*Available online www.jocpr.com*

**Journal of Chemical and Pharmaceutical Research, 2014, 6(6):938-941**

**Research Article**

**ISSN : 0975-7384**
**CODEN(USA) : JCPRC5**

# Measurement of protein content in chestnuts using near infrared spectroscopy

**Liu Jie, Li Xiaoyu*, Wang Wei, Xiao Wu, Zhang Jun and Zhou Zhu**

*College of Engineering, Huazhong Agricultural University, Wuhan, China*

---

**ABSTRACT**

*Protein content in chestnuts is an important parameter for evaluating the fruit quality. In this work, the pilot application of near infrared (NIR) spectroscopy in measuring protein content of chestnuts was implemented. 182 samples were involved and 5 preprocessing methods were compared. PLS regression models developed from the spectra of intact nuts and peeled nuts were established separately. The results shown that, for the peeled chestnuts, the model based on spectra after First Derivative preprocessing performed better than other models with 0.9044 as the correlation coefficient (R) of calibration subset and 0.8029 as the correlation coefficient (R) of validation subset, respectively. For the intact chestnuts, the model established on the spectra after Second Derivative preprocessing got 0.8748 as the correlation coefficient (R) of calibration subset and 0.7324 as the correlation coefficient (R) of validation subset, respectively. The results indicated that NIR spectroscopy was feasibility to measure the protein content of chestnut rapidly and nondestructively.*

**Keywords:** protein; chestnuts; chemometrics; NIR spectroscopy; nondestructive measurement

---

## INTRODUCTION

Chestnut is an important production according to its delicious flavor and abundant nutrients. The main components includes moisture, sugar, starch, protein and various trace elements, while the content of each component is influenced by variety, origin, field management and other factors. Although the protein content is usually 1.6% to 7.2% of the fruit weight, it is an critical basis for assessing the nutritional value and optimize the processing process[1,2]. Therefore, it is necessary to measure the protein content in chestnut. The current measurement mainly relies on the traditional chemical methods which needs sample preparation and is troublesome to operate.

Combined with chemometrics, near infrared spectroscopy has the capability to evaluate the component contain O-H and/or N-H group rapid and nondestructively[3]. Researchers had used it to analysis the animal protein[4],plant protein[5] and the protein in food or fodder[6-8]. However, there is rare report of the application of this method in the protein measurement of target with peel.

In this work, the near infrared (NIR) spectroscopy was pilot used to detect the protein content in chestnut in order to provide an rapid, nondestructive method for chestnut industry.

## EXPERIMENTAL SECTION

### Sample and data collection

The chestnut (Castanea mollissima) samples in this work were yield in Macheng area and the weight of individual fruit ranged from 9.26g to 27.74g. Considering the chestnut growing property, the samples were representative enough for different growing conditions, maturation and weather at harvest time. Prior to experimental tests the

samples were stored under the industry storage conditions (temperature 0 ° C ~ 2 ° C, relative humidity 80% to 90%). The experiment were implemented in 10 days with same time interval 3 months later than harvest considering the perplexing change process of chestnut fruit during storage. Before scanning, the samples were exposed to temperature 26 °C for 2 hours for the reason of temperature equilibrium.

A VECTOR33 NIR spectrometer (Brucker Optics, Ettlingen, Germany) equipped with fiber optic accessories was used to acquire the NIR spectra which were transferred to and stored on a computer ia an AQP card to the spectrometer. During the experiment, the temperature was kept at 26 °C and the spectrometer was warmed up for 1 hour before work to stabilize the light sources. Firstly, the spectrum of standard background was obtained as the basic reference. Secondly, each sample was scanned intact with their flat side facing the probe. Then, the sample was cut parallel to the flat side and the hemispheric part was put into the sample container to be scanned. For both the intact sample and the sample without peel, the probe was completely covered by the nut. The spectrum of each sample was an average of 64 times scan results and collected in reflectance mode in the range of 833nm~2500nm at 1.25×106 nm and stored in absorbance mode.

After spectra acquisition, the kernel of the sample was taken out for reference value evaluation according to the Chinese national standard methods(GB/T 5009.5-2003) and expressed as the percentage of the fresh weight.

**Data analysis**

The partial least square (PLS) regression method was used to establish the relationship between protein content and the spectra which would be a model for measurement of chestnut protein. The optimal latent variables were determined by the lowest root mean square error of cross validation (RMSECV) among the calibration set and the models were validated with the independent validation set. The performance of models were evaluated by the correlation coefficient (R) between predicted values and reference values of the parameters, the root mean square error of calibration (RMSEC) and the root mean and square error of prediction (RMSEP). A high R and low RMSEC and RMSEP means a good model.

Considering the influence of a preprocessing method to the predictive models depends on the characteristics of the spectra, four preprocessing methods, namely first derivation (FD), second derivation (SD), multiplicative scatter correction (MSC) and standard normal variate transformation (SNV) were applied to the spectra data and their function were compared with the raw spectra.

All the data analyses were carried out with the aid of Matlab software (V.7.6, Mathworks, Natick, USA) and the PLS models were developed by using the PLS toolbox (V. 5.20, Eigenvector Research, Inc., Wenatchee, USA).

## RESULTS AND DISCUSSION

**Sample spectra**

Figure 1 presented the raw spectra of one random selected sample in both intact and peeled condition. The spectrum of peeled sample had higher absorbance value than the spectrum of intact sample in the range from 1025 nm to 2500 nm, which possible due to the peel. The two kinds of spectra appeared different in the band between 833 nm~1053 nm and 2000 nm~2040 nm and similar features over the other spectra range. The differences in band of 833 nm~1053 nm maybe were caused by the color or texture about the sample surface while these in band of 2000 nm~2040 nm were lead from the features of peel component. Generally, the characteristic absorption bands associated with protein can be observed from the spectra of both sample conditions.

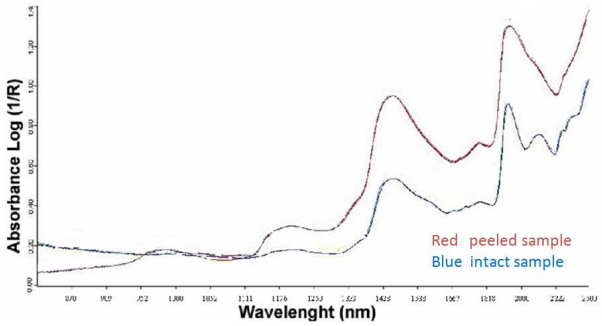


**Fig.1 Spectra of a random selected Chestnut sample**

**Sample set partition**

The protein content of 182 samples was scattered from 3.498% to 6.919% with 5.117% and 0.676 as the average and

standard deviation respectively. The SPXY algorithm was employed to choose one quarter of all samples to be the validation set and the rest samples worked as the calibration set. Table 1 showed the details of protein content and sample set partition. Since the spectra were different between the intact and peeled condition even of a same sample and the SPXY algorithm considered the features of both spectra and the reference value, the samples in each subset for intact and that for peeled chestnuts varied.

**Table 1 Results of Protein Content Value and Sample Set Partition**

| | Sample number | Protein content | | | |
|---|---|---|---|---|---|
| | | Max | Min | Mean | Standard deviation |
| Total | 182 | 6.919% | 3.498% | 5.117% | 0.676 |
| Peeled Calibration | 137 | 6.919% | 3.498% | 5.093% | 0.679 |
| Peeled Validation | 45 | 6.907% | 3.817% | 5.168%, | 0.665 |
| Intact Calibration | 137 | 6.919% | 3.498% | 5.128% | 0.714 |
| Intact Validation | 45 | 6.423% | 3.787% | 5.061% | 0.543 |

**Measuring model of peeled sample**

In the whole spectral range, the prediction models based on the raw and preprocessed spectra of peeled samples were established by PLS regression, respectively and the performance of models were tested by independent validation set samples. The results were presented in table 2. For the model based on raw spectra, the R value of validation was higher than the one of calibration which indicated it was difficult to extract the information of protein from raw spectra directly. The models based on spectra pretreated by SNV and MSC functioned similar with the raw spectra model and the SD model had lowest R value of calibration and validation. The model established from spectra with FD preprocessing had best performance with 0.90, 0.80,0.29 and 0.40 as the R value of calibration, R value of validation, RMSEC and RMSEP, respectively. These results showed the first derivation was the optimized preprocessing method for protein measurement of peeled chestnut using near infrared spectroscopy.

**Table 2 Results of PLS models for protein content base on the spectra of peeled chestnuts**

| Preprocessing method | Calibration | | Validation | |
|---|---|---|---|---|
| | R | RMSEC | R | RMSEP |
| No preprocessing | 0.6140 | 0.54 | 0.7327 | 0.45 |
| First derivation | 0.9044 | 0.29 | 0.8029 | 0.40 |
| Second derivation | 0.5284 | 0.58 | 0.1658 | 0.70 |
| Standard normal variate transformation | 0.6594 | 0.51 | 0.7980 | 0.41 |
| Multiplicative scatter correction | 0.6537 | 0.51 | 0.7924 | 0.41 |

**Measuring model of intact sample**

**Table 3 Results of PLS models for protein content base on the spectra of intact chestnuts**

| Preprocessing method | Calibration | | Validation | |
|---|---|---|---|---|
| | R | RMSEC | R | RMSEP |
| No preprocessing | 0.3697 | 0.72 | 0.3827 | 0.54 |
| First derivation | 0.5455 | 0.64 | 0.6070 | 0.45 |
| Second derivation | 0.8748 | 0.35 | 0.7324 | 0.38 |
| Standard normal variate transformation | 0.5193 | 0.61 | 0.5095 | 0.47 |
| Multiplicative scatter correction | 0.5060 | 0.61 | 0.4813 | 0.47 |

The prediction model of intact sample were established in the same way and the results were generated in table 3. It was found that the information related to protein can be hardly acquired from raw spectra of intact chestnut because the R values of model were less than 0.40. The SNV and MSC model performed better than raw spectra model but the R values were not high enough for practical application. The model based on spectra pretreated by FD appeared over-fitting with a R value of validation higher than the one of calibration. The SD model functioned best with 0.87, 0.73, 0.35and 0.38 as the R value of calibration, R value of validation, RMSEC and RMSEP, respectively. For all the models for intact sample based on the same preprocessing method, the performances were worse than those for peeled sample, which indicated the influence of the nut peel. However, the spectra of intact sample still contained the information of protein in kernel which can be more purified in further study.

## CONCLUSION

In this work, the PLS regression method was applied in the measurement of protein content in chestnut using near infrared spectroscopy and the effect of four preprocessing methods to the model for both peeled and intact sample were compared. The results presented the feasibility of NIR spectroscopy in evaluating the protein content of intact and peeled chestnuts. The correlation coefficients of the optimized models for protein were 0.90 for peeled samples

and 0.87 for intact samples. These results indicated that NIR can work as a nondestructive and rapid means for measuring the protein content of chestnuts with reduced time and labor compared with traditional methods. Further work would addressed more robust and accurate models by intensifying spectra energy, selecting sensitive bands or combining chemometrics methods. The result also would be referenced for developing portable devices offering information for decisions on harvest, market or process, which will benefit the whole chestnut industry.

**Acknowledgements**
The authors really appreciate the financial support provided by Foundation of Ph.D Site in University (20090146110018) and Innovation Foundation of Huazhong Agricultural University (0900206021). The guidance of Prof. Li Peiwu's and the help of Liu Pei during the spectra acquisition are gratefully acknowledges.

## REFERENCES

[1] HS Gao; XD Chang, Storage and Process of Chestnut,1st Edition, Jin Dun Publishing House, Bei Jing, **2004**; 32-40.
[2] EM Pena-Mendez; M Hernandez-Suarez; C Diaz-Romero; E Rodríguez-Rodríguez, *Food Chem.*, **2008**, 107(1): 537-544.
[3] WZ Lu; HF Yuan; GT Xu. Modern Near Infrared Spectroscopy Analytical Technology, 2nd Edition, China Petrochemical Press, Bei Jing, **2000**,6.
[4] YT Liao; YX Fan; F Cheng, *Meat Sci.*,**2010**, 86(4), 901-907.
[5] C Lin; X Chen; L Jian; CH Shi; XL Jin; GP Zhang, *Food Chem.*, **2014**, 162(1), 10-15.
[6] P Ritthiruangdej; RR H Shinzawa; Y Ozaki, *Food Chem.*, **2011**, 129(2), 684-692.
[7] I Ruisánchez; FX Rius; S Maspoch; J Coello; T Azzouz; R Tauler; L Sarabia; MC Ortiz; JA Fernández; D Massart; A Puigdomènech; C García, *Chemometr Intell Lab.*, **2002**,63(2),93-105.
[8] ZY Niu; LJ Han; XO Shu; ZH Yang, *T. Chinese Soc Agr Mach.*, **2005**, 36(5),68-71.