.741	1.54
Ile	1 st der + MSC	8	0.84	0.05	0.75	0.06	0.88	0.03	0.017	0.942	2.00
Val	1 st der + DT	8	0.88	0.06	0.81	0.07	0.78	0.06	0.007	0.278	1.58
Met	2 nd der + BL	11	0.92	0.02	0.76	0.03	0.83	0.03	0.054	0.744	1.60
Glu	2 nd der + BL	9	0.88	0.48	0.76	0.65	0.88	0.41	0.020	0.645	1.99
Ser	2 nd der + DT	10	0.91	0.09	0.80	0.13	0.93	0.09	−0.005	1.150	2.08
Thr	1 st der	8	0.82	0.05	0.71	0.06	0.83	0.04	0.010	0.835	1.50
His	2 nd der + SNV	10	0.88	0.03	0.74	0.05	0.78	0.05	0.002	0.802	1.40

(Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (BL) Baseline correction; (DT) Detrending; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (MSC) Multiple scattering correction; (Phe) Phenylalanine; (Pro) Proline; (R) Pearson correlation coefficient; (RPD) Residual predictive deviation; (SEC) Standard error in calibration; (SECV) Standard error in cross validation; (SEP) Standard error in prediction; (Ser) Serine; (SNV) Standard normalization variable; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine; (1st der) First derivative; (2nd der) Second derivative.

which were better than those in similar research on intact grape berries ($RPD_{Glu} = 1.90$; $RPD_{Ser} = 2.07$) (Shen et al., 2010). The rest of the essential amino acids included Lys, Thr, Leu, and Val. The models of these essential amino acids also had relatively good performance ($R_p = 0.78$ – 0.83 ; $RPD = 1.50$ – 1.58), which were better than the previous studies. Among these, Val and Leu are hydrophobic amino acids, while Lys and Thr are hydrophilic amino acids. Both affect the characteristics of peanut proteins. The rest of the non-essential amino acids included Ala, Gly, Tyr, Pro, and His. Except for His ($RPD = 1.40$), models for these amino acids showed good performance ($R_p = 0.8$ – 0.87 ; $RPD = 1.54$ – 1.72).

Bias in this kind of models is identified as the average differences between the predictive content and the actual content of a compound. A positive bias is regarded that the model overestimated the composition of an individual compound, while a negative value means that it is underestimated. A good model should have a low bias (Shao et al., 2011). To eliminate the mean order of magnitude effect, bias/mean was used instead of bias (Vallone et al., 2019). As shown in Table 2, most established models were slightly over-estimating the content of the amino acids and major components in single peanuts. All the models had low bias/mean (< 0.002) except for Ala, Tyr, Leu, and Met. Meanwhile, those four amino acids had relatively low RPD values. It was further proven that models have more reliable performances leading to high RPD and low bias/mean.

4. Conclusions

Portable NIRS technology coupled with a single detection accessory allowed precise measurements of the major components and amino acids of single peanut kernels. This study highlighted the potential of the proposed portable single kernel instrument to estimate the individual composition of single peanuts. Meanwhile, high-fat peanut varieties and strains usually have relatively low protein and sucrose contents. Ultimately, it allows a better understanding of variation in contents among peanut kernels within and across groups of samples. The methodology is faster and high-throughput compared with conventional methods that require more time and transition of the samples to the laboratories. This offers a great advantage that can be used to control the quality and integrity of peanuts from the beginning of the industrial chain.

Funding

This study was supported by the National peanut industry technology system of China (CARS-13-08B), the Earmarked Fund for Hebei Oil Crop Innovation Team of Modern Agro-industry technology Research System of China, the CAAS-WUR Joint PhD Programme of China (MOE11NL1A20151701N) and the first author received a China Scholarship Council (CSC) scholarship (201903250121).

CRedit authorship contribution statement

Hongwei Yu: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft. **Hongzhi Liu:** Methodology, Investigation, Supervision, Resources. **Sara W. Erasmus:** Supervision, Writing - review & editing. **Simeng Zhao:** Investigation, Methodology. **Qiang Wang:** Conceptualization, Funding acquisition, Project administration, Supervision, Resources, Writing - review & editing. **Saskia M. van Ruth:** Conceptualization, Project administration, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors earnestly acknowledge contributions made by Prof. Xinyou Zhang and other breeders from the National Peanut Industry System (China) offering the peanut samples for this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.indcrop.2020.112956>.

References

- Agelet, L.E., Hurburgh Jr, C.R., 2014. Limitations and current applications of near Infrared Spectroscopy for single seed analysis. *Talanta* 121, 288–299. <https://doi.org/10.1016/j.talanta.2013.12.038>.
- Agelet, L.E., Armstrong, P.R., Clariana, R.I., Hurburgh, C.R., 2012. Measurement of single soybean seed attributes by near-infrared technologies. A comparative study. *J. Agric. Food Chem.* 60, 8314–8322. <https://doi.org/10.1021/jf3012807>.
- Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A., Marocco, A., 2005. Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* 53, 8128–8134. <https://doi.org/10.1021/jf0512297>.
- Bishi, S.K., Kumar, L., Dagla, M.C., Mahatma, M.K., Rathnakumar, A.L., Lalwani, H.B., Misra, J.B., 2013. Characterization of Spanish peanut germplasm (*Arachis hypogaea* L.) for sugar profiling and oil quality. *Ind. Crop. Prod.* 51, 46–50. <https://doi.org/10.1016/j.indcrop.2013.08.050>.
- Bishi, S.K., Lokesh, K., Mahatma, M.K., Khatediya, N., Chauhan, S.M., Misra, J.B., 2015. Quality traits of Indian peanut cultivars and their utility as nutritional and functional food. *Food Chem.* 167, 107–114. <https://doi.org/10.1016/j.foodchem.2014.06.076>.
- Caporaso, N., Whitworth, M.B., Fowler, M.S., Fisk, I.D., 2018. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chem.* 258, 343–351. <https://doi.org/10.1016/j.foodchem.2018.03.039>.
- Carbas, B., Machado, N., Oppolzer, D., Ferreira, L., Brites, C., Rosa, E.A.S., Barros, A., 2020. Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for the determination of nutritional and antinutritional parameters in common beans. *Food Chem.* 306, 125509. <https://doi.org/10.1016/j.foodchem.2019.125509>.
- Chen, K., Li, C., Tang, R., 2017. Estimation of the nitrogen concentration of rubber tree using fractional calculus augmented NIR spectra. *Ind. Crop. Prod.* 108, 831–839. <https://doi.org/10.1016/j.indcrop.2017.06.069>.
- Chopra, R., Folstad, N., Lyons, J., Ulmasov, T., Gallaher, C., Sullivan, L., McGovern, A., Mitacek, R., Frels, K., Altendorf, K., Killam, A., Ismail, B., Anderson, J.A., Wyse, D.L., Marks, M.D., 2019. The adaptable use of Brassica NIRS calibration equations to identify pennycress variants to facilitate the rapid domestication of a new winter oilseed crop. *Ind. Crop. Prod.* 128, 55–61. <https://doi.org/10.1016/j.indcrop.2018.10.079>.
- Chu, X., 2011. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Applications*. Chemical Industry Press, Beijing.
- Dettori, M., Arru, B., Azara, A., Piana, A., Mariotti, G., Camerada, M.V., Stefanelli, P., Rezza, G., Castiglia, P., 2018. In the digital era, is community outrage a feasible proxy indicator of emotional epidemiology? The case of meningococcal disease in Sardinia. *Italy. Int. J. Environ. Res. Public Health* 15, 1512. <https://doi.org/10.3390/ijerph15071512>.
- Dubinina, A., Lehnert, S., Khomenko, O., 2014. Amino acid composition of protein and its biological value in seeds of peanut sorts widen in Ukraine. *Agric. Food* 2, 501–510.
- Faostat, F., 2017. Statistical Databases. <http://www.fao.org/faostat/en/#data/QC>.
- Fearn, T., 2002. Assessing calibrations: sep, rpd, rer and r 2. *Nir News* 13, 12–13.
- Fernandez-Novales, J., Garde-Cerdan, T., Tardaguila, J., Gutierrez-Gamboa, G., Perez-Alvarez, E.P., Diago, M.P., 2019. Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta* 199, 244–253. <https://doi.org/10.1016/j.talanta.2019.02.037>.
- Fox, G., Kelly, A., Sweeney, N., Hocroft, D., 2011. Development of a single kernel NIR barley protein calibration and assessment of variation in protein on grain quality. *J. Inst. Brew.* 117, 582–586. <https://doi.org/10.1002/j.2050-0416.2011.tb00507.x>.
- Fox, G.P., Wu, A., Yiran, L., Force, L., 2013. Variation in caffeine concentration in single coffee beans. *J. Agric. Food Chem.* 61, 10772–10778. <https://doi.org/10.1021/jf4011388>.
- Fraser, D., 2001. Near infra-red (NIR) light penetration into an apple. *Postharvest Biol. Technol.* 22, 191–194.
- GB5009.124, 2016. National Food Safety Standard - Determination of Amino Acid in Foods. Standard Press of China, Beijing.
- GB5009.5, 2016. National Food Safety Standard - Determination of Protein in Foods. Standard Press of China, Beijing.
- GB5009.6, 2016. National Food Safety Standard - Determination of Fat in Foods. Standard Press of China, Beijing.
- GB5009.8, 2016. National Food Safety Standard - Determination of Fructose, Glucose, Sucrose, Maltose and Lactose in Foods. Standard Press of China, Beijing.
- Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., Wang, Q., 2018. Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *J. Integr. Agric.* 17, 1003–1010. [https://doi.org/10.1016/s2095-3119\(18\)61919-7](https://doi.org/10.1016/s2095-3119(18)61919-7).

- Hom, N.H., Becker, H.C., Möllers, C., 2006. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* 153, 27–34. <https://doi.org/10.1007/s10681-006-9195-3>.
- Hourant, P., Baeten, V., Morales, M.T., Meurens, M., Aparicio, R., 2000. Oil and fat classification by selected bands of near-infrared spectroscopy. *Appl. Spectrosc.* 54, 1168–1174. <https://doi.org/10.1366/0003702001950733>.
- Huifang, J., Duan, N., Ren, X., 2006. Specification of Peanut Germplasm Resources. China Agricultural Press, Beijing.
- Jiang, H.Y., Zhu, Y.J., Wei, L.M., Dai, J.R., Song, T.M., Yan, Y.L., Chen, S.J., 2007. Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (*Zea mays* L.). *Plant Breed* 126, 492–497. <https://doi.org/10.1111/j.1439-0523.2007.01338.x>.
- Klevorn, C.M., Dean, L.L., Johanningsmeier, S.D., 2019. Metabolite profiles of raw peanut seeds reveal differences between market-types. *J. Food Sci.* 84, 397–405. <https://doi.org/10.1111/1750-3841.14450>.
- Kutsanedzie, F.Y.H., Hao, L., Yan, S., Ouyang, Q., Chen, Q., 2018. Near infrared chemoresponsive dye intermediaries spectra-based in-situ quantification of volatile organic compounds. *Sens. Actuata B-Chem.* 254, 597–602. <https://doi.org/10.1016/j.snb.2017.07.134>.
- Leng, T., Li, F., Xiong, L., Xiong, Q., Zhu, M., Chen, Y., 2020. Quantitative detection of binary and ternary adulteration of minced beef meat with pork and duck meat by NIR combined with chemometrics. *Food Control* 113, 107203. <https://doi.org/10.1016/j.foodcont.2020.107203>.
- Li, Y., Zhang, R., Qin, X., Liao, Y., Siddique, K.H.M., Prasad, D.M., 2018. Changes in the protein and fat contents of peanut (*Arachis hypogaea* L.) cultivars released in China in the last 60 years. *Plant Breed.* 137, 746–756. <https://doi.org/10.1111/pbr.12621>.
- Lima, A.B.Sd., Batista, A.S., Jesus, J.Cd., Silva, J.d.J., Araújo, A.C.Md., Santos, L.S., 2020. Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control* 107, 106802. <https://doi.org/10.1016/j.foodcont.2019.106802>.
- Liu, N., Parra, H.A., Pustjens, A., Hettinga, K., Mongondry, P., van Ruth, S.M., 2018. Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta* 184, 128–135. <https://doi.org/10.1016/j.talanta.2018.02.097>.
- Natsuga, M., Sue, Y., Ikeda, T., Egashira, H., Akazawa, T., Ura, N., 2007. Development and evaluation of dedicated bench-type near infrared spectrometers to estimate the flavour of green soybeans from single pod measurements. *J. Near Infrared Spectrosc.* 15, 327–332. <https://doi.org/10.1255/jnirs.741>.
- Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol. Technol.* 46, 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>.
- Pérez-Rodríguez, F., Skandamis, P., Valdramidis, V., 2018. *Quantitative Methods for Food Safety and Quality in the Vegetable Industry*. Springer, New York.
- Perry, S.E., Wang, H., 2003. Rapid isolation of *Arabidopsis thaliana* developing embryos. *Biotechniques* 35, 278–282. <https://doi.org/10.2144/03352bm06>.
- Quintelas, C., Mesquita, D.P., Ferreira, E.C., Amaral, A.L., 2019. Quantification of pharmaceutical compounds in wastewater samples by near infrared spectroscopy (NIR). *Talanta* 194, 507–513. <https://doi.org/10.1016/j.talanta.2018.10.076>.
- Radhakrishnan, R., Pae, S.-B., Kang, S.-M., Lee, B.-K., Lee, I.-J., Baek, I.-Y., 2014. An evaluation of amino acid, fatty acid and isoflavone composition in Korean peanut (*Arachis hypogaea* L.) seeds to improve the nutritional quality of breeding lines. *J. Korean Soc. Appl. Bi.* 57, 301–305. <https://doi.org/10.1007/s13765-014-4044-x>.
- Sarvamangala, C., Gowda, M.V.C., Varshney, R.K., 2011. Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). *Field Crops Res.* 122, 49–59. <https://doi.org/10.1016/j.fcr.2011.02.010>.
- Shao, Y., Cen, Y., He, Y., Liu, F., 2011. Infrared spectroscopy and chemometrics for the starch and protein prediction in irradiated rice. *Food Chem.* 126, 1856–1861. <https://doi.org/10.1016/j.foodchem.2010.11.166>.
- Shen, F., Niu, X., Yang, D., Ying, Y., Li, B., Zhu, G., Wu, J., 2010. Determination of amino acids in Chinese rice wine by fourier transform near-infrared spectroscopy. *J. Agric. Food Chem.* 58, 9809–9816. <https://doi.org/10.1021/jf1017912>.
- Shin, E.-C., Craft, B.D., Pegg, R.B., Phillips, R.D., Eitenmiller, R.R., 2010. Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA). *Food Chem.* 119, 1262–1270. <https://doi.org/10.1016/j.foodchem.2009.07.058>.
- 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. *J. Am. Oil Chem. Soc.* 87, 1103–1114. <https://doi.org/10.1007/s11746-010-1589-7>.
- Tallada, J.G., Palacios-Rojas, N., Armstrong, P.R., 2009. Prediction of maize seed attributes using a rapid single kernel near infrared instrument. *J. Cereal Sci.* 50, 381–387. <https://doi.org/10.1016/j.jcs.2009.08.003>.
- Tillman, B.L., Gorbet, D.W., Person, G., 2006. Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. *Crop Sci.* 46, 2121. <https://doi.org/10.2135/cropsci2006.01.0031>.
- Vallone, M., Alleri, M., Bono, F., Catania, P., 2019. Quality evaluation of grapes for mechanical harvest using vis NIR spectroscopy. *Agric. Eng. Int.: CIGR J.* 21, 140–149.
- Wang, Q., 2016. *Peanuts: Processing Technology and Product Development*. Academic Press, Cambridge.
- Wang, Q., 2018. *Peanut Processing Characteristics and Quality Evaluation*. Springer, New York.
- Wang, L., Wang, Q., Liu, H., Liu, L., Du, Y., 2013. Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* 93, 118–124. <https://doi.org/10.1002/jsfa.5738>.
- Wang, Q., Liu, H., Shi, A., Hu, H., Liu, L., Wang, L., Yu, H., 2017. Review on the processing characteristics of cereals and oilseeds and their processing suitability evaluation technology. *J. Integr. Agric.* 16, 2886–2897. [https://doi.org/10.1016/s2095-3119\(17\)61799-4](https://doi.org/10.1016/s2095-3119(17)61799-4).
- Williams, P., Norris, K., 1987. *Near-infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemists, Inc., Eagan.
- Workman Jr, J., Weyer, L., 2012. *Practical Guide and Spectral Atlas for Interpretive Near-infrared Spectroscopy*. CRC press, Boca Raton.
- Xu, S., Wang, J.J., Wei, Y., Deng, W.W., Wan, X., Bao, G.H., Xie, Z., Ling, T.J., Ning, J., 2019a. Metabolomics based on uhplc-orbitrap-ms and global natural product social molecular networking reveals effects of time scale and environment of storage on the metabolites and taste quality of raw pu-erh tea. *J. Agric. Food Chem.* 67, 12084–12093. <https://doi.org/10.1021/acs.jafc.9b05314>.
- Xu, Fan, S., Liu, J., Liu, B., Tao, L., Wu, J., Hu, S., Zhao, L., Wang, Q., Wu, Y., 2019b. A calibration transfer optimized single kernel near-infrared spectroscopic method. *Spectrochim. Acta A* 220, 117098. <https://doi.org/10.1016/j.saa.2019.05.003>.
- Yan, J., van Stuijvenberg, L., van Ruth, S.M., 2019. Handheld Near-Infrared spectroscopy for distinction of extra virgin olive oil from other olive oil grades substantiated by compositional data. *Eur. J. Lipid Sci. Technol.* 121, 1900031. <https://doi.org/10.1002/ejlt.201900031>.
- Ye, M., Gao, Z., Li, Z., Yuan, Y., Yue, T., 2016. Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. *Food Chem.* 190, 701–708. <https://doi.org/10.1016/j.foodchem.2015.05.112>.
- Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbi, R.I., Wang, Q., 2016. Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Anal. Methods* 8, 7482–7492. <https://doi.org/10.1039/c6ay02029a>.
- Yu, H., Wang, Q., Shi, A.M., Yang, Y., Liu, L., Hu, H., Liu, H.Z., 2017. Visualization of protein in peanut using hyperspectral image with chemometrics. *Spectrosc. Spect. Anal.* 37, 853–858. [https://doi.org/10.3964/j.issn.1000-0593\(2017\)03-0853-06](https://doi.org/10.3964/j.issn.1000-0593(2017)03-0853-06).
- Yu, H., Liu, H., Wang, Q., van Ruth, S., 2020. Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Sci. Technol.* 109398. <https://doi.org/10.1016/j.lwt.2020.109398>.
- Zaaboul, F., Raza, H., Chen, C., Liu, Y., 2018. Characterization of peanut oil bodies integral proteins, lipids, and their associated phytochemicals. *J. Food Sci.* 83, 93–100. <https://doi.org/10.1111/1750-3841.13995>.