Establishment and Optimization of Model for Detecting Epidermal Thickness in Newhall Navel Orange

Yande Liu, Yifan Li, Zhiyuan Gong

HAL Id: hal-01614178

https://inria.hal.science/hal-01614178

Submitted on 10 Oct 2017

Establishment and Optimization of Model for Detecting

Epidermal Thickness in Newhall navel orange

Yande Liu[*], Yifan Li , Zhiyuan Gong

(Institute of Electrical Machinery，East China Jiaotong University, 330013)

Abstract : Diffuse transmittance spectra in the near-infrared scope as a prevalent sensitivity method carried out to test epidermal thickness of 'Gannan' navel oranges. In order to lay a good foundation for accurateand rapid online classification, variable sel--ection methods was intervened for navel orange model optimization. In spectral range of 900~1650nm, navel orange in thick skin depth chosen arbitrarily were set up the qualitative models for both calibration and prognostication sets in this experiment. Firstly, different pretreatment methods such as the Savitzky-Golay, the first derivative and so on were compared by PLS modeling results. Then GA and SPA were brought in to improve predictive models. Compared with results, light scattering can be effectively eliminated by the standard normal variate transformation (SNV). Moreover, fewer variables and model optimization were carried out by GA. The supreme calibration model procured with GA-PLS approach had the Rp of 0.864, RMSEP of 0.290, R_C of 0.882 and RMSEC of 0.264. The experiment showed the detection of epidermal thickness of navel orange is completely feasible.

Keywords：Newhall navel orange; Near-infrared diffuse transmittance spectra ; Epidermal thickness; Variable selection

1.Introduction

Navel orange is a kind of comprehensive nutritional food, containing various essential nutrients in human body. A great source of vitamin C and carotenoids can be taken in through eating navel orange. The ripe navel orange is a popular fruit for its benefits, such as: seedless, juicy flesh, tasting good et al. When placed indoor, they can radiate enticing aroma and the rind's color and lustre is charming. [1] Near infrared spectrum detection technology has been used for a wide range of applications, especially in the quality test of thin skin fruit such as apple, peach ,pear et al. [2] In general, the average consumers take cortical thickness as one of the important considerations when they buy orange, water melon and other thick skin fruits. In the process of fruit grading, if the skin is too thick, the fruit can't be treated as optimal fruit. Even if the internal quality of fruit is excellent, the fruit also be degraded, as peel thickness directly influences its edible rate. Currently, the application of near infrared spectral detecting the thickness of the fruit skin also has not been reported. As a consequence, in this paper the navel orange peel thick testing research was discussed. Through a variety of spectral data processing method to establish the

optima navel orange skin thickness detection model of Newhall Navel Orange . The researchers at home and abroad processed a series of researches in near infrared spectrum technology area detecting internal quality of both thin skin fruit and thick skin fruit. [3-6]Taking the application of 3 different wavelength of LED lights to estimate SSC and size of Shuijing pears ，PLS and LS - SVM models were established. The model obtained by PLS was preferable than that of LS – SVM. The R of forecast set was respectively 0.86 and 0.90 for soluble solids content and size.(Liu et al.2010). [7] In order to implement modern management for pear orchards, the technology of NIR was applicated in the quality detection for Dangshan pear. PLS ，LS - SVM and GRNN model was established to portend the SSC of Dangshan pear. Then uninformative variables elimination (UVE) was introduced to simplify the former models. UVE-LSSVM had a great advantage in terms of regression model. (Wang et al.2013). [8] In the wavelength of 1200-2200nm, transmission spectrum of the mangos were collected for measurement during SSC and potential of hydrogen (PH) experiments. Infinite variety of pretreatment techniques was employed to process all wave spectral for the processing efforts. MLR based on PLS was applied to build calibration models. (Shyam et al, 2012). [9] In the wavelength of 400-1000nm, visible/near-infrared spectroscopy of valencia oranges were applied to conduct the SSC and TA experiments. The prediction models especially the model of the fruits taste best characteristic value stand (BrimA) on PLS and PCR are developed. Based on these results, visible/near-infrared spectroscopy technology is a promising and feasible method for detecting the BrimA of valencia oranges. (Jamshidi et al,2012). [10] Analyzing the above scholars' researches comprehensively, relevant testing models are established using fruits in regular shape、smooth surface and high internal quality. [11-18]

   The experiment take Newhall navel oranges which have thick wooden peel and poor uniformity of internal quality as objects of study. After different spectral preprocessing methods and the variable selection method, the raw navel orange spectrum data was to establish the optimal PLS prediction model. Compared with the origin model, it promotes the discriminative and forecasting ability.

2.Materials and methods

2.1 Materials

   One hundred twenty samples of Newhall were procured from a fruit wholesale marketplhce in NanChang (NanChang, China) in January, 2015. The samples were wiped up with distilled water and then naturally dried. All samples were be numbered consecutively and stored in the experimental environment at room temperature and 60% relative humidity (RH) for 24h. Spectral collection were carried out on the next day and SSC measurement performed soon thereafter. Newhall navel oranges were divided by proportion of 3:1in the light of calibration and prediction [19].

2.2 Spectral collection

Portable fruit quality detection device based on Android system consisted mainly of optical module, table computer with spectrum acquisition unit and power supply unit was set up. The optical module (Micro-NIR 1700, JDSU) kept an account of wavelength from 950nm to 2150 nm. The table computer with Atom Z2580 processor displayed the measurement results on a 9.7 inch capacitive touch screen. The JDSU Micro-NIR and table computer are workable under 5V model via 10000 mA   portable power source .

Reference spectral of a white Teflon tile and dark current were gathered before spectral acquisition of test samples. Diffuse reflection spectra of Newhall were gathered at about 20 °C. Acquisition spectrum evenly picked around the equator   equidistantly (approximately 120°). Optical gain had been set to ' low gain ' .
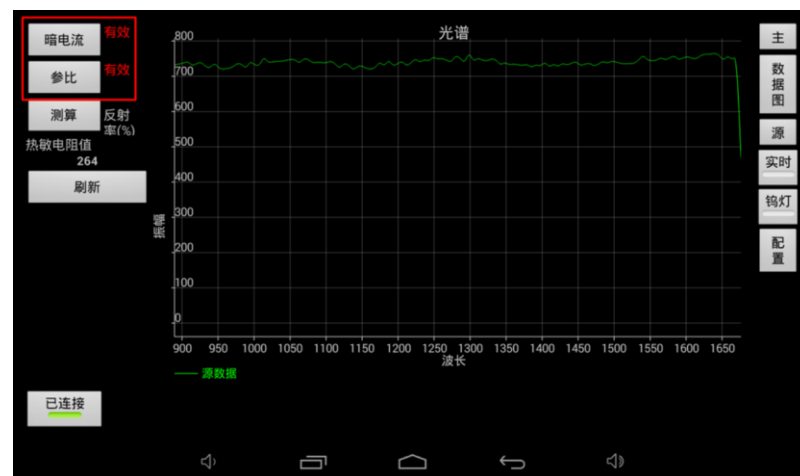

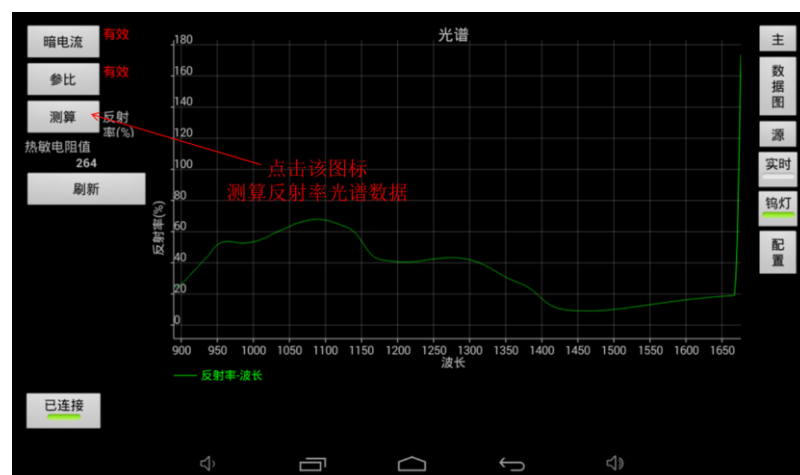Fig. 1   Collecting the effective reference and dark current

Fig. 2   Gathering fruit spectrum data

2.3 Measurement of soluble solid content and skin depth

In the spectral acquisition area, several drops of filter juice extracted by manual compression were used for soluble solid content measurement using the

PR-101αCat refracometer. Observed value of skin depth was conducted after soluble solid content measurement via vernier caliper immediately. When the flesh of the orange was removed from the peel clearly, the peel was to be hold pressure level off to measure the skin depth. The thickness value at three equidistant positions were recorded exactly during the experiment. The average value of the three marked points was taken as the thickness values of navel orange skin.

2.4 Data processing and model evaluation

All of the R、RMSEC and RMSEP together assess the performance of the model built in the experiment. The number of the best principle component achieved when coupled with the RMSEP which had reached the minimum value. [20-22]

In this paper, all the RMSEC and RMSEP were processed by the formula (1) and (2) respectively.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{I_c}(\hat{y}_i - y_i)^2}{I_c - f - 1}} \qquad [1]$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{I_P}(\hat{y}_i - y_i)^2}{I_P - 1}} \qquad [2]$$

Where y_i was the model's actual measured value; $\hat{y}_i$ was model's predictive value; f was used as dependent variables

3. Results and discussion

3.1 Analysis of Newhall navel orange

Fig.3 is the original absorbance spectra chart of 120 ripe Newhall navel orange. Fig.4 is the original absorbance spectra chart of 120 ripe Washington navel orange. It is obviously that two kinds of samples have similar spectrum shape and location of wave crest and trough. Because of the O - H, C - H or N - H stretching vibration[23-24] , there is a obvious absorption peak near the location of 970、1090、1220、1285、1470 nm.
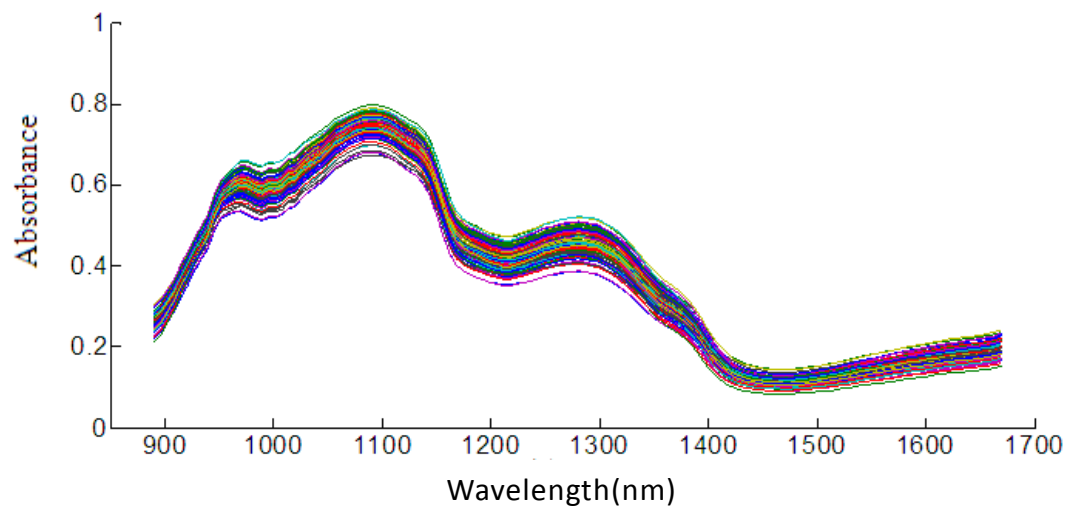
Fig. 3　The original spectra of ripe Newhall navel orange

3.2 Measurement results of soluble solid content and skin depth

The samples of Newhall navel orange had certain representativeness as the sizes of them contain the dimension from small size to large size. Form the table1. available, the skin thickness range of calibration set is 1.98~5.57mm meanwhile the prediction set is 2.63~5.43mm. The data structure make better fit for prediction set to predict using the model calibration set made.

Table 1. Sugar measurement results of Newhall navel orange

Sample sets	Min（°Brix）	Max（°Brix）	Mea（°Brix）	SD （°Brix）	CV（%）
Calibration set	9.70	18.83	14.37	1.97	13.68
Prediction set	10.50	18.63	14.43	1.99	13.81

Table 2. Skin thickness measurement results of Newhall navel orange

Sample sets	Min（°Brix）	Max（°Brix）	Mea（°Brix）	SD （°Brix）	CV（%）
Calibration set	1.98	5.57	3.72	0.57	15.42
Prediction set	2.63	5.43	3.76	0.58	15.58

3.3 Comparison of the results of   spectral preprocessing

Table 3 illustrates that the model PLS established through base line、1st

derivatives 、SNV and MSC improve the model prediction compared with the model built by raw spectral data. The prediction abilities of SNV also with MSC after pretreatment, while SNV had the better result. Take Newhall Navel Orange as research objects to discuss the outcome. The result showed that the Rc reached 0.883 and of prediction set reached 0.815. It is better to see that RMSEC was 0.268 and of PMSEP was 0.333. Therefore, the process of spectra band screening in the further adopted the spectra data preprocessed by SNV.

Table 3   The effect of different pretreatment methods on PLS modeling results for Newhall

preprocessing	calibration set		prediction set	
method	R_C	RMSEC	R_P	RMSEP
Raw	0.881	0.270	0.804	0.343
S-G	0.880	0.271	0.803	0.344
1st derivatives	0.880	0.271	0.808	0.339
Base line	0.880	0.271	0.812	0.337
SNV	**0.883**	**0.268**	**0.815**	**0.333**
MSC	0.882	0.269	0.815	0.333

3.4   Method of spectral band selection

To test and attest the practicability of near infrared diffuse reflection technology to inspect Newhall Navel Orange. In the hope of providing the theoretical   basis for online   detection   through   setting   up   a   corresponding   optimal   prediction model   simultaneously. As a consequence, testing requires not only high accuracy, but also higher detection efficiency. Spectral band selection methods are   needed so as to acquire fewer variables and good prediction ability. Genetic algorithm is an algorithm proposed on basis of imitating the preferred choice of biology and genetic principle. With the advantages of   high efficiency of global searching, keeping the optimal variables, eliminating worse variables, establishing model more conveniently and predicting more accurately. According to the reference[25-27], in the process of GA variable screening of 127 wavelength points, the main parameters are set as follows. Initial value of the group was 30; Adaptive crossover probability and mutation probabilities was set to 0.5 and 0.01 separately; The RMSECV was regarded as fitness function of genetic algorithm. After the iteration reached 100 times, figure 4

illustrated that the minimum value corresponding to the RMSECV was 0.292 when the variable number was 13. For this reason, the 13 variables was the characteristics screened out by GA method. The processed result can be seen from figure 3. The variable frequency was more than 11 times which were chosen as characteristic variable of the model in all 127 variables.
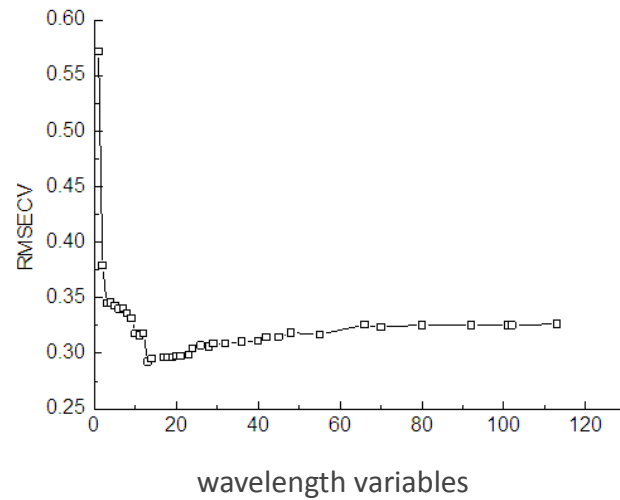

Fig.4   Relation between RMSECV and wavelength variables for epidermal thickness of Newhall navel orange
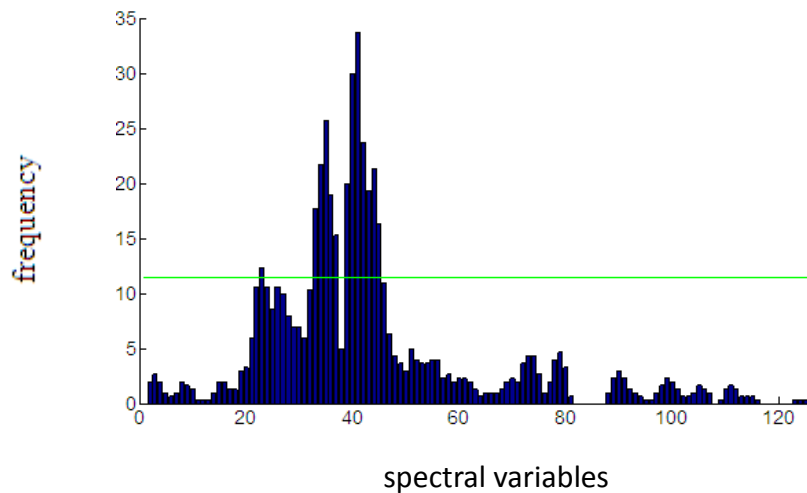

Fig.5 Frequency of spectral variables selection runed by GA

Successive projection algorithm randomly select a variable from spectrum matrix and calculate projection of the variable on the other variables afterwards. This method is based on the minimum RMSECV generally choose the characteristic variables within 20. Calculation procedure and slection of numbers of variables of SPA consult to relevant literature[28-30]. The leading eigenvalue is set to 20 and the least eigenvalue is 1. Figure 6 was the diagram of characteristic variables of screened

out by applying SPA for epidermal thickness of Newhall Navel Orange. There were 6 variables filtered out from the total variables.
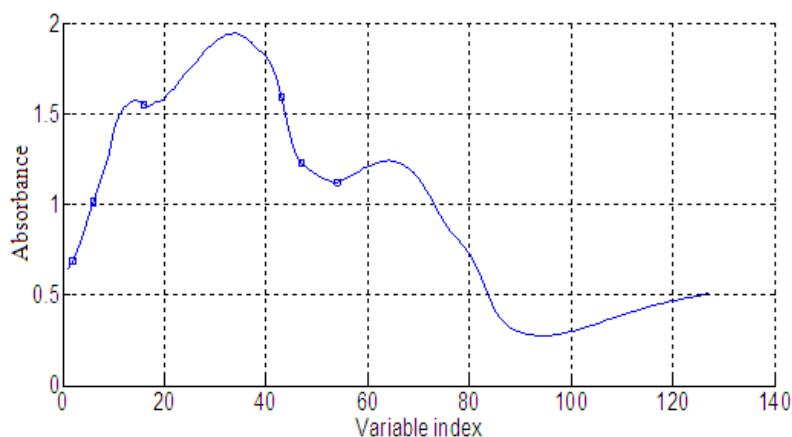

Fig.6   Variables selected by SPA for epidermal thickness of   Newhall   navel orange

3.5 Comparison of models

Table 4 provides a comparison of forecasting results of different forecasting models made by spectral data   processed through method of GA   and SPA which were the method of artificial screening in the multi-variable analysis and full spectral of samples. For the PLS model of Newhall navel orange skin thick in mature stage, research findings shows that GA-PLS models outperformed the SPA-PLS models and full wave band models. SPA-PLS models had the least spectral variables but the ability to predict relatively weakened. The reason may be that some useful variables are eliminated, so that the model precision is lower. It can be observed that GA-PLS models made the variables from 127 down to 13 and improved Rp from 0.815 to 0.815 at the same time. It's very gratifying to see that RMSEP reduced from 0.333 to 0.290. So from the experimental results , GA-PLS models increased the optimization efficiency of   spectral models and produced the best Rp and RMSEP for calibration models. Performance of 30 navel oranges peel thickness in calibration set   had   been   presented   in   figure   7.   Finally practical   data are compared to predicted ones to assess the quality of calibration models[31-30].

Table 4 Modeling results and comparison of different PLS

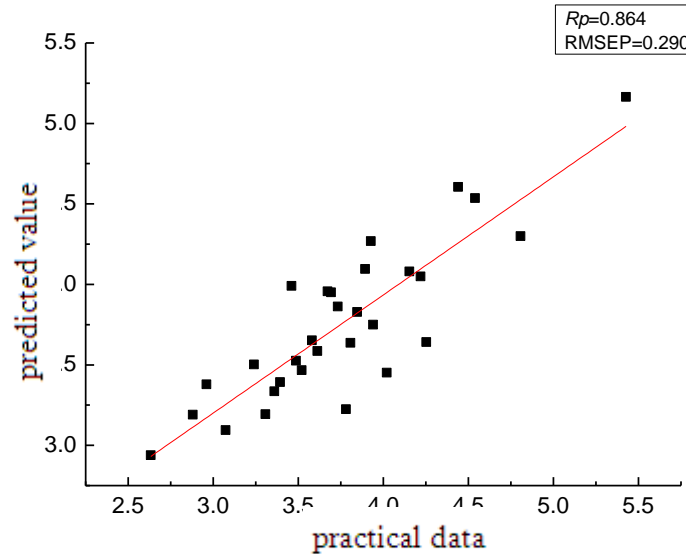| model	variables	R_C	RMSEC	R_P	RMSEP
PLS	127	0.883	0.268	0.815	0.333
GA-PLS	**13**	**0.882**	**0.269**	**0.864**	**0.290**
SPA-PLS	6	0.859	0.293	0.811	0.339

Fig. 7   Comparison of predicted values and measured values of 'Gannan' navel orange in prediction set by GA-PLS mode

4. Conclusion

This study aimed to establish prediction model to detect peel thickness of Newhall navel orange accurately. To make a thorough acquire for the means of simplifying and improving model, comparison of spectral preprocessing methods and the band selection method of GA and SPA were introduced. The predictions suggested that the best model with Rc 、 Rp 、 RMSEP and RMSEC of 0.882 、0.864、0.290 and 0.296 were established through spectral data processed by SNV and spectral variables screened out by GA. Based on these results, SNV was a effective way to greatly reduce the effects of light scattering. Instead of using full spectral data, GA models simplified models and improved the ability to predict with decreasing the extent of the spectral variables .

ACKNOWLEDGMENT

[Reference]

[1] Sun X D, Zhang H L, Liu Y D. International Journal of Agricultural and Biological Engineering, 2009, 2(1): 65.

[2] Jie D F, Xie L J, Rao X Q, et al. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(12): 264.

[3] Sun T, Lin J L, XU W L. Journal of Jiangsu University: Natural Science

Edition, 2013, 34(6):663.

[4] Ma G, Sun T. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(7):170.

[5] Liu Y D, Shi Y, Cai L J, et al. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(9):138.

[6] Cai L J, Liu Y D, Wan C L . Journal of Northwest A&F University: Natural Science Edition, 2012, 40(1):215.

[7] Liu Y D, Peng Y Y, Gao R J, et al. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(11): 338.

[8] M H Wang , W C Guo, Liu H, et al. Journal of Northwest A & F University, : Natural Science Edition, 2013, 41(12):113-119.

[9] Shyam N J, Pranita J. Scientia Honiculturae, 2012, 138: 171-175.

[10] Bahareh J, Saeid M, Ezzedin M, et al. Computers and Electronics in Agriculture, 2012, 85: 64-69.

[11] Greensill, Walsh. Journal of Near Infrared Spectroscopy, 2002,10,27-35.

[12] McGlone , Fraser, et al. Journal of Near Infrared Spectroscopy, 2003,11,323-332.

[13] Saranwong, Sornsrivichai, et al. Journal of Near Infrared Spectroscopy, 2003,11,175-181.

[14] Walsh, Golic, et al. Journal of Near Infrared Spectroscopy, 2003, 12, 141-148.

[15] Gomez , He Y, Pereira A G. Journal of Near Infrared Spectroscopy, 2006, 77, 313-319.

[16] Nicolaï, Beullens, et al. Postharvest Biology and Technology, 2007, 46, 99-118.

[17] Hashimoto et al. US : 6754600B2, 2004 06 22.

[18] Y D Liu, X D Sun, et al. LWT-Food Sci-Technol, 2010, 43: 602-607.

[19] Li H, Wang J X, Xing Z N, et al. Spectroscopy and Spectral Analysis, 2011, 31(2):362-365.

[20] Han D H, Chang D, Song S H, et al. Transactions of the Chinese society for agricultural machinery, 2013, 44(7):174.

[21] Barbin, ElMasry. Food Chemistry .138(2-3), 1162-1171 (2013).

[22] He H. J., Wu D., et al. Food Chemistry .138(2-3), 1162-1171 (2013).

[23] Bahareh J, Saeid M, Ezzedin M, et al. Computers and Electronics in Agriculture. 2012, 85: 64.

[24] Sun X D, Dong X L. Spectroscopy and Spectral Analysis, 2013, 29(14):262.

[25] Goicoechea, Olivieri. Chemometrics and Intelligent Laboratory System. 56(2):73-81 (2001).

[26] W X Zhu, H Jiang, et al. Transactions of the Chinese society for agricultural machinery, 2010, 41(10):129.

[27] S J Zhang, H H Zhang, et al. Transactions of the Chinese society for agricultural machinery, 2012, 43(3):108.

[28] Kamruzzaman, ElMasry, et al. Food Chemistry. 2013, 141(1): 389-396.

[29] Wu D., Sun D. W., et al. Innovative Food Science & Emerging Technologies. 2012,16: 361-372.

[30] Paz, S ánchez, et al. Computers and Electronics in Agriculture, 2009, 69:24-32.

[31] Y H Yun , W T Wang, et al. Analytica Chimica Acta, 2014,10-17.

[32] Y D Liu, R J Gao, et al. Food Bioprocess Technol, 2012, 5: 1106-1112.

[33] Piyamart   J, Yoshinori K, Sumio K. Journal of   Near Infrared Spectroscopy, 2014,22,367-373.

[34] H Jiang, Zhu W X. Food Anal. Methods, 2013,6: 567-577.

[35] Q S Chen , J W Zhao, et al. Analytica Chimica Acta, 2006, 572 : 77-84.