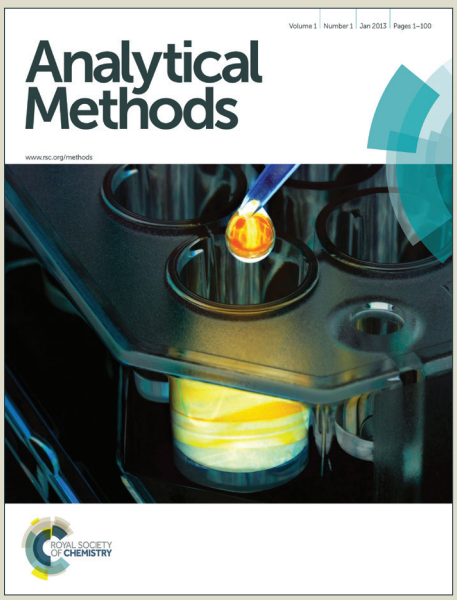
Analytical Methods

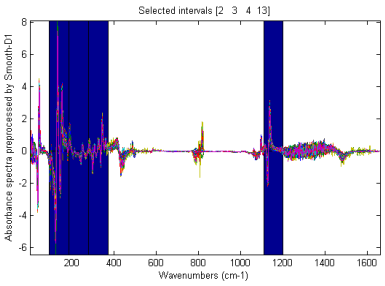
Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Graphical abstract



NIR and MIR combined with chemometrics tools were used to monitor time-related changes during Chinese rice wine fermentation.

Application of FT-NIR spectroscopy and FT-IR spectroscopy to Chinese rice wine for rapid determination of fermentation process parameters

Zhengzong Wu [a,b], Jie Long [a,b], Enbo Xu [a,b], Chunsen Wu [a,b], Fang Wang [a,b], Xueming Xu [a], Zhengyu Jin [a,b] *, Aiquan Jiao [a,b] **

[a] The State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, 1800 Lihu Road, Wuxi 214122, China

[b] Synergetic Innovation Center of Food Safety and Nutrition, Jiangnan University, 1800 Lihu Road, Wuxi 214122, China

*Corresponding authors

Zhengyu Jin

Tel/Fax: +86-510-85913299

E-mail address: jinlab2008@yahoo.com

Aiquan Jiao

Tel/Fax: +86-510-85320225

E-mail address: jiaoaq@jiangnan.edu.cn

Abstract

Effective fermentation monitoring is a growing need during the production of wine due to the rapid pace of change in the industry. Total reducing sugar, pH and amino acid nitrogen (AAN) are three most important process variables indicating the status of Chinese rice wine (CRW) fermentation process. In this study, the potential of near-infrared (NIR) spectroscopy and mid-infrared (MIR) spectroscopy as rapid tools to monitor the evolutions of these three chemical parameters involved in CRW fermentation process was investigated and compared. The results demonstrated that, compared with partial least-squares (PLS) model based on the full spectrum, model based on the spectra intervals selected by synergy interval partial least-squares (SiPLS) algorithm had higher prediction accuracy. In addition, nonlinear models outperformed linear models in predicting fermentation parameters. After systemically comparison and discussion, it was found that for either models developed based on NIR spectra or models developed based on MIR spectra, SiPLS-support vector machine (SiSVM) models obtained the best result with the highest prediction precision. The overall results indicated that it was feasible to monitor the fermentation process of Chinese rice wine using NIR and MIR spectroscopy.

Keywords: Chinese rice wine; Fermentation monitoring; Synergy interval partial least-squares (SiPLS); Support vector machine (SVM); Infrared spectroscopy (IR)

1. Introduction

Nowadays, bio-manufacturing technology is one of the most rapidly developing technologies in the world [1]. However, unavoidable variation in the compositional fluctuation often leads to unstable fermentation process. As a result, there has been a growing interest in developing advanced process analytical technique (PAT) for process control and quality assessment in the last few years. A large number of studies have been conducted on the application of PAT in many fields, including agriculture industry [2], food industry [3], and brewing industry [4]. Moreover, PAT has become a critical process for monitoring parameters in the pharmaceutical industry [5].

Chinese rice wine (CRW), also called yellow wine, is one of the most popular alcoholic beverages in China and other Asian countries with an annual consumption of more than 2 billion liters [6]. In order to assure the quality and consistency of the final CRW products, effective fermentation monitoring at every stage of the fermentation process is essentially important for CRW. In the wineries, CRW fermentation process is mainly monitored by the control of several critical chemical parameters including pH, total reducing sugar and amino acid nitrogen (AAN). Reducing sugars, which are mainly organic sugars, including glucose, fructose and maltose, can reflect the substrate concentration of CRW fermentation mash. The pH value, which is related with the changes of organic acids derived from various yeasts, bacteria and molds in the fermentation mash [7], represents the acidity of the fermentation mash. AAN is an important index reflecting the content of amino acid in CRW. These three parameters are conventionally measured using tedious wet chemical methods which have high accuracies. However, these methods are usually time-consuming and require complex procedures and hazardous chemicals.

As a result, the fermentation process could not be regulated timely, and stuck or sluggish fermentation occurs in CRW winery frequently. These negative effects can be avoided, if such problems are detected early. Thus, there has been a growing need in developing methods which are not only accurate, but also rapid for detecting real-time information of the fermentation process in the latest few years.

Infrared spectroscopy (IR) technique has been proposed as an interesting alternative to wet chemistry in the food industry during the last few years due to its rapidity, easiness and cost effectiveness [8, 9]. IR technique has also been widely used in CRW industry. However, most researches focus on the application of IR technique in the discrimination of CRW from different geographical origins or in the determination of compositional parameters in the final CRW product [10, 11], the application of IR spectroscopy for the control of CRW fermentation process is not yet reported. Although several researches have been conducted on the application of IR spectroscopy in monitoring alcoholic fermentation process, these studies mainly focus on linear regression models (PLS) based on the full spectrum [12, 13], little research exists on nonlinear regression model and calibration model based on the selected spectral regions. For wine samples, there are strong water absorption and a large number of unrelated or collinear spectral variables which could weaken the performances of final models [14]. Therefore, spectral region selection is needed to improve the performance of the final calibration models. With the help of synergy interval partial least-squares (SiPLS), we can focus on the important spectral regions and remove interferences from other regions [15, 16]. In addition, wine mash is a complicated system, in which chemical components are very complex; the correlations between the IR spectra and chemical constituents in the fermentation mash may also be very complex. Linear regression models may

not provide a satisfactory solution to the regression problem for wine samples. In these cases, methods for linear modelling of nonlinear surfaces are needed. Support vector machine (SVM), a promising method, is proposed by Vapnik, and usually applied to fulfill this goal [17].

In this context, the aims of this work were: (1) to investigate the potential of near infrared (NIR) spectroscopy and mid-infrared (MIR) spectroscopy in determining total reducing sugar content, pH and AAN content during the CRW fermentation process, (2) to compare the performances of models based on the efficient spectra intervals and those based on the full-spectrum, (3) to compare the performance of linear regression models and nonlinear regression models, (4) to compare the performances of models based on NIR spectra and models based on MIR spectra.

2. Materials and Methods

2.1 Fermentation trials and sampling

In order to develop robust and stable multivariate models, four most frequently used manufacture processes in CRW industry, namely, enzymatic extrusion pretreatment (EEP) technique, liquefaction pretreatment (LP) technique, two-step fermentation (TSF) technique and simultaneous saccharification and fermentation (SSF) technique, were used in this study. We proceeded with three trials for each kind of process. In total, twelve micro-fermentation trials were performed in this study. The four different manufacture processes were conducted as follows:

2.1.1 EEP fermentation

The rice extrudate (3,750 g, db) was collected in a 15 L jar with 7.5 L water, 3.75 g yeast (*Saccharomyces cerevisiae*, supplied by Zhejiang Nverhong Shaoxing Wine Co. Ltd) and 600-g

wheat Qu (Zhejiang Nv'erhong Shaoxing Wine Co. Ltd.). After agitation, the mash was incubated at 30 ℃ under stable condition for 4 days (main fermentation) and post-fermentation was carried out at 15 ℃ for 16 days.

2.1.2 LP fermentation

The crushed rice powder (3,750 g, db) was mixed with 7.5 L water and 0.328 g thermostable α-amylase (the enzyme had an optimum pH of 6-8, a density of 1.2 g/ml, and an activity of 120 KNU/g), then the mixture was incubated at 100 ℃ for 10 minutes and cooled to room temperature. Then 3.75 g yeast and 600 g wheat Qu were added into the jar, after that it was fermented in the same way as described in EEP fermentation.

2.1.3 TSF fermentation

The steam-cooked glutinous rice (3,750 g, db) was mixed with wheat Qu (600 g) and placed in the jar for 24h, leaving a hole in the middle of the fermentation in order to promote the saccharification of cooked rice, after that 7.5 L water and 3.75 g yeast were added into the jar and fermented in the same way as described above.

2.1.4 SSF fermentation

The steam-cooked glutinous rice (3,750 g, db) was collected in the same jar and 7.5 L water, 3.75 g yeast and 600 g wheat Qu were added into the jar. After that, it was fermented as aforementioned.

For each trial, 50 mL samples were collected after the starting of the fermentation process at thirteen times during CRW fermentation (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 and 20 days). A total of 156 samples were obtained at regular intervals between 0-20 days. Each sample (50 mL) was mixed with 50 μL of 4% sodium fluoride solution to be fixed. Then the samples were centrifuged

for 10 min at 5000×g to separate a clear solution from fermentation mash. The centrifuged supernatant was used for further analysis (spectroscopy collection and chemical analysis).

2.2 Reference measurement

The content of total reducing sugar of the supernatant wine was determined by the 3,5-dinitrosalicylic acid colorimetry (DNS) [7]. ANN content was measured by a titration method which in accordance with the official methods of analysis for Chinese rice wine (GB/T 13662-2008). Briefly, wine samples (10 mL) were mixed with 50 mL distilled water and titrated to end point of pH = 8.2 with 0.1 M NaOH; then, 10 mL formaldehyde solution (37-40%) was added and titrated to pH = 9.2 with 0.1 M NaOH; the volume of the second consumed NaOH was recorded to determine AAN content. Blank test was done with distilled water. The pH was measured using a pH meter (model PHS-4CT, Shanghai Dapu Instrument Co. Ltd., Shanghai, China). These determinations were carried out in triplicate and took the average. All chemicals were of analytical grade.

2.3 Spectra acquisition

2.3.1 NIR spectroscopy

The NIR spectra of the samples were collected in transmission mode using the Antaris II Fourier transform near-infrared spectrometer (Thermo Electron Co., USA), which was equipped with an interferometer, a wide band light source (quartz tungsten halogen, 50W), and an InGaAs detector. The samples were measured in a quartz cuvette with a 1-mm optical path length. Water was used to clean the cuvette to avoid carry over between samples and was dried with the help of a soft tissue paper. The spectral range was 10000-4000 cm^{-1} and the data were measured in the interval of 3.856 cm^{-1} and thus each spectrum has 1557 data points. Sixteen scans were averaged

for each sample spectrum.

2.3.2 MIR spectroscopy

The MIR spectra of the samples were obtained using a FT-IR spectrometer (Nicolet iS10, Thermo Electron Corp., Madison, WI, USA), equipped with an interferometer, a KBr beam splitter and a deuterated triglycine sulphate (DTGS) detector. The samples were scanned in transmission mode. Measurement was carried out using a demountable liquid cell (Pike Technologies, Madison, WI, USA) equipped with two CaF_2 windows. A polyethylene terephtalate spacer provides a 0.025 mm optical path-length. Special attention was taken to avoid air bubbles during the scanning of sample spectra. Water was used to clean the cuvette to avoid carry over between samples and was dried with the help of a soft tissue paper. Background was collected using deionised water as the reference and was taken once in every three sample scans. Spectra were recorded in the range of 4000-800 cm^{-1} with a spectral step of 1.928 cm^{-1} as an average of 16 scans. Each spectrum has 1660 data points.

As the spectrophotometer was sensitive to the change of outer environment condition (temperature and humidity), both the NIR and MIR spectra were collected at controlled temperature (25 ℃) and humidity (60%). Spectra were recorded in triplicate for each sample and the mean spectrum was used in the next analysis.

2.4 Data analysis

Raw IR spectra often contained background information and noises. In order to use the spectral data ‘‘as is’’, in this study, raw IR spectra were first pretreated by four different pre-processing methods, namely, standard normal variate (SNV), multiplicative scattering correction (MSC), first derivative (D1) and second derivative (D2). For MSC, full MSC function

was adopted in this study. For D1 and D2, smoothing points from 3 to 13 and the polynomial orders of 0, 1 and 2 were attempted. The optimal pre-processing method was achieved in accordance with the lowest root mean square error of cross-validation (RMSECV) value based on PLS models. After comparison, the optimal pre-processing method was D1 with seven smoothing points and two degrees of polynomial (Smooth-D1) for MIR spectroscopy, whereas the optimal pre-processing method for NIR was SNV. MIR spectra preprocessed by Smooth-D1 and NIR spectra pretreated by SNV are presented in Fig. 1. They are used for further analysis in the experiment. It is worth mentioning that the zones from 4000 to 4316 cm^{-1} and from 4898 to 5296 cm^{-1} in NIR range were not used in this study owing to the high irreproducibility because of the high absorbance values (above 1.5).

Principal component analysis (PCA) was performed to make a descriptive analysis of data. In addition, partial least squares discriminant analysis (PLS-DA) between the fermentation stages and IR spectra (NIR and MIR) were developed using full-cross validation.

Calibration models were developed with PLS and SVM regression algorithms by relating IR spectra to the reference measurements. Correlation coefficient of calibration set (R^2 (cal)) and root mean square error of cross-validation (RMSECV) were used to evaluate the model fit to the data in the calibration set. The prediction accuracy of the calibration model was tested by the correlation coefficient of prediction (R^2 (pre)) and root mean square error of prediction (RMSEP). The detailed calculations of these indices can be referenced to Ozturk et al. [18]. Additionally, residual predictive deviation (RPD), which is defined as the ratio of standard deviation (SD) of the prediction set to RMSEP [19, 20], was also used in this study to standardize the predictive accuracy. Generally, if an RPD value is greater than 2.43, the model is usable with caution for

most applications. An RPD value between 1.71 and 2.42 is presented to indicate that this model can be used for sample screening. If the value of RPD is lower than 1.70, the model is considered insufficient for prediction of screening purposes [21]. The higher the RPD value, the greater the ability of the model is to predict the chemical compositions.

In order to remove collinear and irrelevant variables in full spectrum regions, SiPLS algorithm was used to select the most important subintervals in this research. The basic principle of this algorithm is as follows: first, the full data interval is subdivided into a number of smaller equidistant subintervals; second, PLS models for all possible combinations of two, three or four spectral intervals were developed with adequate number of latent variables; finally, the RMSECV is calculated for each PLS model based on different combinations of subintervals. The combination of intervals with the lowest RMSECV is chosen to establish the optimal SiPLS model [16].

For SVM model, radial basis function (RBF), a nonlinear function, was selected as kernel function in this study. To enhance the performance of SVM model, two regularization parameters γ and σ^2 were optimized by a two-step grid search technique within the region of (2^{-8}-2^{8}). For each combination of γ and σ^2 parameters, RMSECV was calculated and the optimum parameters that produced the lowest RMSECV were selected.

2.5 Software

All data processing and analysis were implemented in Matlab R2010a (MathWorks, Natick, USA) under Windows XP. Result Software (v. 8.0, Thermo Electron Corp., Madison, WI, USA) was used for raw NIR spectral data acquisition. The OMNIC software (v. 8.0, Thermo Electron Corp., Madison, WI, USA) was used for raw MIR spectral acquisition.

3. Results and Discussion

3.1 Chemical analysis

Fig. 2 shows the temporal variation of the fermentation parameters (total sugar, pH and AAN) in four fermentation trials. As shown in Fig. 2, the fermentation kinetic curves belonging to four different manufacture processes showed similar profiles, which accorded with the results of other authors [7]. During the fermentation of rice wine, carbohydrates were gradually transformed to ethanol and carbon dioxide by the yeast [7]. At the beginning of fermentation, total reducing sugar content was approximately 100 g/L. Then it was sharply decreased in the first 4 days (main fermentation), and slowly decreased in post-fermentation stage. Finally, the content of the residue reducing sugar kept constantly at about 5 g/L (Fig. 2a). It was also reported by other authors [22]. As fermentation process progressed, various acid metabolites, which mainly come from yeasts, bacteria and molds existed in the fermentation mash, were produced [7]. Consequently, a downward trend was observed for pH. As shown in Fig. 2b, pH dropped sharply in the first 3 days and finally remained a constant value of 4.0. For AAN, a steadily rising trend was observed throughout the whole fermentation process which was attributable to the combination effect of the hydrolysis of protein in the rice and the autolysis of yeast cells [23].

The detailed descriptive statistics for the calibration and prediction samples are summarized in Table 1. The CRW samples analyzed in this study showed a relatively wide range, which might due to the continuous changes of chemical components during CRW fermentation and different manufacture techniques applied in this study. The wide-range variation was helpful to develop reliable multivariate models. In addition, high values of the coefficients of variation (CV) confirmed the great variability of the samples used in this research, which contributed to the

applicability of the model established based on these data.

In order to evaluate the relationships among the fermentation parameters analyzed in this study, a correlation analysis was conducted. The correlation matrix was shown in Table 2. A highly significant correlation between ANN and the content of total reducing sugar was found (-0.783). The latter variable was also found to be correlated with pH, but with lower values (0.422). However, there was no significant correlation between pH and AAN (-0.171).

3.2 Spectra analysis

For the sake of clarity, only NIR spectra and MIR spectra belonging to one trial from 0 day to 20 days were presented in Fig. 3. For NIR spectra (Fig. 3a), two strongly absorption bands which were observed at around 6900 cm^{-1} and 5100 cm^{-1} dominated the spectra. These two peaks were related to the first O-H overtone and the combination of stretch and deformation of the O-H group in water, respectively [24]. The absorption band at 4413 cm^{-1} was assigned to C–H combinations and O–H stretch overtones [8]. For MIR spectra (Fig. 3b), the dominating absorption peaks which were observed at around 1500-1710 cm^{-1} and 3005-3655 cm^{-1}, also belonged to water [25], as it was the most abundant components in the fermentation mash. Bands between 1200 and 1500 cm^{-1} were associated to deformations of -CH_2, C-C-H and H-C-O from organic acids [26], alcohols [27] and proteins [28]; whereas peaks between 950 and 1200 cm^{-1} were assigned to the absorption of C-C and C-O stretching [29, 30]. Changes of NIR spectra over time were observed at around 5590 cm^{-1}, associated with various carbohydrates such as sucrose, fructose, and glucose [24], and at 4338 cm^{-1}, explained by the combination of C-H stretching and deformation of C-H from the -CH_2 group of ethanol [31]. While for MIR spectra, the main variations were observed around 1050-1150 cm^{-1}, corresponded to glucose and fructose [32], and at 1042 cm^{-1}, which was

mainly generated by the contribution of C-O stretch from ethanol [32]. In addition, one region related with the time course of fermentation in the MIR spectra was also observed between 2900 and 3000 cm^{-1}, which was attributable to C–H stretch of CH_3 and CH_2 belonging to ethanol [20]. These variations observed in NIR and MIR spectra were marked with arrows in Fig. 3a and b, respectively.

3.3 Principal component analysis

In this study, the samples belonging to different fermentation stages were divided into four classes: 'step 1' (0-3 days), 'step 2' (4-6 days), 'step 3' (7-9 days) and 'step 4' (10-20 days). Fig. 4 showed the scatter plots (PC1-PC2) obtained by applying PCA to NIR and MIR spectra when considering the fermentation stage. Although several samples overlapped with each other, the distinction among samples from different fermentation stages was clear. In order to examine the potential of NIR and MIR to predict the time course (fermentation stages) of CRW fermentation, PLS-DA was performed in this study. The cut-off value was set as 0.5 that was also used in previous studies [33]. The PLS-DA results of the correct classification rates for the calibration set and the prediction set were shown in Table 3. As could be seen, for both NIR and MIR, good performances were obtained. MIR performed slightly better than NIR. The minimum correct classification rates of the calibration set and the prediction set were of 88.9% and 83.3% respectively for NIR and 92.6% and 91.7% respectively for MIR, indicating that it was possible to classify samples belonging to a particularly fermentation stage using NIR or MIR spectroscopy.

3.4 Multivariate analysis

All of the samples were divided into two subsets: the calibration set and the prediction set.

To avoid bias in subset division, the division was made as follows: first, all the spectra data were sorted according to concentration of the specific corresponding chemical component, then three data of every four samples were allocated to the calibration set, and the remaining one was assigned as the prediction data. Finally, the calibration set contained 117 samples and the prediction set contained 39 samples. The mean, standard deviation, range and coefficient of variation for the fermentation parameters measured in collected samples used to develop multivariate models were shown in Table 1. It was observed that the ranges of y-values (total sugar content, pH value and AAN content) in the calibration set covered the entire range in the prediction set, indicating that samples were appropriately distributed in the calibration set and prediction set.

In this study, four different regression models, namely PLS, SiPLS, SVM, SiPLS-support vector machine (SiSVM), were constructed and their results were systemically compared and discussed. These models were established based on NIR spectra preprocessed by SNV and MIR spectra pretreated by Smooth-D1. A summary of the results of four different calibration models developed for each fermentation parameter in NIR and MIR ranges were shown in Table 4. Table 5 showed the specific spectral variables used in different calibration models developed in this study.

3.4.1 Results of PLS model

As shown in Table 4, both for NIR and MIR spectra, the PLS models established based on full-spectral region can be used for quantitative determination of total sugar (based on NIR spectra, R^2 (pre)= 0.8725, RPD= 6.94; base on MIR spectra, R^2 (pre)= 0.9188, RPD= 9.10) and pH (based on NIR spectra, R^2 (pre)= 0.9063, RPD= 4.61; base on MIR spectra, R^2 (pre)= 0.8786,

RPD= 3.30) in fermentation mashes with high accuracy (R^2 > 0.85, RPD > 3). However, the performances for AAN were slightly worse (based on NIR spectra, R^2 = 0.8694, RPD = 1.87; based on MIR spectra, R^2 = 0.8712, RPD = 2.65). It is worth mentioning that with NIR and MIR spectroscopy, the highest RMSEP (%) were obtained for AAN (equal to 22.7 and 16.03, respectively), which might due to the relatively low concentration of AAN in the CRW fermentation mash. This result was in agreement with the research of Rudnitskaya et al., who found that a higher average error of prediction was usually obtained with a lower concentration of organic acids [34]. In addition, a relative high RMSEP (%) were also observed for total sugar with NIR (14.64) and MIR (11.17) spectroscopy, which accorded with the results of other authors [11]. This fact could be due to the wide range of total sugar (Table 4). The RMSEP values are related to the range of reference values, a wide range in chemical components usually results in high RMSEP values, and vice versa [11].

3.4.2 Results of SiPLS model

SiPLS algorithm, the expansion of interval partial least squares (iPLS) algorithm, was implemented to select the most important subintervals to effectively improve the performance of final model in this study. It is proposed by Nørgaard and is an all-possible-subinterval-combination procedure tests based on all possible PLS of all subsets of subintervals [35].

In this work, the number of intervals and PLS factors were optimized by full cross-validation. The whole spectrum region (4320-4894 cm^{-1} and 5300-10000 cm^{-1} for NIR spectra and 800-4000 cm^{-1} for MIR spectra) was divided into 11-25 intervals, and then SiPLS regression models were developed based on different number of intervals divided. The results of

the SiPLS models of total sugar, pH and AAN content in CRW fermentation mashes with the optimal combination of subintervals under different number of subintervals divided in this study was shown in Table 6 and the optimal models were prominent with the bold. Fig. 5 showed the optimal combinations of subintervals selected by SiPLS based on NIR spectra and MIR spectra for the prediction of fermentation parameters.

As could be seen in Table 6 and Fig. 5, among SiPLS models developed based on NIR spectra, for total sugar, the optimal SiPLS model was achieved with the combination of subintervals of [2 3 5 7] from 20 subintervals according to the lowest RMSECV (2.51); for pH, the optimal SiPLS model was achieved with the combination of subintervals of [1 3 5 6] from 20 subintervals; for AAN, the optimal SiPLS model was achieved with the combination of subintervals of [1 3 7 13] from 23 subintervals. Among SiPLS models developed base on MIR spectra, for total sugar, the optimal SiPLS model was achieved with the combination of subintervals of [2 3 4 13] from 20 subintervals; for pH, the optimal SiPLS model was achieved with the combination of subintervals of [4 5 7 8] from 20 subintervals; for AAN, the optimal SiPLS model was achieved with the combination of subintervals of [3 4 5 6] from 23 subintervals.

It was observed that the performances of models based on the efficient spectra intervals selected by SiPLS were much better than those based on the full spectrum (Table 4). Both for SiPLS models developed based on MIR spectra and models based on NIR spectra, the RPD values obtained for the three fermentation parameters analyzed in this study were all higher than 3, indicating that SiPLS models could be used for analytical purpose with high prediction precisions. SiPLS algorithm can provide an overall picture of the relevant information in

different spectral subintervals, thereby focusing on the most important subintervals and eliminating irrelevant information in the spectra. In fact, in NIR spectra region, for total sugar, the 4585-4848 cm^{-1} region was due to O-H stretch overtone and C-H combination bands [36], the 4852-5515 cm^{-1} region was related to the tone combinations, the first overtone of O-H bonds and the second overtone of C=O stretching [13, 14], the 5785-6047 cm^{-1} and 6317-6579 cm^{-1} regions originated from the first overtone of the C–H stretching vibrations [37]; for pH, the 4319-4582 cm^{-1} region was associated with the combination bands of C-H, C=O, O-H and N-H stretch [12, 38], the 6051-6313 cm^{-1} region was due to the first overtone (intermolecular H-bond) from O-H stretching [37]; for AAN, the 4752-5411 cm^{-1} region was related to the combinations of the first overtone of C=O stretch with fundamental N-H in plane bend and the first overtone of N-H in plane bend with fundamental C=O stretch [38], the 6109-6336 cm^{-1} region could be ascribed to the first overtone of N-H bond, the 7497-7725 cm^{-1} region could be attributed to the combinations of the first overtone N-H stretch with fundamental N-H in plane bend and C-N stretch with N-H in plane bend vibrations [38]. In MIR spectra region, for total sugar, the 979-1515 cm^{-1} region was assigned to deformations of CH_2 and H-C-O and stretching of C-C and C-O [9], the 2937-3112 cm^{-1} region was attributed to C-H stretch [11]; for pH, the 1606-1870 cm^{-1} region related with N=O and C=O stretching and N-H bending [39]; for AAN, the 1280-1598 cm^{-1} region was owing to N=O and C=O stretching, bending of N-H and stretching of N=O, N-H and C=O belonging to acids and proteins [26, 39]. Therefore, the optimal spectra intervals selected by SiPLS contained a lot of information related to the corresponding fermentation parameters. As a result, for total sugar and ANN, the models constructed on the optimal spectra intervals performed better than those based on full spectrum (Table 4). However, for pH, the performances of full spectrum PLS models and

SiPLS models were nearly the same (R^2 (pre) of full spectrum PLS models and SiPLS models based on NIR and MIR spectra were 0.9063 and 0.9097, and 0.8786 and 0.8794, respectively). This might due to that while the most important spectra intervals were selected by SiPLS algorithm, some important variables related to pH were removed at the same time. Overall, SiPLS model showed its robustness with superior interpretability in comparison with PLS model based on full-spectral region. In addition, 1370 and 1660 variables existed in the full NIR and MIR spectrum respectively, whereas there were only about 300 variables in the combinations of spectral subintervals selected by SiPLS. The number of wavelength variables decreased by at least 77.7% (for total sugar in MIR spectra range), which simplified the regression models and considerably saved the computation time.

3.4.3 Comparison between the performances of linear regression models and nonlinear regression models

The statistics shown in Table 4 indicated that compared with PLS models, SVM models achieved better performances in predicting the fermentation parameters. Among the three parameters, the predictive precisions of AAN improved the most. For regression models developed based on the NIR spectra, R^2 (pre), RMSEP (%) and RPD of AAN increased from 0.8694, 22.70 and 1.87 in PLS model to 0.8812, 13.97 and 3.04 in SVM model, respectively. For models based on MIR spectra, R^2 (pre), RMSEP (%) and RPD of AAN increased from 0.8712, 16.03 and 2.65 in PLS model to 0.9125, 8.91 and 4.76 in SVM model, respectively. CRW fermentation mash was a complex system, in which a large number of chemical components existed. These chemical compounds included many chemical bonds in fundamental groups including C–H, C=O, N–H, O–H, C-O, etc.; as a result, overtones and combinations of

fundamental vibrations of different chemical bonds occurred in NIR and MIR spectra. Therefore, some latent nonlinear relationship existed between NIR (MIR) spectra and the fermentation parameters.

3.4.4 Comparison between the performances of models developed based on NIR spectra and MIR spectra

For total reducing sugar and AAN, MIR got the better performance with high prediction accuracy, which might due to its smaller light diffusion, higher sensitivity to chemical composition and more specific absorption bands compared with NIR spectra. However, for pH, NIR spectroscopy (in PLS model, R^2 (pre) = 0.9063, RMSEP (%) = 3.33 and RPD = 4.61) outperformed MIR spectroscopy (in PLS model, R^2 (pre) = 0.8786, RMSEP (%) = 4.66 and RPD = 3.30). This could be due to the fact that compared with MIR spectroscopy, light in the NIR region can better penetrate into the chemical matters. These different features of NIR and MIR spectroscopy suggest that one of them can be more adequate to each application.

Among all the four different kinds of models (PLS, SiPLS, SVM, SiSVM), SiSVM model got the best performance with the highest R^2 (pre) and RPD and the lowest RMSEP. For models based on NIR spectra, compared with PLS models using all wavelengths of NIR spectra preprocessed by D1-smooth, R^2 (pre) of SiSVM models increased by 5.60%, 2.64% and 4.89% respectively for total sugar, pH and AAN. For models based on MIR spectra, R^2 (pre) of SiSVM models increased by 4.42%, 4.66% and 8.25% respectively for total sugar, pH and AAN. The scatter plots of reference measurements and NIR and MIR predictions for total sugar, pH and AAN obtained from SiSVM models and PLS models based on full-spectral region were shown in Fig. 6 (NIR) and Fig. 7 (MIR), respectively. The green diagonal represents ideal results, the

closer the points are to this, the better is the model. As could be seen in the plots, compared with PLS models based on the full-spectral region, SVM models (SiSVM models) based on spectral subintervals selected by SiPLS algorithm had a better fitting effect between the predicted and reference data (all of the data points clustered closely to the diagonal lines).

4. Conclusions

The applicability of spectroscopic techniques in NIR and MIR ranges combined with chemometrics was investigated for the prediction of the time-related changes of total sugar, pH and AAN during the CRW fermentation process. The results showed that MIR spectroscopy was superior in determination of total reducing sugar and AAN, while the best results for pH were obtained using NIR spectral range. In developing calibration models, SiPLS showed its incomparable superiority in contrast with classical PLS calibration method. Furthermore, SVM models performed significantly better than PLS models, indicating the correlations between the spectra and the chemical components were inclined to be nonlinear rather than linear. From all the results presented, it can be concluded that the NIR and MIR spectroscopy together with SiPLS and SVM could be utilized as rapid PAT techniques to carry out the monitoring of time-related changes of the main chemical components during CRW winemaking.

Acknowledgements

We are grateful to Chen Chen for technical assistance in operating the FT-IR spectrometer. This study was supported by National 'Twelfth Five-Year' Plan for Science & Technology Support of China (Nos. 2012BAD37B02 and 2012BAD37B06).

References

1. X. Xiao, Y. Hou, J. Du, Y. Liu, Y. Liu, L. Dong, Q. Liang, Y. Wang, G. Bai and G. Luo, *J. Agric. Food Chem.*, 2012, 60, 7830-7835.

2. H. Jiang, Q. Chen and G. Liu, *Process Biochem.*, 2014, 49, 583-588.
3. R. S. Uysal, E. A. Soykut, I. H. Boyaci and A. Topcu, *Food Chem.*, 2013, 141, 4333-4343.
4. Q. Wang, Z. Li, Z. Ma and L. Liang, *Sensors Actuators B: Chem.*, 2014, 202, 426-432.
5. D. Cozzolino, M. Parker, R. G. Dambergs, M. Herderich and M. Gishen, *Biotechnol. Bioeng.*, 2006, 95, 1101-1107.
6. Li, P. Wang, C. Huang, S. Pan and X. Li, *Food Anal. Methods*, 2013, 7, 1337-1344.
7. Z. Jin, X. Xu and h. Li, *Bioprocess Biosystems Eng.*, 2013, 36, 1141-1148.
8. F. Shen, X. Niu, D. Yang, Y. Ying, B. Li, G. Zhu and J. Wu, *J. Agric. Food Chem.*, 2010, 58, 9809-9816.
9. D. Cozzolino and C. Curtin, *Food Control*, 2012, 26, 241-246.
10. H. Yu, Y. Zhou, X. Fu, L. Xie and Y. Ying, *Eur. Food Res. Technol.*, 2007, 225, 313-320.
11. F. Shen, Y. Ying, B. Li, Y. Zheng and J. Hu, *Food Res. Int.*, 2011, 44, 1521-1527.
12. V. Di Egidio, N. Sinelli, G. Giovanelli, A. Moles and E. Casiraghi, *Eur. Food Res. Technol.*, 2010, 230, 947-955.
13. M. Blanco, A. C. Peinado and J. Mas, *Biotechnol. Bioeng.*, 2004, 88, 536-542.
14. Q. Ouyang, Q. Chen, J. Zhao and H. Lin, *Food Bioprocess Technol*, 2012, 6, 2486-2493.
15. L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc.*, 2000, 54, 413-419.
16. J. Cai, Q. Chen, X. Wan and J. Zhao, *Food Chem.*, 2011, 126, 1354-1360.
17. J. Shi, X. Zou, X. Huang, J. Zhao, Y. Li, L. Hao and J. Zhang, *Food Chem.*, 2013, 138, 192-199.
18. B. Ozturk, D. Yucesoy and B. Ozen, *Food Anal. Methods*, 2012, 5, 1435-1442.
19. Y. Bao, F. Liu, W. Kong, D. Sun, Y. He and Z. Qiu, *Food Bioprocess Technol*, 2014, 7, 54-61.
20. S. Fragoso, L. Acena, J. Guasch, M. Mestres and O. Busto, *J. Agric. Food Chem.*, 2011, 59, 10795-10802.
21. P. Williams and K. Norris, *Near-Infrared Technology in the Agricultural and Food In dustries, 2nd, ed.; AACC: St. Paul, MN,* , 2001.
22. G.-W. Choi, H.-W. Kang, Y.-R. Kim and B.-W. Chung, *Biotechnol Bioproc E*, 2008, 13, 765-771.
23. Chen and Y. Xu, *J. Inst. Brew.*, 2013, 119, 71-77.
24. X. Niu, F. Shen, Y. Yu, Z. Yan, K. Xu, H. Yu and Y. Ying, *J. Agric. Food Chem.*, 2008, 56, 7271-7278.
25. D. W. Lachenmeier, *Food Chem.*, 2007, 101, 825-832.
26. A. Edelmann, J. Diewok, K. C. Schuster and B. Lendl, *J. Agric. Food Chem.*, 2001, 49, 1139-1145.
27. C. J. Bevin, A. J. Fergusson, W. B. Perry, L. J. Janik and D. Cozzolino, *J. Agric. Food Chem.*, 2006, 54, 9713-9718.
28. S. Sivakesava, J. Irudayaraj and A. Demirci, *J. Ind. Microbiol. Biotechnol.*, 2001, 26, 185.
29. K. Beullens, D. Kirsanov, J. Irudayaraj, A. Rudnitskaya, A. Legin, B. M. Nicolaï and J. Lammertyn, *Sensors & Actuators B: Chemical*, 2006, 116, 107-115.
30. S. Sivakesava and J. Irudayaraj, *Appl. Eng. Agric.*, 2007, 16, 543-550.
31. R. G. Dambergs, A. Kambouris, I. L. Francis and M. Gishen, *J. Agric. Food Chem.*, 2002, 50, 3079-3084.
32. Urtubia, J. R. Pérez-correa, F. Pizarro and E. Agosin, *Food Control*, 2008, 19, 382-388.
33. J. Weeranantanaphan and G. Downey, *J. Inst. Brew.*, 2010, 116, 56-61.
34. A. Rudnitskaya, D. Kirsanov, A. Legin, K. Beullens, J. Lammertyn, B. M. Nicolaï and J.

Irudayaraj, *Sensors Actuators B: Chem.*, 2006, 116, 23-28.

35. A. L. H. Müller, R. S. Picoloto, P. d. A. Mello, M. F. Ferrão, M. d. F. P. dos Santos, R. C. L. Guimarães, E. I. Müller and E. M. M. Flores, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2012, 89, 82-87.

36. S. Grassi, J. M. Amigo, C. B. Lyndgaard, R. Foschino and E. Casiraghi, *Food Chem.*, 2014, 155, 279-286.

37. Q. Chen, J. Ding, J. Cai and J. Zhao, *Food Chem.*, 2012, 135, 590-595.

38. C. Zhang, N. Xu, L. Luo, F. Liu, W. Kong, L. Feng and Y. He, *Food & Bioprocess Technology*, 2014, 7, 598-604.

39. D. Cozzoino and C. Curtin, *Food Control*, 2012, 26, 241-246.

Figure caption

Fig. 1 SNV preprocessed NIR spectra (a) and Smooth-D1 pretreated MIR spectra (b) of all samples collected during the fermentation processes.

Fig. 2 Overall fermentation kinetic profiles of total sugar (a), pH (b) and AAN (c) of four different processes.

Fig. 3 NIR (a) and MIR (b) spectra of the samples collected at different times (0 to 20 days) in one of the fermentation trials.

Fig. 4 Principal component score plots of the Chinese rice wine samples analysed at different fermentation stages using NIR (a) and MIR (b).

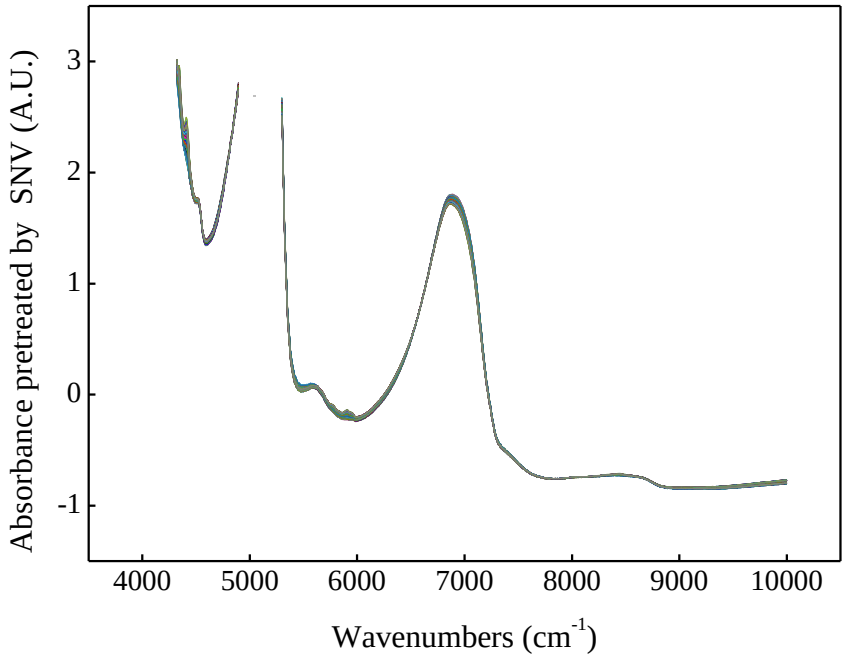
Fig. 5 The optimal subintervals selected by SiPLS based on NIR spectra for the prediction of total sugar (a), pH (b) and AAN (c) and the efficient combination of spectral subintervals selected by SiPLS based on MIR spectra for the prediction of total sugar (d), pH (e) and AAN (f).

Fig. 6 Correlation plots for the prediction of total sugar (a, d), pH (b, e) and AAN (c, f) using the full spectrum PLS models (a, b, c) and the optimal SiSVM models (d, e, f) based on NIR spectra.

Fig. 7 Correlation plots for the prediction of total sugar (a, d), pH (b, e) and AAN (c, f) using the full spectrum PLS models (a, b, c) and the optimal SiSVM models (d, e, f) based on MIR spectra.

Fig. 1

a



b

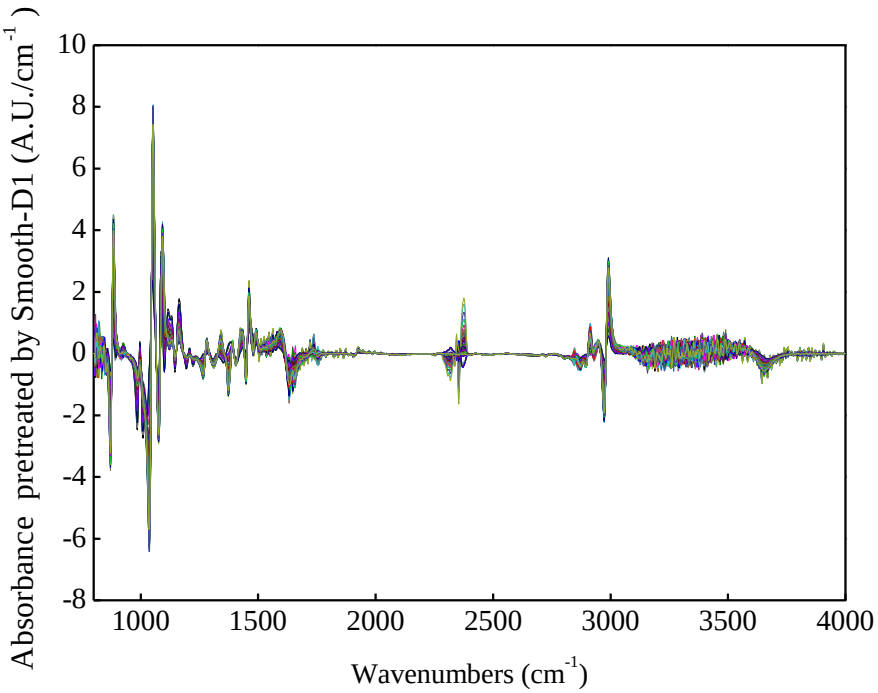


Fig. 2

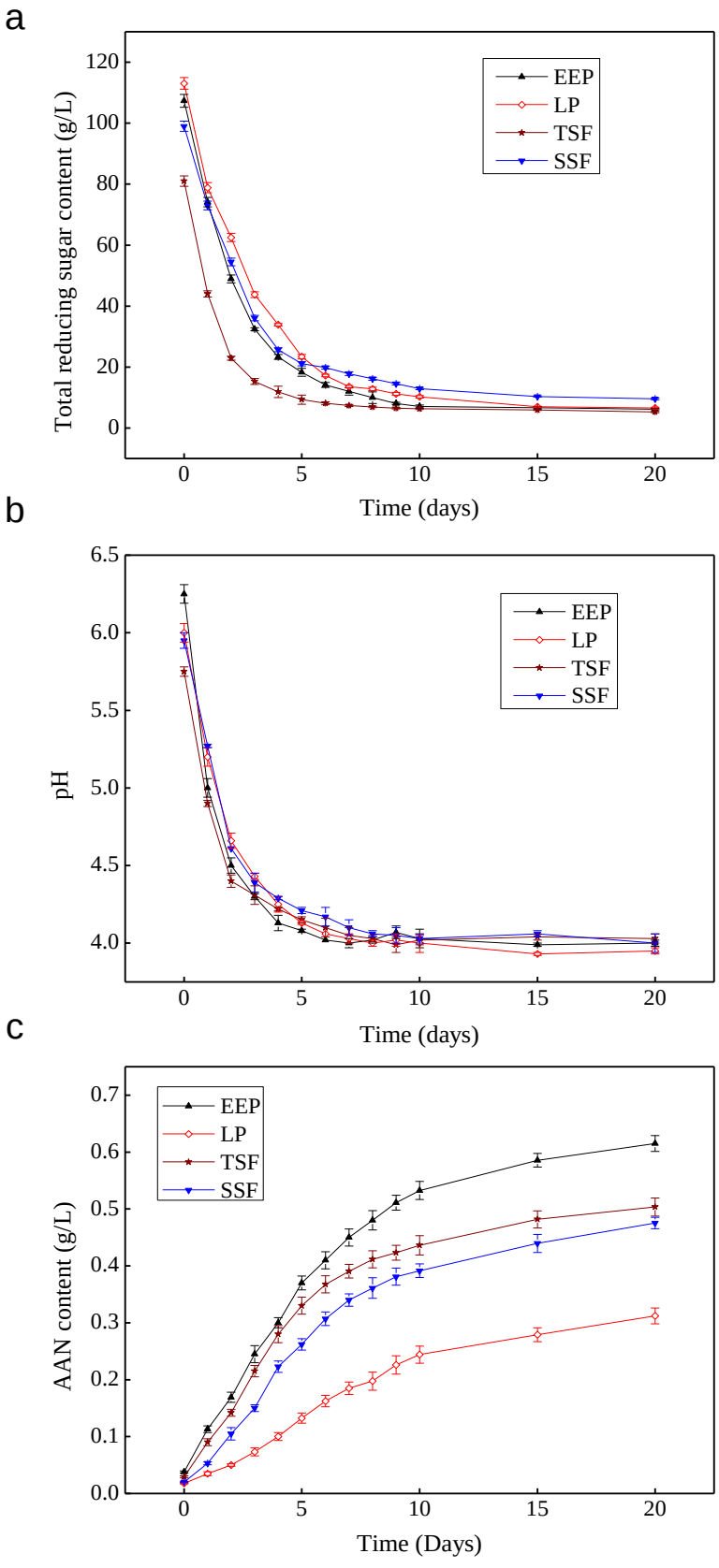


Fig. 3

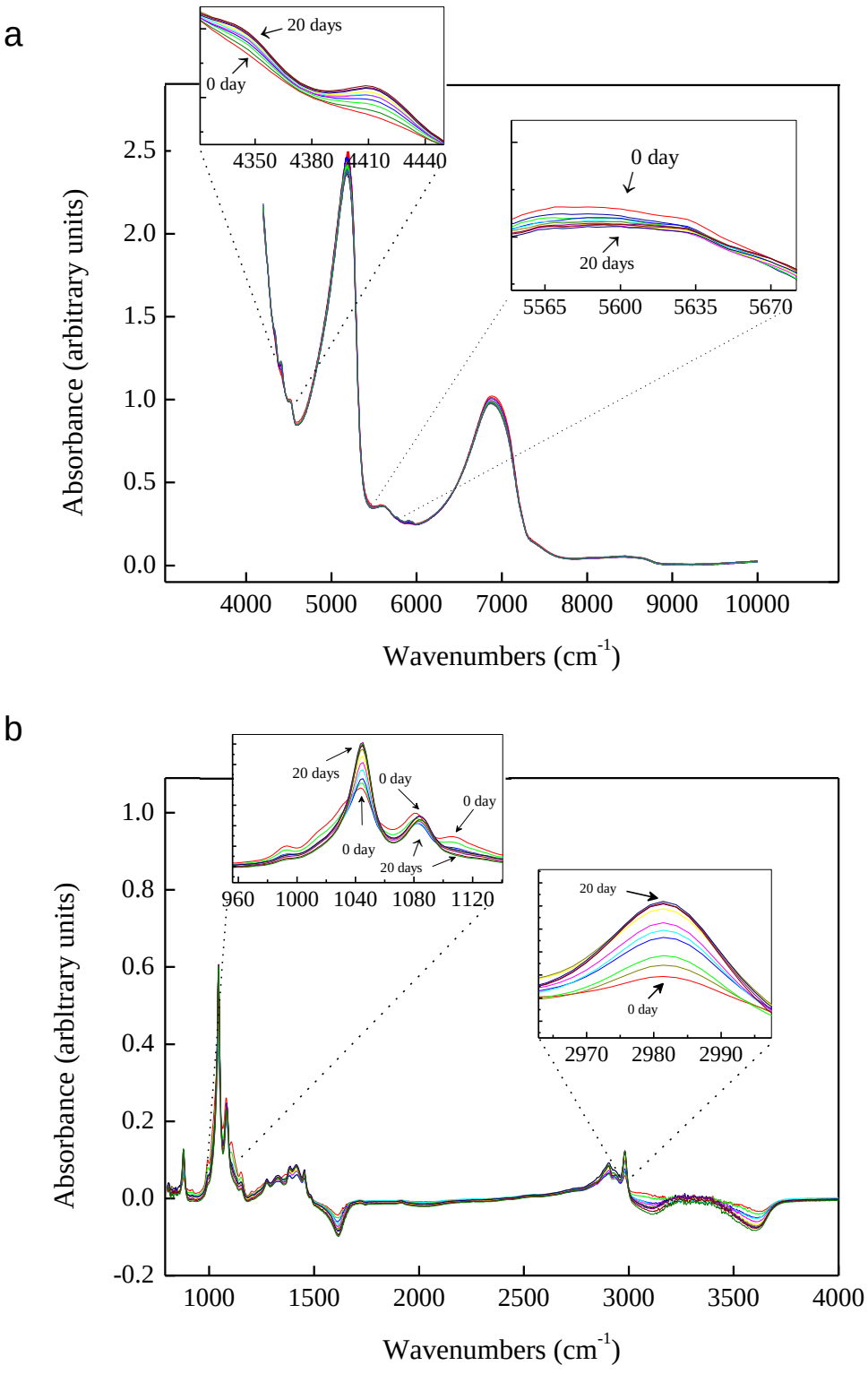


Fig. 4

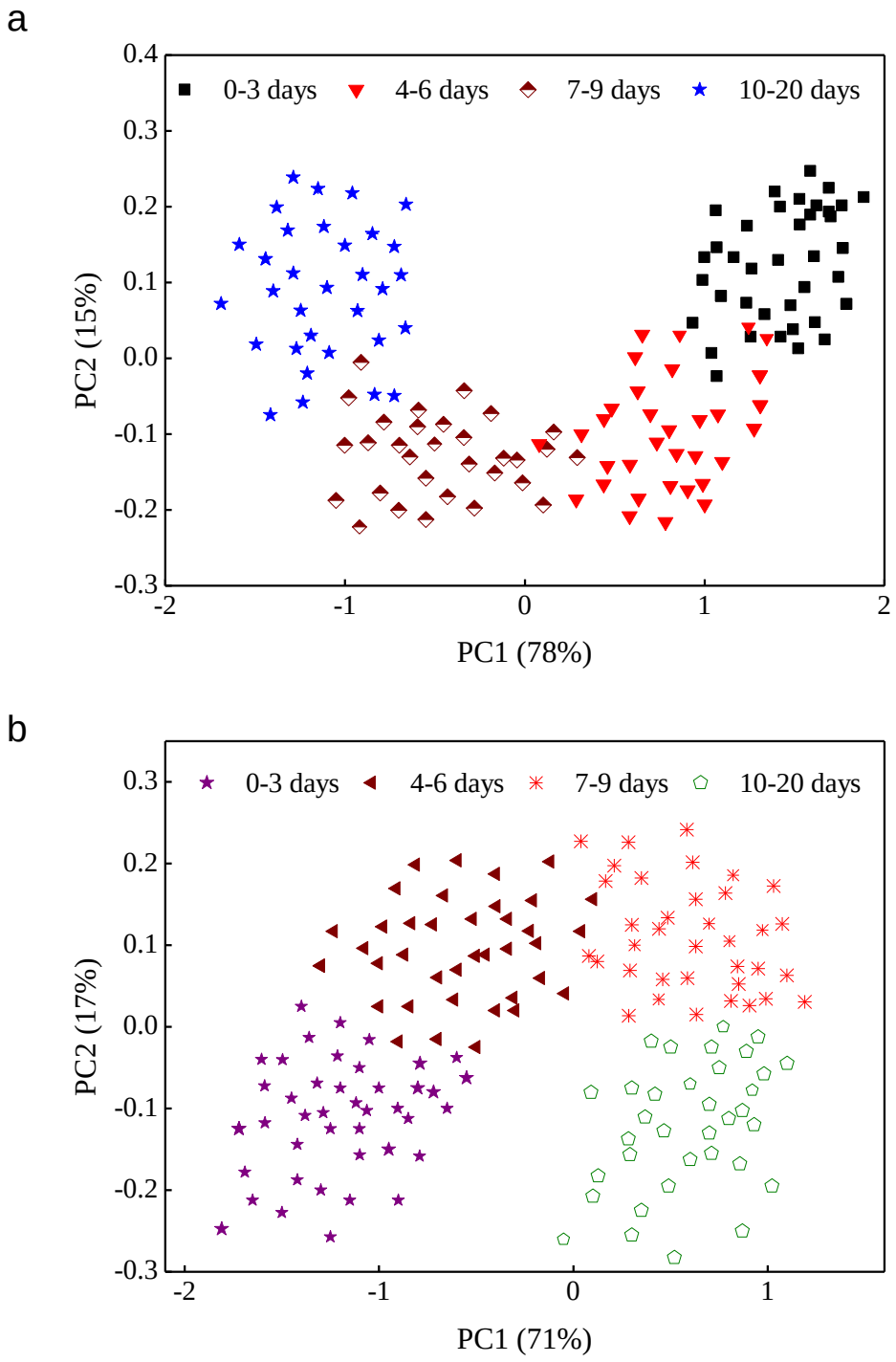


Fig. 5

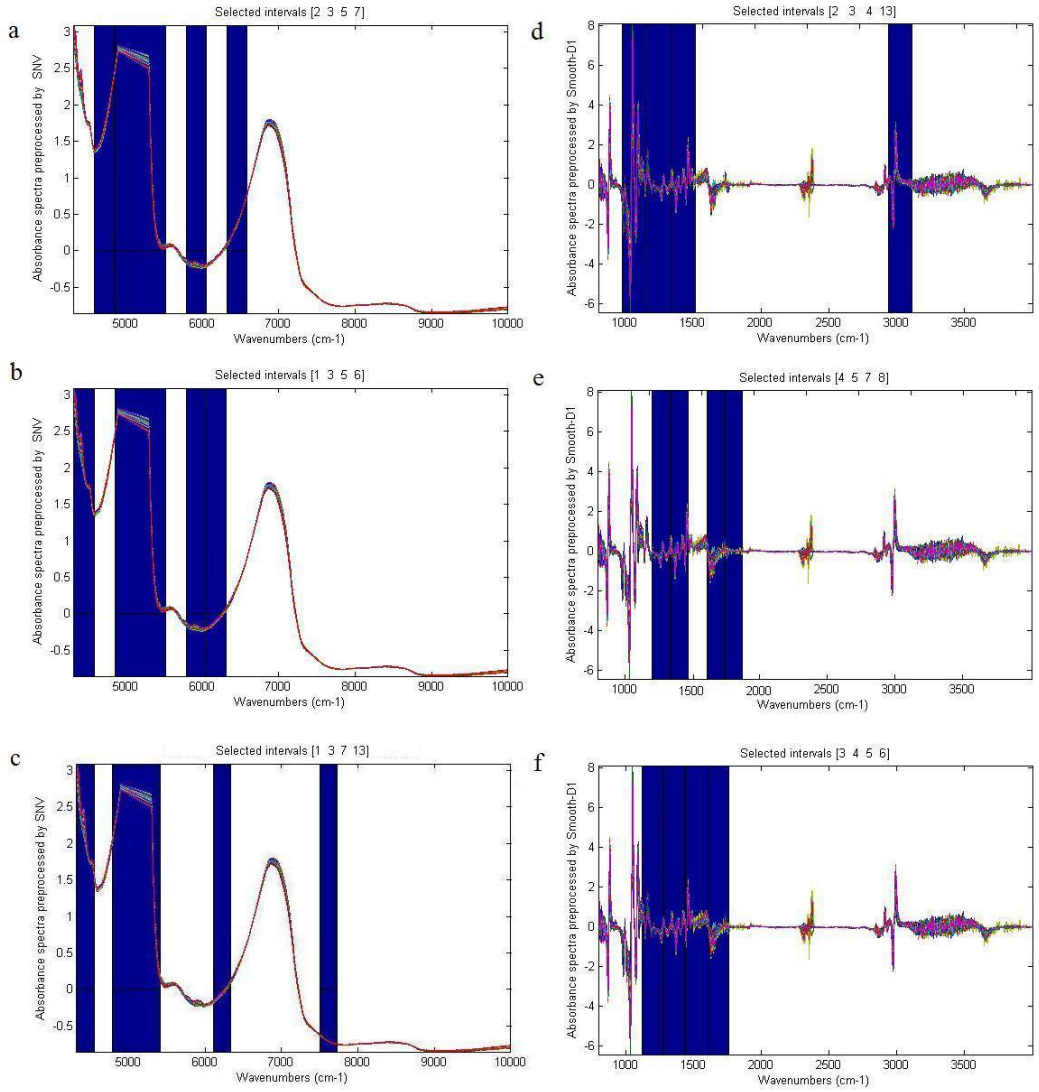


Fig. 6

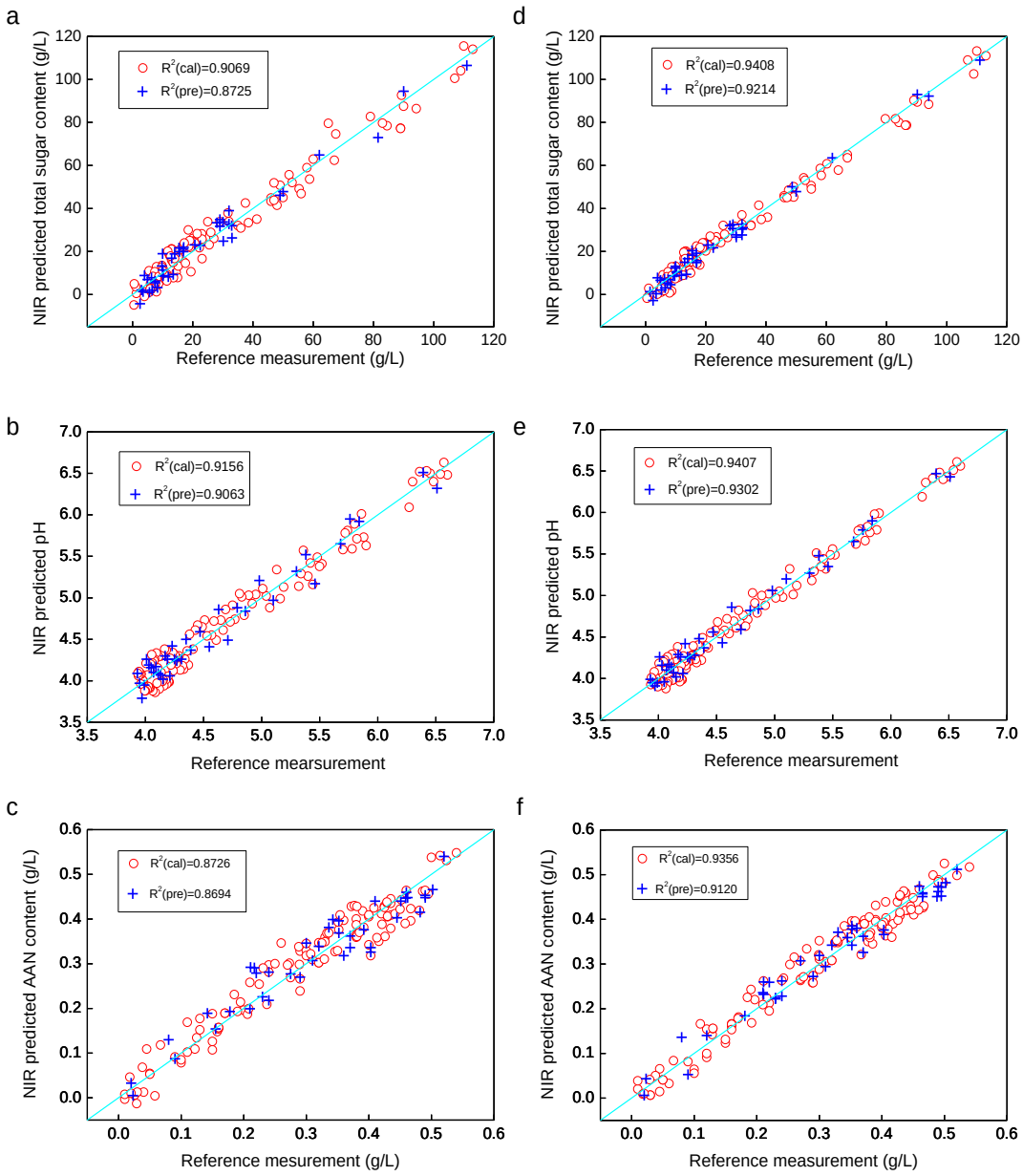


Fig. 7

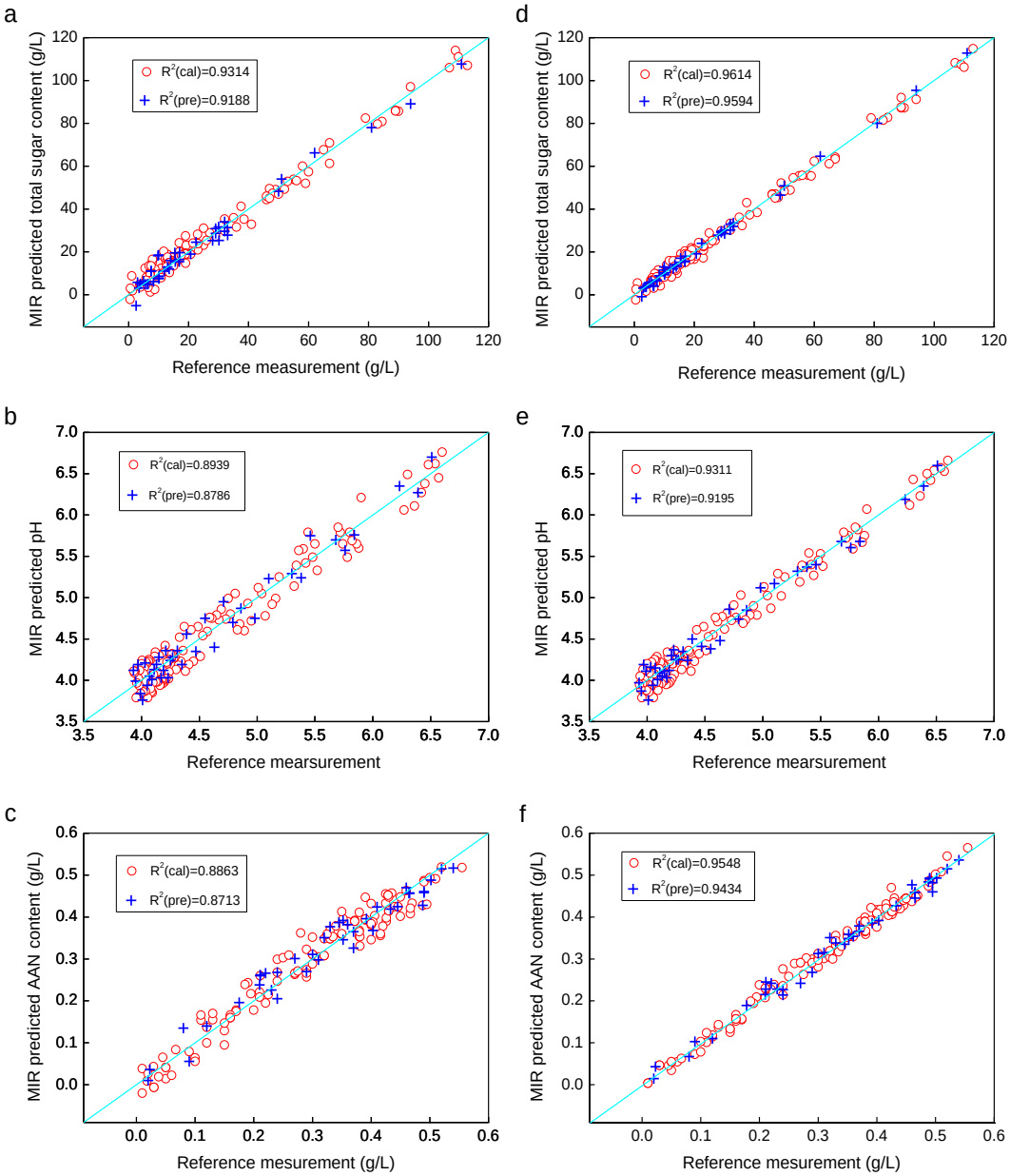


Table 1

Descriptive statistics for the chemical parameters used for the development of ATR-MIR calibration and prediction models.

Fermentation parameters	Units	Subsets	SN[a]	Mean	SD[b]	Range	CV[c]
Total reducing sugar	g/L	Calibration	117	28.31	27.34	0.5-113	96.57
		Prediction	39	24.10	24.49	2.5-111	101.62
pH		Calibration	117	4.70	0.76	3.94-6.6	16.17
		Prediction	39	4.63	0.71	3.94-6.51	15.33
Amino acid nitrogen	g/L	Calibration	117	0.29	0.14	0.01-0.56	48.28
		Prediction	39	0.33	0.14	0.02-0.54	42.42

[a] Sample number

[b] Standard deviation

[c] Coefficient of variation [{SD/mean}*100]

Table 2

Correlation matrix of the analyzed fermentation parameters.

	Total sugar	pH	ANN
Total sugar	1	0.422*	-0.783*
pH	0.422*	1	-0.171
ANN	-0.783*	-0.171	1

* Significant correlations at $p < 0.05$

Table 3

PLS-DA results of the correct classification rates.

	Correctly classified (%)				Average (%)
	Step 1	Step 2	Step 3	Step 4	
NIR					
Calibration	94.4	92.6	88.9	92.6	92.3
Prediction	83.3	88.9	88.9	88.9	89.7
MIR					
Calibration	97.2	92.6	92.6	92.6	94.0
Prediction	91.7	100	88.9	100	94.9

Table 4

Comparison of results based on different calibration models.

				Calibration			Prediction			
	Model	Parameters	variables	R^2	RMSECV	RMSECV (%)	R^2	RMSEP	RMSEP (%)	RPD
NIR	PLS	Total sugar	1370	0.9069	3.73	13.16	0.8725	3.53	14.64	6.94
		pH	1370	0.9156	0.14	2.98	0.9063	0.15	3.33	4.61
		Amino nitrogen	1370	0.8726	0.06	19.40	0.8694	0.07	22.70	1.87
	SiPLS	Total sugar	276	0.9253	3.42	12.09	0.8838	3.18	13.19	7.71
		pH	276	0.9202	0.14	2.93	0.9097	0.15	3.14	4.88
		Amino nitrogen	240	0.9113	0.04	13.10	0.8913	0.04	12.14	3.49
	SVM	Total sugar	1370	0.9278	3.09	10.92	0.8892	2.96	12.27	8.28
		pH	1370	0.9271	0.14	2.87	0.9176	0.12	2.54	6.03
		Amino nitrogen	1370	0.9207	0.03	11.90	0.8812	0.05	13.97	3.04
	SiSVM	Total sugar	276	0.9408	2.15	7.59	0.9214	2.07	8.57	11.84
		pH	276	0.9407	0.10	2.12	0.9302	0.10	2.18	7.06
		Amino nitrogen	240	0.9356	0.03	8.76	0.9120	0.03	9.09	4.67
MIR	PLS	Total sugar	1660	0.9314	2.83	9.98	0.9188	2.69	11.17	9.10
		pH	1660	0.8939	0.17	3.57	0.8786	0.22	4.66	3.30
		Amino nitrogen	1660	0.8863	0.05	16.14	0.8712	0.05	16.03	2.65
	SiPLS	Total sugar	371	0.9497	1.86	6.58	0.9391	1.95	8.08	12.57
		pH	277	0.9003	0.16	3.47	0.8794	0.21	4.45	3.45
		Amino nitrogen	332	0.9275	0.03	9.69	0.9049	0.04	10.67	3.98
	SVM	Total sugar	1660	0.9462	2.26	7.99	0.9406	1.82	7.53	13.48
		pH	1660	0.9214	0.14	2.91	0.9090	0.15	3.30	4.65
		Amino nitrogen	1660	0.9301	0.03	9.21	0.9125	0.03	8.91	4.76
	SiSVM	Total sugar	371	0.9614	1.37	4.82	0.9594	1.68	6.95	14.60
		pH	277	0.9311	0.13	2.74	0.9195	0.11	2.37	6.47
		Amino nitrogen	332	0.9548	0.02	7.90	0.9434	0.03	7.91	5.36

Table 5

Wavelength variables used in different multivariate regression models.

Parameters	Models	Selected wavelength range (cm^{-1})	
		NIR	MIR
Total sugar	PLS	4319-4894；5299-10001	800-4000
	siPLS	4585-4848；4852-5515；5785-6047；6317-6579	979-1157；1159-1336；1338-1515；2937-3112
	SVM	4319-4894；5299-10001	800-4000
	SiSVM	4585-4848；4852-5515；5785-6047；6317-6579	979-1157；1159-1336；1338-1515；2937-3112
pH	PLS	4319-4894；5299-10001	800-4000
	siPLS	4319-4582；4852-5515；5785-6047；6051-6313	1205-1338；1340-1471；1606-1737；1739-1870
	SVM	4319-4894；5299-10001	800-4000
	SiSVM	4319-4582；4852-5515；5785-6047；6051-6313	1205-1338；1340-1471；1606-1737；1739-1870
AAN	PLS	4319-4894；5299-10001	800-4000
	siPLS	4319-4547；4782-5411；6109-6336；7497-7725	1120-1278；1280-1438；1440-1598；1600-1758
	SVM	4319-4894；5299-10001	800-4000
	SiSVM	4319-4547；4782-5411；6109-6336；7497-7725	1120-1278；1280-1438；1440-1598；1600-1758

Table 6

Results of SiPLS multivariate regression models selected optimal spectral subintervals for Total sugar, pH and AAN.

Methods	Number of subintervals	Total sugar			pH			AAN		
		Selected subintervals	R^2	RMSECV	Selected subintervals	R^2	RMSECV	Selected subintervals	R^2	RMSECV
NIR	11	[1 2 3 5]	0.9046	4.8531	[1 2 3 11]	0.9060	0.1471	[1 2 7 9]	0.8797	0.0540
	12	[1 2 3 6]	0.9090	4.5751	[2 3 6]	0.8882	0.1587	[1 2 6 8]	0.8838	0.052[illegible]
	13	[1 2 3 7]	0.9151	4.1737	[2 3 11]	0.8980	0.1528	[1 2 7 11]	0.8826	0.052[illegible]
	14	[2 3 7 11]	0.9058	4.7737	[1 2 3 6]	0.9073	0.1458	[1 3 7 15]	0.9008	0.043[illegible]
	15	[2 3 5 12]	0.9215	3.7192	[3 6]	0.9016	0.1504	[3 12]	0.8969	0.0454
	16	[1 2 6]	0.9244	3.4943	[1 3 6 13]	0.9139	0.1415	[1 4 13 14]	0.9020	0.042[illegible]
	17	[2 5 7 12]	0.9181	3.9707	[3 5 7 14]	0.9116	0.1429	[2 6 7 17]	0.8866	0.0510
	18	[3 7 17]	0.9209	3.7589	[5 6 11 14]	0.9166	0.1397	[6 7 13]	0.9042	0.0416
	19	[3 6 17]	0.9236	3.5559	[1 3 6 8]	0.9192	0.1384	[2 7 8 17]	0.8996	0.044[illegible]
	20	**[2 3 5 7]**	**0.9253**	**3.4236**	**[1 3 5 6]**	**0.9202**	**0.1379**	[2 8 13 16]	0.8984	0.044[illegible]
	21	[2 6 12 17]	0.9227	3.6266	[5 6 11 18]	0.9041	0.1479	[1 7 12]	0.8854	0.051[illegible]
	22	[2 3 5 6]	0.9126	4.3413	[4 5 6 11]	0.9169	0.1396	[2 7 13 21]	0.8996	0.044[illegible]
	23	[2 4 5 7]	0.9096	4.5311	[2 6 19]	0.9157	0.1403	**[1 3 7 13]**	**0.9113**	**0.038[illegible]**
	24	[1 2 3 12]	0.9175	3.7589	[2 6 10 16]	0.8988	0.1521	[2 13 20]	0.8917	0.046[illegible]
	25	[1 2 5 7]	0.9104	4.4869	[2 5 12]	0.9076	0.1455	[2 11 13]	0.8905	0.048[illegible]
MIR	11	[1 2 7 8]	0.9052	2.1856	[1 2 4 7]	0.8805	0.2606	[1 3 6 9]	0.8884	0.039[illegible]
	12	[1 2 3 10]	0.9428	1.9281	[1 3 4 6]	0.8807	0.2576	[3 4 7 11]	0.9119	0.033[illegible]
	13	[3 12]	0.9392	1.9471	[1 2 7 8]	0.8828	0.2445	[5 6 11 12]	0.9045	0.036[illegible]
	14	[1 2 12]	0.9450	1.9093	[1 2 8 13]	0.8823	0.2481	[5 10]	0.9005	0.03[illegible]
	15	[2 4 6 13]	0.9256	2.0366	[3 6 8]	0.8835	0.2410	[1 2 14]	0.9112	0.035[illegible]
	16	[3 4]	0.9294	2.0142	[7 8 10 16]	0.8847	0.2355	[2 5 6 9]	0.9210	0.0317
	17	[3 4 14]	0.9400	1.9423	[2 7 8 14]	0.8883	0.2241	[3 5 7 14]	0.9151	0.034[illegible]
	18	**[2 3 4 13]**	**0.9497**	**1.8636**	[7 10 15]	0.8998	0.1713	[3 6]	0.9229	0.031[illegible]
	19	[1 3 4 10]	0.9151	2.1040	[5 8 19]	0.8914	0.2099	[1 2 4 10]	0.9257	0.030[illegible]
	20	[3 11 13]	0.9470	1.8940	[2 7 15 19]	0.8896	0.2166	**[3 4 5 6]**	**0.9275**	**0.0281**
	21	[4 12 13 18]	0.8963	2.2347	[1 7 10 13]	0.8927	0.2049	[1 3 7 11]	0.9253	0.030[illegible]
	22	[1 2 3 19]	0.9387	1.9497	[5 8 11 18]	0.8924	0.2063	[2 4 6]	0.9232	0.031[illegible]
	23	[3 4 13 21]	0.9065	2.1609	[4 9]	0.8969	0.1869	[2 3 6 19]	0.9228	0.0317
	24	[1 3 4 13]	0.9366	1.9671	**[4 5 7 8]**	**0.9003**	**0.1634**	[6 8 11 12]	0.9239	0.0312
	25	[1 2 13]	0.9238	2.0548	[3 6 7]	0.8849	0.2234	[2 5 6 23]	0.9259	0.0303