

Nondestructive Determination of Apple Internal Qualities Using Near-Infrared Hyperspectral Reflectance Imaging

Jinlei Dong¹ · Wenchuan Guo¹

Received: 2 January 2015 / Accepted: 30 March 2015 / Published online: 11 April 2015
© Springer Science+Business Media New York 2015

Abstract Hyperspectral reflectance imaging technology in near-infrared regions (900–1,700 nm) was used to evaluate soluble solids content (SSC), firmness, moisture content (MC), and pH values of ‘Fuji’ apples during a 13-week storage period. Totally, 167 apples were divided into calibration set (125) and prediction set (42) based on the joint x - y distance sample set partitioning method. Mean spectrum of the regions of interest in the hyperspectral image of each apple was used for analysis. Two typical variable selection methods, i.e., successive projection algorithm (SPA) and uninformative variable elimination (UVE), were applied to extract the characteristic variables from full spectra (FS). The partial least squares (PLS) regression, least squares support vector machine (LSSVM), and backpropagation (BP) network modeling methods were used to establish models to predict SSC, firmness, MC, and pH of apples, respectively. The results showed that the SSC and MC could be predicted exactly by all developed models, and SPA-LSSVM and FS-BP could be used to predict pH value roughly. All models failed to predict firmness. The SPA-LSSVM model had better comprehensive ability in determining SSC, MC, and pH than others, with the correlation coefficient of prediction of 0.961, 0.984, and 0.882 and residual predictive deviation of 3.49, 5.51, and 2.06, respectively. The results demonstrated the feasibility of using near-infrared hyperspectral reflectance imaging technology as a non-invasive method for predicting SSC, MC, and pH of apples simultaneously.

Keywords Apple · Hyperspectral imaging · Soluble solids content · Firmness · Moisture content · pH value

Introduction

Apple is one of the most popular and the most important cash fruits in the world. It originated in Central Asia and Europe (Harris et al. 2002). Now, China is the largest cultivation country of apples, and it exported about 976 thousand tons of apples to the rest of the world in 2012 (National Bureau of Statistic of People’s Republic of China 2013). Sorting is an important process during fruit postharvest processing. At present, apples are usually sorted manually or automatically on the basis of size, shape, color, gloss, and surface defects and decay. However, some internal quality attributes, such as soluble solids content (SSC), firmness, moisture content (MC), and acidity (usually expressed as pH value), which directly contribute to the apples’ unique taste, are essential to meet different tastes of customers (Maniwaru et al. 2014). Standard methods for these quality measurements are mostly destructive, inefficient, or time consuming. Developing nondestructive and efficient methods for sorting apples based on internal qualities is required.

Over the past decades, considerable studies have been carried out to explore nondestructive evaluation methods for fruit qualities based on different technologies, e.g., sonic (Morrison and Abeyratne 2014), electrical (Guo et al. 2010, 2011), and optical technologies (Lu 2003; McGlone et al. 2002; Sun et al. 2009). Among these emerging technologies, optical-based methods demonstrated to have great potential in online application (Mendoza et al. 2011). Near-infrared (NIR) spectroscopy has been applied widely as a nondestructive tool for predicting fruit quality (Liu et al. 2010; McGlone et al. 2002; Zhu et al. 2007). However, NIR spectrum is acquired with relatively small point source, which could not offer much information on samples.

✉ Wenchuan Guo
guowenchuan69@126.com

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

Hyperspectral imaging technology integrates spectroscopic and imaging techniques in one system for providing an informative image of a sample on the spatial and spectral simultaneously (Kamruzzaman et al. 2012). Each hyperspectral image contains a large amount of information in a three-dimensional (3-D) form called “hypercube” which can be analyzed to characterize the object more reliably than the traditional machine vision (Kumar and Mittal 2010) or spectroscopy techniques (Tao and Peng 2014). Hyperspectral imaging technology has been used for predicting internal qualities such as SSC or total sugar content, firmness, MC, and pH of peaches (Lu and Peng 2006), blueberries (Leiva-Valenzuela et al. 2013), strawberries (ElMasry et al. 2007), and bananas (Rajkumar et al. 2012). For apples, Peng and Lu (2008) investigated the firmness and SSC of ‘Golden Delicious’, Mendoza et al. (2011) predicted the firmness and SSC of three varieties of apple, i.e., ‘Jonagold’, ‘Red Delicious’, and ‘Golden Delicious’, and Guo et al. (2014) evaluated the pH value of ‘Fuji’ apples. These studies indicated that hyperspectral imaging technology is feasible to nondestructively predict internal quality attributes of fruits. However, only few reports have been found to investigate these main quality attributes together. Therefore, the work was intended to provide model foundation for online determination of SSC, firmness, MC, and pH values of apples using hyperspectral reflectance imaging technology. A linear model and a nonlinear model were used to establish quality prediction models. Different variable selection methods were applied to extract characteristic variables, and different combinations were investigated. At last, the optimal models for predicting internal quality attributes were determined and proposed.

Materials and Methods

Samples

‘Fuji’ apples were picked from a local orchard located at 34° 21′ north latitude, 108° 10′ east longitude, and at an elevation of 455 m in Yangling, Shaanxi Province, China on October 12, 2013. After the apples were transported to the laboratory, they were kept in polyethylene bags with holes and stored at room temperature (20±2 °C). Measurements were taken initially and at 1-week interval during the 13-week storage period. Before experiment, 10–15 intact apples 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. Totally, 167 apples were used in the study.

NIR Hyperspectral Imaging System

A NIR hyperspectral reflectance imaging system in the spectral rang of 900–1,700 nm was used to acquire the images of

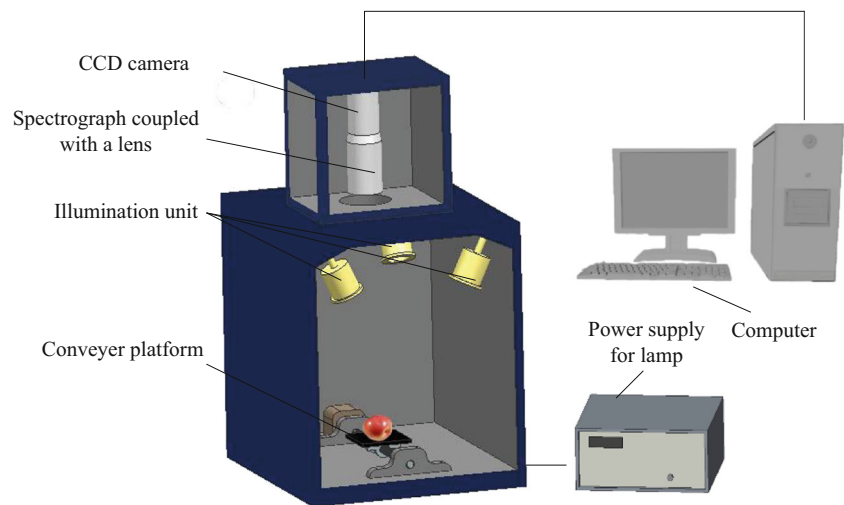
apples. The system mainly consists of a high-performance back-illuminated 8-bit charged couple device (CCD) camera (OPCA05G, Hamamatsu, Japan) coupled with a camera lens, an imaging spectrograph (ImInspector N17E, Spectral Imaging Ltd., Oulu, Finland), an illumination unit equipped with four 100-W halogen lamps at an angle of 45° (HSIA-LS-TAIF, Zolix instruments Co., Ltd., Beijing, China), a conveyer platform (PSA200-11-X, Zolix Instruments Co., Ltd., Beijing, China), data acquisition software (Spectra SENS, Zolix Instruments Co., Ltd., Beijing, China), and a computer. Figure 1 is the schematic diagram of the applied NIR hyperspectral reflectance imaging system. The distance between the CCD camera lens and fruit holding platform was fixed at 65 cm, and the exposure time of the camera was set as 10 ms. Apples were placed on the conveyer platform operated by a 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 (20±2 °C). Each image was acquired as a 3-D image (x, y, λ) which includes 320×256 pixels in spatial dimensional (x, y) and 256 spectral bands from 865.11 to 1,711.71 nm with 3.32-nm interval between contiguous bands in spectral dimensional (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 background by closing the lens cap and white reference using a standard Teflon white board with 99 % reflectance. Two NIR hyperspectral images were obtained at blush side and non-blush side around the equator with 180° interval for each apple. 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 with an interval of 90°. Of which, two spots were on the blush side and non-blush side which were used for hyperspectral image acquisition. The firmness of apples was measured with a GY-3 fruit firmness tester (Beijing THY Science & Technology Co., Ltd, China), whose diameter of the penetrometer tip was 8 mm. After firmness measurements, pulp adjacent to the spot for firmness measurement was put into a garlic press to extract juice for measuring SSC using a digital refractometer (PR-101 α , Atago Co., Ltd., Japan). Then about 3–5 g undamaged pulp adjacent to the each spot for firmness measurement was used to determine moisture content determination for 24 h at 70 °C in an air drying oven (101-1AB, Tianjin Taisite Instrument Co., Ltd., China). Four readings at four spots were recorded for SSC, firmness, and MC. An average of four measurements for SSC, firmness, and MC were calculated, respectively, and were used as the results of the sample.

Finally, the left apple pulp of each sample was cut into pieces and placed in the garlic press to expel juice which was collected in a 10-ml beaker for pH determination with a

Fig. 1 Schematic diagram of NIR hyperspectral imaging system



pH meter (PHSJ-3F, Shanghai Electrical and Instrument Analysis Instrument Co., Ltd., China). Three pH readings were taken for each apple, and their average was used as the pH value of the sample.

Image Processing

To eliminate the differences from camera quantum and physical configuration of hyperspectral imaging system, the raw image R_O , acquired from the imaging system, was calibrated to obtain the corrected reflectance image R using the following equation:

$$R = \frac{R_O - B}{W - B} \times 100 \% \quad (1)$$

where B is background recorded by closing the lens cap, and W is white reference image collected using a standard Teflon white board with 99 % reflectance.

Because of the difference of apple feature and apple size between different samples, a region of interest (ROI) with a size of 40×40 pixels in the center of the apple in each image was manually identified. The average reflectance spectrum of the ROI 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).

Spectral Data Preprocessing

The purpose of spectral preprocessing is to eliminate the effects of noise, distortion, and observational environment and to improve the precision and stability of models. After comparing different preprocessing methods including standard normal variate (SNV), multiplicative scatter correction (MSC), and first and second derivatives, we found that SNV

could offer better quality determination performance than other three methods. Therefore, SNV was used to preprocess the obtained reflectance spectra in this study. SNV is a row-oriented transformation which centers and scales individual spectra. It is capable of correcting multiplicative noise, e.g., particle size influence, scatter interference, and path length effect, arising from the physical structure of samples (Feng and Sun 2013). SNV preprocessing was implemented using Unscrambler (version 9.7, CAMO, Trondheim, Norway).

Sample Division Methods

Sample set partitioning based on the joint x - y distance (SPXY) algorithm, developed by Galvão et al. (2005) and extended from the classic Kennard-Stone (KS) algorithm, encompasses both x - and y -differences in the calculation of inter-sample distances. Since SPXY can cover a multidimensional space efficiently, and can acquire typical sample calibration set and improve the accuracy of model (Zhu et al. 2009), it has been widely used in sample division in spectrum analysis (Shang et al. 2013). In the study, SPXY was used to divide sample sets according to the ratio of 3:1, indicating 125 samples in calibration set and 42 samples in prediction set.

Characteristic Wavelength Selection Strategy

The acquired hyperspectral images usually suffer from the problem of multicollinearity during a multivariate analysis. Some congruent wavelengths are related to similar constituents and consequently contain much of the same information (Liu et al. 2014). Therefore, it is essential to find few characteristic wavelengths that would be most influential on the quality evaluation of the product and to eliminate wavelengths having no discrimination power. The most widely used successive projection algorithm (SPA) and uninformative variable elimination (UVE) were applied here to select

characteristic wavelengths or to eliminate uninformative variables from full spectra.

Successive Projection Algorithm

SPA is a forward variable selection algorithm designed to minimize collinearity problems in multiple linear regression. It employs simple projection operations in a vector space to obtain subsets of variables with minimal collinearity (Zou et al. 2010). The new variables selected by SPA have the maximum projection value on the orthogonal sub-space of the previous selected variables.

The main procedures of SPA are as follows: (1) set the maximum number of variables N to be selected before a start vector is chosen in a space of m -dimensions (where m is the number of original variables); (2) in an orthogonal sub-space, the vector with high projection is selected and becomes the new starting vector; and (3) keep the iteration until N is reached. The final number of selected variables can be determined according to the smallest root-mean-square error of cross-validation (RMSECV), which was used to evaluate the error of each calibration model. It is calculated as follows:

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

where \hat{y}_i is the estimated result for i -th sample when the model is constructed without sampled i , y_i is the measured value of the i -th sample, and n_c is the number of apples in the calibration set.

Uninformative Variable Elimination

UVE is a backward variable removal method based on the stability analysis of the regression coefficients of partial least squares (PLS) models. An advantage compared to other selection methods is that the level to cut the uninformative variables is user-independent and therefore does not present any configuration problems (Centner et al. 1996). It can eliminate uninformative variables by adding artificial random variables to the data set as a reference so that those variables which play a less important role in the model than the random variables could be detected (Cai et al. 2008). By using the variables selected by UVE, an overfitting model could be avoided and the predictive ability could be improved.

Modeling Methods

A linear model (PLS) and two nonlinear models (least squares support vector machine (LSSVM) and backpropagation (BP)

network) were employed to establish determination models for each internal quality parameter.

Partial Least Squares

PLS is a widely used algorithm combined factor analysis with regression. By using PLS, the original independent information (spectral data) is projected onto a small number of latent variables (LVs) to simplify the relationship between the spectral data and the target quality property matrix (Shao and He 2007). In the development of PLS model, full cross-validation was used to validate the quality and to prevent overfitting of calibration model. PLS modeling is often assumed to be influenced by just a few underlying variables called latent variables (Wold et al. 2001a). The LVs were considered as new characteristic vectors of the original spectra to reduce the dimensionality and compress the original spectra data (Li et al. 2013).

Least Squares Support Vector Machine

LSSVM, proposed by Suykens et al. (2002), is a modified algorithm of the standard support vector machine (SVM). It is capable of dealing with linear and nonlinear multivariate calibration and resolving these problems in a relatively fast way. Previous studies showed that SVM could obtain better quality determination performance in the study. Therefore, LSSVM was applied here. Linear kernel, polynomial kernel, and radial basis function (RBF) kernel could be used as kernel function of LSSVM. 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. Therefore, RBF was used as the kernel function of LSSVM. Simplex technique and tenfold cross-validation were applied to find the optimal values of the two important parameters, i.e., regularization parameter (γ) and kernel function parameters (σ^2) of LSSVM. Lssvm toolbox (LSSVM v1.5, Suykens, Leuven, Belgium) was employed for LSSVM regression. The computations and data analysis were preformed in MATLAB (R2013a, MathWorks, Natick, MA, USA).

Backpropagation Network

BP network is one of the most popular and widely used neural networks. It is a feedforward network and can actualize non-linearity mapping between input and output exactly. BP is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. The details of BP algorithm could be found elsewhere (He et al. 2006). In this study, a standard three-layer network, including one input layer, one hidden layer, and one output layer, was established.

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). Besides, residual predictive deviation (RPD) was used to assess the prediction ability of a model. These indices are defined as follows:

$$R_C = \sqrt{\frac{\sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_c} (\hat{y}_i - y_c)^2}} \tag{3}$$

$$R_P = \sqrt{\frac{\sum_{i=1}^{n_p} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_p} (\hat{y}_i - y_p)^2}} \tag{4}$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \tag{5}$$

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (\hat{y}_i - y_i)^2} \tag{6}$$

$$RPD = \text{std}_{\text{prediction}} / RMSEP \tag{7}$$

where \hat{y}_i is the predicted value of an attribute of the i -th apple. y_c and y_p are the mean values of the calibration and prediction sets, respectively. n_p is the number of apples in the prediction set. A good model should have high values of R_C and R_P and low values of RMSEC and RMSEP.

According to Nicolai et al. (2007), an RPD value less than 1.5 indicates very poor prediction, a value between 1.5 and 2.0 means that the model can discriminate low from high values of the response variable, a value between 2.0 and 2.5 indicates that coarse quantitative predictions are possible, and a value between 2.5 and 3.0 or above corresponds to good and excellent prediction accuracy.

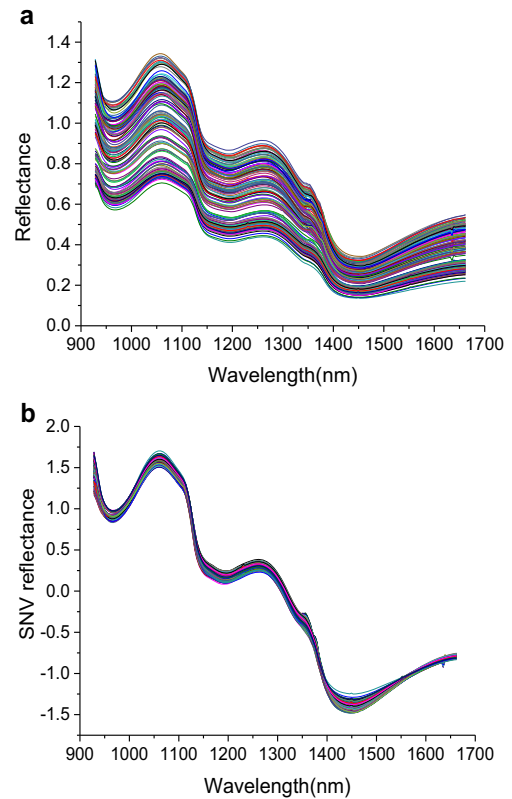


Fig. 2 Original reflectance spectra (a) and spectra after SNV (b) for 167 samples

Results and Discussion

Internal Quality Parameters of Samples

The statistics of SSC, firmness, MC, and pH values of used 167 apples in the calibration and prediction sets are shown in Table 1. Table 1 indicates that for each parameter, its minimum in prediction set was larger than that in calibration set, and its maximum in prediction set was smaller than that in calibration set. The standard deviation of each parameter was 19–26 % of the difference between the maximum and minimum values of that parameter both in calibration set and prediction set. This expresses that the data set contained a

Table 1 The statistics of SSC, firmness, MC, and pH values of 167 samples in the calibration and prediction sets

Quality parameter	Sample set	Amount of samples	Minimum	Maximum	Mean	Standard deviation
SSC (°Brix)	Calibration set	125	8.48	16.60	11.92	2.08
	Prediction set	42	8.55	16.20	11.12	1.86
Firmness (kg/cm ²)	Calibration set	125	3.87	7.43	4.99	0.68
	Prediction set	42	3.97	5.97	4.82	0.43
MC (%)	Calibration set	125	81.03	90.98	87.12	2.49
	Prediction set	42	81.24	90.54	87.67	2.45
pH	Calibration set	125	3.73	4.45	4.101	0.14
	Prediction set	42	3.81	4.31	4.12	0.12

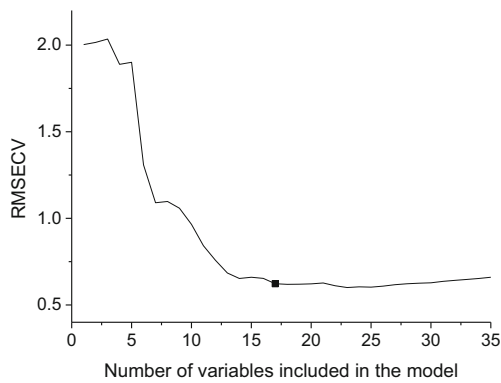


Fig. 3 Changed RMSECV with the number of variables included in SPA for SSC. *Black square* represents the point at which the final number of variables was decided

sufficiently large variation to allow for a meaningful calibration (Savenije et al. 2006). Moreover, each quality parameter in the two sets covers a similar range. All these indicate a rational sample division and a potential ability in developing effective models.

Spectral Features

Since heavy noises existed at both ends of the spectral range, spectrum data below 928.19 nm and above 1,661.91 nm were discarded, leading to spectra within the range of 928.19–1,661.91 nm with a total of 222 bands for further analysis. Figure 2 shows the original reflectance spectra and spectra after SNV for the 167 apple samples between 928.19 and 1,661.91 nm. It can be seen that the trends of spectra for all apples during the 13-week storage were similar to each other, but the reflectance values were different, especially for original spectra.

Near-infrared spectra were sensitive to the concentrations of organic materials, which involved the response of molecular bonds of C–H, O–H, and N–H. Low values of reflectance, i.e., high absorbance, in the region of 950–980 nm were likely attributed to the combination effect of O–H second overtone from carbohydrates and water (Leiva-Valenzuela et al. 2013). Reflectance decreased rapidly at the wavelengths of 1,100–1,170 nm and reached an absorption peak at a wavelength close

to 1,200 nm, which was likely attributed to C–H stretching second overtone from carbohydrates (fructose, sucrose, and glucose). In addition, a strong absorption peak at 1,450 nm due to water absorption bands related to O–H stretching first overtone indicates that moisture dominates the spectral signatures (Kamruzzaman et al. 2012). Similar results were also found in other fruits such as pear (Li et al. 2014), kiwifruit (Liu and Guo 2014), and jujube (Yu et al. 2014) among others.

Characteristic Variable Selection

Characteristic Variables Selected by SPA

The changed RMSECV with the number of variables included in SPA for SSC is shown in Fig. 3. Figure 3 tells that the RMSECV decreased as the number of variables increased. The decrease was great when the number was less than 13. It decreased slowly between 13 and 25 and increased a little above 25. The smallest RMSECV (0.603) was found at 25. Since more variables will increase the complexity of established model, a significance level of $\alpha=0.25$ for *F* test criterion was used to determine the optimal number of variables, as suggested in Galvão et al. (2008). The determined variable number was 17 (RMSECV=0.623). The result was marked with a solid square in Fig. 3. The same method was also used to select the optimal numbers of variables in SPA for firmness, MC, and pH, and the determined numbers were 11 (RMSECV=0.633), 23 (RMSECV=0.586), and 8 (RMSECV=0.093), respectively. The selected variables (wavelengths) are listed in Table 2. These wavelengths were sequenced in the order of importance in the projection produce. Take SSC for instance, wavelength at 1,107.47 nm was the most relevant variable selected by SPA.

Characteristic Variables Selected by UVE

UVE, based on PLS, is assumed to be influenced by some underlying variables, i.e., latent variables (Wold et al. 2001b). The number of variables was selected by UVE changes with the number of LVs. To acquire a more effective SSC prediction model, the number of LVs was set from 1 to 20 and

Table 2 The selected variables (wavelengths) for SSC, firmness, MC, and pH by using SPA

Quality attributes	No. of selected variables	Selected variables (wavelength) (nm)
SSC	17	1,107.47, 1,047.71, 1,339.87, 1,376.39, 1,542.39, 1,618.75, 1,150.63, 1,353.15, 1,373.07, 938.15, 1,422.87, 1,124.07, 1,582.23, 1,011.19, 1,396.31, 1,658.59, 928.19
Firmness	11	1,462.71, 1,359.79, 1,595.51, 1,107.47, 1,300.03, 1,230.31, 1,396.31, 1,373.07, 1,635.35, 971.35, 928.19
MC	23	1,107.47, 954.75, 1,615.43, 1,608.79, 1,060.99, 1,326.59, 1,651.95, 1,383.03, 1,529.11, 1,406.27, 1,592.19, 1,124.07, 1,618.75, 981.31, 938.15, 1,369.75, 1,575.59, 1,147.31, 1,641.99, 948.11, 1,578.91, 1,263.51, 928.19
pH	8	1,356.47, 1,041.07, 1,120.75, 964.71, 1,373.07, 1,200.43, 1,532.43, 1,389.67

was determined by the smallest RMSECV. Figure 4 shows the changed RMSECV with the number of latent variables for SSC. When the number of LVs was 12, the RMSECV reached the minimum (0.659). Therefore, 12 LVs were used in SSC prediction model. Figure 5 illustrates the stability of real variables and random variables for SSC by UVE with 12 LVs. The 222 input spectral variables are at the left of the vertical line, while random variables are at the right side. Two dot lines are the cutoff threshold which is set as the maximum of absolute value among the random variables. The variables, whose stabilities within the cutoff lines, are considered to be uninformative, and the rest variables are selected as the characteristic variables. Finally, 135 variables were selected as characteristic variables for SSC by using UVE.

With the same process, the obtained optimal numbers of LVs for firmness, MC, and pH were 5, 13, and 12, and the selected characteristic variables were 71, 122, and 108, for firmness, MC, and pH, respectively.

Modeling Results

All selected characteristic variables by SPA and UVE as well as the full spectra were used as the inputs to establish PLS, LSSVM, and BP models for determination of each internal quality parameter of apples.

PLS Modeling Results

A critical step in building a robust PLS model is choosing the correct number of LVs, which can avoid establishing an overfitting or underfitting model. Most of the variations can be captured within the first few latent variables/factors, while the remaining latent variables describe random noise or linear dependencies between the wavelengths (ElMasry et al. 2011). Generally, the optimal number of latent variables corresponds to the lowest value of RMSECV (Zou et al. 2007). The number of latent variables for SPA-PLS model of SSC, firmness, MC, and pH was determined by RMSECV as shown in Fig. 6.

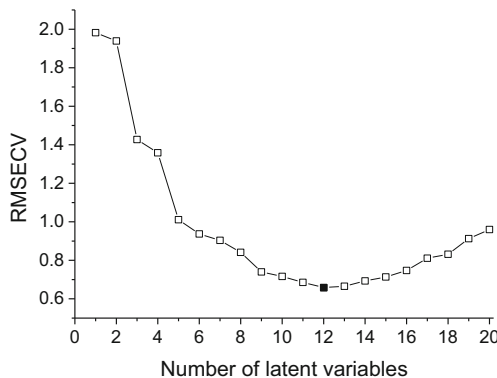


Fig. 4 Changed RMSECV with the number of latent variables in UVE. Black square represents the point at which the final number of variables was decided

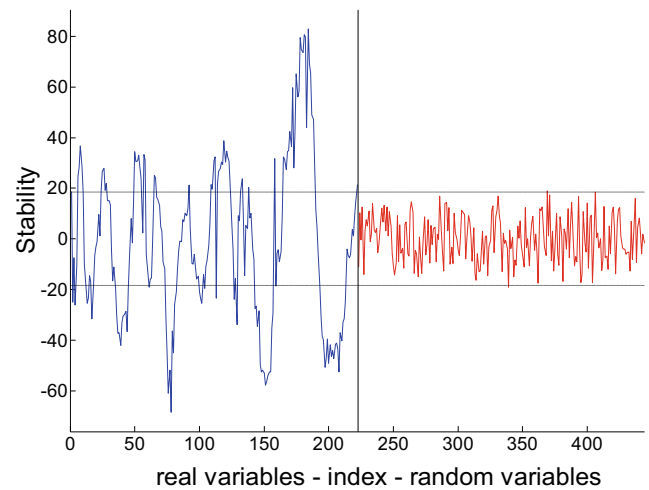


Fig. 5 Stability distribution of each variable for SSC prediction by UVE with 12 LVs. Two horizontal dashed lines show the lower and upper cutoff

Figure 6 indicates that the RMSECV decreased rapidly until the number of LVs reached a value for each quality parameter, then decreased gradually or even kept constant. Since more latent variables may lead to overfitting problem, the optimal numbers of LVs for predicting SSC, firmness, MC, and pH using SPA-PLS were identified as 12, 8, 11, and 6, respectively. The same method was also used to decide the optimal numbers of LVs for each quality parameter when the full spectra (FS) and selected characteristic parameters by UVE were used in building PLS models. The results are shown in Table 3. The calibration and prediction performances by using PLS models for determining SSC, firmness, MC, and pH of used ‘Fuji’ apples were shown in Table 4, respectively. Table 4 shows that SPA-PLS model had the highest R_C (0.962) and the lowest RMSEC (0.573) for SSC prediction. The FS-PLS model had the highest R_P (0.953) and the lowest RMSEP (0.590). That is the SPA-PLS had the best calibration performance and the FS-PLS had the best prediction performance for SSC. The RPD values of the three PLS models for SSC were higher than

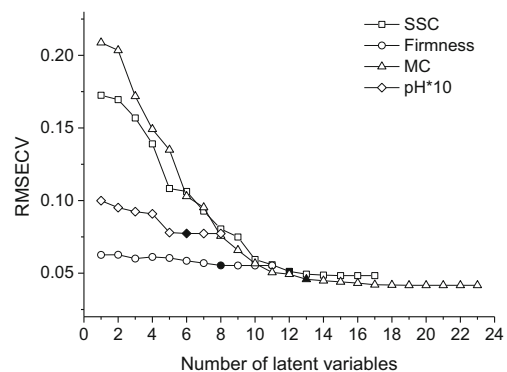


Fig. 6 Changed RMSECV with the number of latent variables in SPA-PLS for SSC, firmness, MC, and pH

Table 3 The determined number of LVs in PLS models and γ and σ^2 in LSSVM models for each quality parameter under different variable selection methods

Variable selection method	SSC			Firmness			MC			pH		
	Number of LVs	γ	σ^2	Number of LVs	γ	σ^2	Number of LVs	γ	σ^2	Number of LVs	γ	σ^2
FS	11	1.03×10^4	217	9	1.99	0.75	11	1.79×10^4	172	9	157	11.7
SPA	12	1.10×10^4	54.9	8	2.83	0.04	11	1.27×10^5	178	6	154	1.33
UVE	9	2.18×10^5	5,090	5	1.26	0.51	12	1.90×10^4	254	8	170	8.38

3.0, meaning all established PLS models had excellent SSC determination ability. FS-PLS had the highest RPD (3.19), a little higher than that of SPA-PLS (3.12), but it used 222 variables, much more than SPA-PLS (17). When there is less use of variables, the model is more efficient. Therefore, it was suggested that SPA-PLS was the best model for SSC determination.

For firmness prediction, FS-PLS model had the highest R_C (0.622) and the lowest RMSEC (0.533), SPA-PLS model had the highest R_p (0.506) and the lowest RMSEP (0.375), but their RPD values were only 1.07 and 1.16, less than 1.5, respectively. This indicates that all PLS models had poor determination ability for firmness.

For MC prediction, SPA-PLS model had the highest R_C (0.974) and the lowest RMSEC (0.565) for MC determination, but had the highest RMSEP (0.566), indicating the stability of SPA-PLS was the worst. UVE-PLS had the highest R_p (0.978), the lowest R_p (0.525), and a better R_C (0.964). FS-PLS and UVE-PLS had the same RPD value (4.71), higher than SPA-PLS (4.38), but much less variables (22) were used in SPA-PLS than that in FS-PLS (222) and UVE-PLS (122). SPA-PLS has much more potential in online determination of MC.

For pH, FS-PLS had the best calibration performance (the highest R_C and the lowest RMSEC), and SPA-PLS had the

best prediction performance (the highest R_p and the lowest RMSEP). The RPD values of all PLS models for pH were between 1.57 and 1.65, a little higher than 1.5, illustrating that all PLS models just can discriminate low values from high values of pH. SPA-PLS has better comprehensive ability than FS-PLS and UVE-PLS in pH determination.

LSSVM Modeling Results

The determined modeling parameters of LSSVM for each quality parameter under different variable selection methods are listed in Table 3. The calibration and prediction performances for SSC, firmness, MC, and pH of 'Fuji' apples by LSSVM models under different characteristic variable selection methods are listed in Table 5.

Table 5 tells that FS-LSSVM model had the highest R_C (0.982) and the lowest RMSEC (0.397), but had the lowest R_p (0.959) and the highest RMSEP (0.550), meaning FS-LSSVM was unstable. SPA-LSSVM and UVE-LSSVM had almost same calibration performance, but SPA-LSSVM had a little poorer prediction performance than UVE-LSSVM. As PLS models, the established LSSVM models had good SSC determination ability since their RPD values were higher than 3.0. For predicting firmness, the RPD values of all LSSVM models were not higher than 1.4, indicating the established

Table 4 Comparison of determination results for different quality attributes of 'Fuji' apples by PLS models

Quality attributes	Variable selection method	Calibration set		Prediction set		RPD
		R_C	RMSEC	R_p	RMSEP	
SSC, °Brix	FS	0.958	0.595	0.953	0.589	3.19
	SPA	0.962	0.573	0.949	0.602	3.12
	UVE	0.955	0.620	0.945	0.620	3.03
Firmness, kg/cm ²	FS	0.6222	0.533	0.378	0.407	1.07
	SPA	0.5870	0.552	0.506	0.375	1.16
	UVE	0.4640	0.604	0.226	0.444	0.98
MC, %	FS	0.967	0.637	0.977	0.526	4.71
	SPA	0.974	0.565	0.975	0.566	4.38
	UVE	0.9644	0.657	0.978	0.525	4.71
pH	FS	0.81743	0.082	0.783	0.074	1.57
	SPA	0.793	0.087	0.806	0.070	1.65
	UVE	0.812	0.083	0.785	0.074	1.58

Table 5 Comparison of determination results for different quality attributes of ‘Fuji’ apples by LSSVM models

Quality attributes	Variable selection method	Calibration set		Prediction set		RPD
		R_C	RMSEC	R_P	RMSEP	
SSC, °Brix	FS	0.982	0.397	0.959	0.550	3.42
	SPA	0.970	0.509	0.961	0.539	3.49
	UVE	0.970	0.510	0.963	0.517	3.64
Firmness, kg/cm ²	FS	0.961	0.257	0.572	0.355	1.23
	SPA	0.974	0.209	0.694	0.314	1.39
	UVE	0.849	0.403	0.603	0.345	1.27
MC, %	FS	0.993	0.295	0.979	0.510	4.86
	SPA	0.985	0.432	0.984	0.450	5.51
	UVE	0.985	0.432	0.980	0.503	4.93
pH	FS	0.965	0.039	0.870	0.062	1.89
	SPA	0.878	0.068	0.882	0.057	2.06
	UVE	0.942	0.049	0.879	0.060	1.94

LSSVM models were incapable of predicting firmness of apples using hyperspectral imaging technology. All LSSVM models for moisture content prediction achieved a good performance both in calibration set and prediction set. The best prediction model was SPA-LSSVM with R_C and RMSEC of 0.985 and 0.432, R_P and RMSEP of 0.984 and 0.450, and RPD of 5.51, respectively. For predicting pH, SPA-LSSVM had the highest R_P (0.882), the lowest RMSEP (0.057), and the highest RPD. But its RPD was 2.06, just higher than 2, indicating SPA-LSSVM could be used to obtain a coarse quantitative prediction for pH.

BP Modeling Results

The parameters which have significant impact on BP model include activation function, learning rate, threshold residual error, etc. A three-layer network was developed with “tansig”

transfer function in the input layer and the hidden layer and “pureline” transfer function in the output layer. The learning rate was set as 0.1, and the threshold residual error was set as 1.0×10^{-4} after several times trials. The BP modeling results for each quality attribute are listed in Table 6.

Table 6 shows that FS-BP performed best for SSC determination, with the highest R_C (0.987) and R_P (0.962), the lowest RMSEC (0.361) and RMSEP (0.535), and the highest RPD. The RPD values of all BP models were higher than 3.0, indicating SSC could be predicted exactly by BP models. All BP models were incapable of predicting firmness since their R_P values were lower than 0.62, and their RPD values were less than 1.3. For MC prediction, the RPD values of the three BP models were higher than 3.0, especially SPA-BP and UVE-BP, whose RPD were higher than 5.0, indicating the BP models could predict moisture content of apples exactly. For pH prediction, the best model was FS-BP with the highest

Table 6 Comparison of determination results for different quality attributes of ‘Fuji’ apples by BP models

Quality attributes	Variable selection method	Calibration set		Prediction set		RPD
		R_C	RMSEC	R_P	RMSEP	
SSC, °Brix	FS	0.987	0.361	0.962	0.535	3.51
	SPA	0.979	0.434	0.955	0.564	3.33
	UVE	0.982	0.412	0.959	0.554	3.39
Firmness, kg/cm ²	FS	0.835	0.428	0.575	0.355	1.23
	SPA	0.794	0.465	0.618	0.342	1.28
	UVE	0.753	0.486	0.518	0.369	1.18
MC, %	FS	0.983	0.558	0.982	0.609	4.07
	SPA	0.984	0.456	0.984	0.481	5.15
	UVE	0.990	0.368	0.983	0.482	5.14
pH	FS	0.922	0.060	0.865	0.058	2.01
	SPA	0.855	0.075	0.853	0.061	1.92
	UVE	0.899	0.065	0.851	0.061	1.90

R_C (0.922) and R_P (0.865) and the lowest RMSEC (0.060) and RMSEP (0.058). But its RPD was 2.01, just higher than 2.0, expressing that FS-BP could be used to obtain a coarse quantitative prediction for pH.

Comprehensive Comparison for Different Models

When the calibration and prediction performances of PLS, LSSVM, and BP models were compared, it was found that LSSVM had better performances than PLS and BP for each investigated quality parameter in most cases. The SSC and MC could be predicted precisely by all developed models, and SPA-LSSVM and FS-BP could be used to predict pH value roughly. However, all models were incapable of predicting firmness. The poor pH and firmness determination capability might lie in the narrow data range. UVE-LSSVM had the highest R_P and RPD, and the lowest RMSEP for SSC prediction. SPA-LSSVM had the highest R_P and RPD and lowest RMSEP not only for moisture content but also for pH prediction.

It is worth noting that variable selection could improve LSSVM model predictive ability but could not always improve predictive ability of PLS and BP model. Compared with UVE, SPA can select more less characteristic variables for each quality parameter. For example, for moisture content, the selected variables by SPA and UVE were 23 and 122, respectively. The number of the later was 5.3 times of the former. Therefore, variable selection methods still have capacity to save time and simplify models. Considering prediction performance and the amount of variables used in model, it is suggested that SPA-LSSVM was the best model for determination of SSC, MC, and pH. The measured values of SSC, MC, and pH against predicted ones in calibration and prediction sets using SPA-LSSVM are shown in Fig. 7.

Comparison with Reported Data

Hyperspectral imaging technology combined with chemometric methods have been used to predict the internal qualities of apples. In predicting SSC, SSC, of ‘Fuji’ apple was predicted by SPA-MLR with R_P^2 of 0.9501 and RMSEP of 0.3087 (Huang et al. 2013). For ‘Golden Delicious’, Peng and Lu (2008) employed modified Lorentzian function for prediction with R_P of 0.883 and standard error of prediction (SEP) of 0.73 %, and for ‘Golden Delicious’, ‘Jonagold’, and ‘Delicious’, $R_P=0.66–0.88$ and $SEP=0.7–0.9$ % (Mendoza et al. 2011). The prediction performance of SSC in this study was better than that obtained by Mendoza et al. (2011) and Peng and Lu (2008) but poorer than that obtained by Huang et al. (2013).

For firmness prediction, PLS model was built for ‘Fuji’ apple with R_P and SEP values of 0.88 and 0.88×10^5 Pa, respectively (Peng et al. 2012). SVM model was built for ‘Fuji’

apple with $R_P=0.6808$ (Zhao et al. 2009). Firmness for ‘Golden Delicious’ was predicted using modified Lorentzian function with $R_P=0.894$ and $SEP=6.14$ N (Peng and Lu 2008), and for ‘Golden Delicious’, ‘Jonagold’, and ‘Delicious’, $R_P=0.84–0.95$ and $SEP=5.9–8.7$ N (Mendoza et al. 2011). Hyperspectral laser-induced fluorescence imaging was used for ‘Golden Delicious’ firmness prediction with R_P of 0.76 (Noh and Lu 2007). The reported results for firmness were better than those obtained here.

As for pH prediction, Guo et al. (2014) established synergy internal partial least squares (siPLS) model based on shortwave infrared hyperspectral imaging (1,000–2,500 nm) for ‘Fuji’ apple with the best R_P of 0.8474 and RMSEP of 0.0398. The

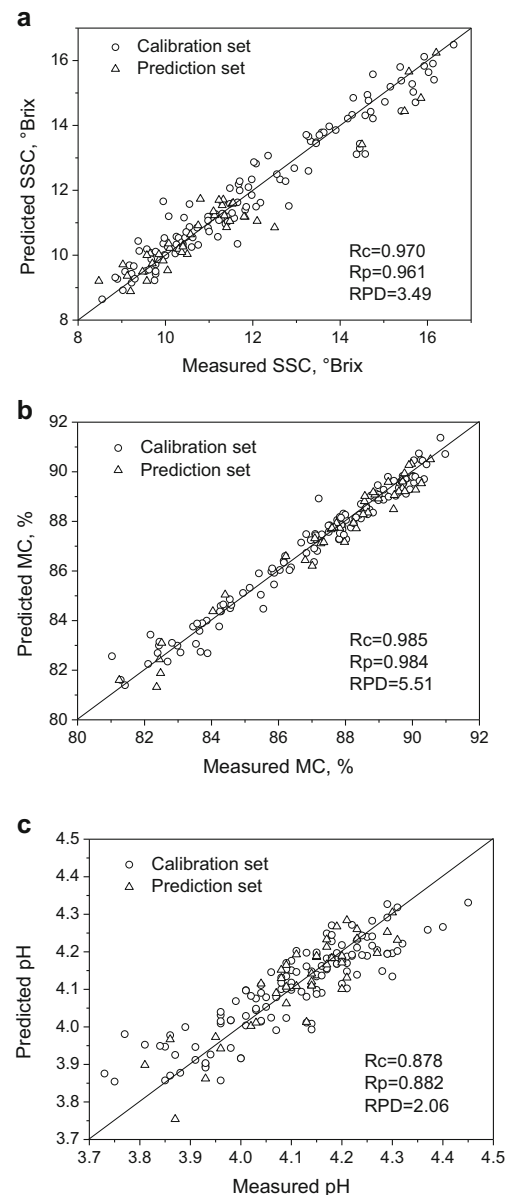


Fig. 7 The measured values of SSC (a), moisture content (b), and pH (c) against predicted ones for calibration and prediction sets using SPA-LSSVM

prediction performance was similar to what was obtained here. At present, no reports have been found to predict moisture content by using hyperspectral imaging technology.

Conclusions

A near-infrared hyperspectral reflectance imaging system (900–1,700 nm) was used to acquire hyperspectral images of ‘Fuji’ apples during the 13-week storage period. PLS, LSSVM, and BP modeling methods were used to establish models for determining internal qualities of apples, and SPA and UVE were used to select characteristic wavelengths or eliminate uninformative variables from full spectra. Some useful information was lost during characteristic variable selection. SPA was more effective than UVE in selecting characteristic variables. The established LSSVM models had better performances than PLS and BP in predicting investigated quality parameters. The best model in predicting SSC, MC, and pH was SPA-LSSVM with the correlation coefficient of prediction of 0.961, 0.984, and 0.882 and residual predictive deviation of 3.49, 5.51, and 2.06, respectively. All models could be used to predict firmness. The study indicates that hyperspectral imaging technology coupled with multivariate data analysis methods can be used to predict soluble solids content and moisture content of apples exactly and to predict pH values roughly. Further researches will be conducted to improve firmness prediction ability.

Acknowledgments This research was supported by a grant from the National Natural Science Foundation of China (Project No. 31171720).

Conflict of Interest Jinlei Dong declares that he has no conflict of interest. Wenchuan Guo declares that he has no conflict of interest. This article does not contain any studies with human or animal subjects.

Compliance with Ethics Requirements Jinlei Dong declares that he has no conflict of interest. Wenchuan Guo has no conflict of interest. This article does not contain any studies with human or animal subjects.

References

- Cai W, Li Y, Shao X (2008) A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom Intell Lab Syst* 90(2):188–194
- Centner V, Massart D-L, de Noord OE, de Jong S, Vandeginste BM, Sterna C (1996) Elimination of uninformative variables for multivariate calibration. *Anal Chem* 68(21):3851–3858
- 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
- ElMasry G, Sun D-W, Allen P (2011) Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Res Int* 44(9):2624–2633
- Feng Y-Z, Sun D-W (2013) Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of *Pseudomonas* loads in chicken fillets. *Talanta* 109:74–83
- 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
- Guo W, Zhu X, Liu Y, Zhuang H (2010) Sugar and water contents of honey with dielectric property sensing. *J Food Eng* 97(2):275–281
- Guo W, Zhu X, Nelson SO, Yue R, Liu H, Liu Y (2011) Maturity effects on dielectric properties of apples from 10 to 4500 MHz. *LWT-Food Sci Technol* 44(1):224–230
- Guo Z, Huang W, Chen L, Peng Y, Wang X (2014) Shortwave infrared hyperspectral imaging for detection of pH value in Fuji apple. *Int J Agric Biol Eng* 7(2):130–137
- Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. *Trends Genet* 18(8):426–430
- He Y, Feng S, Deng X, Li X (2006) Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy. *Food Res Int* 39(6):645–650
- 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
- Kumar S, Mittal GS (2010) Rapid detection of microorganisms using image processing parameters and neural network. *Food and Bioprocess Technology* 3(5):741–751
- 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
- Li J, Huang W, Chen L, Fan S, Zhang B, Guo Z, Zhao C (2014) Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of ‘Ya’ pear. *Food Anal Methods* 7(9):1891–1902
- Liu D, Guo W (2014) Identification of kiwifruits treated with exogenous plant growth regulator using near-infrared hyperspectral reflectance imaging. *Food Anal Methods*. doi:10.1007/s12161-014-9885-8
- Liu Y, Sun X, Ouyang A (2010) Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. *LWT-Food Sci Technol* 43(4):602–607
- Liu D, Sun D-W, Qu J, Zeng X-A, Pu H, Ma J (2014) Feasibility of using hyperspectral imaging to predict moisture content of porcine meat during salting process. *Food Chem* 152:197–204
- Lu R (2003) Detection of bruises on apples using near-infrared hyperspectral imaging. *Transact ASAE* 46(2):523–530
- Lu RF, Peng YK (2006) Hyperspectral scattering for assessing peach fruit firmness. *Biosyst Eng* 93(2):161–171
- Maniwaru P, Nakano K, Boonyakiat D, Ohashi S, Hiroi M, Tohyama T (2014) The use of visible and near infrared spectroscopy for evaluating passion fruit postharvest quality. *J Food Eng* 143:33–43
- McGlone VA, Jordan RB, Martinsen PJ (2002) Vis/NIR estimation at harvest of pre- and post-storage quality indices for ‘Royal Gala’ apple. *Postharvest Biol Technol* 25(2):135–144
- 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

- Morrison DS, Abeyratne UR (2014) Ultrasonic technique for non-destructive quality evaluation of oranges. *J Food Eng* 141:107–112
- National Bureau of Statistic of P. R. China (2013) China statistical year-book. China Statistical Press, Beijing
- Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* 46(2):99–118
- 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 (2008) Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biol Technol* 48(1):52–62
- Peng Y, Li Y, Zhao J, Shan J (2012) Establishment of rapid and non-destructive detection method of apple firmness using hyperspectral images. *J Food Safety Quality* 3(6):667–671
- Rajkumar P, Wang N, Eimasry G, Raghavan GSV, Garipey Y (2012) Studies on banana fruit quality and maturity stages using hyperspectral imaging. *J Food Eng* 108(1):194–200
- Savenije B, Geesink G, Van der Palen J, Hemke G (2006) Prediction of pork quality using visible/near-infrared reflectance spectroscopy. *Meat Sci* 73(1):181–184
- Shang L, Gu J, Guo W (2013) Non-destructively detecting sugar content of nectarines based on dielectric properties and ANN. *Transact Chin Soc Agric Eng* 29(17):257–264 (in Chinese)
- Shao Y, He Y (2007) Nondestructive measurement of the internal quality of bayberry juice using Vis/NIR spectroscopy. *J Food Eng* 79(3):1015–1019
- Sun T, Lin H, Xu H, Ying Y (2009) Effect of fruit moving speed on predicting soluble solids content of ‘Cuiguan’ pears (*Pomaceae pyrifolia* Nakai cv. Cuiguan) using PLS and LS-SVM regression. *Postharvest Biol Technol* 51(1):86–90
- Suykens JA, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J, Suykens J, Van Gestel T. (2002). Least squares support vector machines (Vol. 4): World Scientific.
- Tao F, Peng Y (2014) A method for nondestructive prediction of pork meat quality and safety attributes by hyperspectral imaging technique. *J Food Eng* 126:98–106
- Wold S, Sjöström M, Eriksson L (2001a) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58(2):109–130
- Wold S, Trygg J, Berglund A, Antti H (2001b) Some recent developments in PLS modeling. *Chemom Intell Lab Syst* 58(2):131–150
- Yu K, Zhao Y, Li X, Shao Y, Zhu F, He Y (2014) Identification of crack features in fresh jujube using Vis/NIR hyperspectral imaging combined with image processing. *Comput Electron Agric* 103:1–10
- Zhao J, Chen Q, Vittayapadung S, Chaitep S (2009) Determination of apple firmness using hyperspectral imaging technique and multivariate calibrations. *Transact Chin Soc Agric Eng* 25(11):226–231 (in Chinese)
- Zhu D, Ji B, Meng C, Shi B, Tu Z, Qing Z (2007) The performance of ν -support vector regression on determination of soluble solids content of apple by acousto-optic tunable filter near-infrared spectroscopy. *Anal Chim Acta* 598(2):227–234
- Zhu X, Shan Y, Li G, Huang A, Zhang Z (2009) Prediction of wood property in Chinese Fir based on visible/near-infrared spectroscopy and least square-support vector machine. *Spectrochim Acta A Mol Biomol Spectrosc* 74(2):344–348
- Zou X, Zhao J, Li Y (2007) 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
- Zou X, Zhao J, Povey MJW, Holmes M, Mao H (2010) Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* 667(1–2):14–32