ORIGINAL PAPER

Discrimination Between Shaoxing Wines and Other Chinese Rice Wines by Near-Infrared Spectroscopy and Chemometrics

Fei Shen · Danting Yang · Yibin Ying · Bobin Li ·
Yunfeng Zheng · Tao Jiang

Abstract Shaoxing rice wine (also called Shaoxing wine) is the most well-known Chinese rice wine in China. The common fraudulent practice in the commercialization of Chinese rice wine is to sell wines from different geographical origins under the denomination of Shaoxing rice wine. In this study, the use of near-infrared (NIR) spectroscopy combined with chemometrics as a rapid tool for the discrimination of Chinese rice wine from three geographical origins ("Fujian", "non-Shaoxing", "Shaoxing") has been preliminarily investigated. NIR spectra were collected in transmission mode in the wavelength range of 800–2,500 nm. Discriminant models were developed by principal component analysis (PCA), discriminant analysis (DA), and discriminant partial least-squares analysis (DPLS). The chemical properties of Chinese rice wine were also investigated to find out the difference between samples from three varied origins. The results showed that good classification could be obtained after spectral pre-treatment. The percentage of samples correctly classified by both DA and DPLS methods in calibration and validation set was 97.2% and 100%, respectively. The results demonstrated that NIR could be used as a simple and rapid technique to distinguish Shaoxing wines from non-Shaoxing wines and Fujian wines. To further validate the ability of NIR spectroscopy, more samples should be incorporated to build a more robust model.

F. Shen · D. Yang · Y. Ying (✉)
College of Biosystems Engineering and Food Science,
Zhejiang University,
310029 Hangzhou, China
e-mail: yingyb@zju.edu.cn

B. Li · Y. Zheng · T. Jiang
National Center for Quality Supervision
and Testing of Rice Wine,
312071 Shaoxing, China

Introduction

Chinese rice wine, also named yellow wine, is a sweet, golden wine made mainly from glutinous rice and wheat which has a long history over 2,000 years in China (Wang and Xu 2005). Glutinous rice has higher protein and lower fat contents as compared to non-glutinous rice, and wheat serves as an abundant carbon, nitrogen and microelement source for mold and yeast in fermentation. The unique brewing technology handed down from generations brings rice wine with bright brown color, subtle sweet flavor, and low alcoholicity. It has become increasingly popular in China due to its high nutritional value and medical use function (Que et al. 2006b). Chinese rice wine has many varieties in China according to different geographical origins, such as Fujian catalyzed wine, Zhejiang jiashan wine, Shandong jimo wine, and so on. Among them, Shaoxing rice wine, from which Chinese rice wine originated, is the most famous and representative (Niu et al. 2008). Only rice wine brewed with water from "Mirror Lake" in Shaoxing region can be defined as "Shaoxing wine", and labeled as the product of protected designation of origin (PDO) by Chinese National Standard GB 17946-2000. (GB means national standard; 17946 was the code of the official method; and 2000 was the year in which the method was revised). Until now, only a few wineries were entitled to use the trademark of "Shaoxing wine" in China. Nowadays, the determination of food authenticity and the detection of adulteration are major issues in food control and safety. It has been found that wine quality and flavor

varies greatly with geographical origin. The preference for wine of consumers is also different. Therefore, a rapid and accurate analytical method to determine the origin is essentially required for the prevention of illegal adulteration and the correct value estimation.

However, it is not easy to determine the geographical origin of wine rapidly with existing analytical methods. Currently, wine is often classified according to polyphenolic compounds, volatile compounds, minerals, and trace elements using high-performance liquid chromatography (HPLC) (Beltran et al. 2006), gas chromatography–mass spectroscopy (GC-MS) (Kruzlicova et al. 2009), electronic nose (Aleixandre et al. 2008), atomic absorption spectroscopy (AAS) (Galgano et al. 2008) and etc. These techniques are precise enough to guarantee the authenticity of wine. However, most of them are often sophisticated and time consuming and require tedious and complex processing for wines. Therefore a new cheap, easy, and rapid approach is essentially required to determine the authenticity of wine.

Recently, near-infrared (NIR) spectroscopy has been developed as an alternative to wet chemistry for quantitative and qualitative analysis. The information recorded in NIR spectrum reflects the vibrational responses of chemical bonds such as O–H, C–H, and N–H in materials which is related to its internal component and structure (Blanco and Villarroya 2002). The main advantage of NIR spectroscopy over other traditional methods is that it can be applied without the need for costly, laborious chemical and sensory analysis (Woodcock et al. 2008). In the wine industry, NIR spectroscopy has already been used to predict five enological parameters of Chinese rice wine (Yu et al. 2006), measure volatile aroma compounds in Riesling wine (Smyth et al. 2008), determine oak volatile compounds and ethylphenols in aged red wines (Garde-Cerdan et al. 2010), screen 15 parameters in different types of wines (Urbano-Cuadrado et al. 2004), and analyze elements in wine (Cozzolino et al. 2008). It also has been used to classify Riesling wines from different countries (Liu et al. 2008), discriminate distilled alcoholic beverages, and verification of adulteration (Pontes et al. 2006), and even evaluate sensory attributes of Australian Riesling and Chardonnay wines (Cozzolino et al. 2005).

Multivariate statistical techniques, such as PLS, combined with sensory characteristics and chemical composition data (e.g., trace elements, isotopic ratios, polyphenols) for geographical origin classification of wines has been reported in many literatures (Saurina 2010). However, few studies were focused on the geographical origin classification of Chinese rice wine. Two brands of Chinese rice wine from two origins have been differentiated according to trace elements content by AAS and NIR preliminary (Yu et al.

2007). Further study is still needed to investigate the differences of chemical properties of rice wine from different origins.

Chinese rice wine form three different geographical origins were used in this study, including "Shaoxing" wines; wines produced in Shaoxing region, but not labeled as Shaoxing wine, defined as "non-Shaoxing" wines, and wines from Fujian province in China, defined as "Fujian" wines. The object of this work was to investigate the potential of NIR spectroscopy as a rapid and low-cost tool to discriminate Chinese rice wine from three different origins ("Shaoxing", "non-Shaoxing", and "Fujian") using their spectral properties without depending on chemical composition and sensory characteristics. In addition, the chemical properties of rice wine samples were also investigated and compared to find out the differences between samples from different origins.

Materials and Methods

Samples and Reference Analysis

A total of fifty-four bottles of Chinese rice samples were obtained from National Center for Quality Supervision and Testing of Rice Wine, including 11 "Fujian" wines from 11 different wineries, 20 "non-Shaoxing" wines from 15 wineries (five labels×two bottle replicates and ten labels×one bottle), and 23 "Shaoxing" wines from 16 wineries (seven labels×two bottle replicates and nine labels×one bottle). The samples showed a high diversity of varieties, ages, sensory, and chemical characteristics. The information of marked age of samples is listed in Table 1. Thirty six samples (seven "Fujian", 13 "non-Shaoxing", and 16 "Shaoxing") were selected randomly as calibration set. The remaining 18 samples (four "Fujian", seven "non-Shaoxing", and seven "Shaoxing") were selected as validation set. The calibration set is used to build classification models and validation set is used to validate the performance of calibration models.

In order to investigate the differences of chemical composition between samples, all wines were analyzed for alcohol degree [%, $v\ v^{-1}$], total acidity [g L^{-1}], total sugar [g L^{-1}], non-sugar solid [g L^{-1}], amino acid nitrogen [g L^{-1}] and pH [pH units] using official methods for Chinese rice wine according to Chinese National Standard GB 13662-2000 (Yu et al. 2006).

Spectral Measurement

Before NIR spectra acquisition, all the samples were stored in the laboratory at 20°C for 24 h. Samples were scanned in transmission mode using an FT-NIR spectrometer (Thermo

Table 1 Distribution of Chinese rice wine samples

Origin	Num. of samples of 1-year-aged	Num. of samples of 3-year-aged	Num. of samples of 5-year-aged	Num. of samples of 10-year-aged
"Fujian"	5	3	3	0
"non-Shaoxing"	10	3	6	1
"Shaoxing"	6	8	8	1
Total	21	14	17	2

Electron Corp., Madison, WI, USA), which was equipped with an interferometer, a wide band light source (Quartz Tungsten Halogen, 50 W) and an InGaAs detector. Samples were scanned in a rectangular quartz cuvette of 1 mm path length with air as the reference. Spectra were collected by data acquisition software (OMNIC 6.1a, Thermo Electron Corp., Madison, WI, USA) and saved in absorbance format. Spectra were recorded in the range of 800–2,500 nm. The mirror velocity was 0.9494 cm s^{-1} and the resolution was 8 cm^{-1}. The spectrum of each sample was the average of 32 successive scans.

Spectral Data Preprocessing

NIR spectra were affected by many factors including temperature differences, light scattering, baseline drift, or background noise. As a result, NIR spectra contain not only chemical but also physical information which may be irrelevant and can mask the useful information in the spectra. These unwanted variations can have adverse effects on the development of calibration model. Therefore, it is often necessary to have NIR spectra pre-processed in order to minimize the contribution of physical effects and enhance the chemical information in model development and improvement (Lafargue et al. 2003). In this study, standard normal transformation (SNV), multiplicative scattering correction (MSC), and second derivative were used as spectral data preprocessing methods. SNV and MSC are two commonly used methods for scatter correction (Xu et al. 2008). Second derivative was used to eliminate baseline drifts and enhance small spectral features. To avoid enhancing the noise by second derivative, spectra are often smoothed at first by Savitzky–Golay or Norris derivative algorithm (Liu and He 2009).

Chemometric Methods

All the data analysis was carried out using the commercial chemometric software (TQ Analyst v6.2.1, Thermo Electron Corp., Madison, WI, USA). Principal component analysis (PCA), discriminant analysis (DA), and discriminant partial least squares (DPLS) have been

proven to be effective in many applications (Andre 2003; Lister et al. 2000; Xie et al. 2007), and were therefore used in the present study to classify rice wines from different origins.

Principal Component Analysis

PCA is often the first step of data analysis in order to detect patterns or outliers in the data (Wang et al. 2007). The purpose of PCA is to reduce the dimensionality of spectral data in order to extract main information. It transforms original variables into a few new variables, known as principal components (PCs). The PCs account as much as possible for the variability in the original spectral variables and they are orthogonal to each other to remove overlapped information in original spectra matrix. Each spectrum of the sample has its own PCA score, therefore the sample's information can be expressed by the PCs (Park et al. 2003). By plotting the PCs, one can view the interrelationships between different samples, examining any possible grouping of samples. Also, PCA can provide information of data structure to determine the optimized maximum variance in future modeling (Andre 2003).

Discriminant Analysis

DA is a supervised pattern recognition algorithm and produces a linear decision boundary between classes (Dixon and Brereton 2009). Samples were classified by calculating the Mahalanobis distance of a sample from the mean of gravity of the considered groups by Eq. 1 (Fu et al. 2007). It is suggested that the greater the Mahalanobis distances between two given groups, the greater the spectral differences between them. The Mahalanobis distance is a numerical value and used to classify "unknown" individuals to a given species. The sample was classified by computing its distance from each class center of gravity in Mahalanobis distance. If it is close to the center of gravity of its group, it is "correctly classified". In the case where the distance to the center of gravity of its group is superior to that to the center of gravity of its group, then it is "poorly classified" and will be reassigned to the other group. In this

study, DA was employed to classify wines into three groups based on predefined classes.

$$MD = (x - x_{avg})^T V^{-1} (x - x_{avg}) \qquad (1)$$

where MD is the Mahalanobis distance, x is the score vector of sample ($n \times 1$), x_{avg} is the mean score vector of the sample set ($n \times 1$), V is the score covariance matrix ($n \times n$), ($x - x_{avg}$), T is the transpose of ($x - x_{avg}$).

Discriminant Partial Least-Squares

Leverage diagnostic was applied first to detect the outlier sample and provide information on how much influence each sample had on the method standards (Chen et al. 2008). Once the outlier detection was performed, DPLS model was established for the qualitative analysis of rice wine samples. It is a PLS application for the optimum separation of classes. Each sample was assigned a dummy variable describing the categories as a reference value, an arbitrary number which indicated whether the sample belonged to a particular group or not (Lin et al. 2009). In this case, samples of "Fujian" samples were assigned a numeric of 0, samples of "non-Shaoxing" were assigned 1, and those of "Shaoxing" samples were assigned 2. The DPLS model was then developed by assigned the reference value (dummy variable) for each sample. A sample was considered to be correctly classified if the predicted value lay on the same side of the midpoint of the assign value, 0.5 is the cutoff criteria which is also used in previous studies (Cozzolino et al. 2003; Xie et al. 2007). That is to say the "Fujian" sample was classified correctly if the value was between −0.5 and 0.5; otherwise, the sample was classified

wrong. For "non-Shaoxing" samples with predicted value range from 0.5 to 1.5 and "Shaoxing" sample from 1.5 to 2.5, all were considered to be correctly classified. The optimum number of latent variables (factors or LVs) is determined by the lowest value of predicted residual error sum of squares. The correlation coefficient (r), root mean square error of calibration (RMSEC), and root mean square error of prediction (RMSEP) are often used to evaluate the performance of the model established. It is expected to have ideal models with lower RMSEC and RMSEP as well as higher r.

Results and Discussion

Chemical Analysis

The descriptive statistics for the chemical composition of Chinese rice wine samples are summarized in Table 2. It can be observed that the average value of each parameter of "Shaoxing" wines is the greatest in the three groups, while the standard deviation (SD) value is the smallest which may indicate a stable composition. Chinese rice wine was classified into many varieties according to main component (e.g., total sugar) and winemaking techniques. It is observed that "non-Shaoxing" wines and "Fujian" wines contain more varieties than "Shaoxing" wines which may produce a higher SD value.

In order to compare statistically, t test ($p < 0.05$) was performed on the reference data of three groups. No statistically significant differences could be observed for pH value in the set of wines analyzed. However, statistically significant differences were observed in alcohol and

Table 2 Descriptive statistics for chemical composition of Chinese rice wine samples from three different geographical origins

	"Fujian" (N^a=11)			"non-Shaoxing" (N=20)			"Shaoxing" (N=23)		
	Range	Mean	SD[b]	Range	Mean	SD	Range	Mean	SD
Alcohol [%, $v\,v^{-1}$] *	12.0–15.5	13.58	0.93	13.2–16.8	15.39	1.04	15.3–17.9	16.92	0.70
TA[c] [g L^{-1}]	1.1–6.5	4.78	1.41	3.7–6.0	4.76	0.72	4.9–7.2	5.72	0.50
TS[d] [g L^{-1}]	0.1–54.2	26.76	20.40	14.9–37.4	23.86	5.37	22.1–38.0	26.90	3.33
NSS[e] [g L^{-1}]*	2.4–29.6	19.49	6.99	17.6–33.5	25.15	4.19	27.5–41.1	34.25	3.64
AAN[f] [g L^{-1}]	0.03–1.20	0.65	0.32	0.43–0.94	0.64	0.14	0.60–0.96	0.78	0.10
pH [pH units]	3.2–4.4	3.96	0.34	3.6–4.4	4.01	0.24	3.9–4.4	4.13	0.14

*$p < 0.05$ significant difference

[a] Number of samples
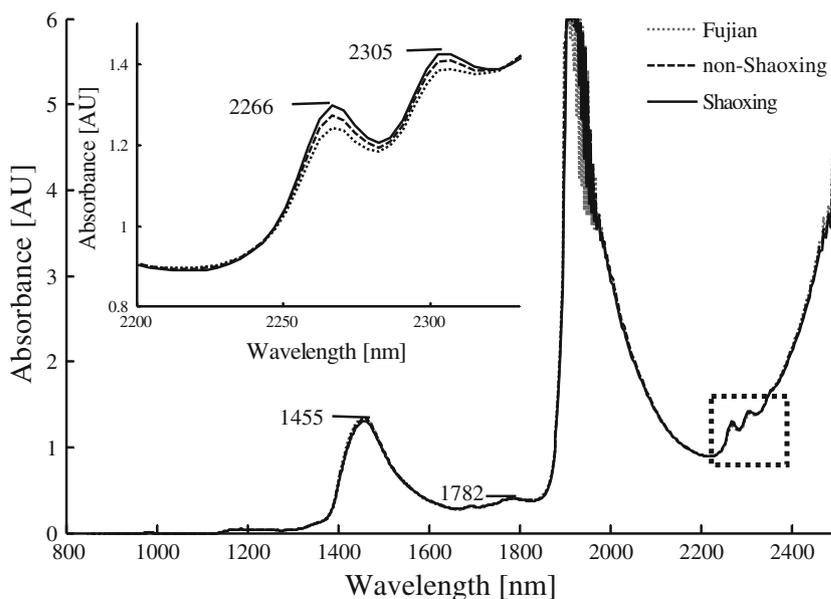
[b] Standard deviation

[c] Total acidity

[d] Total sugar

[e] Non-sugar solid

[f] Amino acid nitrogen

Fig. 1 Average NIR spectra of 11 "Fujian" samples, 20 "non-Shaoxing" samples, and 23 "Shaoxing" samples without any preprocessing
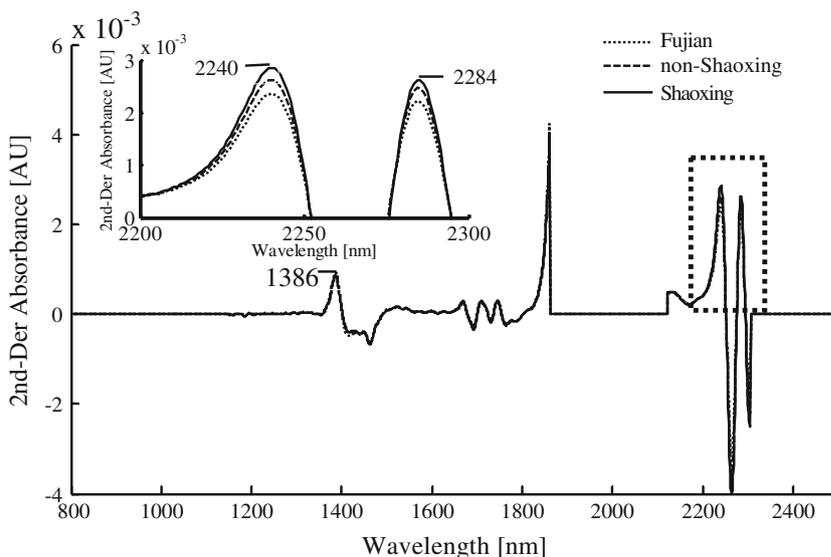

non-sugar solid. The data shows that "Shaoxing" wines contain more alcohol and non-sugar solid than the other groups which indicates differences in raw materials or manufacturing process between them. Because the difference between rice wine samples in NIR spectra belongs to the major components of samples, it is suggested that these indices may play an important role in the discrimination of samples from three different origins.

Spectral Analysis

Figure 1 shows the raw average NIR spectra of "Fujian", "non-Shaoxing", and "Shaoxing" samples. No obvious differences could be observed between the three spectra from visual inspection. The spectra are highly overlapped

except in the region of 1,400–1,600 nm where "Fujian" is a little higher than "non-Shaoxing" and "Shaoxing". In the region of 2,220–2,320 nm, "Shaoxing" is a little higher than "non-Shaoxing" and "Fujian", as shown in Fig. 1. The sequence for the absorbance intensity of the three groups in the range from 2,220 to 2,320 nm is in accordance with that of alcohol content. The main features of the spectra are absorption bands at 1,455 and 1,900–1,950 nm, which are related to the first overtone of the O–H group in water and a combination of stretch of O–H in water and ethanol, respectively. Absorption band at 2,266 nm is likely related with C–H combinations and O–H stretch overtones. Absorption band at 2,305 nm is related with C–H overtones of ethanol and sugars (Dambergs et al. 2002; Niu et al. 2008). Absorption bands at 1,450, 1,790 and 2,266 nm

Fig. 2 Average second derivative spectra of 11 "Fujian" samples , 20 "non-Shaoxing" samples and 23 "Shaoxing" samples (wavelength range of 1,880–2,100 and 2,330–2,500 nm not included)
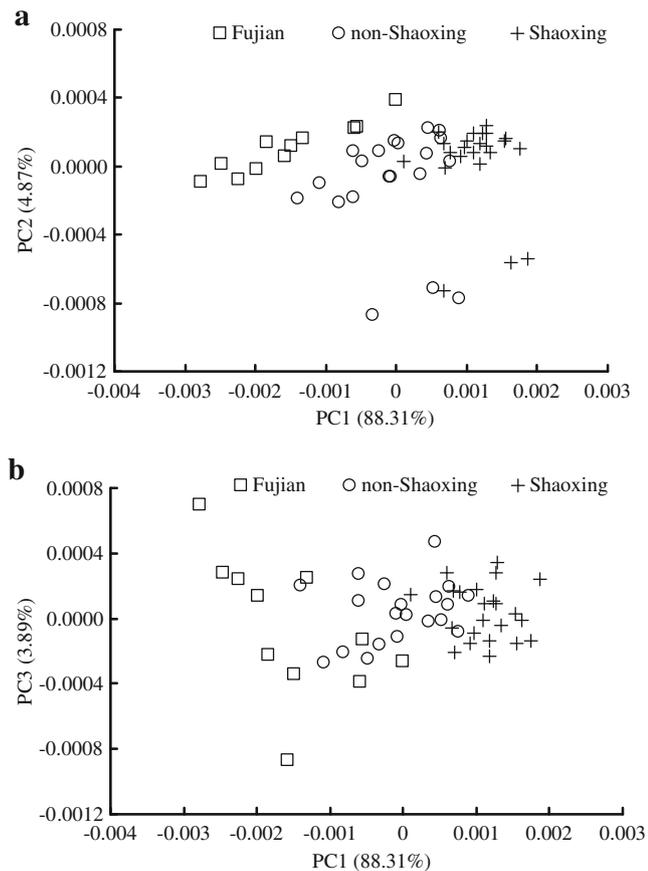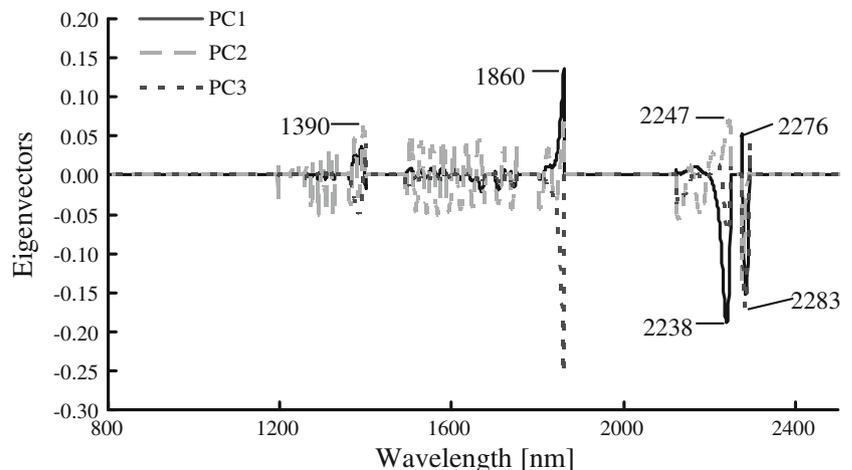
a

b

Fig. 3 Two-dimensional principal component score plot for "Fujian", "non-Shaoxing", and "Shaoxing" samples **a** PC1 vs. PC2; **b** PC1 vs. PC3

The spectra were transformed by second derivative to remove baseline drift and enhance small spectral features. The small spectral differences between the three sample groups were observed at 1,386, 1,668, 1,705, 1,745, 2,240, and 2,284 nm which were shown in Fig. 2. The result demonstrates that the small features of three spectra are enhanced after processing with second derivatives. The classification by NIR spectroscopy is based on the qualitative differences in the spectra.

PCA

PCA was often performed as the first attempt to extract and visualize the main information in multivariate data to example qualitative differences between samples. Different data pretreatments were employed and the second derivative of the spectra which had turned out to be the best data preprocessing method compared to other methods was used in this study. Figure 3a–b shows the two-dimensional score plot (PC1 vs. PC2 and PC1 vs. PC3) derived from second derivative spectra (PC1, PC2, and PC3 explain 88.31%, 4.87%, and 3.89% of the total variance, respectively). From Fig. 3a–b, due to the diversity of Chinese rice wine samples, no obvious groups were observed in the scatter plot. Some samples overlapped with each other, which was similar in the classification of Riesling wines from different countries (Liu et al. 2008). However, a trend of separation could also be detected in Fig. 3a, and the second component showed a limited ability to discriminate between samples of different origins. In Fig. 3a, "Fujian" samples were separated clearly from "non-Shaoxing", and "Shaoxing". The wines from "Shaoxing" were mostly located at the right of the plot and gather to a relative small group except for three samples which might due to different varieties. Several samples from "non-Shaoxing" overlapped with samples from "Shaoxing" group which might due to the similar climate or winemaking

were also reported to be associated with sucrose, fructose, and glucose in fruit juices (Lanza and Li 1984). Two wavelength regions are not used in any further analysis from 1,880 to 2,100 nm and from 2,330 to 2,500 nm because the absorbance value is higher than 1.5 absorbance units, which is beyond the linear response region of the detector.

Fig. 4 Eigenvectors of the first three principal components on the whole sample set analyzed by NIR

technique between them. The overlapped samples from "Fujian" and "non-Shaoxing" groups in Fig. 3b suggested some similarities in the inherent chemical composition. No overlapped sample was observed among "Fujian" and "Shaoxing" samples in Fig. 3 indicated greatest differences between them. Additionally, no obvious separation between samples related to wine age was observed. However, the result obtained by PCA is not perfect. PCA only indicates the visualizing dimension spaces. Therefore, discriminant analysis is utilized for an improved separation.

The eigenvectors for the first three PCs derived from second derivative spectra (the range of 1,880–2,100 and 2,330-2,500 nm not included) which were shown in Fig. 4 were analyzed in an attempt to investigate the bias for the spectral discrimination between samples. PC1 explains 88.31% of the total variances and the highest eigenvectors were found at 1,860 mainly related to the combination of stretch and deformation of O–H group in water, at 2,276 nm relate to CH_2 groups of ethanol. PC2 accounts for 4.87% of the variation, and the highest eigenvectors were found at around 2,247 nm related to C=C and C–H tones of ethanol and sugar. The eigenvector at around 1,390 and 1,490–1,760 nm is associated with O–H stretch of H_2O and NH_2 groups of nitrogen compounds, respectively (Yu et al. 2007). PC3 accounts for 3.89% of the variation, and highest eigenvectors were found at around 2,120–2,300 nm. It is noticed that unlike high eigenvectors found at around 1,400–1,490 nm mainly related to O–H first overtone of H_2O by other authors (Yu et al. 2008; Cozzolino et al. 2003), the result in this study indicated that the region after second derivative transform contributed less for the classification of rice wines from the three origins. The spectral regions mentioned above are associated with water, ethanol, sugar, and other constituents in Chinese rice wine which help to explain the bias for the discrimination of rice wines.

DA

The PCA scores obtained from PCA analysis were used to establish discriminant models. It is suggested to improve the classification result when more PCs are included in developing the model (Liu et al. 2006). Therefore, the first 15 PCs were employed into DA because they covered the most variation (>99.8% of the total variance) contained in the spectral data. The Mahalanobis distance plots of every sample to the centre of gravity of two classes ("Fujian" and "non-Shaoxing" groups, "Fujian" and "Shaoxing" groups as well as "non-Shaoxing" and "Shaoxing" groups) by discriminant analysis are shown in Fig. 5a–c. The diagonal line was used to identify the boundary of two classes. From Fig. 5, it can be seen that the samples were divided into
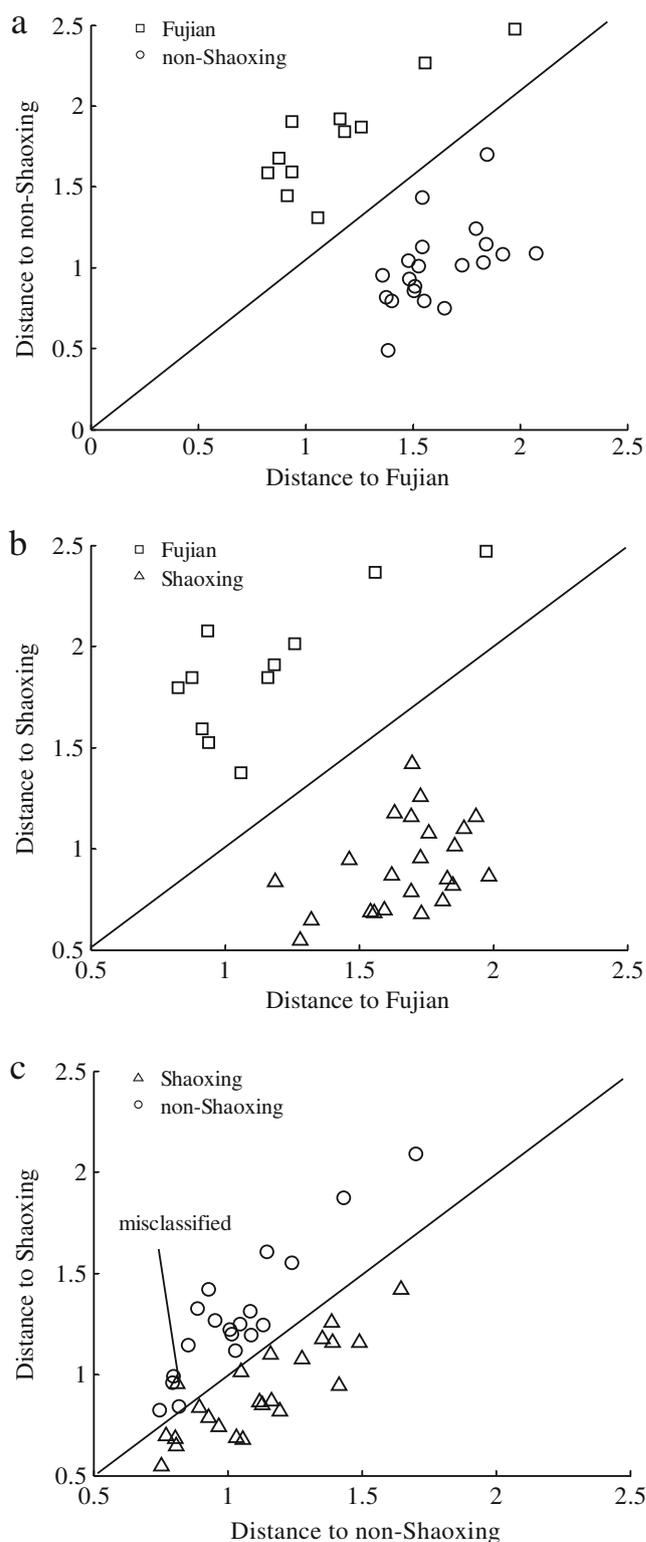

Fig. 5 Pairwise Mahalanobis distance of "Fujian", "non-Shaoxing", and "Shaoxing" samples using discriminant analysis method

three clusters according to origin. In the calibration analysis, the Fujian samples were separated completely from the "non-Shaoxing" and "Shaoxing" groups, as is shown in Fig. 5a and b. Only a sample from "Shaoxing" was misclassified as "non-Shaoxing", as can be seen in Fig. 5c. The Mahalanobis distance of the misclassified sample to the center of gravity of "non-Shaoxing" group is 0.8116, lower than that of "Shaoxing" group (0.9553); then, it is assigned to "non-Shaoxing" group. The percentage of correctly classified rate is 97.2%. In the validation analysis, all the samples were correctly classified. The percentage of correctly classified rate is 100% in validation set. The result indicates that DA method may be suitable for the discrimination of origins of rice wines.

DPLS

In the application of PLS algorithm, it is generally known that the spectral preprocessing methods and the number of PLS factors is critical parameters. Here, DPLS models with one to ten factors were investigated. The number of factors used in each model was determined on the basis of the lowest value of the predicted residual error sum of squares in order to avoid over-fitting. Table 3 shows the statistics for rice wine samples using DPLS models on the raw spectra and spectra with various spectra pretreatments. The result indicates that SNV and MSC processes can increase classification accuracy compared with raw spectra. SNV-treated model with seven factors got the best prediction performance, while the number of factors used in MSC treated model was four. The derivative process can also increase classification accuracy compared with raw spectra, but the classification accuracy is lower than MSC and SNV treatments. The best result was obtained using raw spectra by SNV treatment, with $r=0.9479$, RMSEC value of 0.242, RMSEP value of 0.300. In the validation analysis, no sample was misclassified by DPLS model with raw spectra after SNV treatment and the correctly classified rate is 100%. The result suggests that SNV-treated DPLS model contains enough information for the discrimination of rice wines from the three geographical origins.

The ability of the NIR spectroscopy to identify wine is based on the vibrational responses of chemical bonds (C–H, N–H, and O–H bonds) in NIR region. It is probable that the higher the variability between the samples types in the chemical components, which respond in these regions of the spectra, the better the accuracy of the model (Cozzolino et al. 2003). It is notable that the variation in chemical component of wines from different origins might be related to raw materials as well as winemaking technique used. Shaoxing rice wine is produced from high-quality glutinous rice and wheat koji, which provide sufficient starch, protein and microorganism source for fermentation. Generally, Shaoxing rice wine is more abundant in some chemical components (e.g., non-sugar solid, carbohydrate, amino acid) than rice wines from other origins (Mao et al. 2010). Apart from this, water from "Mirror Lake" served as plentiful mineral source in fermentation which brings with Shaoxing rice wine a preferable mouthfeel and subtle sweet flavor. Fujian wines are mainly produced from rice with "hong Qu" and have their own particular flavor (Que et al. 2006a). The application of HPLC method would be helpful to distinguish small amount of aromatic and volatile compounds between different origins in a further study. The reference data observed in this study shows that alcohol and non-sugar solid may play an important role in the discrimination. However, it is hard to define a single or several compounds that can explain the separation reason between wines due to the nature of the chemometric methods used. Many chemical compounds contained in wines respond in NIR spectra can act as a "fingerprint" for the discrimination of wines. DA and DPLS methods achieved high correct classification rate. The result indicates that differences in rice wines from different origins do exist, and NIR contains information able to discriminate between groups.

Conclusions

In conclusion, the application of NIR spectroscopy to classify Chinese rice wine from different geographical

Table 3 The statistics for the rice wine samples using DPLS models on the raw spectra and spectra with various pre-treatment

Pre-treatment	r	RMSEC	RMSEP	Correctly classified in calibration (%)	Factors
None	0.64765	0.578	0.369	58.8	1
MSC	0.91168	0.312	0.328	91.2	4
SNV	0.94789	0.242	0.300	97.2	7
2nd derivative	0.88919	0.347	0.250	85.3	3
MSC+2nd derivative	0.88808	0.349	0.246	85.3	3
SNV+2nd derivative	0.89491	0.339	0.266	88.2	4

origins is presented in this work. The percentage of samples correctly classified by both DA and DPLS methods in calibration and validation set was 97.2% and 100%, respectively. The results indicate that NIR combined with chemometrics can be used as a simple and rapid tool to distinguish Shaoxing wines from non-Shaoxing wines and Fujian wines without the need for costly and laborious chemical and sensory analysis. However, the work reported here is just a preliminary study and more samples from other varieties and origins will be required to build a more robust model in a further study.

References

Aleixandre, M., Lozano, J., Gutierrez, J., Sayago, I., Fernandez, M. J., & Horrillo, M. C. (2008). Portable e-nose to classify different kinds of wine. *Sensors and Actuators B-Chemical, 131*(1), 71–76.

Andre, M. (2003). Multivariate analysis and classification of the chemical quality of 7-aminocephalosporanic acid using near-infrared reflectance spectroscopy. *Analytical Chemistry, 75*(14), 3460–3467.

Beltran, N. H., Duarte-Mermoud, M. A., Bustos, M. A., Salah, S. A., Loyola, E. A., Pena-Neira, A. I., et al. (2006). Feature extraction and classification of Chilean wines. *Journal of Food Engineering, 75*(1), 1–10.

Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *Trac-Trends in Analytical Chemistry, 21*(4), 240–250.

Chen, Y., Xie, M. Y., Yan, Y., Zhu, S. B., Nie, S. P., Li, C., et al. (2008). Discrimination of Ganoderma lucidum according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Analytica Chimica Acta, 618*(2), 121–130.

Cozzolino, D., Smyth, H. E., & Gishen, M. (2003). Feasibility study on the use of visible and near-infrared Spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry, 51*(26), 7703–7708.

Cozzolino, D., Smyth, H. E., Lattey, K. A., Cynkar, W., Janik, L., Dambergs, R. G., et al. (2005). Relationship between sensory analysis and near infrared spectroscopy in Australian Riesling and Chardonnay wines. *Analytica Chimica Acta, 539*(1–2), 341–348.

Cozzolino, D., Kwiatkowski, M. J., Dambergs, R. G., Cynkar, W. U., Janik, L. J., Skouroumounis, G., et al. (2008). Analysis of elements in wine using near infrared spectroscopy and partial least squares regression. *Talanta, 74*(4), 711–716.

Dambergs, R. G., Kambouris, A., Francis, I. L., & Gishen, M. (2002). Rapid analysis of methanol in grape-derived distillation products using near-infrared transmission spectroscopy. *Journal of Agricultural and Food Chemistry, 50*(11), 3079–3084.

Dixon, S. J., & Brereton, R. G. (2009). Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems, 95*(1), 1–17.

Fu, X., Zhou, Y., Ying, Y., Lu, H., & Xu, H. (2007). Discrimination of pear varieties using three classification methods based on near-infrared spectroscopy. *Transactions of the Asabe, 50*(4), 1355–1361.

Galgano, F., Favati, F., Camso, M., Scarpa, T., & Palma, A. (2008). Analysis of trace elements in southern Italian wines and their classification according to provenance. *Lwt-Food Science and Technology, 41*(10), 1808–1815.

Garde-Cerdan, T., Lorenzo, C., Alonso, G. L., & Salinas, M. R. (2010). Employment of near infrared spectroscopy to determine oak volatile compounds and ethylphenols in aged red wines. *Food Chemistry, 119*(2), 823–828.

Kruzlicova, D., Mocak, J., Balla, B., Petka, J., Farkova, M., & Havel, J. (2009). Classification of Slovak white wines using artificial neural networks and discriminant techniques. *Food Chemistry, 112*(4), 1046–1052.

Lafargue, M. E., Feinberg, M., Daudin, J. J., & Rutledge, D. N. (2003). Detection of heterogeneous wheat samples using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 11*(2), 109–121.

Lanza, E., & Li, B. W. (1984). Application for near-infrared spectroscopy for predicting the sugar content of fruit juices. *Journal of Food Science, 49*(4), 995–998.

Lin, P., Chen, Y. M., & He, Y. (2009). Identification of geographical origin of olive oil using visible and near-infrared spectroscopy technique combined with chemometrics. *Food and Bioprocess Technology*. doi:10.1007/s11947-009-0302-z.

Lister, S. J., Dhanoa, M. S., Stewart, J. L., & Gill, M. (2000). Classification and comparison of Gliricidia provenances using near infrared reflectance spectroscopy. *Animal Feed Science and Technology, 86*(3–4), 221–238.

Liu, F., & He, Y. (2009). Discrimination of producing areas of *Auricularia auricular* using visable/near infrared spectroscopy. *Food and Bioprocess Technology.*. doi:10.1007/s11947-00800174-7.

Liu, L., Cozzolino, D., Cynkar, W. U., Gishen, M., & Colby, C. B. (2006). Geographic classification of Spanish and Australian tempranillo red wines by visible and near-infrared spectroscopy combined with multivariate analysis. *Journal of Agricultural and Food Chemistry, 54*(18), 6754–6759.

Liu, L., Cozzolino, D., Cynkar, W. U., Dambergs, R. G., Janik, L., O'Neill, B. K., et al. (2008). Preliminary study on the application of visible-near infrared spectroscopy and chemometrics to classify Riesling wines from different countries. *Food Chemistry, 106*(2), 781–786.

Mao, L. C., Xu, Y., & Fan, W. L. (2010). Characterization of aroma compounds in Chinese rice wine Qu by solvent-assisted flavor evaporation and headspace solid-phase microextraction. *Journal of Agricultural and Food Chemistry, 58*(4), 2462–2469.

Niu, X. Y., Yu, H. Y., & Ying, Y. B. (2008). The Application of near-Infrared Spectroscopy and chemometrics to classify shaoxing wines from different breweries. *Transactions of the Asabe, 51*(4), 1371–1376.

Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and Gala apples. *Transactions of the Asae, 46*(6), 1721–1731.

Pontes, M. J. C., Santos, S. R. B., Araujo, M. C. U., Almeida, L. F., Lima, R. A. C., Gaiao, E. N., et al. (2006). Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International, 39*(2), 182–189.

Que, F., Mao, L. C., & Pan, X. (2006). Antioxidant activities of five Chinese rice wines and the involvement of phenolic compounds. *Food Research International, 39*(5), 581–587.

Que, F., Mao, L. C., Zhu, C. G., & Xie, G. F. (2006). Antioxidant properties of Chinese yellow wine, its concentrate and volatiles. *Lwt-Food Science and Technology, 39*(2), 111–117.

Saurina, J. (2010). Characterization of wines using compositional profiles and chemometrics. *Trac-Trends in Analytical Chemistry.* doi:10.1016/j.trac.2009.11.008.

Smyth, H. E., Cozzolino, D., Cynkar, W. U., Dambergs, R. G., Sefton, M., & Gishen, M. (2008). Near infrared spectroscopy as a rapid tool to measure volatile aroma compounds in Riesling wine: possibilities and limits. *Analytical and Bioanalytical Chemistry, 390*(7), 1911–1916.

Urbano-Cuadrado, M., de Castro, M. D. L., Perez-Juan, P. M., Garcia-Olmo, J., & Gomez-Nieto, M. A. (2004). Near infrared reflectance, spectroscopy and multivariate analysis in enology - determination or screening of fifteen parameters in different types of wines. *Analytica Chimica Acta, 527*(1), 81–88.

Wang, J. G., & Xu, L. (2005). Features of millet wine and its prospects. *Jiangsu Condiment Subsidiary Food, 22*(6), 5–9.

Wang, L., Lee, F. S. C., & Wang, X. R. (2007). Near-infrared spectroscopy for classification of licorice (Glycyrrhizia uralensis Fisch) and prediction of the glycyrrhizic acid (GA) content. *Lwt-Food Science and Technology, 40*(1), 83–88.

Woodcock, T., Fagan, C. C., O'Donnell, C. P., & Downey, G. (2008). Application of near and mid-Infrared spectroscopy to determine cheese quality and authenticity. *Food and Bioprocess Technology, 1*, 117–129. doi:10.1007/s11947-007-0033-y.

Xie, L. J., Ying, Y. B., Ying, T. J., Yu, H. Y., & Fu, X. P. (2007). Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Analytica Chimica Acta, 584*(2), 379–384.

Xu, L., Zhou, Y. P., Tang, L. J., Wu, H. L., Jiang, J. H., Shen, G. L., et al. (2008). Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. *Analytica Chimica Acta, 616*(2), 138–143.

Yu, H. Y., Ying, Y. B., Fu, X. P., & Lu, H. S. (2006). Quality determination of Chinese rice wine based on Fourier transform near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 14*(1), 37–44.

Yu, H. Y., Zhou, Y., Fu, X. P., Xie, L. J., & Ying, Y. B. (2007). Discrimination between Chinese rice wines of different geographical origins by NIRS and AAS. *European Food Research and Technology, 225*(3–4), 313–320.

Yu, H. Y., Lin, H. R., Xu, H. R., Ying, Y. B., Li, B. B., & Pan, X. X. (2008). Prediction of enological parameters and discrimination of rice wine age using least-squares support vector machines and near infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 56*(2), 307–313.