Journal of Shanghai Jiaotong University (Science), Vol. E-12, No. 5, 2007, 597~601
Article ID: 1007-1172(2007)05-0597-05

NIR Hyperspectral Imaging Measurement of Sugar Content in Peach Using PLS Regression

GUO Feng[1*] (郭 峰), *CAO Qi-xin*[1] (曹其新), *Nagata Masteru*[2], *Jasper Tallada*[2]
(1. Research Institute of Robotics, Shanghai Jiaotong Univ., Shanghai 200240, China;
2. Faculty of Agriculture, Miyazaki Univ., Miyazaki 8892192, Japan)

Abstract: Near infrared (NIR) hyperspectral imaging measurement of sugar content in peach was introduced. NIR spectral images (650~1 000 nm, resolution: 2 nm) of peach samples were captured with developed hyperspectral imaging setup. Partial least square (PLS) regression prediction model was developed to estimate the sugar content in peach; step-wise backward method was utilized to determine optimal wavelength subsets. Experimental results show that the calibration model with optimal wavelength subsets has a correlation coefficient of prediction of 0.97 and a standard error of prediction of 0.19, the prediction accuracy is higher than the calibration model applied over the whole wavelength, which proves that variable selection plays an important role in improving the prediction accuracy of PLS regression model.

Key words: near infrared hyperspectral imaging system; sugar content; partial least square regression
CLC number: TP 242.6 **Document code**: A

Introduction

To satisfy consumers' healthy and safe requirement of agricultural product, internal attributes of fruit: sugar content, acidity, etc., are taken into account in quality judgment process. Traditional chemical internal quality inspection method is slow and complex, and cannot meet the non-destructive requirement for modern fruit inspection. As a non-destructive technique, near infrared (NIR) spectral inspection system causes no damage to tested samples and can analyze multiple internal attributes of fruit simultaneously[1,2].

According to inspection mode, NIR spectral inspection can be divided into reflectance, transmittance and interactance modes[3]. Reflectance mode gets wide application because of its simple setup. According to the number of utilized wavelengths, the inspection system can be divided into multi-spectral device and hyperspectral device. The number of wavelength selected in multi-spectral system is limited, and the NIR spectrum can be acquired by installing specific filters on the lens, which is cheap and easy for realization. Hperspectral device captures spectrum of hundreds of wavelengths and can achieve high prediction accuracy, while the device cost is relatively high. With the decrease of hardware cost and improvement of NIR sensor performance, NIR reflectance hyperspectral imaging system represents the direction of fruit internal quality inspection study in the future. Compared with the spot detection mode of spectroscopy inspection, spectral imaging system can inspect the most part of fruit surface and will provide more complete internal attribute information.

The study of NIR hyperspectral imaging measurement of fruit's internal quality attributes started in 1990s, Lu[4] detected bruises on apples in the spectral region between 900 and 1 700 nm. Polder *et al*[5] analyzed spectral images of five ripeness stages of tomatoes and showed that the spectral images offer more discriminating power than standard RGB images for measuring ripeness. Kim *et al*[6] developed a laboratory-based hyperspectral imaging system (the spectral range is from 430 to 930 nm with spectral resolution of 10 nm), which

Received date: 2006-09-15
* **E-mail**: guofeng@sjtu.edu.cn

was used to conduct food safety and quality research. Mehl *et al*[7] analyzed hyperspectral spectrum to determine three spectral bands capable of separating normal from contaminated apples. These spectral bands were implemented in a multispectral imaging system with specific band pass filters to detect apple contaminations. In China, similar study started in 2002, Liu and Ying[8-10] had developed some NIR spectral systems for predicting sugar content and valid acidity of apples and peaches. Ying *et al*[11] had also evaluated the potential of a Fourier transform near-infrared (FT-NIR) spectrometer with a reflectance fiber optic probe for determination of soluble solids content (SSC) and available acid (VA) in intact peaches. Experimental result proved that FT-NIR could be an easy to facilitate, reliable, accurate, and fast method for non-invasive measurement of peach SSC and VA[11].

The goal of this paper is to present a NIR hyperspectral imaging measurement for sugar content inspection in peach. NIR reflectance spectral images of sample were captured firstly; then partial least square (PLS) regression calibration model was developed, step-wise backward method was utilized to decide the optimal wavelength subsets. After deleting some uncorrelated independent variable subsets, the PLS regression prediction model could achieve high accuracy.

1 Material and Methods

1.1 Peach Sample Preparation

29 peaches were bought from supermarket in Miyazaki Japan, and were stored at 18～20 ℃. The average mass of sample was 265.2 g. Both checks of every peach were divided into four sample grids, so the number of sample grids was 4×2×29=232. 144 grids were designated as calibration set for model development and the remaining 48 grids were designated as prediction set for prediction validation. The grid was marked with black ink on the surface of fruit and its size was 15 mm×15 mm. The spectral image acquirement and sugar content analysis were finished in one day.

1.2 Hyperspectral Imaging Setup

Fruit reflectance images were obtained with hyperspectral imaging setup (in Fig. 1), which was developed by Prof. Nagata (Faculty of Agriculture, Miyazaki University, Japan)[12].

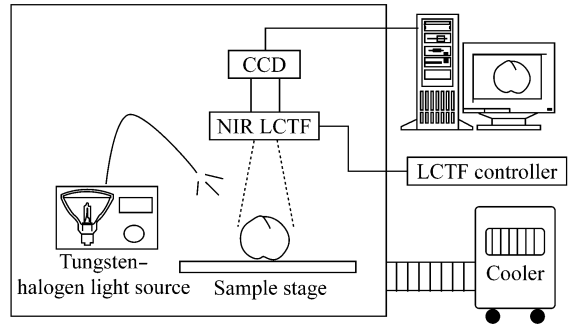

Fig. 1 Diagram of NIR hyperspectral imaging setup

The hyperspectral imaging setup consisted of a CCD camera, liquid crystal tunable filter (LCTF), tungsten-halogen quartz light source and cooler. The camera was a 14-bit Apogee AP2E (Apogee Instruments, Inc., Auburn, Cal.) monochrome camera that had a Nikon *f*/1.2 optical lens. A CRI (Cambridge Research and Instrumentation, Massachusetts, USA) LCTF (VIS-NIR Model (650～1 000 nm, 10 nm bandpass)) was attached on the lens and controlled through specific controller box. The Dolan-Jenner Fiber-Lite PL950 DC (Dolan-Jenner Industries, St. Lawrence, Mass) regulated illuminator had a 150 W 21 V EKE tungsten-halogen quartz light bulb with the IR cut-off filter removed. The fiber optic light guide was placed vertically above the platform. The cooler was used to maintain a relatively uniform temperature within the setup.

Using an in-house Visual Basic for application software to control the Maxim/DL and National Instruments LabView Cris3 programs, spectral images of a set of fruits were automatically acquired from 650 to 1 000 nm at 2 nm interval (176 images). Using the same settings, reference images of a 125 mm×125 mm Spectralon reference panel (99% diffuse reflectance) and dark current images were also acquired with the camera shutter closed.

1.3 Soluble Sugar Content Inspection

After spectral imaging, the samples were taken for soluble solids content (SSC) measurement. 8 sample grids were cut out from every fruit, a 5～

8 mm thickness layer under grid surface was sliced and hand-squeezed to produce juice. The SSC of fruit juice was measured using a digital refractometer Brixmeter RA-410 (Kyoto Electronics Manufacturing Co. Ltd., Japan) and was expressed by Brix%.

2 Data Processing

The hyperspectral fruit images were processed with image processing toolbox, Matlab 6.5.1. Threshold method was used to segment fruit from the 650 nm image, which was designated as mask image for subsequent process. Flat field correction (in Fig. 2) was applied using

$$I_{\text{norm}}(x,y) = \frac{I_{\text{sample}}(x,y) - I_{\text{dark}}(x,y)}{I_{\text{Ref}}(x,y) - I_{\text{dark}}(x,y)} m \quad (1)$$

where $I_{\text{norm}}(x,y)$ is the normalized image pixel value at pixel location (x, y), $I_{\text{sample}}(x,y)$ the sample image pixel value, $I_{\text{Ref}}(x,y)$ the reference image pixel value, $I_{\text{dark}}(x,y)$ the dark frame image pixel value, m the factor assumed equal to 1.0 for white spectral panel.

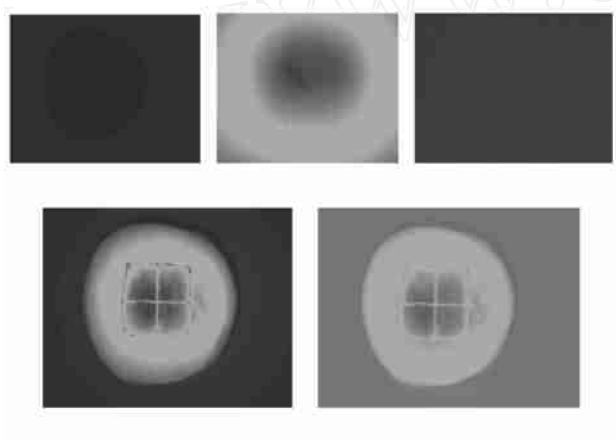

Fig. 2 Flat field correction process of peach image
(a) Mask image (650 nm); (b) reference image (700 nm); (c) dark frame image (700 nm); (d) original sample image (700 nm); (e) normalized sample image (700 nm)

During the image acquirement process, the LCTF changed the bandpass of wavelength from 650 to 1 000 nm with 2 nm resolution for every check of peach. Under every specific wavelength, related original sample image, reference image and dark frame image were captured, then corrected image was built with Eq. (1), the mean value of normalized reflectance was computed in every grid on the peach surface. To every check of peach, four NIR spectra profiles which expressed the relation between wavelength and normalized relative reflectance were generated. The spectra had strong absorptions at around 670 and 980 nm, which were caused by the chlorophyll pigment and water, the value of these two ingredients had high correlation with the ripeness of fruit (in Fig. 3).

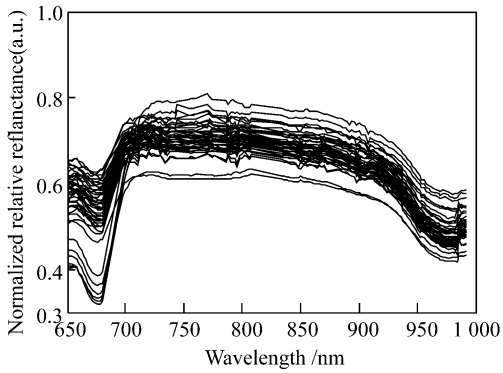

Fig. 3 NIR spectral profile of peach

The NIR spectral data were used for the sugar content prediction model development. To the $N \times K$ matrix of NIR spectral data and $N \times 1$ vector of SSC data (N is the number of samples and K is the number of wavelengths), PLS regression method, which could maximize the covariance between score vector in X-space and a score vector in Y-space, was thought to be effective. A PLS regression analysis was carried out to develop calibration models using Matlab 6.5.1.

3 Subset Selection Method in PLS Regression

In the development of regression model, variable selection method had direct influence on prediction accuracy, Hoskuldsson[13] had pointed out the importance of variable and subset selection in PLS regression. In this paper, step-wise backward method was selected to obtain best wavelengths subsets. Calibration models were compared using the statistics standard error of prediction (SEP) corrected for bias based on

$$\text{SEP} = \sqrt{\sum_{i=1}^{N_p} (Y_i - \hat{Y}_i - \text{bias})^2 / (N_p - 1)} \quad (2)$$

$$\text{bias} = \left(\sum_{i=1}^{N_p} Y_i - \sum_{i=1}^{N_p} \hat{Y}_i \right) / N_p \quad (3)$$

where Y_i is the measured value of sample i, $\hat{Y}_i$ the predicted value of sample i, N_p the number of samples used in the prediction set.

In order to improve the analysis effectiveness and reduce the influence of system error, 176 wavelength variables were divided into 44 variable subsets, each subset contained four wave length variables: {[650 nm, 652 nm, 654 nm, 658 nm], [660 nm, 662 nm, 664 nm, 666 nm], …, [994 nm, 996 nm, 998 nm, 1 000 nm]}.

The procedure of this method was listed below.

(1) PLS regression was applied over current NIR spectrum. Define SEP_O the original SEP calculated from the current NIR spectrum before any subset was deleted from it, the SEP_O of the current prediction model was calculated with Eq. (2).

(2) Single subset was deleted sequentially from current spectrum with exhaustive searching method, PLS regression was applied over the left NIR spectrum, related SEP_i was calculated, i= 1, 2, …, n- 1, n is the number of variable subsets in current spectrum.

(3) Found the minimum SEP_i, compared the SEP_{min} with SEP_o.

(4) If SEP_{min}> SEP_o, current spectrum was decided as best spectrum; else, the subset corresponding with the SEP_{min} would be deleted from current spectrum, SEP_o was replaced with SEP_{min}, went to step (1) and began another new loop.

4 Results and Discussion

PLS regression was applied over the whole NIR spectrum and best NIR spectrum separately to compare the prediction accuracy of sugar content.

PLS regression was applied over the whole NIR spectrum, the main factor of PLS regression model is 12. The whole wavelength calibration model had a correlation coefficient of prediction (R_p) and SEP. The calibration plot for the whole wavelength model was shown in Fig. 4.

Stepwise backward method was utilized to decide the best wavelength subsets, subsets selection result was shown in Tab. 1.

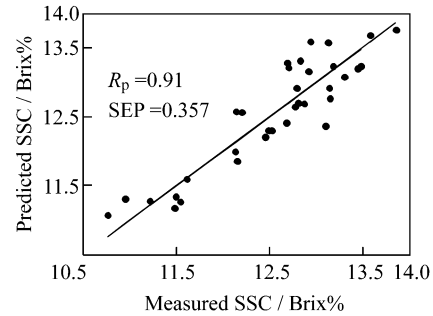

Fig. 4 Calibration plot for the PLS regression model (whole wavelength variables)

Tab. 1 Optimal wavelength subsets selection process and result

Step	Deleted subsets/nm	SEP	R_p
1	746~ 752	0.330 7	0.912 4
2	682~ 688	0.318 1	0.921 2
3	762~ 768	0.304 7	0.926 2
4	994~ 1 000	0.293 0	0.930 7
5	874~ 880	0.286 8	0.934 2
6	690~ 696	0.243 9	0.944 7
7	810~ 816	0.228 5	0.955 9
8	826~ 832	0.224 4	0.958 7
9	778~ 784	0.194 1	0.966 2
10	818~ 824	0.188 2	0.970 1
11	658~ 664	0.196 2	0.968 5

As shown in Tab. 1, the value of SEP reached minimum after 10 steps, PLS regression model achieved the highest accuracy in corresponding spectra. Any deletion of subset from the best wavelength variables would decrease the prediction accuracy.

PLS regression was applied over the optimal NIR spectrum. The optimal wavelength calibration model had a R_p of 0.97 and a SEP of 0.19, and the calibration plot for the optimal wavelength model was shown in Fig. 5. Figures 4 and 5 had shown the high prediction ability for PLS regression method, and the R_p was 0.9 above. During the optimal wavelength determination process, 2 subsets on 650~ 700 nm [682~ 688 nm, 690~ 696 nm], 3 subsets on 700~ 800 nm [746~ 752 nm, 762~ 768 nm, 778~ 784 nm], 4 subsets on 800~ 900 nm [810~ 816 nm, 818~ 824 nm, 826~ 832 nm, 874~ 880 nm], and 1 subset on 900~ 1 000 nm [994~

1 000 nm] were deleted from the original wavelength subsets. The subsets included 670 and 980 nm, which had high correlation with sugar content, were kept during optimization process. The prediction value shown in Fig. 5 showed better colinearity with measurement value than that in Fig. 4, which meant that the prediction value generated on the optimal wavelength subsets had higher calibration correlation than that generated on the original wavelength subsets. The calibration model built from the optimal wavelength subsets had higher prediction accuracy than the calibration model applied over whole wavelength, which proved that variable selection played an important role in improving the prediction accuracy of PLS regression model.

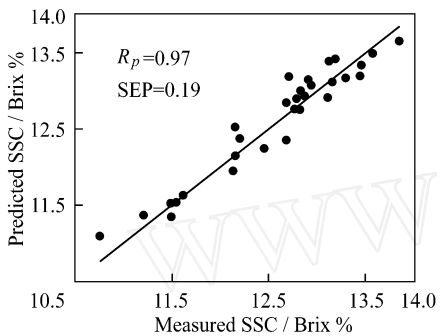

Fig. 5 Calibration plot for the PLS regression model (optimal wavelength variables)

5 Conclusion

NIR spectral technique for non-destructive fruit internal quality inspection had large market potential. With the developed hyperspectral imaging setup, sugar content in peach was analyzed and PLS regression calibration model was developed. The influence of wavelength selection on model accuracy was analyzed. The PLS regression calibration model over optimal wavelength subsets had high accuracy in sugar content prediction.

References

[1] Fu Xing-hu, Fu Guang-wei, Bi Wei-hong. Research of application of NIRS technology in nondestructive detection of quality of fruits[J]. *Infrared*, 2006, **27**(1): 33- 37(in Chinese)

[2] Xu Hui-rong, Ying Yi-bin. Application and prospect of near infrared imaging and spectroscopy analysis on quality inspecting of agricultural products[J]. *Journal of Zhejiang University (Agric & Life Sci)*, 2002, **28**(4): 460- 464(in Chinese)

[3] Abboot J A, Lu R, Upchurch B L, *et al*. Technologies for nondestructive quality evaluation of fruits and vegetables[J]. *Horticultural Reviews*, 1997, **20**(1): 1- 120.

[4] Lu R. Detection of bruises on apples using near-infrared hyperspectral imaging[J]. *Transactions of the ASAE*, 2003, **46**(2): 523- 530.

[5] Polder G, van der Heijeden G W A M, Young I T. Spectral image analysis for measuring ripeness of tomatoes[J]. *Transactions of the ASAE*, 2002, **45**(4): 1155- 1161.

[6] Kim M S, Chen Y R, Mehl P M. Hyperspectral reflectance and fluorescence imaging system for food quality and safety[J]. *Transactions of the ASAE*, 2002, **44**(3): 721- 729.

[7] Mehl P M, Chao K, Kim M, *et al*. Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis[J]. *Applied Engineering in Agriculture*, 2002, **18**(2): 219- 226.

[8] Liu Yan-de, Ying Yi-bin, Fu Xia-ping. Study on predicting sugar content and valid acidity of apples by near infrared diffuse reflectance technique[J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(11): 1793 - 1796(in Chinese)

[9] Liu Yan-de, Ying Yi-bin, Fu Xia-ping, *et al*. Automatic measurement system of fruit internal quality using near-infrared spectroscopy[J]. *Journal of Zhejiang University (Engineering Science)*, 2006, **40**(1): 53- 56(in Chinese)

[10] Liu Yan-de, Ying Yi-bin. Determination of sugar content and valid acidity in honwy peach by infrared reflectance[J]. *ACTA NUTRIMENTA SINICA*, 2004, **26**(5): 400- 402(in Chinese)

[11] Ying Y B, Liu Y D, Wang J P, Fourier transform Near-Infrared determination of total soluble solids and available acid in intact peaches[J]. *Transactions of the ASAE*, 2005, **48**(1): 229- 234.

[12] Nagata M, Tallada J, Kobayashi T, *et al*. NIR hyperspectral imaging for measurement of internal quality in strawberries[C]//ASAE. *Proceedings of ASAE Annual International Meeting, Tampa*. St Joseph M I USA: ASAE Publisher, 2005: Paper Number: 053131 [CDROM].

[13] Hoskuldsson A. Variable and subset selection in PLS regression[J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, **55**(1/2): 23- 38.