African Journal of Biotechnology Vol. 11(32), pp. 8133-8140, 19 April, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.049

Full Length Research Paper

Robust model of fresh jujube soluble solids content with near-infrared (NIR) spectroscopy

Shujuan Zhang*, Yanru Zhao, Wei Guo, Qiaoling Hao, Haihong Zhang, Huamin Zhao and Lihua Yin

College of Engineering, Shanxi Agricultural University, Taigu 030801, China.

Accepted 23 March, 2012

A robust partial least square (PLS) calibration model with high accuracy and stability was established for the measurement of soluble solids content (SSC) of fresh jujube using near-infrared (NIR) spectroscopy technique. Fresh jujube samples were collected in different areas of Taigu and Taiyuan cities, central China in 2008 and 2009. A partial least squares (PLS) calibration model was established based on the NIR spectra of 70 fresh jujube samples collected in 2008. A good calibration result was obtained with correlation coefficient (Rc) of 0.9530 and the root mean square error of calibration (RMSEC) of 0.3951 °Brix. Another PLS calibration model was established based on the NIR spectral of 180 samples collected in 2009; it resulted in the Rc of 0.8536 and the RMSEC of 1.1410 °Brix. It could be seen that the accuracy of established PLS models were different when samples harvested in different years were used for the model calibration. In order to improve the accuracy and robustness of model, different numbers (5, 10, 15, 20, 30 and 40) of samples harvested in 2008 were added to the calibration sample set of the model with samples harvested in 2009, respectively. The established PLS models obtained Rc with the range of 0.8846 to 0.8893 and RMSEC with the range of 1.0248 to 0.9645 °Brix. The obtained results were better than the result of the model which was established only with samples harvested in 2009. Moreover, the models established using different numbers of added samples had similar results. Therefore, it was concluded that adding samples from another harvest year could improve the accuracy and robustness of the model for SSC prediction of fresh jujube. The overall results proved that the consideration of samples from different harvest places and years would be useful for establishing an accuracy and robustness spectral model.

Key words: Near-infrared (NIR) spectroscopy, Huping jujube, soluble solids content (SSC), partial least squares (PLS), accuracy, stability.

INTRODUCTION

Fresh jujube which is mainly distributed in the subtropical and tropical regions of Asia and America, is a tree of the Rhamnaceae family. The fruit of jujube is a profitable fruit, and is much admired for its high nutritional value. Potassium, phosphorus, calcium and manganese are the major mineral constituents in Chinese jujube. It also has high amounts of sodium, zinc, iron and copper. Jujube also contains vitamin c, riboflavin and thiamine. Moreover, it contains 20 times the amount of vitamin C as citrus fruits (Li et al., 2007).The production of jujube fruits with high quality is an important issue for public health. Consumers are always demanding superior quality of jujube products, which in turn has increased the need for enhanced quality monitoring. The traditional quality assurance methods used in the food industry have generally involved human visual inspection, which are however tedious, laborious, time-consuming and inconsistent. With increased demands for jujube products with high quality and safety, it is necessary to realize accurate, rapid and objective

*Corresponding author. E-mail: zsujuan1@163.com.

Abbreviations: PLS, Partial least square; **SSC,** soluble solids content; **NIR,** near-infrared; **RMSEC,** root mean square error of calibration.

quality determination of fresh jujube fruit. Due to the advantages of being rapid and non-destructive, visible and near infrared spectroscopy (Vis-NIR) is widely used for quality measurement of food products (Wu et al., 2008a, 2009a, b, c), including many applications for fruits such as jujube (Wang et al., 2010; Zhang et al., 2010), apple (Ann et al., 2003; Liu et al., 2007; Els et al., 2010), plum (Esmé and Karen, 2010), kiwifruit (Ali et al., 2010; Andrew et al., 1998), mandarin (Liu et al., 2010) and others (Bert et al., 2007). However, before its on-line application for quality inspection in fruit industry, the established spectral calibration models should have the feature of being robust, which can be achieved by considering large datasets with samples from different regions, seasons and operational conditions (Cozzolino et al., 2011).

In recent years, Vis-NIR spectroscopy has been applied for the determination of internal quality characteristics of fresh jujubes. Zhang et al. (2009) applied NIR spectroscopy technique combined with principal component analysis and neural networks for variety identification and prediction of soluble solid content (SSC) of three varieties of fresh jujube collected at an orchard in Taigu city, China. A good prediction identification accuracy of this model of 100% was obtained using the threshold set of ±0.17. Less than 10% of relative deviation was also obtained between predicted value and measured value of the soluble solids content. Later, Wang et al. (2010) compared the abilities of the interactance, reflectance and transmission modes of Vis-NIR spectroscopy in detecting internal insect-infested jujubes from a jujube orchard in Japan. High correct classification rates of 100, 90.0 and 97.0% were achieved using the above three spectral measurement modes, respectively. In another study, Hong et al. (2010) analyzed Vis-NIR spectra and chlorophyll of winter jujubes which were collected from a jujube orchard in Dali (Shanxi, China) using support vector machine method. The best classification precision of 93.3% was obtained for the quality classification of jujube. However, these works of using Vis-NIR spectroscopy for the quality measurement of fresh jujube only used samples harvested in the same year or in the same orchard. Therefore, this research used the samples of Huping jujube from one jujube gardens in Taigu in 2008 and another jujube garden in Taiyuan in 2009 (Shanxi, China).

The objectives of this paper were to investigate the feasibility of using Vis-NIR spectroscopy combined with partial least squares (PLS) for rapid and non-destructive prediction of soluble solids content of jujube samples, and evaluate the improvement of model robustness using samples from different years, different jujube gardens and different quantities.

MATERIALS AND METHODS

The fresh jujube samples were collected from two organic-cultivation jujube orchards in Taigu and Taiyuan (Shanxi, China, in 2008 and 2009) with red peel and no external defects. A total of 70 Huping jujube samples were obtained from a jujube orchard in 2008. The samples were randomly divided into a calibration set of 40 samples and a prediction set of 30 samples. In 2009, another 180 Huping jujube samples were obtained from another jujube orchard, and were randomly divided into a calibration set of 150 samples and a prediction set of 30 samples. The spectroscopy acquisition and the reference measurement of SSC for jujube samples were finished the same day when the samples were obtained.

Collection of spectra

The used Vis-NIR reflectance spectral collecting system comprised a FieldSpec3 spectrophotometer with sample interval of 1 nm and sampling range of 350 to 2500 nm (Analytical Spectral Device, America), a halogen light (14.5 V), a calibration whiteboard and a laptop computer, as illustrated in Figure 1. The scanning frequency is 30 and the resolution is 3.5 nm. Diffuse reflection was used to acquire samples' spectra. Spectrophotometer was fixed above the jujube samples with a distance of 30 mm between probe and the surface of fresh jujubes. Two positions at the equator were chosen to obtain spectra of every sample. To avoid a low signal-to-noise ratio, only wavelength range of 400 to 2500 nm was considered for the spectral analysis. ASD View Spec Pro V 5.0, Unscramble V9.7 (Camo Process AS, Oslo, Norway) and DPS (Data Procession System For Practical Statistics, Zhejiang University, Hangzhou, China) software were adopted to analyze the spectral data.

Reference methods for SSC

The reference SSC of jujube samples was measured destructively using a digital refractometer (PR-101α Palette Sries, Atago Co., Ltd., Tokyo, Japan). Reference values of two equatorial positions were measured, respectively. The two obtained values were further averaged as the reference value of this sample.

Spectral pre-processing and chemometric calibration methods

Spectral pre-processing can sometimes improve the model's performance. Multiplicative scatter correction (MSC) is a transformation method used to compensate for additive and/or multiplicative effects in spectral data (Wu et al., 2008b). In this study, MSC method was used to reduce the effect due to the non-uniform size of samples.

PLS analysis method is a widely used chemometric analysis tool for spectral calibration (Svante et al., 2000; Wu et al., 2011a). PLS method can be used for the establishment of a regression model to perform the prediction of physiological concentrations. PLS shows the fundamental relations between the variable matrix Y (the properties of interest) and the variable matrix X (the spectra). PLS is particularly suited when variables are more than samples, and when there is multicollinearity among X values (Wu et al., 2010). In this study, PLS analysis was based on the spectra preprocessed by MSC.

Model evaluation

The performances of the PLS calibration models were evaluated by comparing the Vis-NIR predicted values and the reference SSC values, in terms of the statistic root mean square error of calibration, root mean square error of prediction and correlation coefficients (Wu

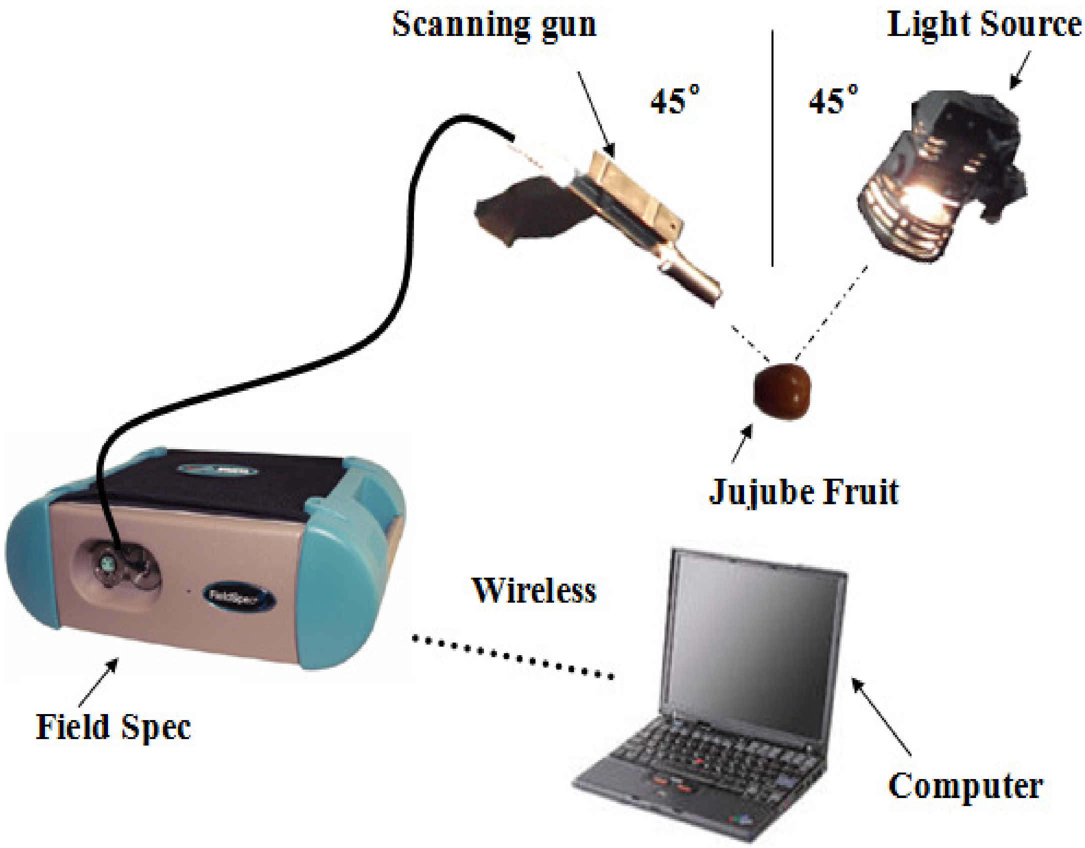

Figure 1. Visible near-infrared measurement equipment.

et al., 2011b), which are defined as follows:

Root mean square error of calibration (RMSEC)

It indicates the calibration error or calibration variance, thus the imprecision (quality) of the calibration model. The smaller the value, the better the model, if the value is too small, it indicates that the process of correction may be over-fitting:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,actual} - y_{i,predicted})^2}{n-1}}$$

where $y_{i,actual}$ is the measured value of calibration set sample, $y_{i,predicted}$ is the prediction value of calibration set sample, and n is the number of observation in calibration set.

Root mean square error of prediction (RMSEP)

This indicates the prediction error or prediction variance, thus the imprecision (quality) of the prediction model. The smaller the value, the better the prediction of the model:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{m}(y_{i,actual} - y_{i,predicted})^2}{m-1}}$$

where $y_{i,actual}$ is the measured value of prediction set sample, $y_{i,predicted}$ is the prediction value of prediction set sample, and m is the number of observation in prediction set.

Correlation coefficients (R) between the predicted and the measured value are calculated for both calibration and prediction set, which as follows:

$$R = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,actual} - y_{i,predicted})^2}{\sum_{i=1}^{n}(y_{i,actual} - \bar{y}_{actual})^2}}$$

where Mahalanobis is the measured value of samples in calibration and prediction sets. $\bar{y}_{actual}$ is the mean of the reference measurement results for all samples in calibration and prediction

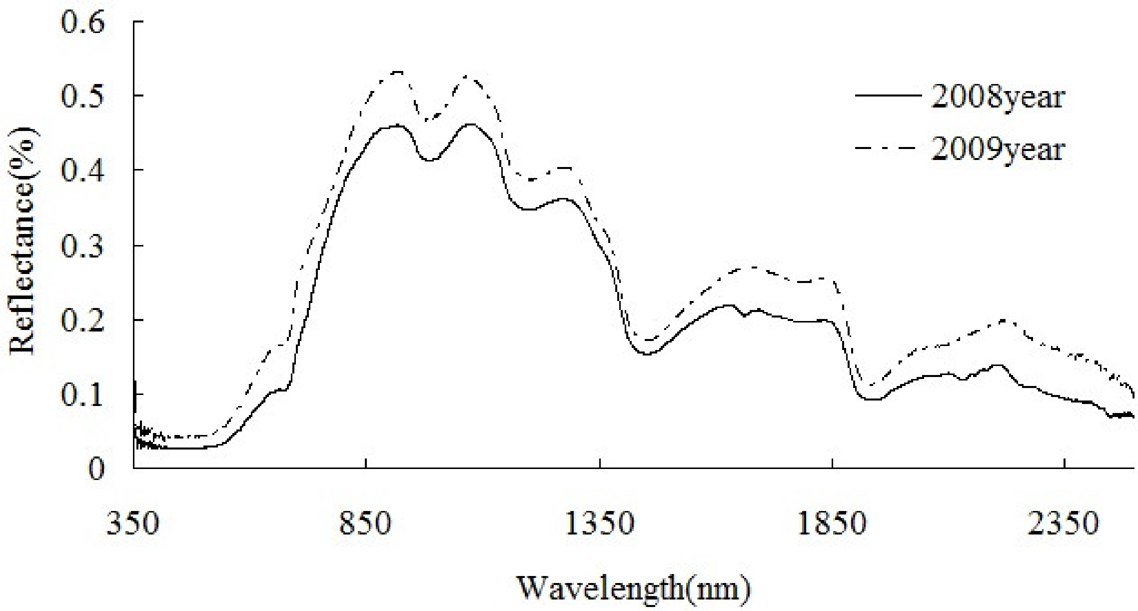

Figure 2. Typical reflectance spectroscopy of Huping jujube samples harvested in 2008 and 2009.

sets, and $y_{i,predicted}$ is the predicted value of samples in calibration and prediction sets. n is the number of samples in calibration and prediction sets.

RESULTS AND DISCUSSION

Analysis of measured Vis-NIR spectra and reference SSC of jujube samples

The typical reflectance spectra of Huping jujube is shown in Figure 2. It could be concluded that the reflectance spectroscopy of the jujube harvested in 2008 has similar trend with that of jujube harvested in 2009. They both have obvious absorption at the wavelengths of 987, 1190, 1450 and 1920 nm. The samples obtained in 2008 and 2009 were 70 and 180, respectively. The statistical values of reference SSC values for comparing the SSC range of the tested samples are shown in Figure 3.

PLS analysis of samples obtained in 2008

After spectral preprocessing of MSC, a PLS model (model-2008) was established based on the full range spectra of 40 jujube samples in the calibration set of 2008. A good result with correlative coefficient (Rc) of 0.9530 and RMSEC of 0.3951 was obtained (Figure 4). In addition, another 30 samples of prediction set was used to validate the performance of the established PLS model, resulting in a good prediction performance with correlative coefficient (Rp) of 0.9267 and RMSEP of 0.3767.

PLS analysis of samples obtained in 2009

Similar to the analysis of samples collected in 2008, another PLS model (model-2009) was established based on the full range spectra of 40 jujube samples in the calibration set of year 2009. On the basis of 150 samples within the calibration set, Rc of 0.8536 and RMSEC of 1.1410 was obtained (Figure 5). Another 30 samples of prediction set was used to validate the performance of the established PLS model, resulting in a good prediction performance with Rp of 0.9040 and RMSEP of 0.9108.

Analysis of PLS models with mixture sample sets

From the above results, it is obvious that the performances of two PLS models, which were established based on the samples collected in 2008 and 2009, respectively, were different. The accuracy of model-2009 is lower than that of model-2008. In order to improve the robustness of PLS model, jujube samples from different growth areas and years were mixed. Considering that the reference SSC values of samples obtained in 2009 are much widely distributed than those of samples collected in 2008 (Figure 3), five new calibration sample sets were generated by adding 5, 10, 20, 30 and 40 samples harvested in 2008 to the calibration set of model-2009, respectively. On the basis of these five new calibration sample sets, five PLS models were established using full range spectra and reference SSC values. The prediction process was conducted using 30 samples in the prediction set of year 2009. Their results are shown in

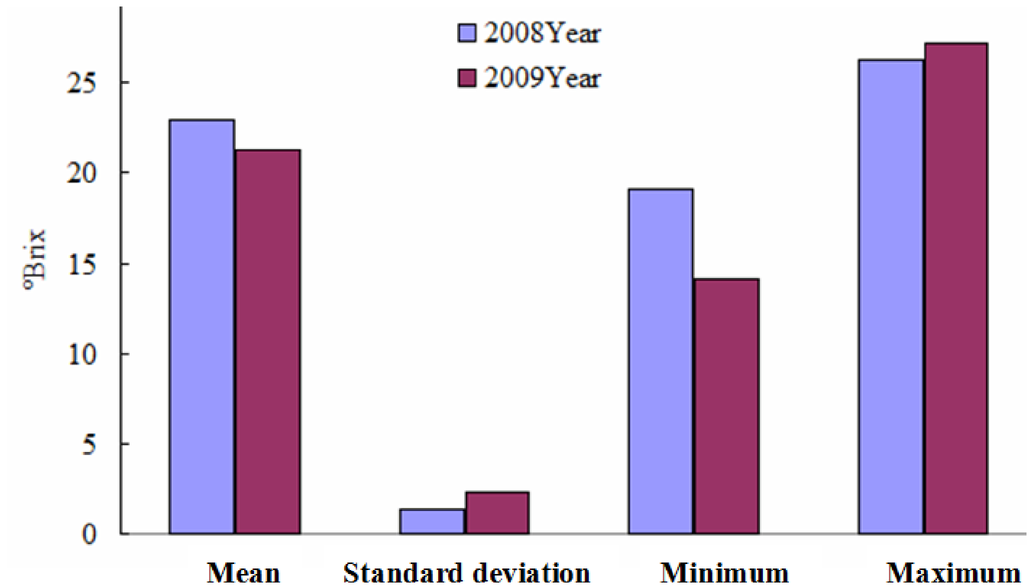

Figure 3. Statistic data of soluble solids content in Huping jujube samples.

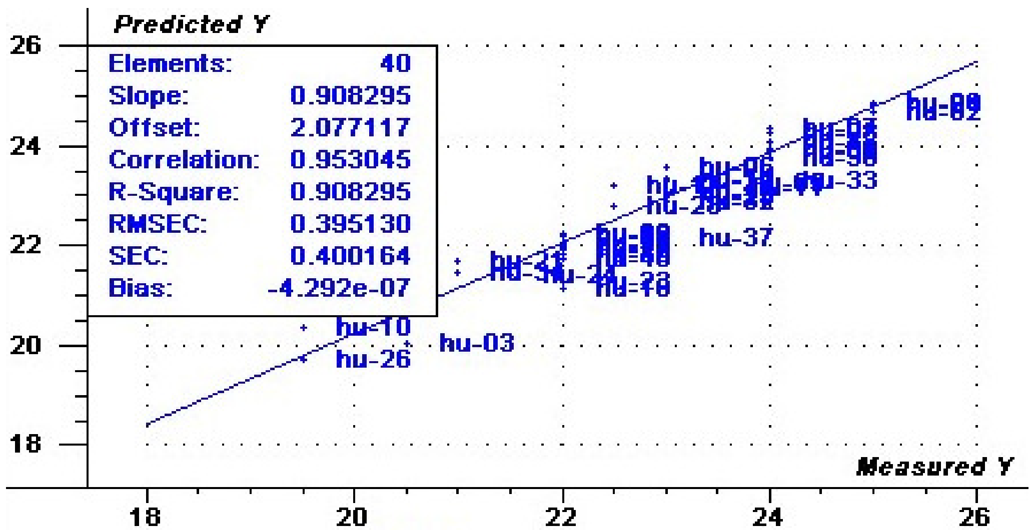

Figure 4. Prediction result of calibration model for determining soluble solids content of Huping jujube in 2008.

Table 1, and the scatter plots of reference values vs. predicted values of different prediction models are shown in Figure 6 (the plot of adding 5 was omitted).

As shown in Table 1, the prediction performances slowly improved with the increment of the number of added samples. The performances of models with added samples were better than that of the model with only samples of year 2009. On the other hand, the prediction ability of our model was stable when the number of added samples changed. The results of models with added samples for predicting 30 samples in the prediction set of year 2009 were similar. The lowest RMSEP of 0.7388 and highest Rp of 0.9391 were obtained when the number of added samples was 20. The results were better than the results of previous work which obtained the RMSEP of between 2.018 and 3.200 (WANG et al., 2010). However, the above research only analyzed the samples from one growth area and harvested in one year. When an external data set was gotten from different cultivars and harvest times, Rp of 0.9120 and RMSEP of 0.68 °Brix was obtained for the SSC determination of pear (Liu, et al., 2008) and another work obtained RMESP of 0.662 °Brix

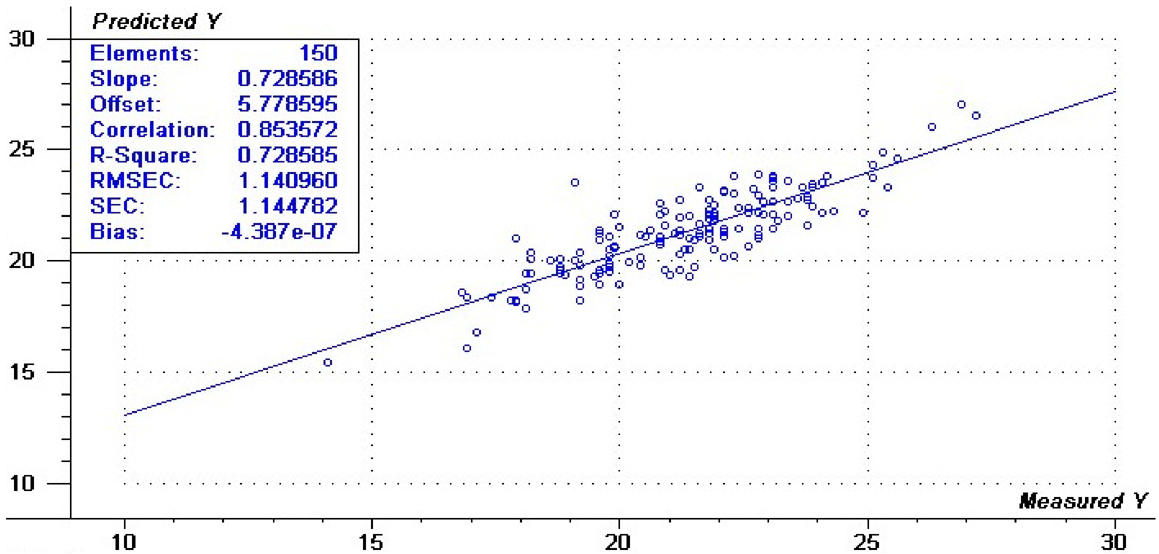

Figure 5. Prediction result of calibration model for determining soluble solids content of Huping jujube in 2009

Table 1. Results of PLS models established by adding different numbers of samples harvested in 2008 to the calibration set of model-2009 for determining soluble solids content of Huping jujube.

Number of added sample	Calibration		Prediction	
	Rc	RMSEC	Rp	RMSEP
5	0.8846	1.0248	0.9183	0.9011
10	0.8851	1.0111	0.9242	0.8798
20	0.8869	0.9951	0.9391	0.7388
30	0.8868	0.9792	0.9360	0.7751
40	0.8893	0.9645	0.9257	0.8661

PLS, Partial least square; root mean square error of calibration; Rc, correlation coefficient; Rp, correlative coefficient.

for predicting SSC in bell pepper (Pathompong et al., 2009). This is the first time to consider external data set for the SSC prediction of fresh jujube based on Vis-NIR spectroscopy. The results proved the possibility of using external data set to improve the accuracy and robustness of PLS model of SSC determination in jujube.

Conclusion

Different sample sets with jujube samples harvested in two orchards in two years of 2008 and 2009 were analyzed to see if the model's accuracy and robustness of determining SSC of fresh jujube using Vis-NIR spectroscopy technique could be improved. When only samples from one year were considered, the PLS model established based on samples from 2009 had better performance than the PLS model with samples from 2008. The location, year and quantity of samples affected the robustness of the model. By adding samples from 2008 to the sample set of 2009, better performances with high accuracy were obtained. When different samples from the year 2008 were added to the sample set of 2009, the models' performances were similar which shows a good robustness of the established models. The results of this work would also be helpful for the quality analysis of other fruits using Vis-NIR spectroscopy technique. In the next work, more samples of Huping jujubes from more years and regions would be considered for the model establishment. Other factors like operational conditions which can influence the establishment of spectral model should also be considered in the near future.

ACKNOWLEDGEMENT

This study was supported by the Doctoral specia l Fund of

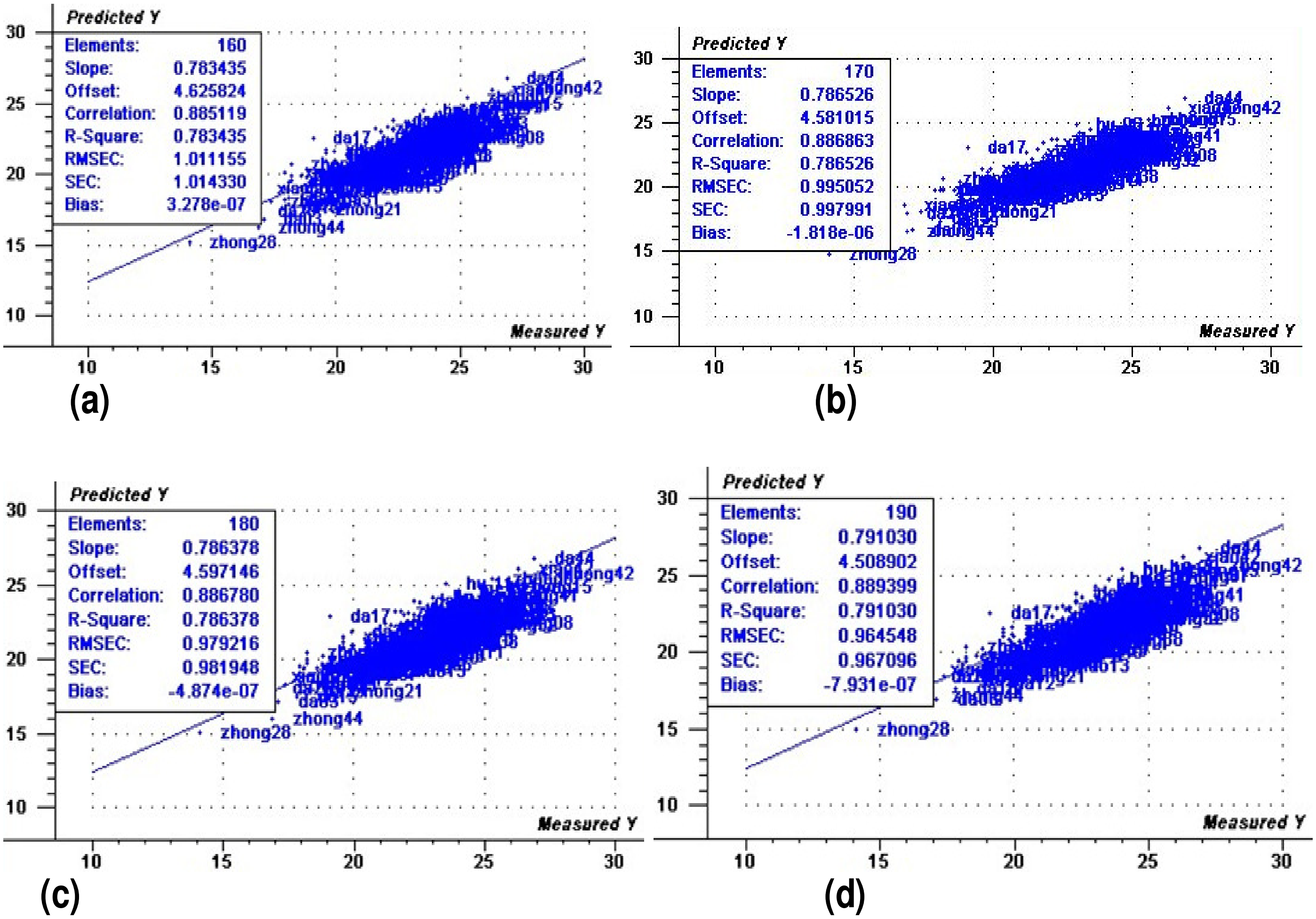

Figure 6. Results of calibration models established by adding different numbers of samples harvested in 2008. Mixed with 10 (a), 20 (b), 30 (c) and 40 (d) samples of year 2008.

Higher Education in China (20101403110003).

REFERENCES

Ali M, Mohammad HA, Ameneh S, Majid S (2010).Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. Biosyst. Eng. 106(3): 295-302.

Ann P, Jeroen T, Bert V, Paul D, Bart MN (2003). Effect of biological variability on the robustness of NIR models for soluble solids content of apples. Postharvest Biol. Technol. 28(2): 269-280.

Andrew McGlone V, Sumio Kawano (1998). Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. Postharvest Biol. Technol. 13(2): 131-141.

Bart MN, Katrien B,Els B, Ann P, Wouter S, Karen IT, Jeroen L (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol. Technol. 46(2): 99-118.

Els B, Anca-Sabina S, Mihai N, Jeroen L, Bart MN, Wouter S (2010).Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. Postharvest Biolo. Technol. 55(3): 133-143.

Esmé DL, Karen IT (2010). Robust prediction models for quality parameters in Japanese plums (*Prunus salicina* L.) using NIR spectroscopy. Postharvest Biol. Technol. 58(3): 176-184.

Li, J W, Fan, LP, Ding, SD, Ding, XL (2007). Nutritional composition of five cultivars of chinese jujube. Food Chem. 103: 454-460.

Liu YD, Ying YB, Fu XP, Lu HS (2007). Experiments on predicting sugar content in apples by FT –NIR Technigue. J. Food Eng. 80(3): 986-989.

Liu YD, Sun XD, Zhang HL, Ouyang AG (2010).Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled divice near infrared spectroscopy. Comput. Electron. Agric. 71 (S1):10-14.

Liu YD, Chen XM, Ouyang AG (2008).Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry. LWT-Food Sci. Technol. 41(9): 1720-1725.

Pathompong P, Els B, Bert EV, Bart MN, Wouter S (2009).Non-destructive measurement of firmness and solube solids content in bell pepper using NIR spectroscopy. J. Food Eng. 94(3-4): 267-273.

Sylvie B, David R, Maryse R , Barbara G, Dominique B, Jean-Marc A, Catherine MGCR (2009). Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. Food Chem. 113(4): 1323-1328.

Wang J, Nakano K,Ohashi S,Takizawa K, He JG (2010).Comparison of different modes of visible and near-infrared spectroscopy for detecting

internal insect infestation in jujubes. J. Food Eng. 101(1): 78-84.
Wu D, He, Y, Feng SJ, Sun DW (2008a). Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. J. Food Eng. 84(1): 124-131.
Wu D, He Y, Feng S (2008b). Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Analytica Chimica Acta. 610(2): 232-242.
Wu D, Chen, XJ, Shi PY, Wang SH, Feng FQ, He Y (2009a). Determination of alpha-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. Analytica Chimica Acta. 634(2): 166-171.
Wu D, He Y, Shi JH, Feng SJ (2009b). Exploring near and midinfrared spectroscopy to predict trace iron and zinc contents in powdered milk. J. Agric. Food Chem. 57(5): 1697-1704.
Wu, D, Wu, HX, Cai, JB, Huang, ZH, He, Y (2009c). Classifying the species of Exopalaemon by using visible and near infrared spectra with uninformative variable elimination and successive projections algorithm. J. Infrared Millim. 28(6): 423-427.
Wu D, He Y, Nie PC, Cao F, Bao YD (2010). Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. Analytica Chimica Acta. 659(1-2): 229-237.
Wu D, Nie PC, He Y, Bao YD (2011a). Determination of calcium content in powdered milk using near and mid-infrared spectroscopy with variable selection and chemometrics. Food Bioprocess Technol. In press, doi:10.1007/s11947-010-0492-4.
Wu, D, Nie, PC, Cuello, J, He, Y, Wang, ZP, Wu, HX (2011b). Application of visible and near infrared spectroscopy for rapid and non-invasive quantification of common adulterants in Spirulina powder. J. Food Eng. 102(3): 278-286.
Zhang SJ, Wang FH, Jie DF (2009).Non-destructive of the fresh jujube varieties and SSC by NIR spectroscopy[J]. Nongye Jixie Xuebao/Transactio. 40(4): 139-142.
Zheng H, Lu HF, Zheng YP, Lou HQ, Chen CQ (2010).Automatic sorting of Chinese jujube (*Zizyphus jujuba* Mill.cv. 'hongxing') using chlorophyll fluorescence and support vector machine. J. Food Eng. 101(4): 402-408.