ORIGINAL PAPER

Visible/Near-Infrared Spectra for Linear and Nonlinear Calibrations: A Case to Predict Soluble Solids Contents and pH Value in Peach

Yongni Shao · Yidan Bao · Yong He

Abstract Two sensitive wavelength (SWs) selection methods combined with visible/near-infrared (Vis/NIR) spectroscopy were investigated to determine the soluble solids content (SSC) and pH value in peaches, including latent variables analysis (LVA) and independent component analysis (ICA). A total of 100 samples were prepared for the calibration ($n=70$) and prediction ($n=30$) sets. Calibration models using SWs selected by LVA and ICA were developed, including linear regression of partial least squares (PLS) analysis and nonlinear regression of least squares-support vector machine (LS-SVM). In the nonlinear models, four SWs selected by ICA achieved the optimal ICA-LS-SVM model compared with LV-LS-SVM and both of them better than linear model of PLS. The correlation coefficients (r_p and r_{cv}), root mean square error of cross validation, root mean square error of prediction, and bias by ICA-LS-SVM were 0.9537, 0.9485, 0.4231, 0.4155, and 0.0167 for SSC and 0.9638, 0.9657, 0.0472, 0.0497, and −0.0082 for pH value, respectively. The overall results indicated that ICA was a powerful way for the selection of SWs, and Vis/NIR spectroscopy incorporated to ICA-LS-SVM was successful for the accurate determination of SSC and pH value in peach.

Keywords Vis/NIR spectroscopy · Peach · Soluble solids content (SSC) · Partial least squares (PLS) · Independent component analysis (ICA) · Least squares-support vector machine (LS-SVM)

Y. Shao · Y. Bao (✉) · Y. He (✉)
College of Biosystems Engineering and Food Science,
Zhejiang University,
Hangzhou 310029, China
e-mail: ydbao@zju.edu.cn
e-mail: yhe@zju.edu.cn

Introduction

Visible and near-infrared reflectance spectroscopy (Vis/NIRS) is a well-established technique for constituent analysis of agricultural and food products as it has many advantages when compared with classical chemical and physical analytical methods (Murray et al. 2001). It has several attractive features including fast analytical speed, ease of operation, and nondestructive measures. The most important one is that it can give the response of the covalent molecular bounds of its corresponding chemical constituents to the spectrum, such as O–H, N–H, and C–H. In recent years, it is regarded as a method for nondestructive sensing of fruit quality. Lammertyn et al. (1998) examined the prediction capacity of the quality characteristics like acidity, firmness, and soluble solid content of Jonagold apples with a wavelength range between 380 and 1,650 nm. Carlini et al. (2000) used visible and near-infrared spectra to analyze soluble solids in cherry and apricot. Lu (2001) evaluated the potential of NIR reflectance for measurement of the firmness and sugar content of sweet cherries. McGlone et al. (2003) used Vis/NIR spectroscopy to analyze mandarin fruit. Pedro and Ferreira (2005) predicted solids and carotenoids in tomato by using NIR.

Recently, variable selection or uninformative variable elimination has attracted more and more attention for the development of multi-component calibrations using spectroscopic techniques. The recently developed methods for variable selection included generalized simulated annealing (Kalivas et al. 1989), genetic algorithm (Jouan-Rimbaud et al. 1995), correlation coefficients and B-matrix coefficients (Min and Lee 2005), latent variables analysis(LVA), x-loading weights (Esbensen 2002; Liu et al. 2007), uninformative variable elimination (Centner et al. 1996), regression coefficient analysis (Liu et al. 2008; Chong and

Jun 2005), independent component analysis (ICA; Hyvärinen et al. 2001; Krier et al. 2008), and so on. Among these methods, ICA has recently attracted broad attention and has been successfully used in many fields, e.g., medical signal analysis, image processing, dimension reduction, fault detection, and near-infrared spectral data analysis (Hyvarinen 1999; Hoyer and Hyvarinen 2000; Hyvarinen and Hoyer 2000; Chen and Wang 2001; Bi et al. 2004).

Various calibration methods have been used to relate near-infrared spectra with measured properties of materials. Principal components regression, partial least squares (PLS), multiple linear regression, and artificial neural networks are the most used multivariate calibration techniques for NIRS (Workman et al. 1996). PLS was usually considered for a large number of applications in fruit and juice analysis and was widely used in multivariate calibration because it takes the advantage of the correlation relationship that already exists between the spectral data and the constituent concentrations. However, PLS is based on linear models, and unsatisfactory results may be obtained when nonlinearity is present (Li and He 2009).

Least-squares support vector machine (LS-SVM) could handle the linear and nonlinear relationships between the spectra and response chemical constituents (Suykens and Vanderwalle 1999; Suykens et al. 2002, Liu and He 2009). Therefore, a new combination of ICA with LS-SVM was proposed as a nonlinear calibration model for quantitative analysis using spectroscopic techniques. The performance of ICA-LS-SVM was evidenced by a case study to determine the soluble solids content (SSC) and pH value of peach, with the purpose of developing a fast and accurate nonlinear model using fewer selected variables for quality estimation of peach.

The objective of this paper is to study the performance of ICA and LVA combined with nonlinear methods (LS-SVM) and compare them with the traditional approach based on linear regression methods (PLS) to predict the SSC and pH value.

Materials and Methods

Sample Preparation

To get dependable prediction equations from Vis/NIRS, it is necessary that the calibration set covers a wide range of fruit samples with enough variability. Two kinds of peaches: Milu peach (from Fenghua of Zhejiang, China) and Hongxianjiu peach (from Shandong, China) used in this experiment were cultivated in 2008, and the maximum, minimum, and average diameters of the peach used in this study were 98, 84, and 87 mm. A total of 100 peaches used for the experiment were purchased at a local market and

stored for 2 days at 20 °C. All samples were divided into calibration sets of 70 samples (35 samples for each kind) and prediction sets of 30 samples (15 samples for each kind) for performing a cross-validation. No single sample was used in calibration and prediction sets at the same time. The performance of different calibration models and the samples in the calibration and prediction sets kept unchanged for all calibration models.

Before measuring the SSC and pH value, peach samples were ground and filtered to juice. The SSC was measured by an Abbebenchtop refractometer (Model: WAY-2S, Shanghai Precision & Scientific Instrument Co. Ltd., Shanghai, China). The refractive index accuracy is ±0.0002, and the °Brix (%) range is 0–95% with temperature correction. pH was measured by a pH meter (SJ-4A, Exact instrument Co., Ltd., Shanghai, China). Both of these measurements were performed immediately after Vis/NIRS measurements.

Spectral Collection and Preprocessing

A Lowell pro-lam interior light source (Assemble/128930) with the Lowell pro-lam 14.5 V Bulb/128690 tungsten halogen, which could be used both in the visible and near-infrared regions, was placed at a distance of 300 mm from the fruit surface, and it never changed to all the 100 peach samples. The reflectance was calculated by comparing the near-infrared energy reflected from the sample with the standard reference. Three reflection measurements (325–1,075 nm) were taken at three equidistant positions around the equator (approximately 120°) of each peach, using a spectrophotometer (FieldSpec Pro FR (325–1,075 nm)/ A110070, Analytical Spectral Devices, Inc. (ASD)) and the RS^2 software for Windows®. Figure 1 shows the schematic diagram of the experimental apparatus for peach. For each reflectance spectrum, the scan number was set to 10 at exactly the same position, and thus the total scan number was 30. After all spectral measurements were performed, the acquired data were properly stored for later use.

Before the calibration stage, the Vis/NIR spectra data were preprocessed. Considering their effects on the measurement accuracy, the first 75 and the last 75 wavelength values were eliminated for all analyses. Thus, all consideration was based on the wavelength range from 400 to 1,000 nm. The Savitzky–Golay smoothing was used to reduce the noise (Savitzky and Golay 1964; Gorry 1990), with a window width of 7 (3–1–3) points. The multiplicative scatter correction (MSC) was used to correct additive and multiplicative effects in the spectra (Helland et al. 1995). The preprocessing was carried out using "The Unscrambler V9.6" (CAMO PROCESS AS, Oslo, Norway), a statistical software package for multivariate calibration.

Fig. 1 The schematic diagram of the experimental apparatus for peach

Spectrometer

50W halogen lamp

Computer

Sample
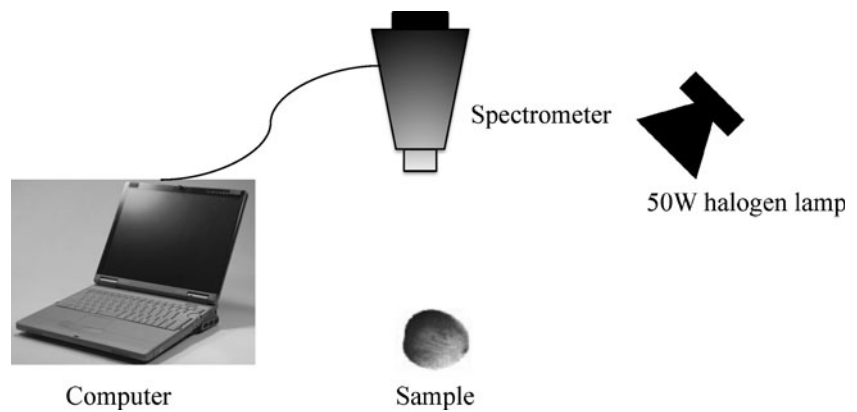
Partial Least Squares Analysis

In the development of PLS model, calibration models were built between the spectra and the SSC and pH values. Full cross-validation was used to evaluate the quality and to prevent over-fitting of calibration models. The optimal number of LVs (LVA) was determined by the lowest value of predicted residual error sum of squares. The prediction performance was evaluated by the correlation coefficients (r_p and r_{cv}), root mean square error of calibration (RMSEC), and root mean square error of cross validation (RMSECV) or prediction (RMSEP).

Independent Component Analysis

ICA is a well-established statistical signal processing technique that aims to decompose a set of multivariate signals into a base of statistically independent components with the minimal loss of information content. The independent components are latent variables, meaning that they cannot be directly observed, and the independent component must have non-Gaussian distributions.

There are lots of algorithms for performing ICA (Hyvärinen et al. 2001; Lee 1998). Among these algorithms, the fast fixed-point algorithm (FastICA) is highly efficient for performing the estimation of ICA, which was developed by Hyvärinen and Oja (2000). FastICA was chosen for ICA and carried out in Matlab 7.0 (The Math Works, Natick, USA).

Least Squares-Support Vector Machine

LS-SVM can work with linear or non-linear regression or multivariate function estimation in a relatively fast way (Suykens and Vanderwalle 1999; Borin et al. 2006). It uses a linear set of equations instead of a quadratic programming problem to obtain the support vectors. The details of LS-SVM algorithm could be found in the literature (Guo et al. 2006; Chen et al. 2007).

In the model development using LS-SVM and radial basis function (RBF) kernel, the optimal combination of gam (γ) and sig^2 (σ^2) parameters was selected when resulting in smaller root mean square error of cross validation. In this study, gam (γ) were optimized in the range of 2^{-1}–2^{10} and 2–2^{15} for sig^2 (σ^2) with adequate increments. These ranges were chosen from previous studies where the magnitude of parameters was optimized (Liu et al. 2009). The grid search had two steps; the first step was for a crude search with a large step size, and the second step was for the specified search with a small step size. The free LS-SVM toolbox (LS-SVM v 1.5, Suykens, Leuven, Belgium) was applied with MATLAB 7.0 to develop the calibration models.

Results and Discussion

Overview of Spectra and Statistic Values of SSC and pH Value

Figure 2 shows the average Vis/NIRS spectral curve for two types of peach. There exist two of the spectra obviously different from others, which were caused by its irregular
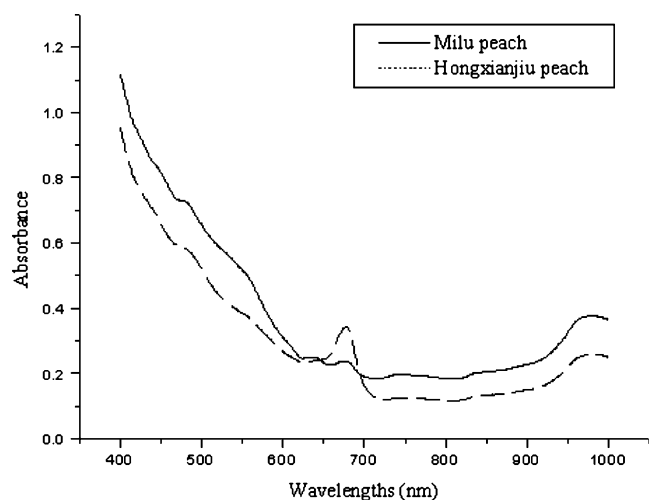

Fig. 2 The average spectral curves for two types of peach samples from wavelengths 400 to 1,000 nm

shape, and this may influence the spectral curve, but to the SSC and pH analysis, it was normal. The range of SSC was from 7.4 to 13.5 °Brix, and the pH value was from 4.08 to 4.82. The trend of spectral curves in Vis/NIR region is similar, and the average spectra for each peach type treated with second derivative to find some peaks and valleys are shown in Fig. 3. There are noises near wavebands 400 and 1,000 nm. The prominent features are the absorption peaks associated with a third overtone stretch of CH and second and third overtones of OH around 700~900 nm; wavelengths below 700 nm were mainly attributed to the color or shape of peach.

PLS Models

PLS model was developed after preprocessing the spectra by Savitzky-Golay smoothing and MSC. Calibration models were built between the spectra and SSC or pH values. Different LVs were applied to build the calibration models, and no outliers were detected in the calibration set during the development of PLS models. The models were used to predict the remaining 30 samples, and the r_p, r_{cv}, RMSECV, RMSEP, and bias were 0.9295, 0.9241, 0.6259, 0.6214, and −0.0291 for SSC and 0.8542, 0.8497, 0.1108, 0.1045, and 0.0426 for pH value, respectively. Figure 4 shows the predicted versus measured charts. The solid line is the regression line corresponding to the correlation between the prediction and reference values.

Sensitive Wavelengths Analysis Based on Regression Coefficients

The sensitive wavelengths reflecting the characteristics of spectra for SSC and pH were obtained based on regression coefficients (Fig. 5). The wavelengths between 700 and
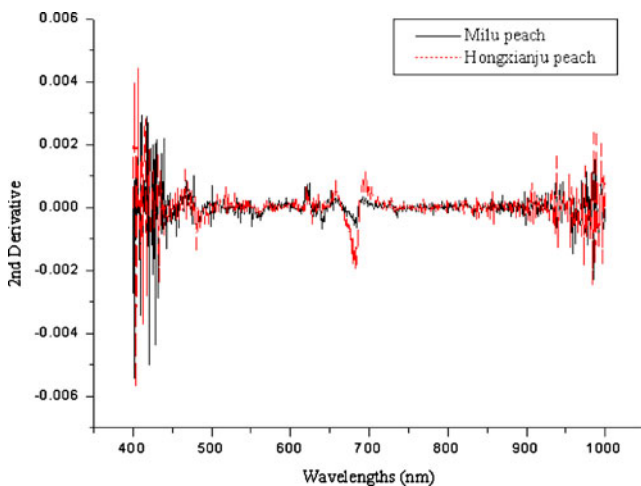

Fig. 4 Vis/NIR prediction results of 30 unknown samples from the PLS model for SSC (**a**) and pH (**b**)

Fig. 3 The average spectral curves for two peach types after second derivative preprocessing from wavelengths 400–1,000 nm

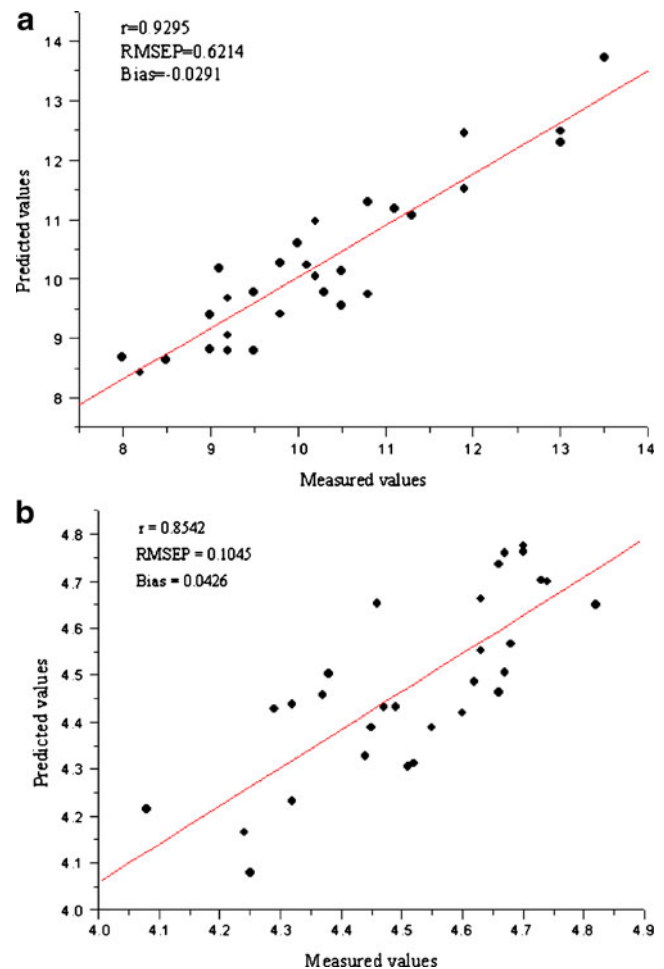950 nm may result from a third overtone stretch of CH and second and third overtones of OH in peach which was referred by Rodriguez-Saona et al. (2001) in their article about rapid analysis of sugars in fruit juices by Fourier transform-NIR spectroscopy. Slobodan and Ozaki (2001) also proposed the detailed band assignment for the short-wave NIR region useful for various biological fluids. From Fig. 5a, we can find that wavelengths of 650~680 and 970~990 nm might be of particular importance for the SSC calibration. The regression coefficients shown in Fig. 5b also have strong peaks and valleys at certain wavelengths such as 685~695 nm, and 910~925 nm may relate to pH value.

However, the absorption wavelengths 650~680 nm could be attributed to the color of peach, and the same effect occurs with pH in the 685~695 nm region. So in our research, to SSC, wavelengths 970~990 nm might be of particular importance, and to pH value, 910~925 nm were better. This finding was similar to the earlier literature; He (1998) found that wavelength of 914 nm was sensitive to
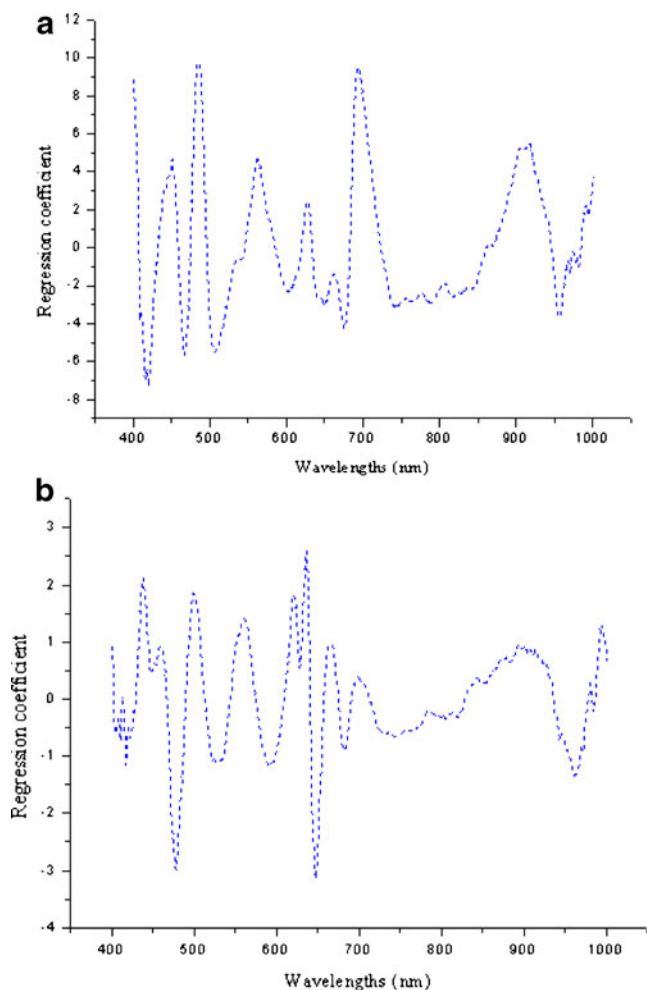
Fig. 5 Regression coefficients with corresponding loadings for SSC (**a**) and pH (**b**)

the SSC of satsuma mandarins, and wavelengths near 900 nm were corresponding to organic acid of oranges.

LV-LS-SVM Models

LVs obtained from PLS were applied as inputs of LS-SVM models to improve the training speed and reduce the training error of Vis/NIR model because the training time increased with the square of the number of training samples and linearly with the number of variables. From the aforementioned analysis of the performance of PLS models, the LVs from the Vis/NIR region were used as new eigenvectors to enhance the features of spectra and reduce the dimensionality of the spectra data matrix. Several LVs were extracted from the spectra of 100 samples. Table 1 shows the explained variance of Y of the first five to nine LVs. The variance of the first five LVs could explain more than 90% of the total variance, and the ninth LV only interpreted an additional 0.814%, which contributed not so much as the other LVs. So it was not necessary for the

consideration of less than five LVs or more than nine LVs. The LS-SVM models with five to nine LVs were developed separately in order to find out the best number of LVs.

Before the LS-SVM calibration model was built, three steps are crucial for the optimal input feature subset, proper kernel function, and the optimal kernel parameters. Firstly, the five to nine LVs obtained from PLS analysis were used as the input data set. Secondly, RBF could handle the nonlinear relationships between the spectra and target attributes. Finally, two important parameters gam (γ) and sig^2 (σ^2) should be optimal for RBF kernel function as aforementioned in multivariate analysis.

The performance of the Vis/NIR models was evaluated by 30 samples in prediction set. With a comparison of the results for calibration and prediction sets, the best performance was achieved with six LVs. The r_p, r_{cv}, RMSECV, RMSEP, and bias for prediction sets were 0.9409, 0.9432, 0.4918, 0.5004, and 0.0242 for SSC and 0.9208, 0.9184, 0.0812, 0.0707, and −0.0120 for pH value, respectively. Figure 6 shows the predicted versus measured charts. The results for calibration and prediction sets showed that LV-LS-SVM models outperformed PLS models.

ICA-LS-SVM Models

ICA was applied for the selection of sensitive wavelength (SWs), which could reflect the main features of the raw absorbance spectra. FastICA was used to the preprocessed spectra data, and the main absorbance peaks and valleys were indicated by the spectra of ICs. The SWs were selected by the weights of the first four ICs, and wavelengths with the highest weights were selected as the SWs. They were 985~992, 905~912, 975~982, and 702~708 nm for SSC, 905~916, 997~1,000, 705~718, and 976~984 nm for pH value. In order to evaluate the performance of SWs, they were applied as the input data matrix to develop the ICA-LS-SVM models. The prediction results showed the r_p, r_{cv}, RMSECV, RMSEP, and bias were 0.9537, 0.9485, 0.4231, 0.4155, and 0.0167 for SSC and 0.9638, 0.9657, 0.0472, 0.0497, and −0.0082 for pH value, respectively. Figure 7 shows the predicted versus measured charts. The ICA-LS-SVM models achieved a better performance compared to the best LV-LS-SVM models both in calibration and prediction sets. Wavelengths around 905~916 and 975~992 nm were close to the third overtone stretch of CH and second and third overtones of OH in peach. Therefore, the selection of SWs was suitable for such situation in the

Table 1 The explained variance of latent variables

LVs	5	6	7	8	9
EV (%)	90.412	93.121	95.438	97.108	97.922

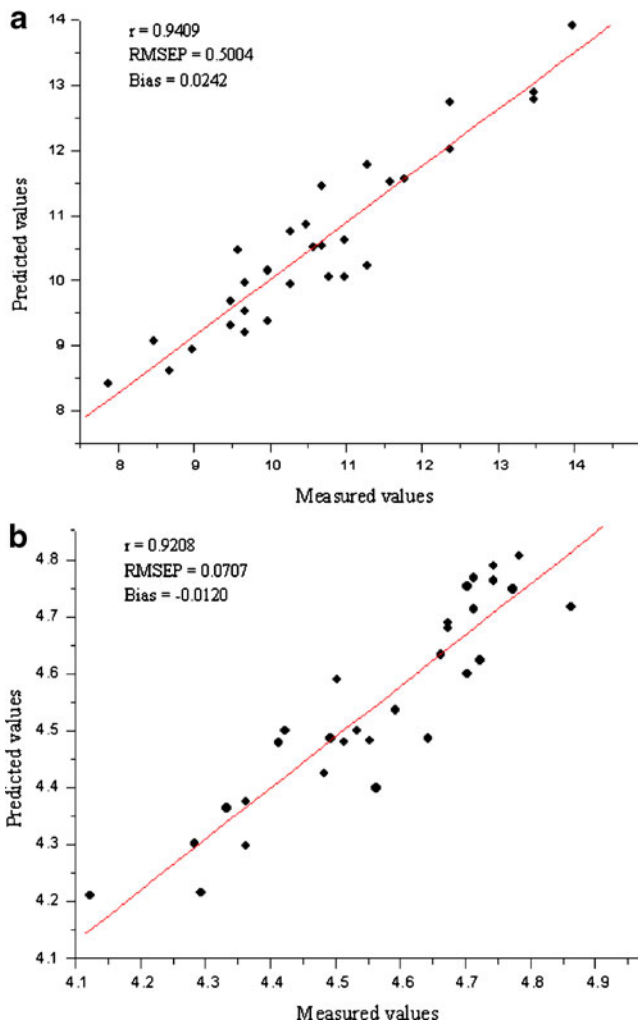*LV* latent variable, *EV* explained variance

Fig. 6 Vis/NIR prediction results of 30 unknown samples from the LV- LS-SVM model for SSC (**a**) and pH (**b**)

present study, and the effectiveness of SWs was also validated.

Analysis of the Results

Compared with the above PLS, LV-LS-SVM, and ICA-LS-SVM models, the nonlinear ICA-LS-SVM model turned out to be the best for prediction of SSC and pH value in peach, and the nonlinear LV-LS-SVM model was better than linear PLS model. The ICA-LS-SVM models had a better performance than the PLS models, and the reason might be that the LS-SVM models took the nonlinear information of the spectral data into consideration, and the nonlinear information had improved the prediction precision. The ICs from ICA were obtained by a high-order statistic which is a much stronger condition than orthogonality, so the SWs selected from ICs were more effective, and it could be very helpful for the

development of portable instrument or real-time monitoring of the peach internal quality.

Conclusions

Vis/NIR spectroscopy was successfully utilized for the determination of SSC and pH value of peach. A new combination of ICA-LS-SVM was proposed with comparison of nonlinear LV-LS-SVM and linear PLS models. ICA-LS-SVM model turned out to be the best for prediction of SSC and pH value in peach, and the nonlinear LV-LS-SVM model was better than linear PLS model. SWs selected from ICs were applied as the input data matrix of ICA-LS-SVM models, and a two-step grid search technique was used for the optimal RBF kernel parameters of (γ, σ^2). The SWs represented most of the features of the original spectra and could replace the whole wavelength region to predict the SSC and pH value of peach. The ICA-LS-SVM models were developed, and the best prediction perfor-
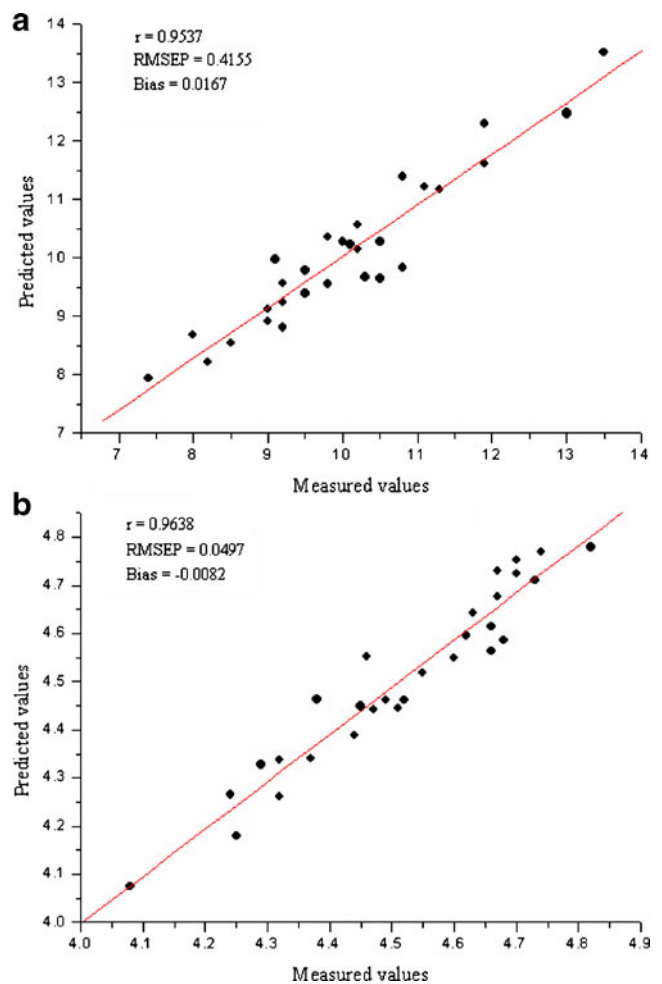

Fig. 7 Vis/NIR prediction results of 30 unknown samples from the ICA-LS-SVM model for SSC (**a**) and pH (**b**)

mance was achieved. The r_p, r_{cv}, RMSECV, RMSEP, and bias by ICA-LS-SVM were 0.9537, 0.9485, 0.4231, 0.4155, and 0.0167 for SSC and 0.9638, 0.9657, 0.0472, 0.0497, and −0.0082 for pH value, respectively. Furthermore, the SWs might be important for the development of portable instruments and online monitoring for commercial applications of SSC and pH value of peach. The overall results demonstrated ICA was powerful for variable selection, and the newly proposed ICA-LS-SVM could be applied as an alternative fast and accurate method for the determination of SSC and pH value of peach.

References

Bi, X., Li, T. H., & Wu, L. (2004). Application of independent component analysis to the IR spectra analysis. *Chemical Journal of Chinese Universities, 25*, 1023–1027.

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica Chimica Acta, 579*, 25–32.

Carlini, P., Massantini, R., & Mencarelli, F. (2000). Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. *Journal of Agricultural and Food Chemistry, 48*, 5236–5242.

Centner, V., Massart, D. L., de Noord, O. E., de Jong, S., Vandeginste, B. M., & Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry, 68* (21), 3851–3858.

Chen, J., & Wang, X. Z. (2001). A new approach to near-infrared spectral data analysis using independent component analysis. *Journal of Chemical Information and Computer Sciences, 41*, 992–1001.

Chen, Q. S., Zhao, J. W., Fang, C. H., & Wang, D. M. (2007). Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy, 66*, 568–574.

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems, 78*, 103–112.

Esbensen, K. H. (2002). *Multivariate data analysis in practice* (5th ed.). Oslo: CAMO Process As.

Gorry, P. A. (1990). General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method. *Analytical Chemistry, 62*, 570–573.

Guo, H., Liu, H. P., & Wang, L. (2006). Method for selecting parameters of least squares support vector machines and application. *Journal of System Simulation, 18*, 2033–2036. 2051.

He, Y. D. F. (1998). *The method for near infrared spectral anlysis*. Beijing: Light Industry of China.

Helland, I. S., Naes, T., & Isaksson, T. (1995). Related versions of multiple scatter correction. *Chemometrics and Intelligent Laboratory Systems, 29*, 233–241.

Hoyer, P. O., & Hyvarinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network-Computation in Neural Systems, 11*, 191–210.

Hyvarinen, A. (1999). Sparse code shrinkage: Denoising of non-gaussian data by maximum likelihood estimation. *Neural Computation, 11*, 1739–1768.

Hyvarinen, A., & Hoyer, P. O. (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation, 12*, 1705–1720.

Hyvarinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

Jouan-Rimbaud, D., Massart, D. L., Leardi, R., & de Noord, O. E. (1995). Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Analytical Chemistry, 67*, 4295–4301.

Kalivas, J. H., Roberts, N., & Sutter, J. M. (1989). Global optimization by simulated annealing with wavelength selection for ultraviolet–visible spectrophotometry. *Analytical Chemistry, 61*, 2024–2030.

Krier, C., Rossi, F., François, D., & Verleysen, M. (2008). A data-driven functional projection approach for the selection of feature ranges in spectra with ICA or cluster analysis. *Chemometrics and Intelligent Laboratory Systems, 91*, 43–53.

Lammertyn, J., Nicolay, B., Ooms, K., De Semedt, V., & De Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the ASABE, 41*, 1089–1094.

Lee, T. W. (1998). *Independent component analysis: Theory and application*. Boston: Kluwer.

Li, X. L., & He, Y. (2009). Evaluation of least squares support vector machine regression and other multivariate calibrations in determination of internal attributes of tea beverages. *Food and Bioprocess Technology*. doi:10.1007/s11947-008-0101-y.

Liu, F., & He, Y. (2009). Discrimination of producing areas of Auricularia auricula using visible/near infrared spectroscopy. *Food and Bioprocess Technology*, doi:10.1007/s11947-008-0174-7.

Liu, F., He, Y., Wang, L., & Pan, H. M. (2007). Feasibility of the use of visible and near infrared spectroscopy to assess soluble solids content and pH of rice wines. *Journal of Food Engineering, 83*, 430–435.

Liu, F., He, Y., & Wang, L. (2008). Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Analytica Chimica Acta, 615*, 10–17.

Liu, F., Jiang, Y. H., & He, Y. (2009). Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determine soluble solids content of beer. *Analytica Chimica Acta, 635*, 45–52.

Lu, R. (2001). Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Transactions of the ASABE, 44*, 1265–1271.

McGlone, V. A., Fraser, D. G., Jordan, R. B., & Kunnemeyer, R. (2003). Internal quality assessment of Mandarin fruit by vis/NIR spectroscopy. *Journal of Near Infrared Spectroscopy, 11*, 323–332.

Min, M., & Lee, W. S. (2005). Determination of significant wavelengths and prediction of nitrogen content for citrus. *Transactions of the ASABE, 48*, 455–461.

Murray, I., Aucott, L., & Pike, I. H. (2001). Use of discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fishmeal with meat and bone meal. *Journal of Near Infrared Spectroscopy, 9*, 297–311.

Pedro, A. M. K., & Ferreira, M. M. C. (2005). Nondestructive determination of solids and carotenoids in tomato products by

near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry, 77*, 2505–2511.

Rodriguez-Saona, L. E., Fry, F. S., McLaughlin, M. A., & Calvey, E. M. (2001). Rapid analysis of sugars in fruits by FT-NIR spectroscopy. *Carbohydrate Research, 336*, 63–74.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*, 1627–1639.

Slobodan, S., & Ozaki, Y. (2001). Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment. *Analytical Chemistry, 73*, 64–71.

Suykens, J. A. K., & Vanderwalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters, 9*, 293–300.

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

Workman, J. R. J. J., Mobley, P. R., Kowalski, B. R., & Bro, R. (1996). Review of chemometrics applied to spectroscopy: 1985–98, Part 1. *Applied Spectroscopy Reviews, 31*, 73–124.