

Full Length Research Paper

Quality determination of fruit vinegars using visible/near infrared spectroscopy and least squares-support vector machine

Fei Liu, Wenwen Kong and Yong He*

College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China.

Accepted 15 March, 2012

Soluble solids content (SSC) and pH are two important quality parameters of fruit vinegars. Visible and near infrared (VIS/NIR) spectroscopy was employed to determine SSC and pH of fruit vinegars based on partial least squares (PLS) analysis and least squares-support vector machine (LS-SVM). 300 vinegar samples were prepared. PLS models were developed with different preprocessing methods including no treatment, smoothing way of Savitzky-Golay, standard normal variate, 1st- and 2nd-derivatives. Simultaneously, certain selected latent variables (LVs) were used as LS-SVM inputs according to their explained variance. Finally, LS-SVM models with RBF kernel were developed compared with PLS models. The raw spectral data showed the best performance. The best LS-SVM models were achieved with 7 LVs for SSC and 6 LVs for pH, and LS-SVM outperformed all PLS models. The correlation coefficient (r), RMSEP and bias for validation set were 0.980, 0.667 and -0.043 for SSC, whereas 0.992, 0.040 and -0.006 for pH, respectively. The overall results indicated that VIS/NIR spectroscopy combined with LS-SVM could be used as a rapid alternative method for the prediction of SSC and pH of fruit vinegars. These results would be helpful for the process monitoring during the fermentation of fruit vinegars.

Key word: Visible and near infrared spectroscopy, fruit vinegar, soluble solids content, pH, partial least squares analysis, least squares-support vector machine.

INTRODUCTION

Fruit vinegars are mainly made of different kinds of fruits and their residuals by traditional fermentation and modern food processing techniques. Fruit vinegars have some physiological functions because fruit vinegars are rich in organic acid, amino acid, vitamins, mineral substances and so on (Chang et al., 2005). The soluble solids content

(SSC) and pH are two important internal quality parameters of fruit vinegars. SSC is the main reflection of organic sugars such as glucose, sucrose and fructose. The pH is a measure of its acidity such as acetic, lactic, citric, malic acids and so on. The SSC and pH influence the taste of fruit vinegars and reflect the stage of fermentation. In some countries, the quality evaluation of vinegars had been implemented by chemical analysis or sensory analysis such as pyrolysis-mass spectrometry (Anklam et al., 1998; Lipp et al., 1998), ion-selective electrodes (Lapa et al., 1995), gas chromatography

*Corresponding author. E-mail: yhe@zju.edu.cn. Tel/Fax: +86-571-88982143.

(Cocchi et al., 2004, 2006; Durante et al., 2006; Tesfaye et al., 2004), electronic nose (Anklam et al., 1998; Zhang et al., 2006), atomic absorption spectrum (Oliveira et al., 2007) and so on. Some new detection techniques were also applied for the identification and quantification of the main organic components of vinegars by high resolution ^1H NMR spectroscopy (Caligiani et al., 2007). However, these methods were time consuming, laborious and costly and not convenient enough for the fermentation process monitoring. Therefore, a new and fast detection method was quite necessary for the quality determination of fruit vinegars. Nowadays, near infrared spectroscopy is widely used as alternatives to wet chemistry procedures for determining concentrations of major classes of chemical compositions. Near infrared spectroscopy has unrivalled characteristics with combination of high speed, accuracy, simplicity and low cost. Hence, near infrared spectroscopic technique is widely applied for quantitative and qualitative analysis in industries such as agriculture, pharmaceuticals, food, textiles, cosmetics and polymer production industry (Yan et al., 2005; Liu et al., 2011). In the quality evaluation of vinegars, near infrared spectroscopy had been applied for the determination of aging of vinegar during storage (Casale et al., 2006), prediction of organic acids and other quality parameters of wine vinegar (Sáiz-Abajo et al., 2006), the determination of reducing sugars (Fu et al., 2005), total procyanidins (García-Parrilla et al., 1997), ethanol and acetic acid (Yano et al., 1997). Liu et al. (2008) applied visible and short-wave near infrared spectroscopy to evaluate the quality parameters of rice vinegar. Fan et al. (2011) studied the classification of fermented vinegar and blender vinegar using near infrared spectroscopy. However, there were few reports on the determination of soluble solids content (SSC) and pH of fruit vinegars using visible and near infrared (VIS/NIR) spectroscopy.

The objectives of this study were: 1) to investigate the feasibility of using visible and near infrared spectroscopy to determine the soluble solids content and pH of fruit vinegars; 2) to obtain the optimal partial least squares (PLS) models with comparison of different preprocessing methods including smoothing way of Savitzky-Golay, standard normal variate, 1st and 2nd-derivatives; 3) to obtain the optimal number of latent variables and develop the least squares-support vector machine (LS-SVM) regression models for the prediction of SSC and pH of fruit vinegars.

MATERIALS AND METHODS

Fruit vinegar samples

In this study, five varieties of fruit vinegars were obtained in local

market. They were made from different fruits including aloe, apple, lemon, peach and plum. All these fruit vinegars were fermented and produced by Taiwan Pai Chia Chen Brewery & Foods Co., Ltd. All the fruit vinegar samples were stored in the laboratory with a constant temperature of $25 \pm 1^\circ\text{C}$ for more than 48 h to equalize the temperature. 60 samples for each variety and a total of 300 samples were prepared for spectral analysis. 225 fruit vinegar samples (45 samples for each variety) were randomly selected for the calibration set, whereas the remaining 75 samples (15 for each variety) were applied for the validation set. No single sample was used in calibration and validation sets at the same time. Before the spectral scanning, the fruit vinegars were homogenized by putting the bottles upside down for several times to achieve the same concentration of fruit vinegar in the same bottle.

Spectral acquisition and reference methods for SSC and pH

In this study, a handheld FieldSpec Pro FR (325–1075 nm)/A110070 spectroradiometer with Trademarks of Analytical Spectral Devices, Inc. (Analytical Spectral Devices, Boulder, USA) was applied for the spectral scanning. The field-of-view (FOV) of the spectroradiometer is 25° . The light source consists of a Lowell pro-lam interior light source assemble/128930 with Lowell pro-lam 14.5 V bulb/128690 tungsten halogen bulb that could be used both in visible and near infrared region (325 to 1075 nm). The transmission mode was applied for this experiment. The light source was placed at a height of approximately 300 mm above the sample. The space between the light source and sample was depended on the energy of light source and what kind of sample was detected. The energy of light source could be adjusted according to the standard curve of spectroradiometer. Some adjustments of light source energy and space had been done before 300 mm was settled. The probe was under the sample, and the distance between the sample and probe was 50 mm and this value was settled by the spectroradiometer. Fruit vinegar sample was placed in a cuvette with a 2 mm light path length. The transmission spectra were measured from 325 to 1075 nm at 1.5 nm intervals with an average reading of 30 scans for each spectrum. For each sample, three transmission spectra were scanned and the average spectrum of these three transmission spectra was used as the spectra data of this sample. All spectral data were stored in a personal computer and processed using the RS³ software for Windows (Analytical Spectral Devices, Boulder, USA) designed with a graphical user interface.

The reference value of SSC was measured by an Abbebenchttop 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 to 95% with temperature correction. The pH was measured using a pH meter (Model: PHS-4CT, Shanghai Dapu Instrument Co. Ltd., Shanghai, China), with the accuracy of 0.001 pH units.

Preprocessing methods

In this study, calibration models with different preprocessing methods were compared to achieve the best prediction performance. Firstly, the transmission spectra were transformed into ASCII format by using the ASD ViewSpecPro software (Analytical Spectral Devices, Boulder, USA). Three transmission spectra for each sample were averaged into one spectrum and then the averaged spectrum was transformed into absorbance spectrum by $\log(1/T)$.

The preprocessing was implemented by "The Unscrambler V 9.6" (CAMO Process AS, OSLO, Norway). Secondly, the raw spectra and some preprocessed spectra were obtained for the development of calibration models. The preprocessing methods included smoothing way of Savitzky-Golay (SG) (Gorry, 1990), standard normal variate (SNV) (Barnes et al., 1989), 1st-derivative (D1) and 2nd-derivative (D2) (Chu et al., 2004). Thirdly, the first and last 75 nm was removed from the spectral data, and only the region of wavelengths (400 to 1000 nm) was used for calibrations and validations to avoid low signal-to-noise ratio.

Partial least squares analysis

Partial least squares (PLS) analysis is a widely utilized multi-analysis and regression method for the spectroscopy analysis. PLS analysis can be applied to develop a calibration model to progress the prediction of chemical components of fruit vinegars. PLS considers simultaneously the spectra data matrix X and the target chemical component matrix Y (Cen et al., 2006). PLS uses the chemical concentration information during the decomposition process, so the spectra containing higher constituent concentrations can be weighted more heavily than those with low concentrations. In the development of PLS model, full cross-validation was used to validate the quality and to prevent overfitting of calibration model. Hence, PLS models could be developed for the determination of SSC and pH of fruit vinegars. In this paper, PLS analysis was also used as a way to extract the latent variables (LVs) of the spectral data. The LVs could be used as new eigenvectors to present the important information of the original spectra, reduce the dimensionality and compress the original spectra data. The explained variance of LVs could explain the variance of the original spectra data to the chemical constituents. According to the explained variance, certain PLS factors could be selected as the inputs of least squares-support vector machine (LS-SVM) to develop the calibration models. Then, LS-SVM models could be applied for the quality assessment of fruit vinegars.

The predictive capability of models was evaluated by the following standards: correlation coefficient (r), root mean square error of calibration (RMSEC) or prediction (RMSEP), bias, slope and offset. The values of correlation coefficient and RMSEP were the main evaluation indices in this paper. The slope and bias should be taken into consideration for distinguishing systematic errors and studying the correlation between the reference and VIS/NIR models. Generally, a good model should have high correlation coefficients, low RMSEC, RMSEP and bias values, and the slope values should be closely to the value 1.

Least squares-support vector machine

Least squares-support vector machine (LS-SVM) is a state-of-the-art learning algorithm and has a good theoretical foundation in statistical learning method. LS-SVM is capable of dealing with linear and nonlinear multivariate analysis and resolving these problems in a relatively fast way (Vapnik, 1995; Suykens and Vandewalle, 1999; Suykens et al., 2002). Moreover, 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 (QP) problems to obtain the support vectors (SVs). SVM embodies the structural risk minimization (SRM) principle instead of traditional empirical risk minimization (ERM) principle to

avoid overfitting problems. The details of LS-SVM algorithm could be found in the literatures (Vapnik, 1995; Guo et al., 2006; Chen et al., 2007). The LS-SVM regression model can be expressed as:

$$y(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (1)$$

Where $K(x, x_i)$ is the kernel function, x_i is the input vector, α_i is Lagrange multipliers called support value, b is the bias term. All the calculations were performed using MATLAB 7.0 (The Math Works, Natick, USA). 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

Spectral features and statistics of SSC and pH

Figure 1a shows part of raw absorbance spectra of five varieties of fruit vinegars. Figure 1b to d shows the preprocessed absorbance spectra with combination of SG and SNV, combination of SG, SNV and D1, and combination of SG, SNV and D2, respectively. The trends of the spectral curves were quite similar and there were no obvious peaks or valleys in the visible and near infrared spectral region. Table 1 shows the statistics of SSC and pH of fruit vinegar samples. A relatively wide range of SSC and pH were covered in the calibration and validation sets due to different varieties of fruit vinegars. This wide range of SSC and pH was helpful for the development of calibration models.

PLS models and selection of latent variables

After the transform of the transmission spectra into absorbance spectral data, the PLS models with different preprocessing methods were developed for the determination of SSC and pH of fruit vinegars. Four kinds of preprocessing combinations were applied including no preprocessing, combination of SG with 3 segments and SNV, combination of SG, SNV and 1st-derivative, and combination of SG, SNV and 2nd-derivative. Twenty latent variables were calculated for each kind of preprocessing methods and the optimal numbers of latent variables were different by different preprocessing models. After the computation, different calibration models were developed for the determination of SSC and pH of fruit vinegars. The performance of the models was validated by 75 samples in validation set. The prediction results for calibration and validation sets by different PLS models are shown in Table 2. The results indicated that the PLS

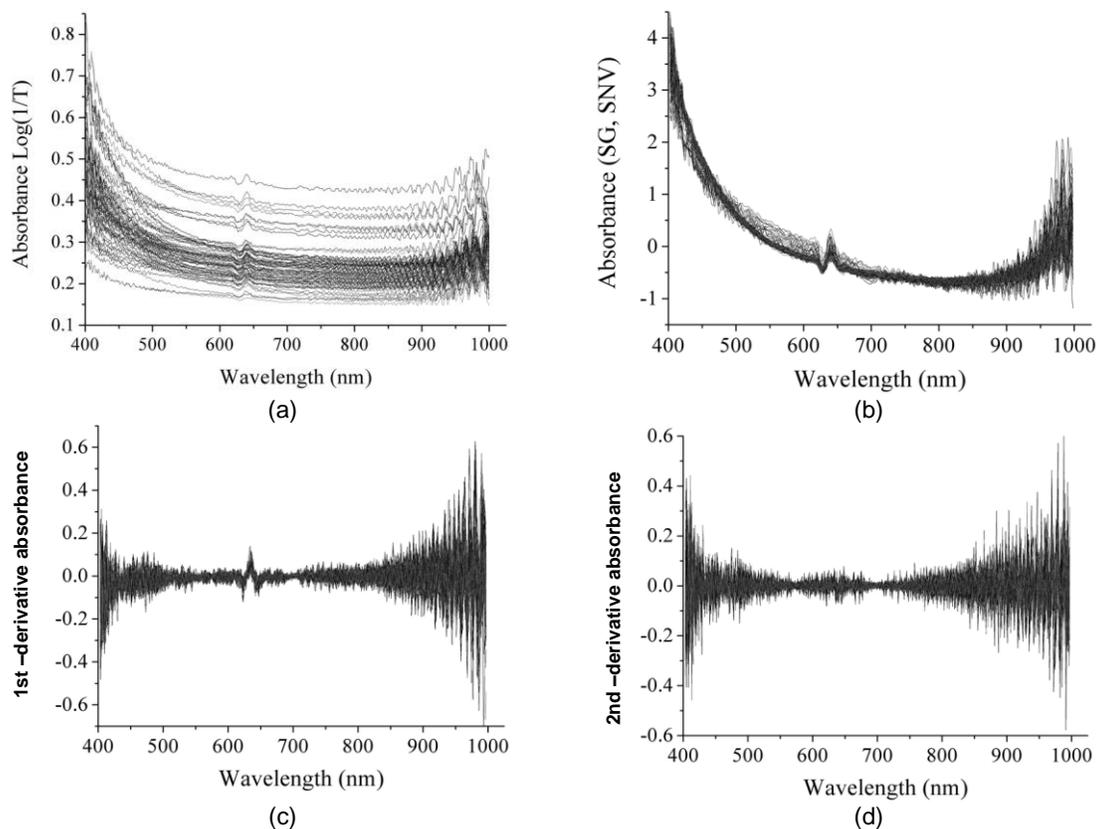


Figure 1. Part of raw absorbance spectra (a), preprocessed spectra with SG and SNV (b), SG, SNV and 1st-derivative (c) and SG, SNV and 2nd-derivative of fruit vinegars (d).

Table 1. The statistics of SSC and pH of fruit vinegars in calibration and validation sets.

Constituent	Data set	Sample no.	Range	Mean	Standard deviation
SSC (°Brix)	Calibration	225	40.0-49.6	44.0	3.277
	Validation	75	40.3-49.5	44.0	3.296
	All samples	300	40.0-49.6	44.0	3.276
pH (pH units)	Calibration	225	2.896-3.735	3.520	0.313
	Validation	75	2.898-3.733	3.520	0.313
	All samples	300	2.896-3.735	3.520	0.312

models with raw spectra achieved the optimal performance for both SSC and pH. The correlation coefficient (r), RMSEP and bias for validation set were 0.966, 0.849 and 0.044 for SSC, whereas 0.974, 0.073 and 0.004 for pH, respectively. The reason for the better performance of raw spectra than those of SG, SNV, 1st and 2nd-derivatives spectra was that the noise of the raw

spectra was enlarged by these preprocessing methods. This could be seen in Figure 1. Many noises are shown at the beginning and the end of the spectra in Figure 1c and d.

The preprocessed spectra enhanced no features of the raw spectra. This situation was consistent with the literature for the vintage year determination of rice wine

Table 2. Prediction results of SSC and pH by PLS models with different preprocessing methods.

Preprocessing	LVs ^f	r ^g	RMSEC/RMSEP ^h	Bias	Slope	Offset
Calibration-SSC (°Brix)						
None ^a	7	0.976	0.716	-3.4×10 ⁻⁸	0.952	2.109
SG ^b , SNV ^c	6	0.956	0.957	-1.0×10 ⁻⁷	0.914	3.769
SG, SNV, D1 ^d	6	0.968	0.817	-1.1×10 ⁻⁷	0.938	2.746
SG, SNV, D2 ^e	3	0.952	1.003	-2.0×10 ⁻⁷	0.906	4.143
Validation-SSC (°Brix)						
None	7	0.966	0.849	0.044	0.921	3.507
SG, SNV	6	0.948	1.043	-0.006	0.915	3.756
SG, SNV, D1	6	0.942	1.115	0.019	0.944	2.472
SG, SNV, D2	3	0.950	1.023	-0.010	0.875	5.489
Calibration-pH (pH units)						
None	6	0.978	0.065	3.5×10 ⁻⁸	0.956	0.155
SG, SNV	8	0.967	0.799	2.0×10 ⁻⁸	0.934	0.231
SG, SNV, D1	9	0.964	0.083	6.4×10 ⁻⁹	0.929	0.249
SG, SNV, D2	6	0.931	0.114	5.5×10 ⁻⁸	0.866	0.471
Validation-pH (pH units)						
None	6	0.974	0.073	0.004	0.898	0.362
SG, SNV	8	0.956	0.092	0.008	0.911	0.321
SG, SNV, D1	9	0.894	0.140	-0.006	0.826	0.607
SG, SNV, D2	6	0.851	0.167	0.004	0.814	0.658

^a None: no preprocessing; ^b SG: smoothing way of Savitzky-Golay; ^c SNV: standard normal variate; ^d D1: 1st-derivative; ^e D2: 2nd-derivative; ^f LVs: latent variables; ^g r: correlation coefficient; ^h RMSEC/RMSEP: root mean square error of calibration or prediction.

(Yu et al., 2007). The scatter plots of predicted versus reference values in validation set are shown in Figure 2a for SSC and Figure 2b for pH. The solid line was the fitting line corresponding to the ideal unity correlation between the predicted and reference values. In Figure 2, there were only four levels corresponding to the five different varieties of fruit vinegars. The reason could be explained as follows. The mean values of SSC were 46.1, 40.7, 43.1, 49.3 and 41.1 °Brix for aloe, apple, lemon, peach and plum vinegars, respectively. The SSC of apple and plum were quite similar and this small difference could not be distinguished in Figure 2a. Hence, the samples of apple and plum were mixed together and only four levels are shown in Figure 2a, but actually these four levels of SSC were corresponding to five different varieties of fruit vinegars. For the similar reason in Figure 2b, the mean values of pH were 3.727, 3.613, 2.902, 3.681 and 3.678 pH units for aloe, apple, lemon, peach and plum vinegars, respectively. The pH values of peach and plum were quite similar and the samples were mixed together in Figure 2b,

but actually these four levels of pH were corresponding to five different varieties of fruit vinegars. During the analysis of PLS, twenty LVs were calculated for SSC and pH of 300 samples and certain LVs would be selected as input data set of the least squares-support vector machine (LS-SVM). Although, the whole spectral wavelength region (400 to 1000 nm) could be applied as the inputs, the training time using LS-SVM increased with the square of the number of training samples and linearly with the number of variables (dimension of spectra) (Chauchard et al., 2004).

In order to reduce the computational time, certain selected latent variables were used as the inputs data according to their explained variance by PLS analysis. The LVs could explain most of the spectral variances and represent the main information of the raw spectra to the chemical constituents. The explained variance of certain LVs for SSC and pH are shown in Tables 3 and 4. It indicated that the first seven LVs could explain 93.4% of the total variance for SSC, the next LV only contributed

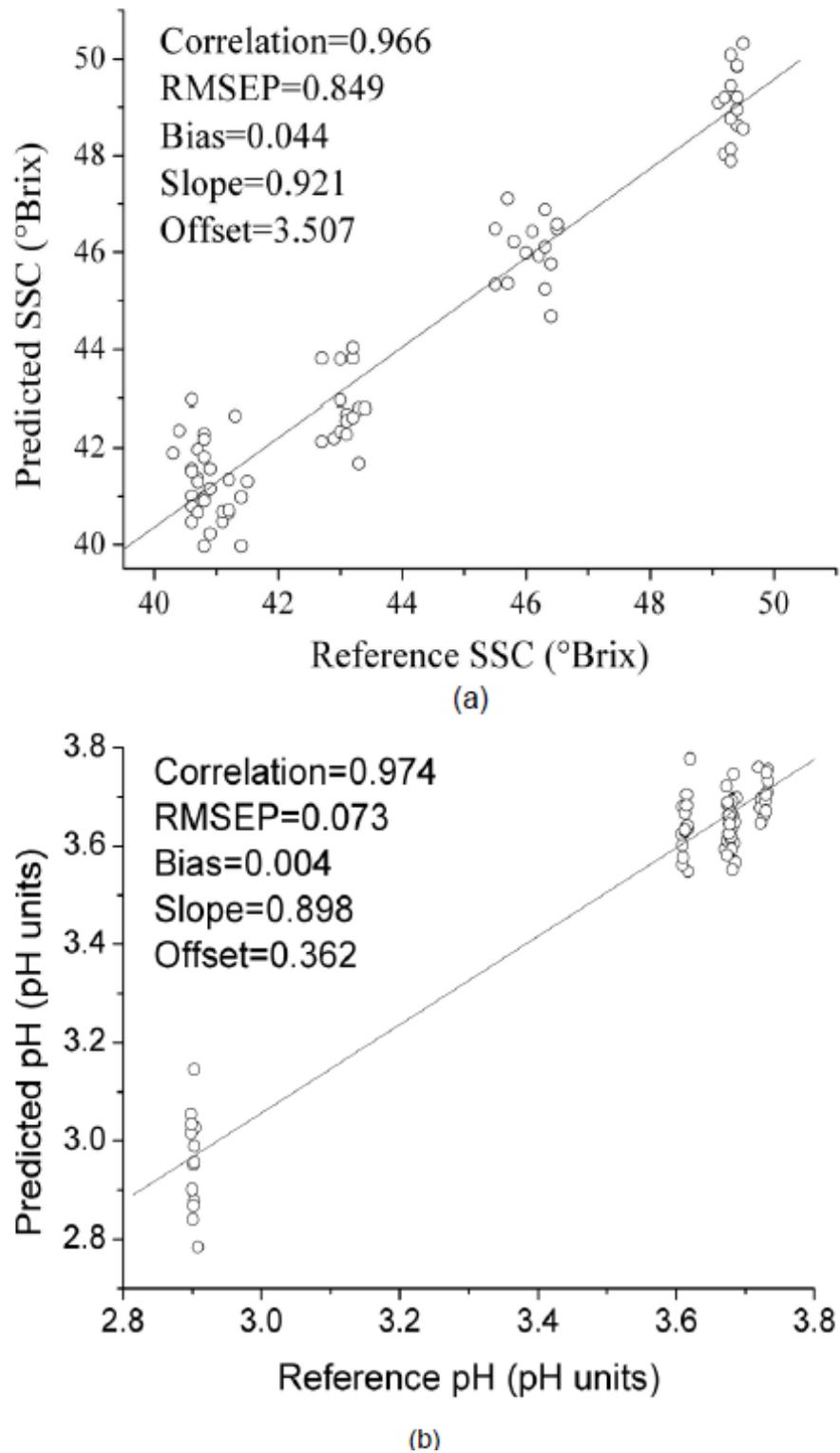


Figure 2. Predicted versus reference values of SSC (a) and pH (b) by PLS models with raw spectral data.

Table 3. The explained variance of certain latent variables by PLS analysis.

Parameter	SSC				pH			
	5	6	7	8	5	6	7	8
Latent variable								
Explained variance (%)	90.6	92.2	93.4	93.9	92.3	94.1	94.9	/

Table 4. Prediction results of SSC and pH by LS-SVM models with different latent variables.

LVs ^a	(γ , σ^2) ^b	r ^c	RMSEC/RMSEP ^d	Bias	Slope	Offset
Calibration-SSC (°Brix)						
6	(195.1, 4.2)	0.999	0.096	8.0×10^{-8}	0.997	0.121
7	(70.7, 6.6)	0.999	0.169	-3.8×10^{-9}	0.993	0.308
8	(130.1, 7.7)	0.999	0.111	-7.9×10^{-9}	0.996	0.171
Validation-SSC (°Brix)						
6	(195.1, 4.2)	0.977	0.708	-0.061	0.948	2.229
7	(70.7, 6.6)	0.980	0.667	-0.043	0.928	3.132
8	(130.1, 7.7)	0.973	0.772	-0.133	0.931	2.930
Calibration-pH (pH units)						
5	(376.6, 19.3)	0.995	0.030	-4.1×10^{-9}	0.986	0.050
6	(273.4, 10.9)	0.999	0.014	2.1×10^{-9}	0.996	0.015
7	(337.9, 21.3)	0.998	0.017	-9.9×10^{-11}	0.994	0.020
Validation-pH (pH units)						
5	(376.6, 19.3)	0.987	0.050	0.004	0.976	0.088
6	(273.4, 10.9)	0.992	0.040	-0.006	0.987	0.040
7	(337.9, 21.3)	0.990	0.043	-0.001	0.968	0.112

^a LVs: latent variables; ^b (γ , σ^2): the parameters of LS-SVM regression models; ^c r: correlation coefficient; ^d RMSEC/RMSEP: root mean square error of calibration or prediction.

0.5% of total variance which was less than 1% and contributed not so much as the first seven LVs. In order to obtain the optimal number of LVs for SSC, different LVs (6, 7 and 8 LVs) were used as the inputs of LS-SVM for comparison. For the same explanation, the seventh LV explained an additional 0.8% of the total variance, and LVs (5, 6 and 7 LVs) were applied as the inputs of LS-SVM for the pH prediction. Therefore, the optimal number of LVs for LS-SVM calibration could be achieved by the comparison of prediction performance and the computational time could be reduced by LS-SVM models.

LS-SVM models

Before the application of LS-SVM, three crucial problems were required to solve, including the optimal input data set, proper kernel function and the optimal LS-SVM

parameters. The optimal inputs had been settled by using the aforementioned LVs. The commonly used kernel functions were linear, polynomial, radial basis function (RBF) kernel and multi-layer perceptron (MLP). Currently, there was no systematic methodology for the selection of kernel function. However, compared with the aforementioned kernel functions, RBF kernel as a nonlinear function was a more compacted supported kernel and able to reduce the computational complexity of the training procedure. Simultaneously, RBF kernel could handle the nonlinear relationships between the spectra and target attributes and give a good performance under general smoothness assumptions (Wang et al., 2003). Thus, RBF kernel was recommended as the kernel function of LS-SVM in this paper. There were two significant parameters to be decided in the LS-SVM model. The regularization parameter γ determined the tradeoff between minimizing the training error and

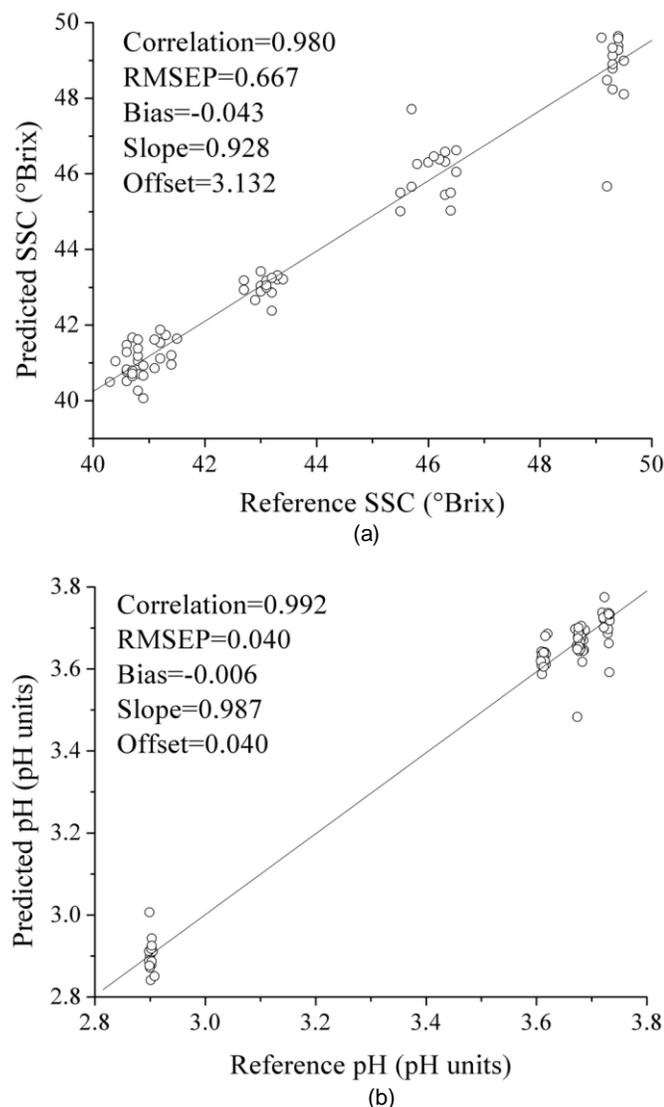


Figure 3. Predicted versus reference values of SSC (a) and pH (b) by LS-SVM models.

minimizing model complexity. The parameter σ^2 (σ^2) of RBF kernel function was the bandwidth and implicitly defines the nonlinear mapping from input space to some high dimensional feature space. In order to obtain the optimal combination of (γ, σ^2) , a two-step grid search technique was employed with leave-one-out cross validation to avoid overfitting problems. The ranges of γ and σ^2 within $(10^{-3}$ to 10^3) were set based on experience and previous researches (Guo et al., 2006; Chen et al., 2007; Belousov et al., 2002). Grid search tries values of each parameter across the specified search range using

geometric steps.

The first step 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. After the process of grid search, the optimal combination of (γ, σ^2) would be achieved for the LS-SVM models. The aforementioned LVs were used as the input matrix to develop LS-SVM model for the prediction of SSC and pH. The performance was validated by 75 samples in the validation set. Different optimal combinations of (γ, σ^2) were obtained for the prediction of SSC and pH. The prediction results are shown in Table 4. After comparison of prediction performance for both calibration and validation sets by aforementioned evaluation standards, the best LS-SVM models were achieved by 7 LVs for SSC and 6 LVs for pH. This was consistent with the numbers of LVs used in PLS models. The correlation coefficient (r), RMSEP and bias for validation set were 0.980, 0.667 and -0.043 for SSC, whereas 0.992, 0.040 and -0.006 for pH, respectively. The results indicated that LS-SVM models outperformed PLS models. The reason might be that there were some latent nonlinear useful information in the spectral data, and PLS only dealt with the linear relationships between the spectra data and chemical compositions. However, LS-SVM with RBF kernel could make use of the nonlinear information of spectra data to build the regression model. The results were in agreement with the literature described that LS-SVM was better than PLS models (Liu and He, 2007a; Thissen et al., 2004). The scatter plots of predicted versus reference values in validation set are shown in Figure 3a for SSC and Figure 3b for pH. Compared with Figure 2, the samples were distributed more closely to the fitting line.

The results indicated that the LS-SVM regression method was better than PLS regression method, and LS-SVM had a better capability for the prediction of SSC and pH of fruit vinegars. Simultaneously, the prediction results by LS-SVM were also better than those described similar studies. Liu et al. (2007b) determined the SSC ($r = 0.95$) and pH ($r = 0.94$) of rice wines using VIS/NIR spectroscopy and PLS analysis. Urbano-Cuadrado et al. (2004) predicted reducing sugars ($r = 0.844$), pH ($r = 0.905$) and other parameters in different types of wines by using NIR reflectance spectroscopy and PLS regression method. Shao et al. (2007) predicted the acidity of bayberry juice ($r = 0.945$) using VIS/NIR spectroscopy and back propagation neural network. Liu and He (2007a) applied VIS/NIR spectroscopy to determine the soluble solids content ($r = 0.959$) and pH ($r = 0.973$) of cola beverage. Therefore, the results indicated that VIS/NIR spectroscopy combined with LS-SVM could be utilized as an excellent prediction method for the determination of SSC and pH of fruit vinegars. These results would be

helpful for the process monitoring during the fermentation of fruit vinegars.

Conclusion

VIS/NIR spectroscopy combined with LS-SVM regression method was successfully utilized for the determination of SSC and pH of fruit vinegars. The PLS models were developed for the determination of SSC and pH with different preprocessing methods including no treatment of raw spectra, combination of SG and SNV, combination of SG, SNV and 1st-derivative, and combination of SG, SNV and 2nd-derivative. The optimal prediction performance was achieved by raw spectral data with 7 LVs for SSC and 6 LVs for pH. Simultaneously, certain LVs were selected as the inputs of LS-SVM models according to the values of explained variance. Different LS-SVM models were developed with RBF kernel function and a two-step grid search techniques. The prediction results indicated that LS-SVM models outperformed PLS models. The optimal number of LVs for LS-SVM was 7 for SSC and 6 for pH, and this was consistent with PLS models. The correlation coefficient, RMSEP and bias for validation set were 0.980, 0.667 and -0.043 for SSC, whereas 0.992, 0.040 and -0.006 for pH, respectively.

The overall results indicated that VIS/NIR spectroscopy combined with LS-SVM method could be applied as an alternative way for the determination of SSC and pH of fruit vinegars. The results might be useful for the process and online monitoring of vinegar fermentation.

ACKNOWLEDGEMENTS

This work was supported by 863 National High-Tech Research and Development Plan (2011AA100705), Natural Science Foundation of China (31071332) and Zhejiang Provincial Natural Science Foundation of China (Z3090295) and the Fundamental Research Funds for the Central Universities.

REFERENCES

- Anklam E, Lipp M, Radovic B, Chiavaro E, Palla G (1998). Characterization of Italian vinegar by pyrolysis-mass spectrometry and a sensor device ('electronic nose'). *Food Chem.*, 61: 243-248.
- Barnes R, Dhanoa M, Lister J (1989). Standard normal variable transformation and detrending of near infrared diffuse reflectance spectra. *Appl. Spectrosc.*, 43: 772-777.
- Belousov AI, Verzakov SA, von Frese J (2002). Applicational aspects of support vector machines. *J. Chemometr.*, 16: 482-489.
- Caligiani A, Acquotti D, Palla G, Bocchi V (2007). Identification and quantification of the main organic components of vinegars by high resolution ^1H NMR spectroscopy. *Anal. Chim. Acta*, 585: 110-119.
- Casale M, Sáiz-Abajo MJ, González-Sáiz JM, Pizarro C, Forina M (2006). Study of the aging and oxidation processes of vinegar samples from different origins during storage by near-infrared spectroscopy. *Anal. Chim. Acta*, 557: 360-366.
- Cen HY, He Y, Huang M (2006). Measurement of soluble solids contents and pH in orange juice using chemometrics and VIS/NIRS. *J. Agric. Food Chem.*, 54: 7437-7443.
- Chang RC, Lee HC, Ou ASM (2005). Investigation of the physicochemical properties of concentrated fruit vinegar. *J. Food Drug Analysis*, 13: 348-356.
- Chaucharad F, Cogdill R, Roussel S, Roger JM, Bellon-Maurel V (2004). Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics Intell. Lab. Syst.*, 71: 141-150.
- Chen QS, Zhao JW, Fang CH, Wang DM (2007). Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrosc. Acta Pt A*, 66: 568-574.
- Chu XL, Yuan HF, Lu WZ (2004). Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. *Prog. Chem.*, 16: 528-542.
- Cocchi M, Durante C, Foca G, Manzini D, Marchetti A, Ulrici A (2004). Application of a wavelet-based algorithm on HS-SPME/GC signals for the classification of balsamic vinegars. *Chemometrics Intell. Lab. Syst.*, 71: 129-140.
- Cocchi M, Durante C, Grandi M, Lambertini P, Manzini D, Marchetti A (2006). Simultaneous determination of sugars and organic acids in aged vinegars and chemometric data analysis. *Talanta*, 69: 1166-1175.
- Durante C, Cocchi M, Grandi M, Marchetti A, Bro R (2006). Application of N-PLS to gas chromatographic and sensory data of traditional balsamic vinegars of Modena. *Chemometrics Intell. Lab. Syst.*, 83:54-65.
- Fan W, Li HD, Shan Y, Lv HY, Zhang HX, Liang YZ (2011). Classification of vinegar samples based on near infrared spectroscopy combined with wavelength selection. *Anal. Methods*, 3(8): 1872-1876.
- Fu XG, Yan GZ, Chen B, Li HB (2005). Application of wavelet transforms to improve prediction precision of near infrared spectra. *J. Food Eng.*, 69: 461-466.
- García-Parrilla MC, Heredia FJ, Troncoso AM, González AG (1997). Spectrophotometric determination of total procyanidins in wine vinegars. *Talanta*, 44: 119-123.
- Gorry PA (1990). General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. *Anal. Chem.*, 62: 570-573.
- Guo H, Liu HP, Wang L (2006). Method for selecting parameters of least squares support vector machines and application. *J. Syst. Simul.*, 18: 2033-2036.
- Lapa RAS, Lima JLFC, Pérez-Olmos R, Pilar Ruiz M (1995). Simultaneous automatic potentiometric determination of acidity, chloride and fluoride in vinegar. *Food Control*, 6: 155-159.
- Lipp M, Radovic BS, Anklam F (1998). Characterisation of vinegar by pyrolysis-mass spectrometry. *Food Control*, 9: 349-355.
- Liu F, He Y, Wang L, Pan HM (2007b). Feasibility of the use of visible and near infrared spectroscopy to assess soluble solids content and pH of rice wines. *J. Food Eng.*, 83: 430-435.
- Liu F, He Y, Wang L (2008). Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. *Anal. Chim. Acta* 610: 196-204.
- Liu F, He Y (2007a). Use of visible and near infrared spectroscopy and least squares-support vector machine to determine soluble solids content and pH of cola beverage. *J. Agric. Food Chem.*, 55: 8883-8888.
- Liu F, Jin ZL, Naeem MS, Tian T, Zhang F, He Y, Fang H, Ye QF, Zhou

- WJ (2011). Applying near near-infrared spectroscopy and chemometrics to determine total amino acids in herbicide-stressed oilseed rape leaves. *Food Bioprocess Technol.*, 4(7): 1314-1321.
- Oliveira SR, Oliveira AP, Neto JAG (2007). Tungsten permanent chemical modifier with co-injection of $\text{Pd}(\text{NO}_3)_2 + \text{Mg}(\text{NO}_3)_2$ for direct determination of Pb in vinegar by graphite furnace atomic absorption spectrometry. *Food Chem.*, 105: 236-241.
- Sáiz-Abajo MJ, González-Sáiz JM, Pizarro C (2006). Prediction of organic acids and other quality parameters of wine vinegar by near-infrared spectroscopy. A feasibility study. *Food Chem.*, 99: 615-621.
- Shao YN, He Y, Mao JY (2007). Quantitative analysis of bayberry juice acidity based on visible and near-infrared spectroscopy. *Appl. Optics* 46: 6391-6396.
- Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J (2002). Least Squares Support Vector Machines. Singapore: World Scientific Publishing.
- Suykens JAK, Vandewalle J (1999). Least squares support vector machine classifiers. *Neural Process. Lett.*, 9: 293-300.
- Tesfaye W, Morales ML, Benítez B, García-Parrilla MC, Troncoso AM (2004). Evolution of wine vinegar composition during accelerated aging with oak chips. *Anal. Chim. Acta* 513: 239-245.
- Thissen U, Pepers M, Üstün B, Melssen WJ, Buydens LMC (2004). Comparing support vector machines to PLS for spectral regression applications. *Chemometrics Intell. Lab. Syst.* 73: 169-179.
- Urbano-Cuadrado M, De Castro MDL, Pérez-Juan PM, García-Olmo J, Gómez-Nieto MA (2004). Near infrared reflectance spectroscopy and multivariate analysis in enology determination or screening of fifteen parameters in different types of wines. *Anal. Chim. Acta*, 527: 81-88.
- Vapnik VN (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
- Wang WJ, Xu ZB, Lu WZ, Zhang XY (2003). Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing*, 55: 643-663.
- Yan YL, Zhao LL, Han DH, Yang SM (2005). *The Foundation and Application of Near Infrared Spectroscopy Analysis*. Beijing: China Light Industry Press.
- Yano T, Aimi T, Nakano Y, Tamai M (1997). Prediction of the concentrations of ethanol and acetic acid in the culture broth of a rice vinegar fermentation using near-infrared spectroscopy. *J. Ferment. Bioeng.*, 84: 461-465.
- Yu HY, Ying YB, Sun T, Niu XY, Pan XX (2007). Vintage year determination of bottled Chinese rice wine by VIS-NIR spectroscopy. *J. Food Sci.*, 72: 125-129.
- Zhang QY, Zhang SP, Xie CS, Zeng DW, Fan CQ, Li DF, Bai ZK (2006). Characterization of Chinese vinegars by electronic nose. *Sens. Actuators B*, 119: 538-546.