CrossMark

Nondestructive Determination of Soluble Solids Content of 'Fuji' Apples Produced in Different Areas and Bagged with Different Materials During Ripening

Jinlei Dong[1] · Wenchuan Guo[1] · Zhuanwei Wang[1] · Dayang Liu[1] · Fan Zhao[1]

© Springer Science+Business Media New York 2015

Abstract To investigate the feasibility of hyperspectral imaging technique in nondestructive determination of soluble solids content (SSC) of fruits produced in different places and bagged with different materials during ripening, the near infrared hyperspectral reflectance images were acquired on 196 'Fuji' apples picked from four orchards in different areas and bagged with polyethylene film or light-impermeable paper. Mean reflectance spectrum from the regions of interest in the hyperspectral image of each apple was extracted. Standard normal variate (SNV) was used to eliminate the effect of instrument and environment on spectra. The sample set partitioning based on joint x–y distances method was applied to divide the samples into calibration set and prediction set as the ratio of 3:1. Successive projection algorithm (SPA) and uninformative variable elimination (UVE) method were used to select effective wavelengths (EWs) from the full spectra. Partial least squares (PLS), least squares support vector machine (LSSVM), and extreme learning machine (ELM) were used to develop SSC determination models. The results showed that 24 and 122 EWs were selected by SPA and UVE, respectively. The selection of EWs was helpful to SSC determination performance improvement. The optimal SSC prediction model was LSSVM based on selected EWs by SPA, with the correlation coefficient and root-mean-square error of prediction set of 0.878 and 0.908 °Brix, respectively. This study indicates that hyperspectral imaging technique could be used to determine SSC of intact apples produced in different places and bagged with different materials during ripening.

Introduction

Sugar content or soluble solids content (SSC, mostly sugars) is one of the most important fruit quality properties appealed to consumers. Traditional method used to measure fruit SSC requires samples from the internal tissues, and is carried out by using digital refractometer or Abbe refractometer. However, it is a destructive measurement. Developing a rapid, nondestructive method on measuring fruit SSC is welcome by fruit producers, processors, and distributors. Near infrared (NIR) spectroscopy is a fast and nondestructive analysis technique, and has been widely used for quality analysis of fruits, such as apples (Zou et al. 2007a), pears (Li et al. 2013), and peaches (Shao et al. 2011). However, NIR spectra were usually obtained on one point of a sample. Big difference in SSC values in different parts of a fruit, for example, higher than 2 °Brix difference in an apple (Zou et al. 2007b), causes the NIR spectroscopy could not offer the general SSC value of a whole fruit.

Hyperspectral imaging technique is a newly developed rapid detection technique. The acquired hyperspectral image is in a three-dimensional form, called *hypercube*, which has spectral and spatial information together (Kamruzzaman et al. 2012). This technique has potential in identifying and quantifying chemical constituents as well as their location or spatial distribution simultaneously. At present, the hyperspectral imaging technique has been used in fruits quality analysis including surface defects detection and internal qualities

✉ Wenchuan Guo
guowenchuan69@126.com; wencg915@sina.com

[1] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

determination. In SSC determination, apples (Huang and Lu 2010; Lu 2007; Mendoza et al. 2011; Noh and Lu 2007; Peng and Lu 2008), pears (Hong et al. 2007), peaches (Cen et al. 2012), strawberries (ElMasry et al. 2007), and blueberries (Leiva-Valenzuela et al. 2013) have been investigated. These studies have shown that it is feasible to predict SSC of fruits by using hyperspectral imaging technique.

China ranks the first in the world and Shaanxi Province ranks the first in China in apple production and acreage (Bai et al. 2012). Since the apple qualities are influenced by geographical location, climate, planting technique, etc. (Wei et al. 1998), the apples produced in different places have different internal qualities. On the other hand, bagging apples is a general technique used in apple planting to reduce production loss caused by insect and disease, and to improve apple appearance quality. Polyethylene film and light-impermeable paper which have different physicochemical characteristics are the two kinds of usually used bag materials in apple bagging treatment during fruit ripening in China (Wen and Ma 2006). Different bag materials cause the apples to have different internal qualities (Huang et al. 2009). Developing a universal model used to predict SSC of apples produced in different areas and bagged with different materials will be helpful to develop SSC detector with wide applications. Few studies have been reported on SSC prediction of apples produced in different areas and treated with different bagging materials. Therefore, the objectives of this study were to investigate the feasibility of hyperspectral imaging technique combined with chemometrics to detect the SSC values of intact postharvest 'Fuji' apples bagged with light-impermeable paper or polyethylene film and picked from four different orchards in Shaanxi Province, and to offer the optimal SSC determination model.

Materials and Methods

Samples

Matured 'Fuji' apples were picked from four different orchards in Shaanxi Province, China. Two orchards are located at Yangling District, Xi'an, one is located at Liquan County, and another is located at Baishui County. The apples picked from the two orchards located in Yangling District were bagged with polyethylene film, and other apples were bagged with light-impermeable paper during apple ripening. After the apples were transported to a laboratory, 10–12 apples were stored in a polyethylene bag with air holes at room temperature (22±2 °C). Measurements were taken initially and at 1-week interval during the 5-week storage period. Before the experiment, about 10 intact apples from each orchard were washed with tap water to remove any foreign materials on surface and wiped dry. Then, they were marked and used for

acquiring hyperspectral images and for SSC measurement. In total, 196 apples were used in the study, including 49 apples from each orchard. That is, 98 apples were bagged with polyethylene film, and the other 98 apples were bagged with light-impermeable paper during apple ripening. Table 1 lists the orchard location, bag materials, and the sample number of apples used in the study.

NIR Hyperspectral Imaging System

A NIR hyperspectral reflectance imaging system (Fig. 1) in the spectral range of 900–1700 nm was used to acquire the hyperspectral images of apples. The system mainly consists of a high-performance 8-bit charged couple device (CCD) camera (OPCA05G, Hamamastu, Japan) coupled with a camera lens, an imaging spectrograph (Imspector N17E, Spectral Imaging Ltd., Oulu, Finland), an illumination unit equipped with four 100 W halogen lamps at angle of 45° (HSIA-LS-TAIF, Zolix instruments Co., Ltd., Beijing, China), a translation platform (PSA200-11-X, Zolix Instruments Co., Ltd., Beijing, China), a dark box, a stepper motor, a motor controller, a computer and data acquisition software Spectra SENS (Zolix Instruments Co., Ltd., Beijing, China). The distance between the CCD camera lens and translation platform was fixed at 65 cm, and the exposure time of the camera was set as 10 ms. Apples were placed on the translation platform operated by the stepper motor moving at a speed of 20 mm/s. NIR hyperspectral image acquisition was finished by Spectral SENS and carried out at room temperature (22±2 °C). Each image was acquired as a three-dimensional image (x, y, λ) which includes 320×250 pixels in spatial dimension (x, y) and 256 spectral bands from 865.11 to 1711.71 nm with 3.32 nm interval between contiguous bands in spectral dimension (λ).

Procedures

To obtain repeatable readings, the NIR hyperspectral reflectance imaging system was turned on and kept in a standby condition for at least 1 h. Then, the system was calibrated with white reference using a standard Teflon white board with 99 % reflectance and dark reference by closing the lens cap. One NIR hyperspectral image was obtained at blush side, and another image was obtained at the opposite of the blush side. After completion of the NIR hyperspectral image acquisition on each apple, a peeler was used to remove the peel on four spots in the equatorial region of the apple with an interval of 90°. Two spots were on the blush side and non-blush side which were used for hyperspectral image acquisition. The pulp at each spot was put into a garlic press to squeeze juice for measuring SSC using a digital refractometer (PR-101α, Atago Co., Ltd., Japan). One reading was recorded for one

Table 1 Statistics for the SSC values of apples

Orchard location	Bag materials	Total				Calibration set			Prediction set		
		Num. of samples	Min. /°Brix	Max. /°Brix	Mean±SD[a] /°Brix	Num. of samples	Min. /°Brix	Max. /°Brix	Num. of samples	Min. /°Brix	Max. /°Brix
Yangling	Polyethylene film	49	9.0	14.0	11.3±1.3 a[b]	37	9.0	14.0	12	9.5	13.1
Yangling		49	9.4	16.2	13.1±2.1 b	37	9.4	16.1	12	10.5	16.2
Baishui	Light-impermeable paper	49	8.1	17.7	14.7±1.8 c	37	8.1	17.7	12	12.8	16.1
Liquan		49	9.6	16.5	13.7±1.6 b	37	9.6	16.5	12	11.9	16.0

[a] SD represents standard deviation

[b] Represents a column followed by different letters are significantly different at $p < 0.05$

spot. The mean of four readings at four spots was used as the SSC value of the apple.

Image Extraction

To eliminate the differences in camera quantum and physical configuration of hyperspectral imaging systems, the original image X_O was calibrated with dark reflectance image D and with white reflectance image W. The corrected image X was calculated using Eq. (1).

$$X = \frac{X_O - D}{W - D} \times 100\% \qquad (1)$$

A region of interest (ROI) with the size of 40 pixel× 40 pixel in the center of each apple image was manually identified. The average reflectance spectrum of the ROIs in two images of each apple was calculated and used for further analysis. The calculation was carried out using Environment for Visualizing Images software (ENVI 4.8) software (ITT Visual Information Solutions, Boulder, CO, USA).
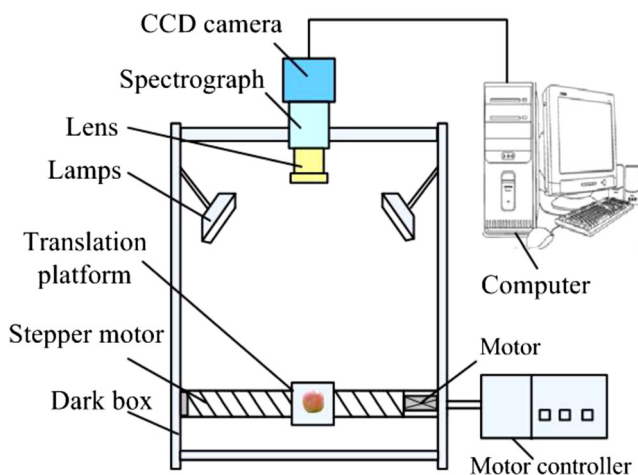

Fig. 1 The schematic diagram of applied NIR hyperspectral imaging system

Spectral Preprocessing

To reduce or even eliminate the effects on images arising from instrument and measurement environment, the spectral pre-processing methods, such as standard normal variate (SNV), multiplicative scatter correction, first and second derivatives were used to preprocess spectra. It was found that SNV could offer better SSC determination performance than other three methods. Therefore, SNV was used to preprocess spectra here. SNV preprocessing was carried out using the Unscrambler 9.7 (CAMO, Trondheim, Norway).

Sample Division

The method used for data partitioning was the sample sets partitioning method based on the joint x–y distances (SPXY), developed by Galvão et al. (2005) and extended from the classic Kennard-Stone (KS) algorithm. Contrasted with KS algorithm, SPXY considered spectra and reference values together. In this study, the 49 samples picked from each orchard were divided into calibration set and prediction set according to the ratio of 3:1, that is 37 apples in calibration set and 12 apples in prediction set. Totally, the calibration set and the prediction set had 148 (74 apples bagged with light-impermeable paper and 74 apples bagged with polyethylene film) and 48 apples (24 apples bagged with light-impermeable paper and 24 apples bagged with polyethylene film), respectively.

Effective Wavelengths Selection

An immense amount of spectral data is generated during hyperspectral imaging acquisition. Some spectral bands are highly correlated, and some contiguous spectral bands may contain redundant information. Therefore, it is essential to find the most influential effective wavelengths (EWs) on quality evaluation. In this study, successive projections algorithm (SPA) and uninformative variable elimination (UVE) were applied to select EWs from the full spectra (FS).

Successive Projections Algorithm

SPA is a forward variable selection technique specifically designed to improve the condition of multiple linear regression by minimizing collinearity effects in the calibration set (Galvão et al. 2008). After a wavelength is chosen, another new wavelength is added each time until the preset wavelength number is reached. The selected new wavelength has the maximum projection value on the orthogonal sub-space of the previous selected variables. The optimal number of wavelengths could be determined based on the smallest root-mean-square error of calibration set (RMSEC) in multiple linear regression calibration. The detailed description of SPA can be found in the literature (Galvão et al. 2008).

Uninformative Variable Elimination

UVE is a variable selection method based on the stability analysis of the regression coefficients of partial least squares (PLS) models. A stability value S and *cutoff* threshold are used to evaluate the dependability of each variable. S is defined as:

$$S_i = \frac{\text{mean}(b_i)}{\text{std}(b_i)}; \ i = 1, \ 2, \dots, m \tag{2}$$

where S_i is the stability of the ith variable of calibration sample, b_i denotes the regression coefficient of the ith variable of calibration sample. Mean(b_i) and std(b_i) are the mean and standard deviation of b_i, respectively, and m is the number of input variables. In this study, m is the number of used wavelengths to analyze, and it was 211 (see Section 3.2). An artificial random variables matrix with the same size as the spectral matrix is added to the original data set to compute their stability. The variables whose stabilities are less than a threshold are regarded as uninformative and are eliminated. The *cutoff* threshold is calculated as:

$$\text{cutoff} = k \times \max(\text{abs}(S_{\text{random}})) \tag{3}$$

where k is the contribution coefficient, and S_{random} expresses the stability of random variables matrix. More details on the UVE process can be found elsewhere (Centner et al. 1996).

Modeling Methods

Partial Least Squares

PLS regression is a widely utilized multianalysis and regression algorithm in present chemometric analysis. In PLS regression, an orthogonal on the basis of latent variables (LVs) is constructed one by one in such a way that they are oriented to along directions of maximal covariance between the spectral matrix and the variable matrix (Guidetti et al. 2010). The LVs were determined using leave-one-sample-out cross-

validation (LOOCV) method combined with the smallest root-mean-square error of cross-validation (RMSECV). The calculation of RMSECV is as follows (Chen et al. 2012):

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_{\backslash i} - y_i\right)^2}{n}} \tag{4}$$

where y_i is the measured SSC value of ith sample, $\hat{y}_{\backslash i}$ is the estimated SSC value for sample i when the model is constructed with sample i removed, n is the number of samples in calibration set. In this study, n is 148.

Least Squares Support Vector Machine

LSSVM is a modified algorithm of the standard support vector machine. It is capable of dealing with linear and nonlinear multivariate analysis and resolving these problems in a relatively fast way. The LSSVM regression model can be expressed as follows (Li et al. 2013):

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{5}$$

where $y(x)$ is the predicted values of calibration set, $K(x, x_k)$ is the kernel function, x_k is the input vector, α_k is the Lagrange multiplier called support value, b is bias, and N is the number of samples in calibration set. In this study, $y(x)$ is predicted SSC values and N is 148. Linear kernel, polynomial kernel, and radial basis function (RBF) kernel could be used as kernel function. Contrasted with linear and polynomial kernel functions, RBF kernel function is more able to reduce the computational complexity of training procedure and could handle the nonlinear relationships between the spectra and target attributes (Liu and He 2009). The procedure of LSSVM was performed in Matlab 2013b (Math Works, Natick, MA, USA) based on the free LSSVM toolbox (LSSVM v1.5, Suykens, Leuven, Belgium).

Extreme Learning Machine

ELM is a learning algorithm for single-hidden layer feedforward networks proposed by Huang et al. (2006). It can provide good generalization performance at an extremely fast speed. The main ELM algorithm can be summarized as follows: (1) assign the number of nodes in the hidden layer and randomly generate parameters of weights (w_i) and bias (b_i) for hidden nodes; (2) choose a suitable activation function and calculate the hidden layer output matrix H; (3) calculate the weights β of the output layer as: $\beta = H^+ T$, where H^+ is the Moore-Penrose generalized inverse of the hidden layer output matrix H, T is the target matrix.

Model Assessment

The model performances were evaluated by correlation coefficient of calibration set (R_c), correlation coefficient of prediction set (R_p), root-mean-square error of calibration set (RMSEC), and root-mean-square error of prediction set (RMSEP). A good model should have high values of R_c and R_p, and low values of RMSEC and RMSEP. Moreover, the number of input variables of the established model should be as low as possible.

Results and Discussion

Statistics of Samples

The statistics for the SSC values of total apples and apples in calibration and prediction sets, which were produced in different areas and bagged with different materials are listed in Table 1. Also, the results of the analysis of variance (ANOVA) on SSC of the total apples of four orchards are listed, too. The results show that even the apples produced in the same area (Yangling) and bagged with same material (polyethylene film), their SSC values still had significant differences at the significance level of 5 %. Although the apples produced in different areas (Yangling and Liquan) and bagged with different materials (polyethylene film and light-impermeable paper), their SSC values had not significant difference. The sample division results shows that the SSC values in the calibration set cover the ranges those in the prediction set, indicating the sample division was reasonable.

Spectral Features

Since heavy noises existed at both ends of the spectral range, spectrum data below 928.19 nm and above 1625.39 nm were discarded, leading to spectra within the range of 928.19–1625.39 nm with 211 bands for further analysis. Figure 2 shows the spectra preprocessed by SNV for the 196 apples between 928.19 nm and 1625.39 nm. It was found that the first absorption peak was in the region of 960–980 nm, which was likely attributed to the combination effect of second overtone of bond O–H from carbohydrates and water (Leiva-Valenzuela et al. 2013). The second absorption peak was at about 1200 nm, which was likely attributed to C–H stretching second overtone from carbohydrates (fructose, sucrose, and glucose). Another strong absorption peak found at 1450 nm was due to water absorption bands related to O–H stretching first overtone, indicating moisture dominates the spectral signatures (Cen and He 2007).
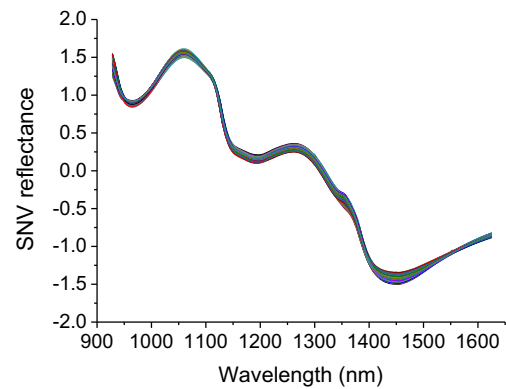

Fig. 2 The spectra preprocessed by SNV of 196 samples

Linear Correlation Between Reflectance Value and SSC

Figure 3 shows the linear correlation coefficient between the reflectance value and SSC. It was noted that there was negative linear relationship between the values of reflectance and SSC with coefficient higher than −0.41 and lower than −0.30, and the wavelengths at the absorption peaks had high correlation to SSC. However, no single wavelength was strongly correlated with SSC, indicating the necessity to use more wavelengths or even full spectra for SSC prediction. Therefore, chemometrics was used to select EWs, and some modeling methods were used to build nonlinear SSC determination models.

Effective Wavelengths Selection

Effective Wavelengths Selected by SPA

Figure 4 shows the calculated RMSEC at different number of selected EWs by SPA from 1 to 40 with an interval of 1. The RMSEC decreased with the number of selected EWs when the number was lower than 33, and the smallest RMSEC (0.839 °Brix) was found at 33 variables (Fig. 4). Since more variables included in model will decrease model calculation speed, the number of selected EWs was determined according to an *F*
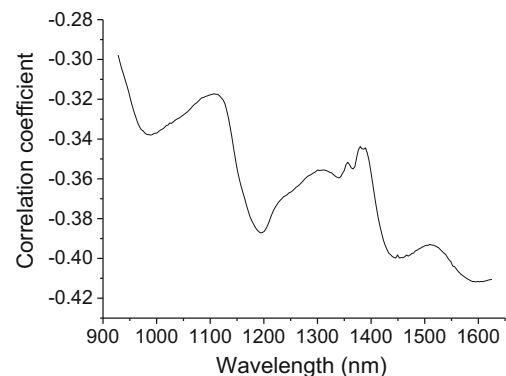

Fig. 3 Linear correlation coefficient between SSC and reflectance value at each wavelength from 928.19 to 1625.39 nm
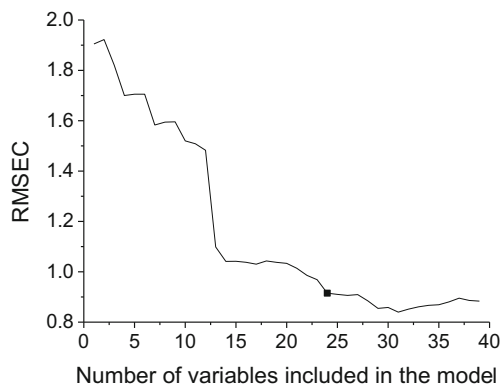
Fig. 4 Changed RMSEC with increasing number of variables in SPA

Fig. 6 Stability distribution of each variable for SSC predication by UVE with 4 LVs in PLS

test criterion with a significance level $\alpha=0.25$, suggested by Galvão et al. (2008). Finally, 24 EWs (RMSEC=0.915 °Brix) were selected. They are 928.2, 938.2, 961.4, 964.7, 977.0, 974.7, 994.6, 1011.2, 1017.8, 1021.2, 1077.6, 1097.5, 1104.2, 1120.8, 1353.2, 1376.4, 1396.3, 1466.0, 1549.0, 1569.0, 1572.3, 1575.0, 1578.9, and 1598.8 nm.

Effective Wavelengths Selected by UVE

Since the variables selected by UVE were influenced by the number of LVs in PLS, the RMSECV was calculated when the LVs was set from 1 to 20 with an interval of 1, shown in Fig. 5. When the number of LVs was 4, the RMSECV was the smallest (2.091 °Brix). So, the optimal number of LVs was 4 in this study. According to the stability value S of random variables in Eq. (3), the *cutoff* threshold in Eq. (4) was calculated as 28.46. Figure 6 shows the stability of each wavelength in the spectra for SSC prediction by UVE with four latent variables. The input wavelengths were at the left of the vertical line, while random wavelengths were at the right side. The *cutoff* threshold of UVE is indicated by dashed lines. The wavelengths whose stabilities were within the *cutoff* lines were treated as uninformative and eliminated. Finally, 122 EWs were selected by UVE. The EWs were mainly distributed in the wavelength bands of 928.19–951.43 nm, 1021.15–
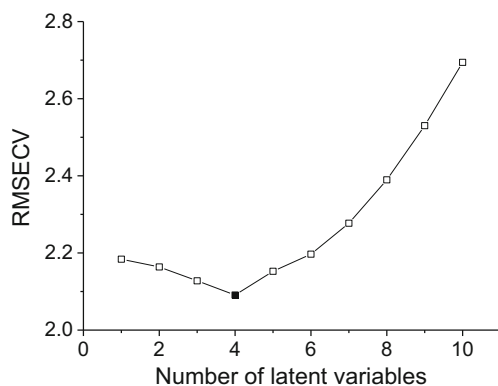
1080.91 nm, 1163.91–1207.07 nm, 1246.91–1339.87 nm, 1456.07–1539.07 nm, and 1562.31–1625.39 nm.

Modeling Results

PLS Modeling Results

LOOCV method combined with RMSECV was used to determine the optimal number of LVs in SSC determination model of PLS. The calculated RMSECV values at different number of LVs from 1 to 20 with an interval of 1 at different EWs selection methods are shown in Fig. 7. The determined numbers of LVs were decided by the smallest RMSECV, shown in Table 2. Table 3 lists the SSC determination results using PLS at different EWs selection methods. The PLS model based on selected EWs by UVE (UVE-PLS) had the highest R_p (0.863) and the lowest RMSEP (0.956 °Brix), but lower R_c (0.744) and higher RMSEC (1.460 °Brix). SPA-PLS had the highest R_c (0.752), the lowest RMSEC (1.441 °Brix), higher R_p (0.861) and lower RMSEP (0.971 °Brix). FS-PLS had the poorest calibration and prediction performance. Since the EWs selected by SPA was only 24, much less than 122 selected by UVE, SPA-PLS was regarded as the best PLS model in apple SSC prediction.
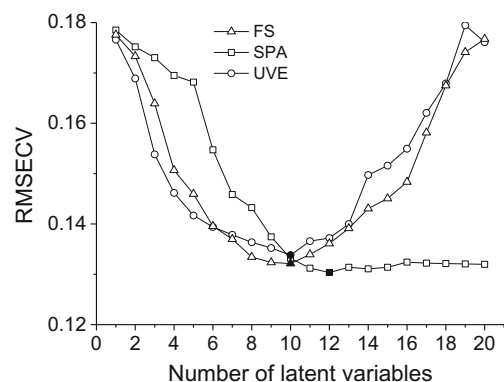

Fig. 5 Changed RMSECV with the number of LVs in PLS of UVE

Fig. 7 Changed RMSECV with the number of LVs in PLS model at different EWs selection methods

Table 2 Parameters of PLS, LSSVM, and ELM

Variables selection methods	PLS	LSSVM		*ELM*
	Number of LVs	γ	σ^2	Number of hidden layer nodes
FS	10	2.44×10^5	5.33×10^4	50
SPA	12	2.74×10^6	4.13×10^3	40
UVE	10	1.75×10^7	6.61×10^5	45

LSSVM Modeling Results

Determining the proper kernel function and the best kernel parameters are two crucial problems that need to be solved before LSSVM modeling. In this study, radial basis function (RBF) kernel was used as the kernel function. Regularization parameter γ and the RBF kernel function parameter σ^2 are two key parameters. γ determined the trade-off between minimizing the training error and minimizing model complexity. σ^2 was the bandwidth parameter. Small σ^2 produces a large number of regressors and may lead to overfitting. On the contrary, large σ^2 could simplify the model, but may lead to lower accuracy (Wu et al. 2008). Simplex technique and tenfold cross-validation were applied to find the optimal parameter values. The determined parameter values are listed in Table 2. The SSC determination performance of LSSVM at different EWs selection methods are listed in Table 3. Obviously, SPA-LSSVM had the highest R_c (0.791) and R_p (0.878), and the lowest RMSEC (1.339 °Brix) and RMSEP (0.908 °Brix), indicating SPA-LSSVM was the best LSSVM model. UVE-LSSVM was superior to FS-LSSVM in prediction performance, but poorer than FS-LSSVM in calibration performance.

ELM Modeling Results

The activation function used here was *sigmoid* function. The number of hidden layer nodes was obtained by a trial and error method (Lan et al. 2010). It was set from 1 to 100, and increased by 5 from 1. The optimal number was determined based on the lowest RMSEC, shown in Table 2. Since the initial weights were generated randomly, in order to overcome the problem of unstable results, each ELM model at each EWs selection methods was repeated 1000 times. The average R_c, R_p, RMSEC, and RMSEP in 1000 times repetition were calculated and listed in Table 3. Table 3 shows that SPA-ELM had the highest R_c (0.802) and R_p (0.816), and the lowest RMSEC (1.304 °Brix) and RMSEP (1.121 °Brix), indicating SPA-ELM was the best ELM model. Although FS-ELM had higher R_c (0.783) and lower RMSEC (1.358 °Brix), it had the lowest R_p (0.769) and the highest RMSEP (1.267 °Brix), meaning the stability of FS-ELM was poor. UVE-ELM had the worst calibration performance, but had better prediction performance than FS-ELM. Since the number of hidden layer nodes of SPA-ELM was 40, less than 45 of UVE-ELM and 50 of FS-ELM, it can be said that SPA-ELM is capable of improving calculation speed.

Comparison

When comparing the performance of different models at different EWs selection methods for determining SSC of apples, it was found that the PLS and LSSVM models had better prediction performance than ELM, and the EWs selected by SPA could offer better calibration and prediction performance than FS and UVE. This illustrates that EWs is helpful in extracting useful information from full spectra. In all developed models, SPA-LSSVM had the highest R_p and the lowest

Table 3 SSC determination results of PLS, LSSVM, and ELM models based on different variables selection methods

Modeling method	Variables selection method	Calibration set		Prediction set	
		R_c	RMSEC/°Brix	R_p	RMSEP/°Brix
PLS	FS	0.739	1.472	0.839	1.051
	SPA	0.752	1.441	0.861	0.971
	UVE	0.744	1.460	0.863	0.956
LSSVM	FS	0.767	1.404	0.852	0.995
	SPA	0.791	1.339	0.878	0.908
	UVE	0.765	1.409	0.863	0.960
ELM	FS	0.783	1.358	0.769	1.267
	SPA	0.802	1.304	0.816	1.121
	UVE	0.774	1.383	0.791	1.191

RMSEP, as well as higher R_c and lower RMSEC. Moreover, the number of selected EWs by SPA was 24, much less than 122 by UVE and 211 by FS. It can be concluded that SPA-LSSVM was the best SSC prediction model in this study.

To investigate the effect of bag materials on SSC prediction of 'Fuji' apples, the SSC of the apples bagged with polyethylene film and light-impermeable paper were predicted using SPA-LSSVM, respectively. It showed that R_p and RMSEP of the apples bagged with polyethylene film were 0.919 and 0.865 °Brix , respectively, and the values of the apples bagged with light-impermeable paper were 0.860 and 0.951 °Brix, respectively. Obviously, the prediction performance for the apples bagged with polyethylene film was better than those bagged with light-impermeable paper. The reason that causes this different prediction performance and the correction of this result will be investigated in more varieties of apples in the future.

Comparing to other reported studies on SSC prediction for apples based on hyperspectral imaging technique, the obtained prediction performance of SPA-LSSVM here was superior to the results reported by Huang and Lu (2010) (R_p=0.822, RMSEP=0.78), by Mendoza et al. (2011) (R_p=0.66–0.88 and SEP=0.7–0.9), and by Noh and Lu (2007) (R_p=0.66, SEP= 1.19), but poorer than that obtained by Lu (2007) (R^2=0.79, SEP=0.72) for 'Gold Delicious' apple, by Peng and Lu (2007) (R_p=0.816, SEP=0.92) and by Huang et al. (2013) (R^2=0.9501, RMSEP=0.3087). Although the apple used in the study was of 'Fuji' variety only, different production areas and different bagging materials caused the apples had much different SSC values, which restrict the improvement of model prediction performance. Contrasted with the model developed for the samples produced in one area or bagged with the same material during ripening, the developed models considering production areas and bag materials should be used for more apples. Further research is needed to establish a more universal model considering more production areas and different bag materials for different varieties of apples.

Conclusions

NIR hyperspectral reflectance images were acquired from 900 to 1700 nm on 196 'Fuji' apples picked from four orchards (49 from each orchard) and bagged with polyethylene film and light-impermeable paper, and the SSC of apples were measured with digital refractometer. According to the sample division method of SPXY, 148 apples were used for calibration and other 48 samples were used for prediction. Twenty-four and 122 EWs were selected by SPA and UVE from the full spectra with 211 bands from 928.19 and 1625.39 nm, respectively. The EWs selected by SPA could offer better calibration and prediction performance than FS and UVE. ELM had the best calibration performance while LSSVM had the best prediction performance. Generally, LSSVM models were

superior to PLS and ELM models in apple SSC prediction. The optimal SSC prediction model was SPA-LSSVM with R_c, R_p, RMSEC, and RMSEP of 0.791, 0.878, 1.339 °Brix, and 0.908 °Brix, respectively. This study demonstrates that hyperspectral imaging technique is capable of predicting SSC values of apples produced in different areas and bagged with different materials during on-tree ripening.

Compliance with Ethics Requirements

Conflict of Interest Jinlei Dong has no conflict of interest. Wenchuan Guo declares that she has no conflict of interest. Zhuanwei Wang, Dayang Liu and Fan Zhao have no conflict of interest. This article does not contain any studies with human or animal subjects.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

Bai S, Bi J, Wang P, Gong L, Wang X (2012) Fruit quality analysis of different apple varieties. Food Sci 33(17):68–72

Cen H, He Y (2007) Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends Food Sci Technol 18(2):72–83

Cen H, Lu R, Mendoza F, Ariana D (2012) Assessing multiple quality attributes of peaches using optical absorption and scattering properties. Trans ASABE 55(2):647–657

Centner V, Massart DL, de Noord OE, de Jong S, Vandeginste BM, Sterna C (1996) Elimination of uninformative variables for multivariate calibration. Anal Chem 68(21):3851–3858

Chen Q, Ding J, Cai J, Zhao J (2012) Rapid measurement of total acid content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. Food Chem 135(2):590–595

ElMasry G, Wang N, ElSayed A, Ngadi M (2007) Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. J Food Eng 81(1):98–107

Galvão RKH, Araujo MCU, José GE, Pontes MJC, Silva EC, Saldanha TCB (2005) A method for calibration and validation subset partitioning. Talanta 67(4):736–740

Galvão RKH, Araújo MCU, Fragoso WD, Silva EC, José GE, Soares SFC, Paiva HM (2008) A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. Chemom Intell Lab Syst 92(1):83–91

Guidetti R, Beghi R, Bodria L (2010) Evaluation of grape quality parameters by a simple Vis/NIR system. Trans ASABE 53(2):477–484

Hong T, Qiao J, Wang N, Ngadi MO, Zhao Z, Li Z (2007) Nondestructive inspection of Chinese pear quality based on hyperspectral imaging technique. Trans CSAE 23(2):151–155 (in Chinese)

Huang M, Lu R (2010) Optimal wavelength selection for hyperspectral scattering prediction of apple firmness and soluble solids content. Trans ASABE 53(4):1175–1182

Huang G, Zhu Q, Siew C (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501

Huang C, Yu B, Teng Y, Su J, Shu Q, Cheng Z, Zeng L (2009) Effects of fruit bagging on coloring and related physiology, and qualities of red Chinese sand pears during fruit maturation. Sci Hortic 121(2):149–158

Huang W, Li J, Chen L, Guo Z (2013) Effectively predicting soluble solids content in apple based on hyperspectral imaging. Spectrosc Spectr Anal 33(10):2843–2846 (in Chinese)

Kamruzzaman M, ElMasry G, Sun D-W, Allen P (2012) Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. Anal Chim Acta 714:57–67

Lan Y, Soh YC, Huang G-B (2010) Constructive hidden nodes selection of extreme learning machine for regression. Neurocomputing 73(16–18):3191–3199

Leiva-Valenzuela GA, Lu R, Aguilera JM (2013) Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. J Food Eng 115(1):91–98

Li J, Huang W, Zhao C, Zhang B (2013) A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. J Food Eng 116(2):324–332

Liu F, He Y (2009) Application of successive projections algorithm for variable selection to determine organic acids of plum vinegar. Food Chem 115(4):1430–1436

Lu R (2007) Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images. Sens & Instrumen Food Qual 1(1):19–27

Mendoza F, Lu R, Ariana D, Cen H, Bailey B (2011) Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. Postharvest Biol Technol 62(2):149–160

Noh HK, Lu R (2007) Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. Postharvest Biol Technol 43(2):193–201

Peng Y, Lu R (2007) Prediction of apple fruit firmness and soluble solids content using characteristics of multispectral scattering images. J Food Eng 82(2):142–152

Peng Y, Lu R (2008) Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. Postharvest Biol Technol 48(1):52–62

Shao Y, Bao Y, He Y (2011) Visible/near-infrared spectra for linear and nonlinear calibrations: a case to predict soluble solids contents and pH value in peach. Food Bioprocess Technol 4(8):1376–1383

Wei Q, Li J, Shu H (1998) Relationships between fruit quality and ecological factors in apple. J Shandong Agric Univ 24(4):532–536 (in Chinese)

Wen Y, Ma F (2006) An advance in application and study of apple bagging technology in China. J Northwest A and F Univ Nat Sci Ed 34(2):100–104 (in Chinese)

Wu D, He Y, Feng S, Sun D-W (2008) 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

Zou X, Zhao J, Huang X, Li Y (2007a) Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models. Chemom Intell Lab Syst 87(1):43–51

Zou X, Zhao J, Li Y (2007b) Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of 'Fuji' apple based on BiPLS and FiPLS models. Vib Spectrosc 44(2):220–227