Potable NIR spectroscopy predicting soluble solids content of pears based on LEDs

Yande Liu*, Wei Liu, Xudong Sun, Rongjie Gao, Yuanyuan Pan, Aiguo Ouyang,
School of Mechatronics Engineering, East China Jiaotong University, Changbei Open and Developing District, Nanchang, 330013, China

ABSTRACT

A portable near-infrared (NIR) instrument was developed for predicting soluble solids content (SSC) of pears equipped with light emitting diodes (LEDs). NIR spectra were collected on the calibration and prediction sets (145:45). Relationships between spectra and SSC were developed by multivariate linear regression (MLR), partial least squares (PLS) and artificial neural networks (ANNs) in the calibration set. The 45 unknown pears were applied to evaluate the performance of them in terms of root mean square errors of prediction (RMSEP) and correlation coefficients (r). The best result was obtained by PLS with RMSEP of 0.62°Brix and r of 0.82. The results showed that the SSC of pears could be predicted by the portable NIR instrument.

Keywords: Near-infrared spectroscopy, Artificial neural network, Soluble solids content, Partial least squares, Portable, LEDs

1. INTRODUCTION

Pear is one of major fruits in our daily life. Brix is one of the major factors that affect taste in pears. Traditionally, pear quality is determined quantitatively by using destructive testing methods, which are both waste and time-consuming.
consuming. Some papers have been published on the NIR analysis of brix value in pears \cite{1}. Even if these instruments are highly precise, their application to field research, such as monitoring of chemical changes of developing pears on trees, are limited by their large size and weight. Fortunately, the availability of low cost miniaturized spectrometers has opened up the possibility of portable devices which can be used for brix prediction. However, before these technology can be successfully transferred to the fruit industry and especially, implemented on a large scale, further research is required into solve these problems: the device must be low cost, be quick-response and deliver robust performance.

Presently, there are several commercially available portable NIR instruments offered for determining Brix value \cite{2-3}, such as \cite{4}. Temma et al. showed that the portable NIR instrument developed by their research centre had excellent potential in determining Brix value of intact apples and it is being successfully implemented in various fruits, such as apples \cite{5}; cherries \cite{6}; mangoes \cite{7}; peaches \cite{8}. However, the illuminant of a portable NIR instrument when compared to a research one is still not stable. The high-power light emitting diode (LED) is a semiconductor device usually adapted on portable photometers (LED-photometers) as radiation source because it confers some advantages for optical instruments as simplicity, easy operation, light stability and low power and cost \cite{9, 10}.

The objective of this study was to research the feasibility of rapid measurement of the SSC of intact pear fruit with NIR instrument equipped with LEDs. Investigate the performance for determining SSC of pear quality nondestructively and discusses the developed high-power LED spectrometer and presents the test results characterizing the developed NIR module.

\textbf{2. MATERIALS AND METHODS}

2.1. Samples preparation and destructive analyses

A total of samples of 190 ‘crystal’ pears (from Shandong province, China) were purchased from a local market. The pears were stored two days at 20ºC and 60% relative humidity before spectra collection. Immediately after spectra collection, each pear was mixed using a juice centrifuge and the obtained juice was filtered using a filter paper. Corresponding to each pear, the SSC value of filtered juice was determined using a hand-held refractometer.
(LB32T, MinRui, Guangdong, China) and expressed in °Brix. 190 samples were used to analyze and develop Vis-NIR models in laboratory. In terms of the testing samples, as selected by fairly normally distributing around the mean value, 45 pears were used for prediction set. The remaining 145 samples were used for calibration set. To ensure the adaptability of the calibration models, the samples with high and low values were put in the calibration set and the other samples selection inside each group were performed manually.

2.2. Spectral Collection

As shown in figure 1, the portable NIR instrument consisted of a hand-held probe, a detector, system microcontroller, display unit and data processing software. The handheld probe consisted of a Plano mirror, a fiber-optics, four LEDs. The Plano mirror is located in the center of the round metal bottom cover. It can take more rays of light in the fiber-optics. The length of the fiber-optics is 2 meters and the fiber core diameter is 400 um. Four different types of 3W high power LEDs (FZ-P003ORDX-XXXP, Ark Lighting, Guangdong, China) made up the light source, centered at the wavelengths 620, 850, 880, and 940 nm. The LEDs were arranged in a round metal bottom cover equably [7]. The length of the handle is 120 mm, minor diameter is 40 mm and major diameter is 76 mm [8]. A convex in the center of metal bottom cover was able to secure and tight attachment the surface of pears, thus the reflected light could be prevented and the pure diffuse reflection spectrum is detected by the probe and transmit the optical fiber to the spectrometer. Diffuse transmission spectra were collected using a fiber spectrometer equipped with a 3648-element linear silicon CCD array detector (TCD1304AP, Toshiba, Japan), which covers a spectral range of 345–1040nm at a sampling interval of 1.5nm. The system microcontroller adopted a single board computer (PCM-9361, Advantech, China). The measurement result was displayed on a LCD screen. Seen from Figure 1, all modules, except for the fiber-optics probe, were housed in a mater case.
The digitalized absorbance data send by the control unit was displayed on the screen. Therefore, the parameters can be set just one touching pen \[ ^9 \].

One hundred and ninety samples of pears were used to register the blank signal. The measurements with the portable NIR spectrometry were performed in a room temperature. The round metal bottom cover was tight attachment the surface of the sample. Adjusted the pear and let the center of the round metal bottom cover aim the fruit shoulder (form the fruit stem axis). When the light sources, spectrometer and energized, the diffuse reflection spectra data automatically on the LCD screen. A dark scan was carried out, and a white Teflon tile was used for white calibration at the beginning of each measurement session. For each fruit, spectra were taken at four equidistant positions around the equator (approximately 90°). Then connect the portable NIR spectrometry to a computer with the USB port and the spectra data can be transmitted by USB wire to the computer. The data that were processed was finally used to verify feasibility of the portable NIR spectrometry.

2.3. Data Analysis

Before the calibration step, the spectra in reflectance(R) were converted to absorbance (log (1/R)) values to obtain linear correlations of the NIR spectra with the SSC measurement values. Then, the log (1/R) spectra were pre-treated by using multiplicative scatter correction (MSC). Chemometric analysis was performed using the Unscrambler software v9.5 (CAMO AS, Trondheim, Norway) and Matlab 7.0 (The MathWorks, USA). The Unscrambler program was used to establish partial least squares (PLS) model and regression, multivariate linear regression (MLR) model and the Matlab 7.0 (The MathWorks, USA) was used for developing the ANNs models. PLS were carried out to perform linear models of prediction between spectral data and the values obtained from the destructive tests. In order to get efficient and quick models, the optimal number of latent variables for the PLS models was obtained by using the leave-one-out cross-validation technique.

ANNs are the most widely used mathematical algorithms for overcoming non-linearity in calibration model which were more and more widely applied during the past several years \[ ^{10-11} \]. Back-propagation training algorithms are usually preferred for content applications and were used in this study \[ ^{12} \]. ANNs comprise several layers. PCA was performed firstly to extract information from the whole spectra region as principle components, which were used to be the neurons of network input layer.

The statistic correlation coefficient \( r \), root mean squares error of calibration (RMSEC) and root mean squares error of prediction (RMSEP) were used to evaluation of the fit of a mathematical model to the experimental data.
Bias denotes the average of the differences between estimated and measured values. The correlation coefficient of validation is calculated using Eq. (1) from the samples for validation. \( r \), RMSEP, Bias and RMSEC are defined as follows:

\[
    r = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sqrt{\sum_{i=1}^{n} (\bar{y} - y_m)^2}} \\
    RMSEC, RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \\
    \text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)
\]

Where, \( \hat{y}_i \) = predicted value of the i-th observation; \( y_i \) = measured value of the i-th observation; \( n \) = number of observations in the calibration or prediction set; \( y_m \) = mean value of the calibration or the prediction set.

3. RESULTS AND DISCUSSION

3.1. SSC distribution

The mean SSC was 10.3 ºBrix. The portable device was applied on the analysis of 190 samples of pears in order to determinate the standard deviations (SD) of measurements. In this study, the SSC measurements (n=190) were approximately normally distributed around the mean (max=12.7 ºBrix, min=8.2 ºBrix). An overview of SSC distributions of pears samples in the calibration and validation sets were presented in Table 1.

| Items               | Quantity | Range(ºBrix) | Mean(ºBrix) | SD(ºBrix) |
|---------------------|----------|--------------|-------------|-----------|
| Calibration set     | 145      | 8.2-12.7     | 10.4        | 1.05      |
| Validation set      | 45       | 8.6-12.3     | 10.1        | 0.98      |

3.2. Spectral pre-processing analyses

The NIR spectra often require mathematical pretreatment to remove interference of specula radiation, such as smooth, multiplicative scatter correction (MSC) and first derivative. The pretreatment methods were attempted to eliminate the interference by smooth, MSC and first derivative, respectively. For results in Table 2, the RMSEP, RMSECV and R were very close by different pretreatment methods. This indicated that the major information in pear spectra could be reserved while noise was removed by 3-average smoothing. Consequently, the PLS
methods were used to develop the multivariate calibration models, where twenty factors (latent variables) were necessary to produce reasonable results for the spectra with 3-average smoothing spectral pretreatments, respectively.\[14\].

| Pre-processing         | RMSEC | RMSEP | R   |
|------------------------|-------|-------|-----|
| None                   | 0.54  | 0.60  | 0.85|
| 3-average smoothing    | 0.58  | 0.58  | 0.88|
| MSC                    | 0.59  | 0.64  | 0.87|
| First-derivative       | 0.53  | 0.60  | 0.85|

3.3. The establishment of models

PLS was used to build the prediction models, because PLS have the potential to estimate not only component concentrations but also chemical and physical properties from their infrared spectra. Due to its simplicity and small volume of calculations, it is used to analyze different data. PLS method itself is a linear method of data analysis. Fig. 2(a) and (b) were the calibration and prediction models. From the plots, the low values were over estimated, while high values were under estimated, and the least squares solution was pulled to the mean. This was known as the “Dunne-effect” for regression on normally distributed data. Hence, PLS was unsuitable for dealing with the nonlinear phenomenon.

In the above discussion of the prediction results using PLS models, no consideration has been given to the contribution of the individual individual wavelength. This was because the PLS model applied linear transforms using the entire wavelength data. As a result, it was difficult to ascertain which individual wavelength was directly related to SSC. However, it would be helpful to examine how SSC levels are related to individual wavelengths, so that a better understanding of spectroscopy of pear, and SSC may be obtained. For that reason, an MLR model was built.

After PLS processing of the 190 samples, cross validation revealed the existence of six LVs. The cumulative reliabilities of the first six LVs accounted for 87% of X and 31% of Y (Fig. 3), suggesting a possible association between the sensitive individual wavelengths and SSC. By choosing spectral wavebands with the highest absolute loading weight values in the spectral region (800–980), the optimal wavelengths were: 812, 825, 831, 846, 856, 872, 891, 904, 918, 925, 954, 966 and 979nm. Since the individual wavelengths might be particularly important for reducing sugar content, all these wavelengths were selected in the loading weight curves, whether or not they were the real fingerprint spectra. The performance of these wavelengths was then evaluated. MLR model was built based on the above sensitive wavelengths and the calibration and prediction results of the MLR model are shown.
Fig. 4(a) and (b). The model displayed high correlation ($r^2 = 0.87, 0.78$), SEC (0.58), and SEP (0.64), respectively, for calibration and prediction.

In the BP-ANN method, BP-ANN consisting of three layers (input layer, hidden layer and output layer) was established. The tansig function was used for the transfer function between the input layer and the hidden layer, and the prelim function for that between the hidden layer and the output layer. PLS was performed firstly to extract information from the whole spectral region as factors (principle components), which were then used to be the neurons of network input layer. At the beginning of a training run, the parameters of ANNs including input nodes, hidden nodes, learning coefficient, momentum and number of iterations were initialized with optional values \cite{15}. Being in operation process, the modifications of the network input nodes (4), hidden nodes (5), learning coefficient (0.12) and momentum (0.05–0.50) were selected by the back-propagation of the errors and the degree of approximation. Because there was only one active compound in established models, the output layer contained one neuron. Figure 5 showed the measured and predicted plots of brix for the BP-ANN model. It is clear that the ANNs model obtained better result of $r = 0.83$ and RMSEP = 0.57\textdegree Brix.

Fig. 2. PLS calibration (a) and prediction (b) models
Fig. 3. Loading weights of the first six LVs of the PLS model.

Fig. 4. MLR calibration (a) and prediction (b) models.
3.4. **Comparison of calibration methods**

The quality of the calibration was quantified by SEC, SEP and r. A good model should have a low SEC, a low SEP, a high r but also a small difference between SEC and SEP [16-17]. The PLS model, based on 3-average smoothing spectral pretreatment, could be obtained the best of prediction results with $r=0.88$ and $SEP=0.58^\circ Brix$.

The plot of relationship between the values of the SSC measurement and the PLS model prediction, with MSC spectral pretreatment, is shown in table 3. Scatter plots between actual and predicted Brix values of the validation sample set are also shown in Figure 4. The result showed that the PLS spectra using wavelength region from 800 to 950nm more than the ANNs and MLR equation.

| Methods | Wavelength | Calibration | Prediction |
|---------|------------|-------------|------------|
|         |            | $R_c$ | SEC | $R_v$ | SEP |
| PLS     | 800-980    | 0.87  | 0.51  | 0.82  | 0.62  |
| MLR     | 800-980    | 0.87  | 0.58  | 0.78  | 0.64  |
| ANN     | 800-980    | 0.88  | 0.51  | 0.83  | 0.57  |

### 4. **CONCLUSION**

In this paper, performance of PLS, MLR and ANNs were compared in developing calibration models for brix prediction in pears by the portable near infrared device. ANNs was applied to improve prediction performance by correcting the nonlinearities which limited the performance of the classical linear method. Compared to PLS, the performance of ANNs obtained better result, with higher $r$ and lower RMSEP for brix value. The portable NIR spectrometry uses a NIR high-power LED emitting as portable detector whose noise was kept at low levels by a control system based on thermistor and thermoelectric cooling. Thus, the portable NIR spectrometry showed to be an economical and viable alternative for SSC analysis of pears. This paper successfully demonstrates the application of NIR spectroscopy using the spectral region from 800 nm to 950 nm to the quantitative measurement of pears in PLS models study.

### ACKNOWLEDGEMENTS

This study was supported by the Natural Science Foundation of Jiangxi Province (2008GQN0029, 2007GZN0266), National Science and Technology Support Plan (2008BAD96B04), Special Science and
Technology-Support Program for Foreign Science and Technology Cooperation Plan (2009BHB15200) and Technological expertise and academic leaders training plan of Jiangxi Province (2009DD00700).

REFERENCES

1. T. Temma, K. Hanamatsu, and F. Shinoki, “Development of a portable near infrared sugar-measuring instrument,” J. Near Infrared Spectrosc., vol. 10, pp. 77–83, 2002.

2. Y.B. Ying, Y.D. Liu, J.P. Wang, X.P. Fu, and Y.B. Li, “Fourier transform near infrared determination of total soluble solids and available acid in intact peaches,” Transactions of the ASAE, vol. 48 (1), pp. 229-23, 2005.

3. Y.D. Liu, Y.B. Ying, and H.Y. Jiang, “Rapid determination of maturity in apple using outlet detection and calibration model optimization,” Transactions of the ASAE, vol. 49 (1), pp. 91-95, 2006.

4. T. Temma, K. Hanamatsu and F. Shinoki, J. Near Infrared Spectrosc. 10, 77 2002.

5. Chemleds, Telcom Devices datasheet.

6. B.A. Matveev, et al., Mid-infrared (3–5 mm) LEDs as sources for gas and liquid sensors, Sensors and Actuators B 38–39 (1997) 339–343.

7. Pascal Chalus, Serge Walter, and Michel Ulmschneider, “Combined wavelet transform–artificial neural network use in tablet active content determination by near-infrared spectroscopy,” Analytica Chimica Acta, vol. 591, PP. 219–224, 2007.

8. Jouko Malinen, Markku Kaˇnsaˇkoski, Raimo Rikola, and Colin G. Eddison “LED-based NIR spectrometer module for hand-held and process analyser applications,” Sensors and Actuators B, vol. 51, pp. 220–224, 1998.

9. Edvaldo da Nóbrega Gaião et al, “An inexpensive, portable and microcontrolled near infrared LED-photometer for screening analysis of gasoline,” Talanta, vol. 75, PP. 792–796, 2008.

10. F. Despagne, D.L. Massart, Analyst .157–178,174–180, 1998.

11. D. Perez-Marin, A. Garrido-Varo, J.E. Guerrero, Talanta 72 28–42, 2007.

12. B.N. Peshlov, F.E. Dowell, F.A. Drummond, and D.W. Donshue, “Comparison of three near infrared spectrophotometers for infestation detection in wild blueberries using multivariate calibration models,” J. Near Infrared Spectrosc, vol. 17, PP. 203–212, 2009.
13. Lingzhi Zhao, Ying Dou, Ye Guo, Yonghong Chen, and Yulin Ren, “Artificial neural networks for non-destructive determination of acetylspiramycin powder by short-wavelength NIR spectroscopy,” Vibrational Spectroscopy, vol. 47, pp. 21–25, 2008.

14. R. Lu, and D. Ariana, “A near-infrared sensing technique for measuring internal quality of apple fruit,” Applied Engineering in Agriculture, vol. 18 (5), pp. 585_590, 2002.

15. Yongni Shao, Yong He, Antius H. Gomez, Anni G. Pereir, Zhengjun Qiu, and Yun Zhang, “Visible/near infrared spectrometric technique for nondestructive assessment of tomato ‘Heatwave’ (Lycopersicum esculentum) quality characteristics,” Journal of Food Engineering, vol. 81, pp. 672-678, 2007.

16. Y.D. Liu, Y.B. Ying, X.P. Fu, and H.S. Lu, “Experiments on predicting sugar content in apples by FT-NIR technique,” Journal of Food Engineering, vol. 80, pp. 986_989, 2007.

17. B.M. Nicolaï, K.I. Theron, J. Lammertyn, and Kernel, “PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple,” Chemometrics and Intelligent Laboratory Systems, vol. 85 (2), pp. 243-252, 2007.