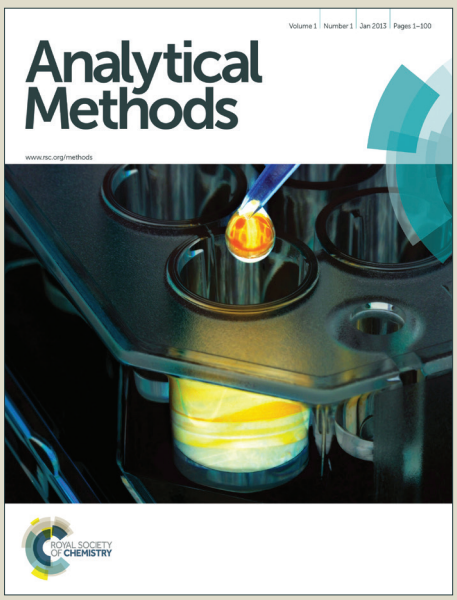
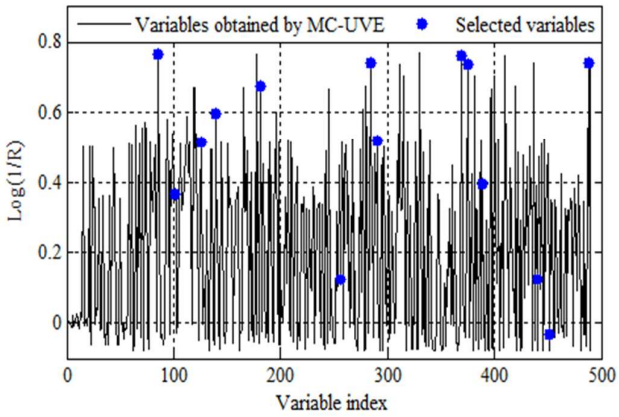
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods



Graphical Abstract: A newly combination (MC-UVE-SPA) of Monte Carlo-uninformative variable elimination and successive projections algorithm was proposed to select the most effective variables.

A Combination Algorithm for Variable Selection to Determine Soluble Solids Content and Firmness of Pear

Jiangbo Li[a,b], Chunjiang Zhao[a,*], Wenqian Huang[a], Chi Zhang[a], Yankun Peng [b]

[a] *Beijing Research Center of Intelligent Equipment for Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China*

[b] *College of Engineering, China Agricultural University, Beijing 100083, China*

Abstract

Informative variable (or wavelength) selection plays an important role in quantitative analysis by visible and near infrared (Vis/NIR) spectroscopy. In this study, a newly combination of Monte Carlo-uninformative variable elimination (MC-UVE) and successive projections algorithm (SPA) was proposed to select most effective variables. The selected variables were used as the inputs of least squares-support vector machine (LS-SVM) to build the MC-UVE-SPA-LS-SVM models for determining the soluble solids content (SSC) and firmness of pear. The conventional PLS models were also developed for comparison. The results indicated that calibration models built using MC-UVE-SPA-LS-SVM on 14 and 17 effective variables achieved the optimal performance for two internal quality indices comparing with full-spectrum PLS, MC-UVE-PLS, MC-UVE-LS-SVM and MC-UVE-SPA-PLS models by balancing between model accuracy and model complexity. The correlation coefficient (*r*) and root mean square error of prediction (RMSEP) and RPD for prediction set were 0.9486, 0.3244, 3.1598 and 0.8955, 1.1077, 2.2469 for SSC and firmness, respectively. The overall results indicated that Vis/NIR spectroscopy incorporated to MC-UVE-SPA-LS-SVM could be applied as an alternative fast and accurate method for the determination of SSC and firmness of pear, nondestructively. The effective variables might be important for the development of portable instruments and online

* Corresponding author. Tel.: +86 010 51503411.
E-mail address: zhaocj@nercita.org.cn.

monitoring for quality of pear.

Keywords: Vis/NIR spectroscopy, Pear, Internal quality, Least squares-support vector machine, Variable selection

1. Introduction

Visible and near infrared (Vis/NIR) spectroscopy is a fast, easy-to-use and non-destructive analytical technique.[1] Nowadays, it has been widely employed as alternatives to develop chemistry procedures for qualitative and quantitative analysis in agro-food products fields.[2-4] Multivariate calibration methods play crucial roles in Vis/NIR spectral analysis because Vis/NIR spectra are typically consisted of broad, weak, non-specific and overlapped band which there may be some irrelevant variables for multivariate calibration.[5] However, the used spectral data sets may have thousands of variables because the modern spectroscopy instrumentations usually have a high resolution. Thus, the calibration process is time-consuming and not convenient to fulfill the high speed feature of spectroscopy in industrial applications.[6] Moreover, the full spectral region may include wavelengths which contribute more collinearity, redundancies and noises than relevant information to models.[7] Thus, constructing a robust model using selected characteristic variables to improve modeling efficiency and prediction accuracy is crucial.[8,9] Also, variable selection can make the model simpler and better interpretation and improve the modeling and prediction rate of model. Hence, some variable selection methods, such as stepwise regression analysis,[10] generalized simulated annealing,[11] interval PLS,[12] genetic algorithms,[13] x-loading weights[14], regression coefficients,[15] wavelet transforms,[16] successive projections algorithm (SPA)[17] and correlation coefficient method,[18] have been developed to obtain a parsimonious model for the qualitative and quantitative analysis.

Among these methods, SPA uses simple projection operations to select variables with minimum of collinearity. It is a novelty variable selection algorithm and has also been employed to effectively solve the collinearity problems. Di Nezio et al.[17] used the SPA to determine directly and simultaneously five phenolic compounds in sea water, and found that MLR-SPA was potential for solving complex analytical problems. Liu et al.[19] applied the SPA to determine the

effective wavelengths of acetic, tartaric and lactic of plum vinegar. Different calibration models were compared. The results indicated that SPA-LS-SVM achieved the optimal performance for three acids comparing with full-spectrum PLS, SPA-MLR, SPA-PLS, RC-PLS and RC-LS-SVM. Balabin and Smirnov used the iPLS, MWPLS and SPA approaches for variable selection in PLS regression, they found that these algorithms decrease the prediction error of PLS regression by 8±5%.[20] Xu et al.[9] compared four different variable selection methods. Calibration model built using successive projection algorithm-multiple linear regression combined with GA (GA-SPA-MLR) on 18 selected wavelengths (2% of the total number of variables) exhibited higher coefficient of determination (R^2=0.880) and root mean square error of prediction (RMSEP=0.459^{o}Brix) for the validation set. However, two main problems may be existent: one is variables selected by SPA may be with low signal-noise ratio (S/N) or useless information for establishment of model which can affect prediction performance of model,[7] the other is SPA operation is time-consuming when the full spectra were considered. Therefore, it might be possible to improve the performance of calibration model if informative variables with low S/N were first obtained before SPA was performed. Uninformative variable elimination (UVE) can be very suitable for this aim. Elimination of uninformative variables can effectively remove useless information of spectral data and predigest calibration model.[21] UVE is a method for variable selection based on an analysis of regression coefficients. This technology has been widely applied in analytical chemistry.[22-25] Although good results were obtained, this method is time-consuming when it meets a large data set. And, there is a compelling problem for the UVE method because the regression coefficients are obtained by leave-one-out method in UVE process. The leave-one-out method in the multivariate data analysis usually results in over-fitting which leads to an under-estimation of true predictive error.[26]

In this paper, Monte Carlo cross-validation strategy proposed by Picard and Cook[27] is first introduced to UVE-PLS instead of leave-one-out strategy. Then, a new combination of Monte Carlo-uninformative variable elimination (MC-UVE) with SPA called MC-UVE-SPA is proposed for variables selection of NIR spectral data to determine the soluble solids content and firmness of pear. Soluble solids content and firmness are important quality attribute of pear.

Good-quality pears develop a buttery, juicy texture, following softening related to changes in cell wall structure.[28] Charting SSC and firmness as a function of postharvest storage time could provide valuable information for commercial decision-making, since fruit sold to the consumer must meet given quality standards based on these parameters.[29] Nowadays, Vis/NIR spectroscopy has been used to assess the internal quality of fruits such as apple,[30,31] orange,[32] kiwifruit,[33] intact banana and mango,[34] banana,[35,36] olive,[37,38] apricot fruit,[39] jujube,[40] date fruit,[41] etc. Near spectroscopy also had been applied to determine the SSC and firmness of pear.[9,42-44] However, up to our knowledge, few studies focused on the incorporated MC-UVE and SPA methods for variables selection of pear or other any fruit types.

In addition, among currently developed calibration models for quantitative analysis of pear, the most used models are linear analysis methods such as multiple linear regression (MLR), partial least squares (PLS) and principle component regression (PCR). These methods only discussed the linear relationship between the spectra and chemical components, whereas, there might be latent nonlinear information which is existed in the spectral data and related to the chemical constituents. In especial, some studies also found nonlinear models were better than linear models for prediction of internal quality of some fruits.[41,45-48] The reason might be that some latent nonlinear information was existed in the spectral data and related to the chemical constituents. Therefore, study on linear and nonlinear models was significative for specific fruit in order to develop the more robust prediction model. In addition, when fruit quality was online and fast estimated for the industrial application, some nonlinear factors such as fruit movement speed, change of temperature, change of fruit size, etc. might be included in the obtained original spectra. Although some preprocessing methods can decrease their effects, they could not be completely eliminated. Therefore, development of nonlinear models could be more helpful for online and fast measurement of fruit quality. In order to make advantage of the nonlinear information as well as the most linear information, a new and promising machine learning method,[49] least squares-support vector machine (LS-SVM) has been applied to develop the calibration models in the present work. Herein, a new combination of MC-UVE-SPA and LS-SVM is proposed for the determination of soluble solid content (SSC) and firmness of pear using Vis/NIR spectra. Combining both MC-UVE-SPA and LS-SVM was thought to be a

powerful calibration method using the selected relevant variables as well as the linear and nonlinear spectral information.

The objective of this paper is (1) to confirm the effective variables of SSC and firmness of pear by newly proposed combination of MC-UVE and SPA (2) to build the PLS linear calibration models and LS-SVM nonlinear calibration models based on selected variables; and (3) to achieve the best calibration models for predicting the SSC and firmness of pear after the comparison of PLS and LS-SVM calibration models.

2. Materials and Methods

2.1 Fruit samples

A total of three varieties of pears were obtained in local market named Cuiguan, Huanghua and Qingxiang. All of these pears were commonly variety in China. The equatorial diameter range of pears was 70-80 mm, and all samples were individually numbered and stored in standard refrigeration at 0℃ and 90% relative humidity (RH). Before the measurement, samples were taken out from cold storage accommodation and placed under room condition (20℃, 60% relative humidity) for more than 2 days to have an equalization room temperature to reduce the effect on the prediction accuracy by the temperature of samples.[50,51] 240 samples (80 for each variety) were prepared as calibration set and prediction set. Seventy samples were selected randomly from each variety and a total of 210 pear samples were used in the calibration set, whereas, the remaining 30 samples (10 for each variety) were selected as the validation set for estimating the performance of models. In order to compare the performance of different calibration models, the samples in the calibration and prediction sets would keep unchanged for all models.

2.2 Vis/NIR reflectance spectroscopy collection and preprocessing

The experimental system for testing pear fruit included a spectrometer (QualitySpec®Pro (350-1800 nm), Analytical Spectral Devices, Inc., USA) with an external fiber-optic cable installed at the high intensity contact probe, a Si detector for 350-1000 nm, an InGaAs detector for 1000-1800 nm, tungsten halogen lamp with 12V Bulb/A350610. A

white Teflon tile was used for white calibration before measurement. The angle between the incident light source and the detector fiber was set to 45°. The fiber cable delivered the collected optical energy into the spectrometer, where it was projected onto a holographic diffraction grating. The grating was separated and reflected the wavelength components for independent measurement by the detectors. The measurement system was arranged in reflectance mode for collecting Vis/NIR diffuse reflectance spectra from the peel and flesh of pear fruit. Pears were placed steadily upon the fruit holder, with the stem-calyx axis horizontal. The reflectance spectra from 350 to 1800 nm were measured at 1 nm interval with an average reading of 10 scans for each spectrum. Three separate spectral measurements were made at three marked locations on each sample around the equator (120°) of the fruit, and the average spectrum of these three measurements was used for calibration model. Finally, all spectral data were stored in a computer for further analysis.

Because the acquired data contain background information and noises besides sample information, it is very necessary to preprocess spectral data for obtaining reliable, accurate and stable calibration models.[52] In this study, the reflectance spectra were firstly transformed into ASCII format by using the ASD ViewspecPro software (Analytical Spectral Devices, Boulder, USA). Then, three spectra for each sample were averaged into one spectrum and transformed by log(1/R), where R = reflectance, into absorbance spectrum by using "The Unscrambler v9.7" software (CAMO PRECESS AS, Oslo, Norway). Finally, different preprocessing methods including multiplicative scatter correction (MSC), standard normal variate transformation (SNV), smoothing way of moving average, Savitzkye-Golay (SG), first derivative and second derivative were performed. After some trials and computation, it was found that the smoothing way of moving average and SNV were optimal for building the PLS models, whereas the first-derivative for development of LS-SVM models. Smoothing was necessary to optimize the signal-to-noise ratio.[53] SNV and first-derivative were effective for light scatter correction and reducing the baseline shift.[54] Therefore, these three preprocessing methods were used in subsequent study.

2.3 SSC and firmness analysis

The real SSC and firmness (compression test) values were determined using traditional destructive tests. The firmness by compression of diameter height fruit 3% was used as a general term to describe the mechanical properties of the fruit. The measurements were carried out with a 3.5 mm diameter Magness-Taylor (MT) probe, which was attached to a fruit sclerometer (Model: gY-1, hangzhou Huier Instrument, Co. Ltd., Hangzhou, China) with accuracy of ±0.1 N. The skin was removed thinly and the MT probe was forced into the peeled tissue. The force required to compress a fruit by about 3% of its diameter was recorded at an operating speed of 10 mm min^{-1}.[55] The maximum force was used as the measurement of fruit firmness. Next, about 1.0 mL filtered juice was then taken using a manual fruit squeezer (Model: HL-56, Shanghai, China) from the same position for SSC measurement with a digital refractometer (Model: PR-101α, Atago Co, Ltd, Tokyo, Japan). The refractive index accuracy is ±0.1 and the Brix (%) range is 0.0–45.0% with temperature correction. In terms of each sample, these two quality parameters were measured from three different marked positions where spectral measurement had just been performed, respectively. For further calculations, the averaged values for SSC and firmness values from three different locations of each fruit was stored and used as reference for proposed Vis/NIR spectrometric method, respectively.[56] In addition, the standard error of laboratory (SEL) represents the precision of a laboratory method. A statistical definition was given from these three measurements for each sample using the following equation: [57]

$$SEL = \sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{r}(y_{ij} - \bar{y}_j)^2}{n(r-1)}} \quad (1)$$

where n is the number of sanples, y_{ij} is the jth measurement of the ith sample, $\bar{y}_j$ is the reference method mean value of all there measurements of the ith sample and r is the number of measurements.

2.4 Monte Carlo-Uninformative Variable Elimination (MC-UVE)

The Monte Carlo method is a powerful and widely used technique for analyzing complex multi-variable problems.[58,59] It is a stochastic technique and investigates problems based on the use of random numbers and probability statistics. In this work, MC method is used to acquire stability of each variable. A large number of PLS models with

different calibration samples selected by the MC technique are produced firstly. Then, the stability of the corresponding coefficient is calculated by using the regression coefficients of these models. The MC technique has advantages of reducing dependence on single model and evaluating the reliability of each variable credibly to judge the remaining or rejection of them. In our study, a combination of MC and UVE (MC-UVE) is used for variable selection in Vis/NIR spectral data. It uses the stability defined in UVE method to evaluate the reliability of each variable, but the stability values are obtained through the Monte Carlo method replacing the leave-one-out procedure in UVE. Moreover, instead of adding random noise variables to the original data matrix as in UVE method to estimate the cutoff threshold, the wavelengths to be selected are determined directly by their stability. The retained variables were used as the inputs of PLS and LS-SVM models in order to compare the performance between linear model and nonlinear model. However, more variables may also be obtained by using MC-UVE. Therefore, the retained variables would be used to further analyze by using the SPA. (Note that more detailed information on UVE could be found in the literature.[21,60]

2.5 Successive projections algorithm (SPA)

SPA is a forward selection method, which starts with one wavelength, and then incorporates a new one at each iteration, until a specified number N of wavelengths is reached. The main purpose of this algorithm is to select wavelengths whose information content is minimally redundant. The main points are summarized here. The main steps of SPA can be summarized as follows, assuming that the first wavelength $k(0)$ and the number N are given:[61-64]

Step 0: Before the first iteration (n=1), let $\boldsymbol{x}_j = j$th column of $\mathbf{X}_{cal}$; j=1, 2, …, J;

Step 1: Let S be the set of wavelengths which have not been selected yet. That is, S={j such that $1 \leqq j \leqq J$ and $j \notin$ {$k(0)$, …, $k(n\text{-}1)$}};

Step 2: Calculate the projection of $\boldsymbol{x}_j$ on the subspace orthogonal to $\boldsymbol{x}_{k(n-1)}$ as

$$\mathbf{P}\mathbf{x}_j = \mathbf{x}_j - (\mathbf{x}_j^T \mathbf{x}_{k(n-1)})\mathbf{x}_{k(n-1)}(\mathbf{x}_{k(n-1)}^T \mathbf{x}_{k(n-1)})^{-1} \quad (2)$$

for all $j \in S$,where $\mathbf{P}$ is the projection operator;

Step 3: Let $k(n)$=arg(max||$\mathbf{P}\boldsymbol{x}_j$||, $j \in S$);

Step 4: Let $\boldsymbol{x}_j$= $\mathbf{P}\boldsymbol{x}_j$, $j \in S$;

Step 5: Let n=n+1. If n<N go back to Step 1;

End: The resulting wavelengths are {$k(n)$; n=0, … , N-1}.

The optimal initial variable and number of variables can be determined on the basis of the smallest root mean square error of validation (RMSEV) in validation set of MLR calibration. More details on the steps involved in SPA application can be found in previous studies. SPA was initially proposed by Araújo et al.[61] as a technique for variable selection for multiple linear regression (MLR). Nowadays, however, SPA method has found its way with different modelling techniques such as SPA-PLS,[19,65-67] SPA-least square support vector machine (LS-SVM),[19,25,68,69] SPA-artificial neural network (ANN),[70,71] SPA-support vector machine (SVM),[72] SPA linear discriminate analysis (LDA).[73,74] These studies indicated that SPA was a very good algorithm of variable selection and also can be used by combining with different multivariate calibration methods to build the linear and nonlinear models in special research field. In this work, the variables selected by the SPA were also used as the inputs of the PLS and LS-SVM. A graphical user interface for the SPA (GUI_SPA) is available at http://www.ele.ita.br/~kawakami/spa/.

2.6 Least squares-support vector machine (LS-SVM) analysis

LS-SVM, a state-of-the-art statistical learning method, is capable of dealing with linear and nonlinear multivariate analysis and resolving these problems in a relatively fast way.[75] Moreover, support vector machine (SVM) is capable of learning in high-dimensional feature space with fewer training data. It employs a set of linear equations instead of quadratic programming problems to obtain the support vectors. SVM embodies the structural risk minimization principle instead of traditional empirical risk minimization principle to avoid overfitting problems. The LS-SVM regression model can be expressed as:

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \qquad (3)$$

where $K(x, x_k)$ is the kernel function, x_k is the input vector, α_k is the Lagrange multiplier called support value, and b is the bias. The frequently-used kernel function $K(x, x_k)$ includes linear kernel, nonlinear kernel and radial basis function (RBF) kernel. $K(x, x_k)$ must follow Mercer's condition and perform the linear and nonlinear mapping. After the comparison of the above functions, we found that RBF kernel could handle the nonlinear relationship between the spectra and target attributes and gave a good performance. So, RBF kernel function was used in this study and defined as follows:

$$K(x, x_k) = \exp(-\|x_k-x\|^2/(2\sigma^2)) \quad (4)$$

where $\|x_k-x\|$ represents the distance between input vector and threshold vector, and σ is a width vector. All the calculations on LS-SVM were performed using Matlab 2008a (The Math Works, Natick, MA, USA). The free LS-SVM v1.5 toolbox (Suykens, Leuven, Belgium) was applied to develop the calibration models.

Before developing the LS-SVM calibration model, proper kernel function and the optimal model parameters firstly need to be obtained. Regularization parameter gam (γ) and RBF kernel function parameter sig2 (σ^2) are very important parameters in the modeling. To a great extent, these two parameters determine the learning ability, prediction ability and generalization ability of LS-SVM.[49] Gam (γ) is used to maximize model performance (on training) and minimize model complexity. Large gam(γ) implies little regularization, and thus a more nonlinear model. Sig2 (σ^2) influences the number of neighbors in the model. And large sig2 (σ^2) means more neighbors in the model which leads to a more nonlinear model.[76] In this paper, a two-step grid search technique with leave-one-out cross-validation was conducted to find the optimal parameter values gam (γ) and sig2 (σ^2). The first step of grid search was for a crude search with a large step size and the second step for the specified search with a small step size. The selection of the initial value has no effect on results because search process is a traverse process. Therefore, the initial values of gam (γ) and sig2 (σ^2) were set to 0.01 in this study.

2.7 Model evaluation standard

The performance of the calibration model was evaluated in terms of the root mean square error of calibration (RMSEC) in the calibration process for the calibration set. Root mean square error of prediction (RMSEP) and residual predictive deviation (RPD) were used to evaluate the performance of the prediction in the prediction process. The correlation coefficient (r) including r_{cal} (calibration) and r_{pre} (prediction) were used in all the processes.

RMSEC and RMSEP are defined as follows:

$$RMSEC = \sqrt{\frac{1}{n_c}\sum_{i=1}^{n_c}(y_{i,predicted} - y_{i,actual})^2} \quad (5)$$

$$RMSEP = \sqrt{\frac{1}{n_p}\sum_{i=1}^{n_p}(y_{i,predicted} - y_{i,actual})^2} \quad (6)$$

where $y_{i,predicted}$ and $y_{i,actual}$ is the predicted value and measured value of the i-th observation in the calibration set and prediction set, respectively. n_c, *and* n_p are the number of observations in the calibration and prediction set, respectively. The RMSECV was measured in form of leave-one-out cross-validation, which means that the model was fitted to n_c-1 training samples and used to predict the one left out for all n_c different left out samples.

The correlation coefficient (r) is defined as follows:

$$r = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_{i,actual} - y_{i,predicted})^2}{\sum_{i=1}^{n}(y_{i,actual} - \bar{y}_{i,actual})^2}} \quad (7)$$

where $y_{i,predicted}$ is the predicted value of the i-th observation, $y_{i,actual}$ is the measured value of the i-th observation, $\bar{y}_{i,actual}$ is the mean value of the calibration or prediction set, n is the number of observations in the data set.

RPD is the standard deviation of reference data for the prediction samples divided by the standard error of prediction (SEP) and provides a standardization of the SEP.[80] An RPD between 1.5 and 2 means that the model can discriminate low from high values of the response variable; a value between 2 and 2.5 indicates that coarse quantitative predictions are possible, and a value between 2.5 and 3 or above corresponds to good and excellent prediction accuracy, respectively.[4]

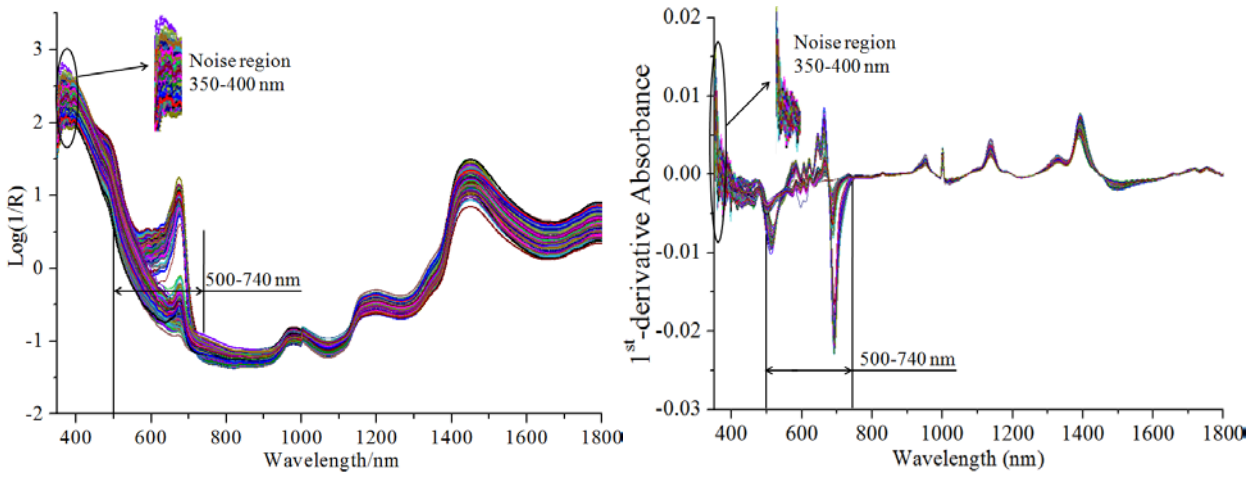
Generally, a good model should have higher r_{cal}, r_{pre} and RPD, lower RMSEC and RMSEP, but also a small

difference between RMSEC and RMSEP values.

3. Results and discussion

3.1 Overview of spectra and statistic values of SSC and firmness

The absorbance spectra of 240 pears by smoothing way of moving average and SNV preprocessing, and first-derivative preprocessing are shown in Fig.1(a) and (b), respectively. In the spectral curves, the trends of spectra were similar. There are only some differences in the wavelength range of 500-740 nm. This may be due to differences of peel color of different batches of pears. There is an absorbance at 680 nm which may be related to peel color of pear. The spectra curves are relatively flat and low except for 960 nm from 750 nm to 1100 nm. There is the absorption peak, associated with second overtone of band O-H.[56,77] There two strong absorption peaks at 1190 nm and 1450 nm, associated with a combination of second overtone of band C-H and overtones of band O-H in H_2O, and first overtone of bond O-H, respectively.[78] The first-derivative spectra shown in Fig.1(b) effectively reduced the baseline shift and enhanced the features and characteristics of spectra. In the middle part, the great jump and drop of the curves is due to the slope changes of the original spectra. Seen from the Fig.1, it could be also noticed that spectra introduced considerable noises at the beginning of the wavelength range. Therefore, the first 50 nm were removed from the spectral data, and only the spectral region (400-1800 nm) was employed for building the models.



(a) (b)

Fig.1. The Vis/NIR preprocessed absorbance spectra of pears. (a) smoothing way of moving average and SNV processing, and (b) first-derivative processing.

Table 1 showed that the statistic values of SSC and firmness of pear in the calibration and prediction sets. Thus, range, mean, standard deviation (S.D.) and number of samples for the calibration and prediction sets are summarized in the table, in addition to standard error of laboratory (SEL). As can be seen, the SSC and firmness measurements of 240 samples were fairly normally distributed around the mean values (10.81°Brix and 9.08 N), with standard deviations of 1.056 and 2.366, respectively. The SSC and firmness values in the calibration sets and prediction sets covered a large enough range. Moreover, the range of calibration sets is bigger than prediction sets for each quality index. The table 1 also shows the similarity between the descriptive statistics of calibration and prediction sets. These features are helpful to develop a good model. In addition, because the spatial variation of internal constituents may be existed for any quality parameter (SSC or firmness) in fruit, the precision of analyses (SEL) was acceptable (Table 1).

Table 1. Real measurement results of firmness and SSC of pear in the calibration and prediction sets.

Quality indices	Units	Sets	S. N.[a]	Min.	Max.	Mean	S. D.[b]	SEL
SSC	°Brix	Total samples	240	8.5	13.3	10.81	1.056	0.225
		Calibration set	210	8.5	13.3	10.81	1.072	0.231
		Prediction set	30	9.2	12.7	10.79	0.955	0.196
Firmness	N	Total samples	240	2.3	16.8	9.08	2.366	0.383
		Calibration set	210	2.3	16.8	9.13	2.347	0.352
		Prediction set	30	3.8	13.6	8.78	2.515	0.365

[a] S. N. shows sample number. [b]S. D. shows standard deviation.

3.3 Variables selection based on MC-UVE

Fig.2(a) and (b) show the stability of each variable in the spectral region 400-1800 nm for the SSC and firmness by MC-UVE method, respectively. In the figures, the dot lines represent the cutoff, which is determined by a number *N* of the informative variables. Variables between dot lines will be eliminated, and the variables whose stability lies out of the dot lines are used for further analysis. In order to more clear analyze the stability distribution of each variable, one spectral curve (red line) that is randomly selected from sample set is also shown in Fig.2(a) and (b). Seen from figures, it is easy found that effective variables (wavelengths) were distributed in Vis/NIR spectral region. How to determine the number of retained variables is very important because number of selected variables decides the stability and accuracy of the model. If the number of retained variables is too large, some uninformative variables may be contained in developed model and make its performance poor. On the contrary, if the number of retained variables is too small, the robustness and accuracy of model may be affected due to the loss of informative.

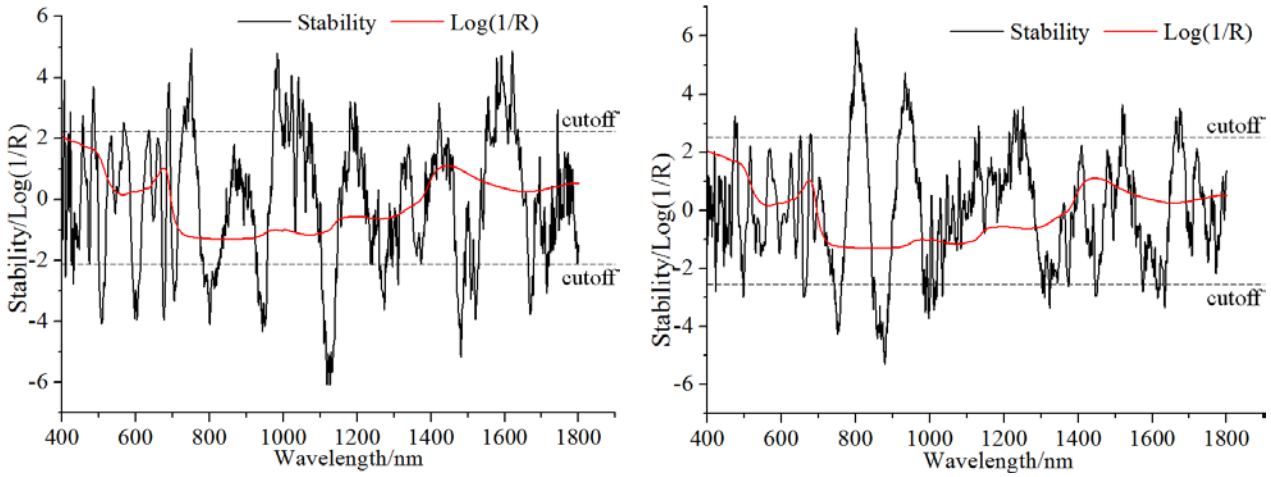


Fig.2. The stability distribution of each variable for prediction of the SSC (a) and firmness (b) by MC-UVE.

For each variable number *N*, a PLS model is developed and the model is then used to predict the sample set. Therefore, the change relationship between the RMSEP values and variable number *N* is investigated and shown in Fig.3(a) and (b), respectively. With a close observation, the RMSEP is large at the beginning, and then sharply decreased from *N*=1 to *N*=115 for SCC and firmness, respectively, although a little variation in the variable range from *N*=75 to *N*=100 and from *N*=30 to *N*=65 in Fig.3(a) and (b), respectively. Afterwards, RMSEP slowly decreases with the

variable number *N*, and when *N* is 490 and 240, the optimal values of RMSEP (0.3639 and 1.2377) are obtained for SSC and firmness prediction. Next, when *N* is bigger than 490 and 240, values of RMSEP have a little increase with increase of variable number. Above statement indicates that variable number selected to model will affect the prediction performance of model. Therefore, *N*=490 and *N*=240 are used for further analysis.

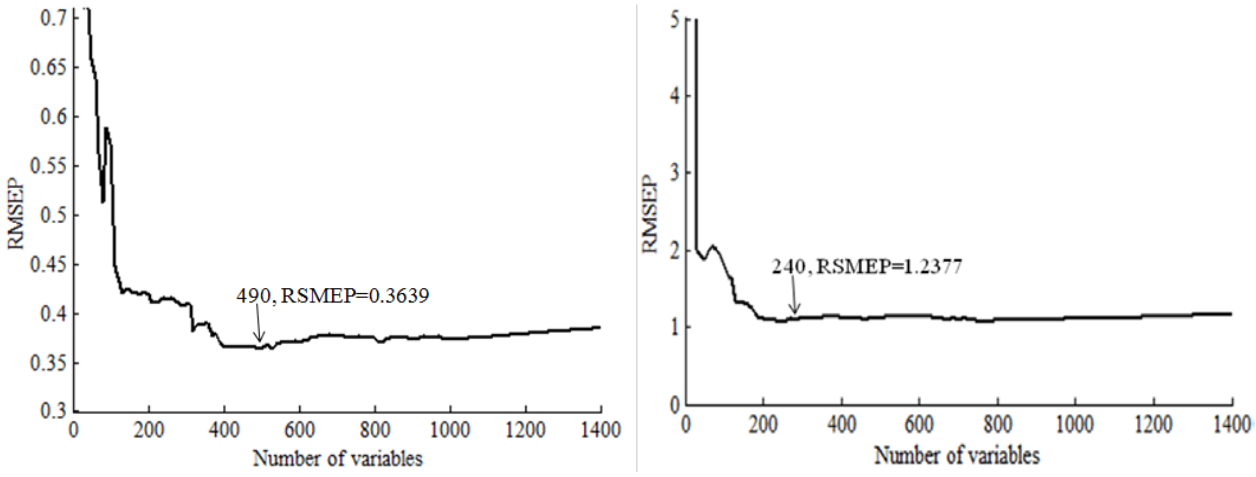


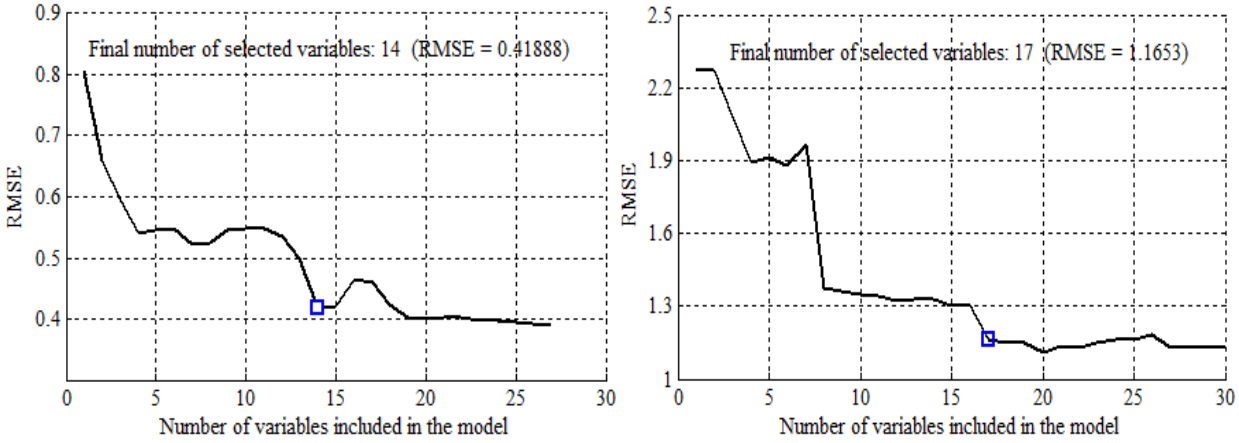
Fig.3. The change of RMSEP with number of selected variables for prediction of the SSC (a) and firmness (b) by MC-UVE.

3.4 Variables selection based on SPA

As analyzed above, although the variable number *N* of model is decreased from 1400 to 490 and 240 for SSC and firmness by MC-UVE method respectively, the *N* is still large for the development of portable instruments and online monitoring SSC and firmness of pear. In addition, although some uninformative variables in spectra data are eliminated, some collinear variables are still existent. These collinear variables will contain a large number of redundant information. Therefore, SPA was used for further variable selection for the determination of two internal quality indices of pear. During performing the SPA, a SPA-MLR procedure was applied for the calculation of a sequence of root mean square error (RMSE) values using the selected variable subsets. This process confirmed the achievement of the optimal number of selected variables with an optimal RMSE value, and this RMSE value was not significantly larger than the minimum RMSE value.

Fig.4(a) and (b) show the RMSE screen plots for the number of selected variables obtained by applying SPA. As

can be seen, the RMSE values sharply declined as SPA selected the first four variables. The RMSE values then gradually descended with increasing number of selected variables. When the 14 and 17 variables were selected (shown in open square markers) for MLR analysis, the RMSE reached its optimal values (0.41888°Brix and 1.1653 N) for prediction of SSC and firmness, respectively. The overall trends of RMSE curves for SSC and firmness are still descent with increasing number of selected variables although some fluctuations are also existent in descent.

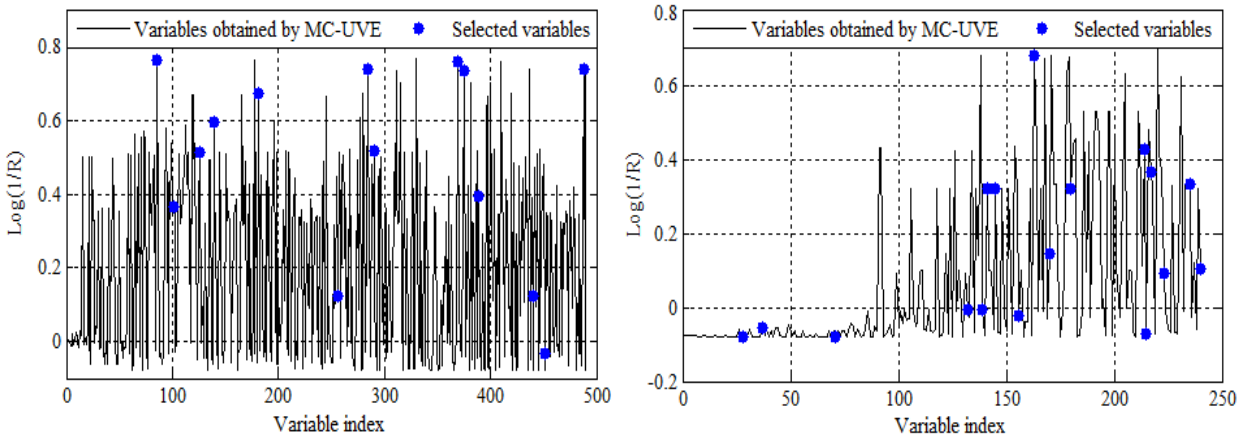


(a) (b)

Fig.4. RMSE plots obtained for SSC (a) and firmness (b) by SPA.

The selected variables for prediction of SSC and firmness (solid blue circles) are shown Fig.5(a) and (b), respectively. In the figures, the curves represent 490 and 240 variables obtained by MC-UVE for SSC and firmness of pear. Although less variables were obtained and shown in Fig.5, the exactly variables (effective wavelengths) corresponding to the variable index could not be get in the figures. Therefore, these effective wavelengths corresponding to raw spectra are also shown in Fig.6. The variables (vertical solid line) above and below raw spectra curves represent the effective variables (wavelengths) for prediction of firmness and SSC. As can be seen in Fig.6, the selected 14 variables for prediction of SSC included 406, 690, 1580, 503, 483, 702, 422, 1470, 409, 431, 1549, 1312, 726 and 1634 nm, and the selected 17 variables for prediction of firmness included 817, 934, 811, 1012, 987, 1664, 1673, 954, 473, 1324, 1632, 1523, 784, 1574, 1246, 1610 and 1214 nm. These selected variables by SPA are sequenced in order of relevance. The selected variables indicated that wavelengths at 406 nm and 817 nm were the most relevant

wavelengths of 14 and 17 selected variables for the prediction of SSC and firmness of pear, respectively. Comparing the selected variables for two internal indices, the variables for SSC and firmness were not the same. This might indicate that different chemistry feature is needful for analysis of SSC and firmness. In addition, it was also easy found (see Fig.6) that only one wavelength included in the Vis region (400-780 nm), whereas16 wavelengths in the NIR region (781-1800 nm) for firmness. This indicated that NIR spectra, especially, wavelengths regions in the 781-1150 nm, 1200-1350 nm and 1300-1700 nm were more important in the development of model for firmness prediction of pear. In terms of SSC prediction, 9 wavelengths distributed in the Vis region, whereas, 5 wavelengths in the NIR region. It indicated that the color variances had certain indirect and latent relationship with the chemical compositions such as the SSC of pear. Therefore, it should not be considered that only certain wavelength region (such as only Vis region or only NIR region) was used the prediction of SSC of pear. Comprehensive analysis based on Vis/NIR spectra is helpful for development of reliable model for predicting the SSC.



(a) (b)

Fig.5. The selected variables (shown in circle markers) for SSC (a) and firmness (b) by SPA.

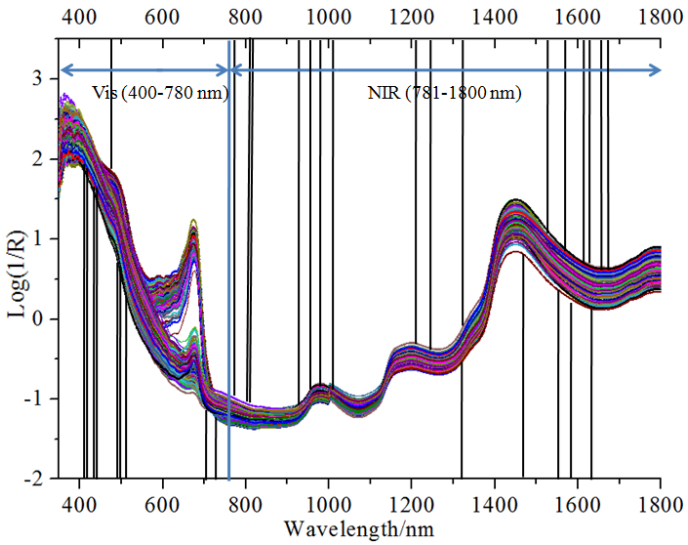


Fig.6. The selected variables (vertical solid line) corresponding to raw spectra for SSC (below spectra) and firmness (above spectra).

3.5 Calibration models and prediction performance

To more effectively assess the performance of informative variables (wavelengths) obtained by MC-UVE and SPA, the full-spectrum PLS models were developed without variable elimination for the prediction of SSC and firmness of pear. Moreover, PLS and LS-SVM models based on MC-UVE and MC-UVE-SPA were also developed to compare the performance of linear and nonlinear models for predicting the internal attributes in pear. In this work, the optimal combinations of (γ, σ^2) are achieved with (3.0×10^4, 1.1×10^3) and (5.9×10^4, 7.0×10^2) by MC-UVE-SPA-LS-SVM model, whereas (3.2×10, 2.8×10^4) and (2.1×10, 1.7×10^3) by MC-UVE-LS-SVM model for SSC and firmness, respectively.

The performance was confirmed by the samples in the prediction set. Different variables including full-spectrum, 490 and 240 variables by MC-UVE, and 14 and 17 variables by SPA were used as input matrixes to develop PLS and LS-SVM models. The prediction results are shown in Table 2. As can be seen, in terms of full-spectrum, the r_{pre}, RMSEP and RPD were 0.9173, 0.4221 and 2.5113 for SSC, while 0.8740, 1.1502 and 2.0579 for firmness, respectively. Results indicated that coarse quantitative predictions are possible for SSC and firmness by full-spectrum in this work.[4] However, all models based on MC-UVE and MC-UVE-SPA for SSC and firmness also yielded good results comparing with full-spectrum PLS models for both calibration and prediction sets. These results indicated the proposed variable selection methods (MC-UVE and SPA) are effective for determination of SSC and firmness of pear. Compared between

models by MC-UVE and models by MC-UVE-SPA, models with MC-UVE performed slightly better than models with MC-UVE-SPA for prediction of SSC and firmness, respectively. However, it is also worth mentioning that models with MC-UVE-SPA also preformed an acceptance result considering the largely reduced number of variables.

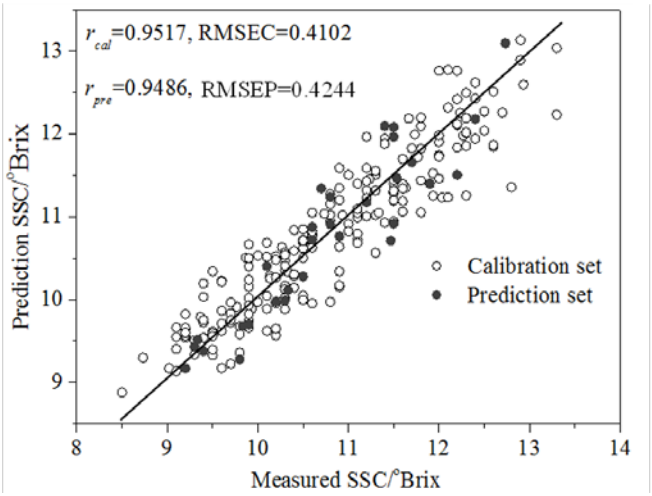
Table 2. The prediction results of SSC and firmness of pear by different models.

Parameters	Variable selection methods (models)	Variable	Calibration set		Prediction set		
			r_{cal}[a]	RMSEC	r_{pre}[b]	RMSEP	RPD
SSC (°Brix)	None (PLS)	1400	0.9490	0.3325	0.9173	0.4221	2.5113
	MC-UVE (PLS)	490	0.9409	0.3632	0.9420	0.3639	2.9796
	MC-UVE (LS-SVM)	490	0.9621	0.3208	0.9587	0.3239	3.5159
	MC-UVE-SPA (PLS)	14	0.9335	0.4011	0.9301	0.4162	2.7225
	MC-UVE-SPA (LS-SVM)	14	0.9517	0.3156	0.9486	0.3244	3.1598
Firmness (N)	None (PLS)	1400	0.9094	0.9821	0.8740	1.1502	2.0579
	MC-UVE (PLS)	240	0.8906	1.0713	0.8701	1.2377	2.0289
	MC-UVE (LS-SVM)	240	0.9226	0.8915	0.9035	0.9931	2.3332
	MC-UVE-SPA (PLS)	17	0.8807	1.1092	0.8616	1.1885	1.9701
	MC-UVE-SPA (LS-SVM)	17	0.9081	1.0352	0.8955	1.1077	2.2469

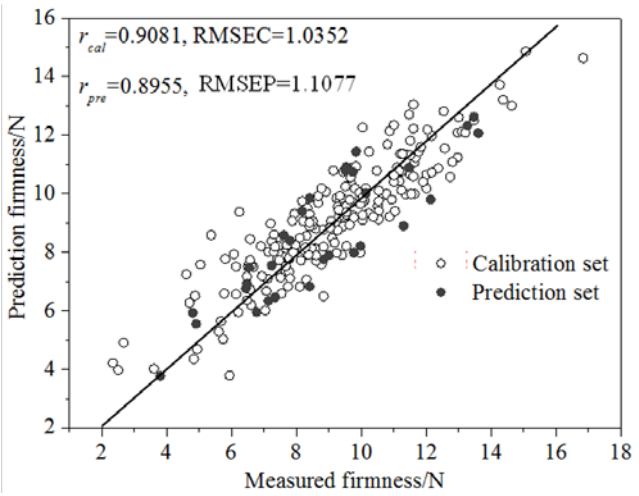
[a] r_{cal} represents correlation coefficient of calibration set. [b] r_{pre} represents correlation coefficient of prediction set.

It is also be seen from Table 2 that all LS-SVM models for two indices yielded better results than PLS models for calibration and prediction sets. The reason could be that LS-SVM could make advantage of the latent nonlinear information of spectral data which contributed a better prediction performance, and PLS only dealt with the linear relationships between the spectra and chemical compositions. In our study, the superiority of LS-SVM for handling nonlinear information was not very distinct might be due to little nonlinear information in spectral data or small samples

number used. However, the overall results still suggested that LS-SVM was a powerful analysis tool to quantity SSC and firmness of pear. Moreover, some recent studies also have found nonlinear models were better than linear models for prediction of internal quality of some fruits including pear fruit.[45-48,72] So, some latent nonlinear information might be existed in the spectral data and relate to the chemical constituents, especial for online and fast assessment of fruit quality. Compared with all the models shown in Table 2, LS-SVM models combining with MC-UVE-SPA variables selection method achieved the optimal results among all developed models for two indices both in calibration and in prediction sets by balancing between model accuracy and model complexity. Therefore, selection of effective variables (wavelengths) by MC-UVE-SPA was suitable for such situation in the present study. The r_{pre}, RMSEP and RPD were 0.9486, 0.3244 and 3.1598 for SSC, while 0.8955, 1.1077 and 2.2469 for firmness, respectively. The results showed that good prediction accuracy for SSC values and coarse quantitative prediction for firmness, respectively.[4] However, firmness prediction by MC-UVE-SPA-LS-SVM model was more effective than full spectrum LS-SVM or PLS models. Therefore, in this work, the effective variables could represent most of the features and characteristics of the whole spectral region, and could be applied instead of the whole wavelength region to predict the SSC and firmness of pear. Fig.7(a) and (b) show the scatter plots of predicted versus measured SSC and firmness obtained by LS-SVM combining both MC-UVE and SPA methods, respectively. The solid line is the regression line corresponding to the ideal, unity correlation between the predicted and measured values.



(a)



(b)

Fig.7. Measured vs. predicted values for SSC (a) and firmness (b) prediction by the LS-SVM calibration models combining both MC-UVE and SPA method.

The prediction results for the SSC obtained in this study are slightly superior or similar to those obtained by Ying and Liu[42] in pear with r=0.9109, RMSEP=0.395 and RPD=2.4235 for SSC, Paz et al.[29] in 'Blanquilla' pear with r^2=0.76 (coefficient of determination), SECV=0.59 and RPD=2.0413 for SSC and r^2=0.79, SECV=5.33 and RPD=2.1822 for firmness, Blanke[31] in 'Lucas' pear with r^2=0.54 and RPD=1.4744 for firmness. On the other hand, better results also have been found in pear with RPD=3.43 for SSC, and RPD=2.28 for firmness by Jiang and Zhu.[44] In addition, it is worth mentioning that a relatively good result also has been reported by Xu et al.[9] in Chinese royal pear with RPD=2.8867 for SSC for online application. Compared to the research results from other fruit types, the prediction results obtained by MC-UVE-SPA-LS-SVM model in this study is superior to those in Huang et al.[47] for SSC prediction of mulberry fruit (r = 0.8367, RMSEP=1.1721, RPD=1.8232) and in Zhang et al.[79] for SSC prediction of nanfeng mandarin orange (r = 0.8577, RMSEP =0.7113, RPD=1.9450) and in Blanke[31] for apple with RPD=1.7678 for firmness and RPD=2.7735 for SSC as well as in Møller et al.[30] for apple (r^2 = 0.80, RPD=2.2380 for SSC). However, the result in this study is also slightly inferior to the results obtained by Jamshidi et al.[32] for Valencia oranges (r = 0.96, RMSEP=0.33, RPD=3.5714 for SSC) and the results obtained by Mireei et al.[41] for firmness prediction of 'Mazafati' date fruit (r^2 = 0.90, RMSEP=1.3, RPD=3.1623) and similar to the results obtained by Shao et al.[46] for SSC of peach (r=0.9485, RMSEP=0.4155 and RPD=3.1568) and Ragni et al.[33] for firmness of 'Hayward' kiwifruit (r^2=0.806 and RPD=2.2704). Although the prediction accuracies in our study were not optimal compared with all the above models might due to appropriately influenced by instrumentation differences and different measurement modes such as reflectance and transmission, the application of LS-SVM combining both MC-UVE and SPA has been proved to be effective in this specific case, and the application of MC-UVE-SPA-LS-SVM supplied a new way for variables selection and further potential applications in other field.

4. Conclusions

The Vis/NIR spectroscopy spectrometric technique was developed to measure the SSC and firmness of pear fruit in the wavelength range of 400-1800 nm. The prediction results for PLS and LS-SVM models based on MC-UVE and SPA methods were compared. MC-UVE was proposed to eliminate the uninformative variables in the full-spectrum and SPA was also proposed as a new powerful way for further characteristic variables selection, and the new developed combination of MC-UVE-SPA-LS-SVM achieved the optimal prediction performance for SSC and firmness of pear comparing with full-spectrum PLS, MC-UVE-PLS, MC-UVE-LS-SVM and MC-UVE-SPA-PLS models by balancing between model accuracy and model complexity. The prediction results by MC-UVE-SPA-LS-SVM were that the *r*, RMSEP and RPD for prediction set were 0.9486, 0.3244, 3.1598 and 0.8955, 1.1077, 2.2469 for SSC and firmness, respectively. The overall results indicated that Vis/NIR spectroscopy incorporated to MC-UVE-SPA-LS-SVM could be applied as an alternative fast and accurate method for the determination of SSC and firmness of pear, nondestructively.

Acknowledgements

The authors gratefully acknowledge the financial support provided by Young Scientist Fund of National Natural Science Fundation of China （Project No. 31301236）, China Postdoctoral Science Foundation (Project No. 2012M520193) and Postdoctoral Science Foundation of Beijing of China (Project No. 2013ZZ-70).

References

1. M. S. Day and F. R. B. Fearn, *Laboratory Practice*, 1982, **31**, 328-330.
2. W. B. Wang and J. Paliwal, *Sens. & Instrumen. Food Qual.*, 2007, **1**, 193-207.
3. H. B. Huang, H. Y. Yu, H. R. Xu and Y. B. Ying, *J. Food Eng.*, 2008, **87**, 303-313.
4. B. M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron and J. Lammertyn, *Postharvest Biol. Tec.*, 2007, **46**, 99-118.

5. O. Berntsson, L. G. Danielsson, M. O. Johansson and S. Folestad, *Anal. Chim. Acta*, 2000, **419**, 45-54.

6. F. Liu, Y. He, and L. Wang, *Anal. Chim. Acta*, 2008, **610**, 196-204.

7. S. F. Ye, D. Wang and S. G. Min, *Chemom. Intell. Lab. Syst.*, 2008, **91**, 194-199.

8. M. Arakawa, Y. Yamashita and K. Funatsu, *J. Chemom.*, 2011, **25**, 10-19.

9. H. R. Xu, B. Qi, T. Sun, X. P. Fu and Y. B. Ying, *J. Food Eng.*, 2012, **109**, 142-147.

10. R. F. Kokaly and R. N. Clark, *Remote Sens. Environ.*, 1999, **67**, 267-287.

11. J. H. Kalivas, N. Roberts, and J. M. Sutter, *Anal. Chem.*, 1989, **67**, 2024-2030.

12. L. Nøgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc.*, 2000, **54**, 413-419.

13. J. R. Delphine, D. L. Massart, R. Leardi and O. E. De Noord, *Anal. Chem.*, 1995, **67**, 4295-4301.

14. Y. D. Liu, Y. B. Ying, X. P. Fu and H. S. Lu, *J. Food Eng.*, 2007, **80**, 986-989.

15. D. Wu, Y. He and S. Feng, *Anal. Chim. Acta*, 2008, **610**, 232-242.

16. B. K. Alsberg, A. M. Woodward, M. K. Winson, J. J. Rowland and D. B. Kell, *Anal. Chim. Acta*, 1998, **368**, 29-44.

17. M. S. Di Nezio, M. F. Pistonesi, W. D. Fragoso, M. J. C. Pontes, H. C. Goicoechea, M. C. U. Araujo, B. Fernández and S. Beatriz, *Microchem. J.*, 2007, **85**, 194-200.

18. M. Min and W. S. Lee, *T. ASAE*, 2005, **48**, 455-461.

19. F. Liu and Y. He. *Food Chem.*, 2009, **115**, 1430-1436.

20. R. M. Balabin and S. V. Smirnov, *Anal. Chim. Acta*, 2011, **692**, 63-72.

21. V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste and C. Sterna, *Anal. Chem.*, 1996, **68**, 3851-3858.

22. D. Chen, X. G. Shao, B. Hu and Q. D. Su, *Anal. Chim. Acta.*, 2004, **511**, 37-45.

23. R. Put, M. Daszykowski, T. Baczek and H. Y. Vander, *J. Proteome Res.*, 2006, **5**, 1618-1625.

24. J. Polański and R. Gieleciak, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 656-666.

25. D. Wu, H. X. Wu and J. B. Cai, *J. Infrared Millim. W.*, 2009, **28**, 423-427.

26. H. A. Martens and P. Dardenne, *Chemom. Intell. Lab. Syst.*, 1998, **44**(1-2), 99-121.

27. R. R. Picard and R. D. Cook, *J. Am. Stat. Assoc.*, 1984,**79**, 575-583.

28. H. Murayama, D. Sekine, Y. Yamauchi, M. Gao, W. Mitsuhashi and T. Toyomasu, *J. Exp. Bot.*, 2006, **57**, 3679-3686.

29. P. Paz, M. T. Sánchez, D. Pérez-Marín, J. E. Guerrero and A, Garrido-Varo. *Comput. Electron. Agri.*, 2009, **69**, 24-32.

30. S. M. Møller, S. Travers, H. C. Bertram and M. G. Bertelsen, *Eur. Food Res. Technol.*, 2013, **237** 1021-1024.

31. M. M. Blanke, *Erwerbs-Obstbau*, 2013, **55**, 19-24.

32. B. Jamshidi, S. Minaei, E. Mohajerani and H. Ghassemian, *Comput.Electron. Agr.*, 2012, **85**, 64-69.

33. L. Ragni, C. Cevoli, A. Berardinelli and F. A. Silaghi, *J. Food Eng.*, 2012, **109**, 32-37.

34. P. P. Subedi and K. B. Walsh, *Postharvest Biol. Tec.*, 2011, **62**, 238-245.

35. C. Y. Liew and C. Y. Lau, *Int. Food Res. J.*, 2012, **19**, 751-758.

36. P. Jaiswal, S. N. Jha and R. Bharadwaj, *Sci. Hortic-amsterdam*, 2012, **135**, 14-22.

37. J. A. Cayuela and M. C. P. Camino, *Eur. J. Lipid Sci. Tech.*, 2010, **112**, 1209-1217.

38. A. Morales-Sillero, V. M. Fernández-Cabanás, LCasanova, M. R. Jiménez, M. P. Suárez and P. Rallo, *J. Food Eng.*, 2011, **107**, 99-106.

39. C. Camps and D. Christen, *LWT-Food Sci. Technol.*, 2009, **42**, 1125-1131.

40. J. Wang, K. Nakano and S. Ohashi, *LWT-Food Sci. Technol.*, 2011, **44**, 1119-1125.

41. S. A. Mireei, S. S. Mohtasebi and M. Sadeghi, *Int. J. Food Prop.*, DOI:10.1080/10942912.2012.678533.

42. Y. B. Ying and Y. D. Liu, *J. Food Eng.*, 2008, **84**, 206-213.

43. T. Sun, H. J. Lin, H. R, Xu and Y. B. Ying, *Postharvest Biol. Tec.*, 2009, **51**, 86-90.

44. H. Jiang and W. X. Zhu, *Food Anal. Methods*, 2013, **6** 569-577.

45. F. Cao, D. Wu and Y. He, *Comput.Electron. Agr.*, 2010, **71**, S15-S18.

46. Y. N. Shao, Y. D. Bao and Y. He, *Food Bioprocess Tech.*, 2011, **4**, 1376-1383.

47. L. X. Huang, D. Wu, H. F. Jin, J. K. Zhang, Y. He and C. F. Lou, *Biosyst. Eng.*, 2011, **109**, 377-384

48. J. B. Li, W. Q. Huang, C. J. Zhao and B. H. Zhao, *J. Food Eng.*, 2013, **116**, 324-332.

49. J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, *Least squares support vector machines. Singapore*. World Scientific, Singapore, 2002.

50. S. Kawano, H. Abe and M. Iwamoto, *J. Near Infrared Spec.*, 1995, **3**, 211-218.

51. D. Cozzolino, L. Liu, W. U. Cynkar, R. G. Dambergs, L. Janik, C. B. Colby and M. Gishen, *Anal. Chim. Acta.*, 2007, **588**, 224-230.

52. H. Cen and Y. He, *Trends Food Sci. Tech.*, 2007, **18**, 72-83.

53. J. H. Perkin, B.Tenge and D. E. Honigs, *Atom. Spectrosc.*, 1988, **43**, 575-603.

54. R. J. Barnes, M. S. Dhanoa and S. J. Lister, *Appl. Spectrosc.*, 1989, **43**, 772-777.

55. G. Fan, J. W. Zha, R. Dub and L. Gao, *J. Food Eng.*, 2009, **93**, 416-420.

56. Y. D. Liu, X. D. Sun and A. G. Ouyang, *LWT-Food Sci. Technol.*, 2010, **43**, 602-607.

57. J. Workman and M. Mark, *Spectrosc*, 2006, **21**, 19-24.

58. L. H. Wang, S. L. Jacques and L. Q. Zheng, *Comput. Meth. Prog. Bio.*, 1995, **47**, 131-146.

59. M. Marseguerra and A. Zoia. *Ann. Nucl. Energy*, 2006, **33**, 1396-1407.

60. W. S. Cai, Y. K. Li and X. G. Shao, *Chemom. Intell. Lab. Syst.*, 2008, **90**, 188-194.

61. M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvã, T. Yoneyama, H. C. Chame and V. Visani, *Chemom. Intell. Lab. Syst.*, 2001, **57**, 65-73.

62. S. F. C. Soares, A. A. Gomes, A. R. Galvao, M. C. U. Araujo, and R. K. H. Galvao, *Trends Anal. Chem.*, 2013, **42**, 84-98.

63. A. D. Gomes, R. K. H. Galvao, M. C. U. de Araujo, G. Veras and E. C. da Silva, *Microchem. J.*, 2013, **110**, 202-208.

64. H. M. Paiva, S. F. C. Soares, R. K. H. Galvao and M. C. U. Araujo, *Chemom. Intell. Lab. Syst.*, 2012, **118**, 260-266.

65. D. Wu, H. Shi, S. Wang, Y. He, Y. Bao and K. Liu, *Anal. Chim. Acta.*, 2012, 726, 57-66.

66. N. Goudarzi and M. Goodarzi, *Anal. Methods*, 2010, **2**, 758-764.

67. A. G. Ouyang and J. Liu, *Meas. Sci. Technol.*, doi:10.1088/0957-0233/24/2/025502.

68. F. Liu, Z. L. Jin, M. S. Naeem, T. Tian, F. Zhang, Y. He, H. Fang, Q. F. Ye and W. J. Zhou, *Food Bioprocess Technol.*, 2011, **4**,

1314-1321.

69. D. Wu, X. J. Chen, X. G. Zhu, X. C. Guan and G. C. Wu. *Anal. Methods*, 2011, **3**, 1790-1796.

70. N. Goudarzi, M. Goodarzi, M. C. U. Araujo and R. K. H. Galvao, *J. Agr. Food Chem.*, 2009, **57**, 7153-7158.

71. M. Khanmohammadi, A. B. Garmarudi, K. Ghasemi, S. Garrigues and M. Guardia, *Microchem. J.*, 2009, **91**, 47-52.

72. M. Goudarzi, M. Goodarzi, M. Arab Chamjangali and M. H. Fatemi, *Chinese Chem. Lett.*, 2013, **24**, 904-908.

73. M. Ghasemi-Varnamkhasti, S. S. Mohtasebi, M. L. Rodriguez-Mendez, A. A. Gomes, M. C. U. Araújo and R. K. H. Galvão, *Talanta*, 2012, **89**, 286-291.

74. E. D. T. Moreira, M. J. C. Pontes, R. K. H. Galvão and M. C. U. Araújo, *Talanta*, 2009, **79**, 1260-1264.

75. Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

76. T. Coen, W. Saeys, H. Ramon and J. De Baerdemaeker, *J. Chemometr.*, 2006, **20**, 184-192.

77. H. Shinzawa, P. Ritthiruangdej and Y. Ozaki, *Appl. Spectrosc.*, 2011, **5**, 549-556.

78. E. Lanza and B. W. Li, *J. Food Sci.*, 1984, **49**, 995-998.

79. L. Zhang, X. Long, M. H. Liu and J. Li, (2012). *Advanced Materials Research*, 2012, **361**-**363**, 1634-1637.

80. P. Williams and K. Norris, *Near-Infrared Technology in the Agricultural and Food Industries*, Amer Assn of Cereal Chemists, St. Paul, MN, 2001.