Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Prediction of valid acidity in intact apples with Fourier transform near infrared spectroscopy*

LIU Yan-de (刘燕德)[1,2], YING Yi-bin (应义斌)[†1], FU Xia-ping (傅霞萍)[1]

([1]*School of Biosystems Engineering & Food Science, Zhejiang University, Hangzhou 310029, China*)

([2]*College of Engineering, Jiangxi Agriculture University, Nanchang 330045, China*)

[†]E-mail: ybying@zju.edu.cn

Received June 1, 2004; revision accepted Nov. 4, 2004

Abstract: To develop nondestructive acidity prediction for intact Fuji apples, the potential of Fourier transform near infrared (FT-NIR) method with fiber optics in interactance mode was investigated. Interactance in the 800 nm to 2619 nm region was measured for intact apples, harvested from early to late maturity stages. Spectral data were analyzed by two multivariate calibration techniques including partial least squares (PLS) and principal component regression (PCR) methods. A total of 120 Fuji apples were tested and 80 of them were used to form a calibration data set. The influences of different data preprocessing and spectra treatments were also quantified. Calibration models based on smoothing spectra were slightly worse than that based on derivative spectra, and the best result was obtained when the segment length was 5 nm and the gap size was 10 points. Depending on data preprocessing and PLS method, the best prediction model yielded correlation coefficient of determination (r^2) of 0.759, low root mean square error of prediction (RMSEP) of 0.0677, low root mean square error of calibration (RMSEC) of 0.0562. The results indicated the feasibility of FT-NIR spectral analysis for predicting apple valid acidity in a nondestructive way.

Key words: Apples, Nondestructive prediction, FT-NIR, Valid acidity, Multivariate analysis
doi:10.1631/jzus.2005.B0158 **Document code:** A **CLC number:** S132

INTRODUCTION

Consumers' acceptance of fresh or processed apples is the ultimate goal of apple breeders, food scientists and supermarket managers. Internal quality assessment has focused on two major objectives: removal of fruit with internal defects and taste selection. Three major parameters including sugar content, acidity and firmness have to be taken into account to determine the internal quality and the taste of an apple. Near infrared spectroscopy has been used to measure several properties in a wide range of products, e.g. soluble solids in apple juice (Ventura *et al*., 1998) and in peaches (Peiris *et al*., 1997), dry matter in onions (Birth *et al*., 1985), potatoes (Dull *et al*., 1989) and whole dates (Dull *et al*., 1991), internal quality in peaches, nectarines (Slaughter, 1995), raisins (Huxsoll *et al*., 1995), intact apples (Moons *et al*., 1997; Lammertyn *et al*., 1998; 2000; Peiris *et al*., 1997; Lu and Ariana, 2002). All of the above researches were based on NIR spectroscopic instrumentation. However, there is little reported in literature about FT-NIR method used for determining the interior quality of intact fruit. Peirs *et al*.(2002) compared Fourier transform (FT-NIR) spectroscopy with dispersive near infrared spectroscopy, and the instrument stability, the light penetration depth and the predictive capacity of some quality characteristics when both instruments were compared. Based on the results, they concluded that FT-NIR reflectance spectroscopy is an interesting alternative for standard dispersive instruments for non-destructive quality evaluation. Some researchers (Belton *et al*., 1995; Luis *et al*.,

* Projects (Nos. 30370371 and 60468002) supported by the National Natural Science Foundation of China

2001) used FT-NIR transmission to measure the main components of fruit juices.

The overall objective of this research is to examine the possibility of the FT-NIR method for predicting the valid acidity in intact apples. The research focuses on the evaluating the prediction performance of calibration models with different data preprocessing. The specific objectives of the research are to:

1. Examine the potential of the FT-NIR spectroscopy as a nondestructive method to measure the valid acidity of apple.

2. Investigate the influence of different data preprocessing on the prediction performance.

3. Determine the optimal wavelength range for developing the calibration model.

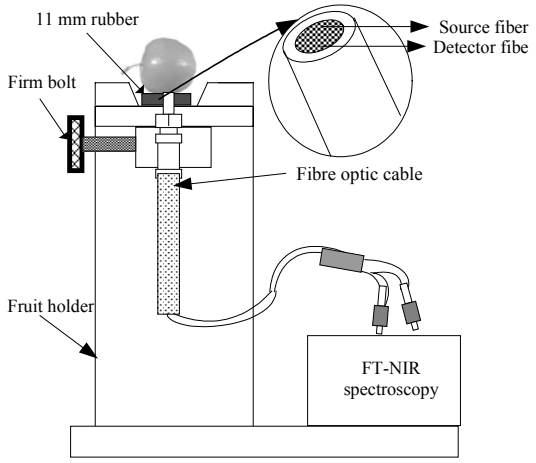

Fig.1 Schematic diagram of the setup for FT-NIR measurement of apple fruit in the wavelength of 800–2619 nm

MATERIALS AND METHODS

Fruit

The 120 apples Shandong Fuji used in the experiment were purchased at a local auction and stored for 2 d at 25 °C and 68% relative humidity to equilibrate. Eighty apples were used for the calibration models, and the remaining 40 samples were used for prediction models. From each apple, four spectra were taken at the equator position with bifurcated optical configuration. The valid acidity was measured at the same position with a pH meter (SJ-4A, Exact instrument Co., Ltd., Shanghai, China). Each apple's valid acidity was the mean value of four measurements made by extracting juice from the four positions for the spectra acquisition.

FT-NIR measurements

A commercial spectrometer (Nexus FT-NIR, Nicolet, USA) was used for this research. This device was equipped with an interferometer, an InGaAs detector, and a wide band light source (50 W) quartz halogen to provide reflectance measurements. A bifurcated optical fiber and a fruit holder as the main accessories were used for the experiment. Apples were placed steadily on the fruit holder, with the stem-calyx axis horizontal. For each apple, 4 diffuse reflectance spectra were taken at four equidistant positions along the equator. In the head of the bifurcated cable, the source and detector fibers were situated randomly (Fig.1).

The light was guided to the sample by source fiber, and from the sample with the detector fibers to Nexus FT-NIR spectrometer, with spectral range of 800–2619 nm and the interferograms (64) were co-added followed by strong Beer-Norton apodization. The mirror velocity was 0.9494 cm/s and the resolution was 16 cm^{-1}. In order to avoid surface reflectance and ensure subsurface penetration of the light into the apple flesh, the bifurcated optical probe was placed at a 75° angle to the level (Liu and Ying, 2003). All spectra were first converted into relative reflectance by dividing the measured intensity from a standard reference spectrum (obtained by placing a 22 mm diameter, 19 mm high cylindrical Teflon block directly above the fruit holder) was measured after every ten fruits during the course of the experiments. Fig.1 shows FT-NIR measurement scheme. In order to obtain enough sensitivity to measure the diffuse reflectance of the intact fruit, He-Ne laser emitting monochromatic light at a specific wavelength was used to adjust the scale. Each spectrum was recorded as $\log(1/R)$, where R=reflectance, by averaging 128 scans and for each apple a mean spectrum was calculated by averaging the spectra collected at the four positions around the apple.

Data preprocessing and analysis

Data acquisition and spectra storage were achieved with a PC running specially developed software OMNIC6.1 of Nicolet Com. USA. To compare the influence of the preprocessing on the

prediction performance of the calibration models, a number of pretreatment options including Savitzky-Golay filter and the Norris derivative filter were investigated. Once these preprocessing procedures were completed, partial least squares (PLS) and principal component regression (PCR) were used to develop calibration models for predicting the valid acidity. The optimal factors used to develop the calibration models were based on the root mean square error of cross validation. The root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and the ratio of data set standard deviation to RMSEP (SDR) were used to judge the success and accuracy of the model. An SDR value of 3 is generally considered the minimum for any useful sorting purposes (McGlone *et al*., 2002). The RMSEC, RMSEP and SDR were calculated as follows:

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

$$RMSEP = \sqrt{\frac{1}{I_p - 1}\sum_{i=1}^{I_p}(\hat{y}_i - y_i - bias)^2} \quad (2)$$

$$SDR = \frac{SD}{RMSEP} \quad (3)$$

where

$\hat{y}_i$=predicted value of the ith observation;

y_i=measured value of the ith observation;

I_c=number of observations in calibration set;

I_p=number of observations in prediction set;

SD=prediction set standard deviation.

Valid acidity measurement

The valid acidity of the apple juice was determined at 25 °C with a pH meter (SJ-4A, Exact instrument Co., Shanghai, China). The apple juice was expressed after spectrum acquisition; a cylindrical core of flesh was removed with 2 cm cork borer from the location of the FT-NIR measurement.

RESULTS AND DISCUSSION

Selection of wavelength ranges

The 800−2619 nm spectra were measured by the FT-NIR spectrophotometer. Based on the suggested region for TQ analyst by v.6.2.1 software, different wavelength ranges with different prediction performance were obtained, which included three effective ranges: 967−2619 nm, two NIR range 967−1375 nm and 1375−2619 nm known to contain typical absorption bands.

The calibration set was selected with the aim of providing strong calibration for valid acidity by maximizing the variability among sample compositions and obtaining a wide range of spectra to avoid outliers in the validation set. Table 1 shows mean, median, maximum, minimum values and standard deviations of valid acidity. Selection of the optimum wavelength range for best predictive model was done by PLS analysis by TQ analyst v.6.2.1 (Nicolet, Co., USA). The prediction results of different wavelengths range are presented in Table 2.

Table 1 Descriptive statistics for sample measurements of valid acidity

Items	Calibration	Prediction
No. of sample	80	40
Mean	3.41	3.40
Maximum	4.03	3.93
Minimum	3.24	3.25
Median	3.38	3.38
[a]*SD*	0.134	0.136
[b]*CV* (%)	3.92	4.01

[a]*SD*: standard deviation; [b]*CV*: coefficient of variation

Table 2 Results of original spectrum (Log(1/*R*)) with different wavelength ranges by PLS method

Wavelength range (nm)	No. of factors	r^2	*RMSEC*	*RMSEP*
967−2619	4	0.752	0.0681	0.0679
967−1375	8	0.729	0.0522	0.0760
1375−2619	5	0.780	0.0642	0.0690

r^2: Correlation coefficient of determination

The optimum wavelength range for best predictive model was at 967−2619 nm range with a relative high correlation coefficient of determination (r^2=0.752), low RMSEP (0.0679), low RMSEC (0.0681) and also small difference between RMSEP and RMSEC, and was therefore selected as the optimum wavelength range from Table 2.

Influences of data preprocessing

The valid acidity model was built using the wavelength range of 967–2619 nm by means of PLS method and the influences of the preprocessing on the prediction quality of the model can be seen in Tables 3, 4 and 5. The gap size *n* had influenced on the prediction performance of the model. Eight models with different gap size are compared in Table 3. Models 4, 6 and 7 revealed that there was an optimal gap size over which the second derivative was calculated. Model 7 had high r^2 (0.831), large number of factors (7), low RMSEP (0.0670), RMSEC (0.0562), and small difference between RMSEP and RMSEC and was therefore selected as the best model. Models

Table 3 Prediction performances with Norris derivate filter between valid acidity and second derivate spectra using different gap size (segment length=5 nm)

Model	Gap size	No. of factors	r^2	*RMSEC*	*RMSEP*
1	1	6	0.793	0.0623	0.0717
2	3	5	0.785	0.0635	0.0697
3	5	5	0.790	0.0627	0.0693
4	7	6	0.815	0.0573	0.0689
5	8	4	0.779	0.0644	0.0703
6	9	6	0.816	0.0588	0.0660
7	10	7	0.831	0.0562	0.0670
8	12	4	0.781	0.0640	0.0705

r^2: Correlation coefficient of determination

Table 4 Prediction performances with Norris derivate filter between valid acidity and second derivate spectra using each segment length (gap size=10 points)

Model	Segment length	No. of factors	r^2	*RMSEC*	*RMSEP*
1	3	6	0.837	0.0553	0.0694
2	5	7	0.831	0.0562	0.0677
3	7	4	0.781	0.0640	0.0703
4	9	4	0.783	0.0638	0.0698
5	11	4	0.784	0.0636	0.0691
6	13	4	0.787	0.0633	0.0689
7	15	4	0.790	0.0628	0.0689
8	17	3	0.786	0.0634	0.0666
9	19	3	0.788	0.0631	0.0689
10	21	4	0.797	0.0617	0.0665

r^2: Correlation coefficient of determination

Table 5 Prediction performances with Savitzky-Golay filter between valid acidity and second derivate spectra using each smoothing point (polynomial order=2)

Model	Smoothing point	No. of factors	r^2	*RMSEC*	*RMSEP*
1	3	6	0.625	0.0839	0.109
2	5	10	0.761	0.0670	0.0933
3	7	9	0.785	0.0634	0.0823
4	9	6	0.750	0.0634	0.0821
5	11	6	0.786	0.0634	0.0723
6	13	4	0.762	0.0667	0.0749
7	15	4	0.773	0.0653	0.0735
8	17	5	0.786	0.0634	0.0693
9	19	5	0.786	0.0634	0.0699
10	21	5	0.786	0.0633	0.0711

r^2: Correlation coefficient of determination

2 and 3 and Models 1 and 6 had the different prediction results with the same number of factors. Models 5 and 8 had the same prediction results with different number of factors. Models with a low number of factors had a relatively high RMSEP and a relatively high RMSEC. Model 5 and model 8 illustrated this.

There was influence of segment length *n* on the prediction performance of the model and ten models of different segment length are compared in Table 4. Models 2, 8 and 10 revealed that there was an optimal segment length over which the second derivative was calculated. Model 2 had high r^2 (0.831), large number of factors (7), low RMSEP (0.0677) and low RMSEC (0.0562), and small difference between RMSEP and RMSEC and was therefore selected as the best model. Models 1, 5 and 9 had the same prediction results with different number of factors. Models 3 and 4 had a relatively high RMSEP (0.0703, 0.0698) and a relatively high RMSEC (0.0640, 0.0638), respectively.

The smoothing point *n* had influence on the prediction performance of the model and ten models with different smoothing point are compared in Table 5. Models 8, 9 and 10 revealed that there was an optimal smoothing point over which the second derivative was calculated. Model 8 had high r^2 (0.786), large number of factors (5), low RMSEP (0.0693), low RMSEC (0.0634), and relatively small difference between RMSEP and RMSEC and was therefore selected as the best model. Models 3 and 4 yielded the same prediction results with different number of factors. Models 1 and 2 yielded the worst prediction results. Models with low number of factors had relatively high RMSEP and relatively high RMSEC. Models 5, 6 and 7 illustrated this. From Tables 3, 4 and 5, it was concluded that models based on smoothing spectra were slightly worse than models based on derivative spectra. In the Norris derivative filter, the best result was obtained when the segment length was 5 nm and the gap size was 10 point. Therefore, the first and the second derivative treatments were performed under the derivative conditions mentioned above.

Influence of spectra data preprocessing

Based on the above results, calibration models were built in the wavelength range of 967–2619 nm for original, first and second derivative spectra using PLS and PCR analysis. Original spectral data were transformed to first and second derivate spectral data by Norris derivative filter type (segment length=5 nm, gap size=10 points). Results of calibration and prediction for prediction valid acidity of intact apples are shown in Tables 6 and 7.

The second derivative spectra yielded the best results with high correlation coefficient of determination (r^2=0.759), low RMSEP (0.0677), low RMSEC (0.0562), SDR value of 1.80 and small difference between RMSEP and RMSEC. The original spectra and the first derivative spectra had the same prediction performance with different number of factors.

Table 7 shows that different mathematical treatments had influence on the prediction performance of the model using PCR method. Model with original spectra showed the best results with relatively high correlation coefficient of determination (r^2=0.755),

Table 6 Results of PLS analysis with different spectral treatments in the wavelength range of 967–2619 nm

Spectrum	Number of factors	Calibration		Validation		
		r^2	*RMSEC*	r^2	*RMSEP*	*SDR*
Log(1/*R*)	4	0.740	0.068	0.753	0.0679	1.62
*D*log(1/*R*)	4	0.781	0.064	0.748	0.0688	1.61
D^2log(1/*R*)	7	0.832	0.056	0.759	0.0677	1.80

D: Derivate spectra; *R*: Reflectance

Table 7 Results of PCR analysis with different spectra treatments in the wavelength range of 967–2619 nm

Spectrum	Number of PC	Calibration		Validation		
		r^2	*RMSEC*	r^2	*RMSEP*	*SDR*
Log(1/*R*)	6	0.771	0.0656	0.755	0.0671	1.72
*D*log(1/*R*)	10	0.794	0.0621	0.736	0.0700	1.63
D^2log(1/*R*)	7	0.780	0.0642	0.747	0.0689	1.62

PC: Principle Component; *D*: Derivate spectra; *R*: Reflectance

low RMSEP (0.0671), low RMSEC (0.0656), SDR value of 1.72, and small difference between RMSEP and RMSEC, and was therefore chosen as the best model. Models with the first and second derivate spectra were slightly worse than the models with original spectra.

Choice of the best prediction model

For valid acidity prediction of intact apples, different calibration models were calculated by using PLS and PCR calibration technique. The preprocessing of the spectra and the number of factors were taken into consideration.

In order to find the best model, three criteria were considered: First, a good model should have low RMSEP, low RMSEC, high correlation coefficient and small difference between RMSEP and RMSEC (Lammertyn *et al*., 1998); Second, a relatively low number of factors was desirable in order to avoid inclusion of signal noise in the modeling. The minimum prediction residual error sum of square (PRESS) was used to determine the optimal number of factors; Third, multiple calibration techniques influenced the result as well. PLS models were better than PCR models (Tables 6 and 7). Therefore, model with second derivative spectra by PLS showed the best results in both calibration and prediction results.

The correlation between the measured and predicted valid acidity values for the best model are shown in Fig.2 and the correlation coefficient between measured and predicted values for validation set was equal to 0.869 and the RMSEP was 0.0671.

CONCLUSIONS

The above results indicate that it has good prediction performance for the best calibration models is in the wavelength of 967–2619 nm by PLS method. A relationship was established between diffused reflectance spectra and valid acidity using PLS and PCR calibration technique. Data preprocessing influences the performance of the calibration models. The models based on smoothing spectra were slightly worse than the models based on derivative spectra and the best result was obtained when the segment length was 5 nm and the gap size was 10 points. The second derivative spectra showed the best results with high correlation coefficient of determination (r^2=0.759), low RMSEP (0.0677), low RMSEC (0.0562) and small difference between the RMSEP and the RMSEC by PLS method. The original spectra and the first derivative spectra had the same prediction performance with different number of factors. This research indicates that it is feasible to develop a nondestructive technique for nondestructive prediction apple valid acidity by FT-NIR method.

ACKNOWLEDGEMENTS

The authors sincerely thank Lu Hui-shan and Jiang Xue-song for their helps in this experiment.

References

Belton, P.S., Kemsley, E.K., McCann, M.C., Ttofis, S., Wilson, R.H., Delgadillo, I., 1995. The identification of vegetable

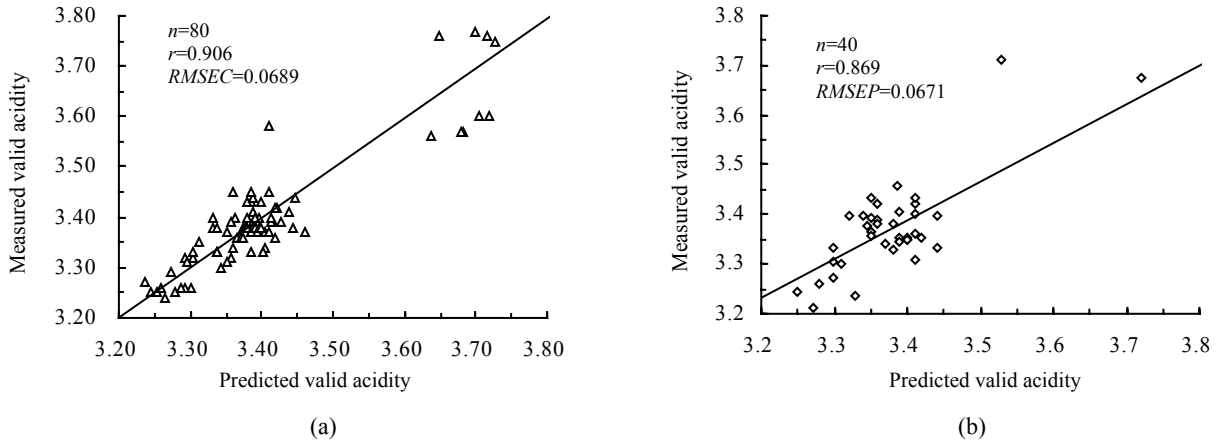

Fig.2 Correlation plot for the measured and the predicted valid acidity values with second derivative treatment. (a) Calibration; (b) Validation

and matter using Fourier transform infrared spectroscopy. *J. Food Chem.*, **54**:437-441.

Birth, G.S., Dull, G.G., Renfore, W.T., Kays, S.J., 1985. Nondestructive spectrophotometric determination of dry matter in onions. *J. Am. Soc. Hort. Sci.*, **110**:297-303.

Dull, G.G., Birth, G.S., Leffler, R.G., 1989. Use of near infrared analysis of nondestructive measurement of dry matter in potatoes. *J. Amer. Potato*, **66**:215-225.

Dull, G.G., Birth, G.S., Leffler, R.G., Zaltzman, A., Schmmilovitch, Z., 1991. The near infrared determination of moisture in whole date. *HortScience*, **26**(10): 1303-1305.

Huxsoll, C.C., Bolin, H.R., Mackey, B.E., 1995. Near infrared analysis potential for grading raisin quality and moisture. *J. Food Sci*., **60**:176-180.

Lammertyn, J., Nicolay, B., Ooms, K., DeSemedt, V., De Baerdemaeker, J., 1998. Non-destructive measurement of acidity, soluble solids and firmness of Jonagold apples using NIR-spectroscopy. *Trans. of ASAE*, **41**(4): 1089-1094.

Lammertyn, J., Peirs, A., De Baerdemaeker, J., Nicolai, B., 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Boil. and Technol.*, **18**:121-132.

Liu, Y., Ying, Y., 2003. A study and fiber sensing technique used for fruit interior quality inspection. *J. Sen. Actuator*, **16**(2):170-174.

Lu, R., Ariana, D.A., 2002. Near-infrared sensing technique for measuring internal quality of apple fruit. *Trans. of ASAE*, **18**:585-590.

Luis, E.R., Fredrick, S.F., Michael, A.M., Elizabeth, M.C., 2001. Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *J. Carbohydrate Research.*, **336**:63-74.

McGlone, V.A., Robert, B.J., Martinsen, J.P., 2002. Vis/NIR estimation at harvest of pre- and post-storagequality indices for 'Royal Gala' apple. *Postharvest Bio. and Technol*., **25**:135-144.

Moons, E., Dardenne, A.P., Sindic, M., 1997. Nondestructive Visible and NIR Spectroscopy for the Determination of Internal Quality in Apple. *In*: Proceedings from the Sensors for Non-destructive Testing International Conference and Tour. Orlando, FL, NRAES (Northeast Reg. Agric. Eng. Serv.), Ithaca, NY, p.122-132.

Peiris, K.H.S., Dull, G.G., Leffler, R.G., Kays, S.J., 1997. Nondestructive Determination of Sugar Content of Peach by near Infrared Spectroscopy. *In*: Proceedings of Conference on 'Sensors for Non-destructive Testing'. Orlando, FL, NRAES, Ithaca, NY, p.77-87.

Peirs, A., Scheerlinck, N., Touchant, K., Nicolai, B.M., 2002. Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems Engineering*, **81**(3):305-311.

Slaughter, D.C., 1995. Nondestructive determination of internal quality in peaches and nectarines. *Trans of ASAE*. **38**(2):617-623.

Ventura, M., Jager, A.D., Putter. H., Roelofs, F.P.M., 1998. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biology and Technology*, **14**:21-27.