Determination of ethyl octanoate in Chinese liquor using FT-NIR spectroscopy

Liu, J., Dong, X., Han, S., Xie, A., Li, X., Li, P., Xu, B., and Luo, D.

1College of Food and Bioengineering, Henan University of Science and Technology, Luoyang, China
2Henan Engineering Research Centre of Food Material, Luoyang, China

Abstract

To quantitatively detect ethyl octanoate in Chinese liquor, Fourier-transform near-infrared (FT-NIR) spectroscopy was performed in the present work, with 162 Chinese liquor samples selected from Luoyang Dukang Distillery. The chemical values of ethyl octanoate were determined by gas chromatography (GC), and spectral data from 12,000 to 4000 cm⁻¹ were collected. The calibration model was established with partial least squares (PLS) regression, and then validated using internal cross-validation. The predictability of the model was further confirmed by the validation set as external validation. After comparing the effects of the models set up with sample data under different pre-processing methods, the model was built within the spectral region of 6101.7 - 5449.8 cm⁻¹; based on the SNV pre-processing method which was selected as the optimal model. The coefficient of determination (R²) for cross-validation of the model was 0.9507, and the corresponding root mean square errors of cross-validation (RMSECV) was 3.91 mg L⁻¹. The R² for external validation was 0.9537, and the root mean square errors of prediction (RMSEP) was 3.62 mg L⁻¹. The results demonstrated that using NIR spectroscopy to determine ethyl octanoate in Chinese liquor is feasible and can achieve satisfactory results.

Keywords

ethyl octanoate, Chinese liquor, FT-NIR spectroscopy, partial least squares

Introduction

Chinese liquor is a traditional spirit (Liu and Sun, 2018) distilled from sorghum, wheat, corn, rice, glutinous rice, and other grains; and goes through the processes of cooking, saccharifying, fermenting, distilling, storing, and blending. During liquor fermentation and distillation, a large number of flavour compounds including alcohols, acids, esters, phenols, ketones, acetals, nitrogenous compounds, and sulphur compounds are produced (Zheng et al., 2016). Although these compounds account for a small portion of liquor, they play an essential role in liquor flavour and quality. To produce high-quality liquor with appealing flavour and strong aroma, it is necessary to determine the most appropriate concentration ratio among these flavour compounds. Therefore, investigating flavour-producing substances plays a critical role in flavour type innovation and process improvement of liquor. Ethyl caproate is generally considered the main component of strong-aroma type liquor; its aroma intensity is much higher than that of ethyl acetate and ethyl lactate, and is second only to ethyl caproate. Therefore, it is important to achieve rapid determination of ethyl octanoate in Chinese liquor (Wang et al., 2014).

A typical method for analysing flavour compounds in liquor is gas chromatography (GC) or its combination with other techniques, including stir bar sorption extraction (SBSE) coupled with gas chromatography-mass spectrometry (GC-MS) (Fan et al., 2011), gas chromatography-olfactometry (GC-O) and GC-MS (Fan et al., 2012; Gao et al., 2014), solid-phase micro extraction (SPME) and GC-MS (Wang et al., 2015), SBSE, thermal desorption system (TDS), and GC-MS (Niu et al., 2015), GC-flame photometric detection (FPD) (Niu et al., 2017), and headspace (HS)-SPME and GC-pulsed flame photometric detection (FPFD) (Sha et al., 2016). Although these methods are accurate and sensitive, the implementation procedures are cumbersome and time-consuming, and they cannot be applied to rapid detection of liquor in the liquor industry. To meet actual production needs, a simple and fast online analysis technology is required. In such a case, near-infrared
spectroscopy (NIRS) is a suitable choice. Near infrared spectroscopy is a rapid, convenient, non-destructive, and environmentally friendly technology, widely used for simultaneous detection of multiple components in the food industry (Adedipe et al., 2016; Bernhard et al., 2016; Zhong and Qin, 2016). NIRS has also greatly contributed to the alcoholic beverage industry, where it is applied for rapid determination of the parameters in the fermentation process of alcoholic beverages (Grassi et al., 2014; Wu et al., 2015a; 2015b), for classification and identification of liquor (Fei et al., 2012; Chen et al., 2014; Li et al., 2014), and for analyses of conventional physical and chemical indicators of liquor (Lorenzo et al., 2009; Martelovidal and Vázquez, 2014; Ye et al., 2014; Ouyang et al., 2015). However, there have been few reports on the determination of ethyl octanoate in Chinese liquor by NIRS.

In the present work, Dukang liquor, a strong-aroma type liquor, was selected to study the content of ethyl octanoate, which was determined by GC-flame ionisation detector (FID). Then, the near-infrared (NIR) spectrum of liquor was scanned and analysed, and the mathematical relationship between the chemical value and spectrum was established by the partial least squares (PLS) method. The feasibility of rapid determination of ethyl octanoate in liquor by NIRS was explored.

**Materials and methods**

**Materials and chemicals**

A total of 162 Chinese liquor samples were obtained from Luoyang Dukang Holdings Ltd. (Henan province, China) as the original distillate. The samples were collected from different fermentation pools, and at different distillation phases in the liquor production line. All liquor samples were stored at 4°C until analysis. Ethyl octanoate of the GC standard (purity ≥ 99%) was purchased from Aladdin Bio-Chem Technology Co., Ltd (Shanghai, China), and absolute ethanol was from Kermel Chemical Reagent Co., Ltd (Tianjin, China).

Ethyl octanoate (0.1 mL) was transferred into a 100 mL volumetric flask and diluted with 600 mL L⁻¹ ethanol to obtain standard stock solution (1 mL L⁻¹). A volume of 10 mL of calibration standards with 600 mL L⁻¹ ethanol was prepared by the dilution of 0.1, 0.25, 0.5, 1, 2, and 3 mL standard stock solution at six gradient concentrations, respectively.

**Gas chromatography**

The GC analysis was carried out using an Agilent 7890A gas chromatography system, equipped with a flame ionisation detector (Agilent Technologies, Inc., USA). Separation was performed with a silica capillary column AT.LZP-930 (25 m × 0.53 mm × 1 μm; Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences). Nitrogen (99.999% purity) was employed as the carrier gas, and maintained at a flow rate of 30 mL min⁻¹, by being injected for a volume of 1 μL with a split ratio of 10:1. The injector and detector temperatures were set at 220 and 250°C, respectively. The chromatographic program was set at 75°C, held for 3 min, and then raised to 210°C at a step of 10°C min⁻¹. The total GC run time was 16.5 min.

**NIR spectrum acquisition**

All spectra from the liquor were acquired using a Fourier transform near-infrared (FT-NIR) spectrometer VECTOR33 (Bruker Corporation, Germany). Before measurements, the FT-NIR instrument was pre-heated for 30 min. After the instrument passed the test, air was used as the reference, and the quartz cuvette was selected with the optical path of 1 mm, and background scanning was carried out at first. Based on the characteristics of the liquor samples, the transmission mode was used to collect the spectra. The ambient temperature was equilibrated at 25 ± 2°C for spectrum acquisition. The spectral scanning range was 12,000 - 4,000 cm⁻¹, the instrument resolution was 8 cm⁻¹, and each sample was obtained by an average of 32 scans. The sample spectra were analysed using OPUS 8.1 software (Bruker Optics Inc., Germany).

**Spectrum processing**

Raw spectra of the samples should be pre-processed to reduce external interference, minimise baseline variation, and increase spectral differences. In order to achieve the highest accuracy and reliability of the prediction model, different pre-processing methods were applied, including straight line subtraction (SLS), standard normal variate (SNV), min-max normalisation (MMN), multiplicative scatter correction (MSC), first derivative (D1), and second derivative (D2) (Egidio et al., 2010). Besides, spectral region selection for optimal modelling has a profound impact on the quality of the multivariate calibration model. The NIR full-spectra data have a large volume and contain much interference information unrelated to the prediction target, and the characteristic spectral region needs to be filtered. The OPUS software has a set of five NIR frequency regions for the optimised models: 9,400 - 7,500, 7,500 - 6,100, 6,100 - 5,450, 5,450
By combining these pre-processing methods and spectral regions, the optimal band and best pre-processing method can be identified (Callado et al., 2018).

**Establishment of the models and evaluation standards**

Partial least squares (PLS) regression was utilised to establish a mathematical model for predicting the content of ethyl octanoate. The PLS regression is based on simultaneous decomposition of the spectral matrix (X) and concentration matrix (Y), which gradually extracts the components (usually referred to as PLS factors) from the data, and verifies the significance of the model until it meets the requirements. The PLS regression algorithm can use part or all of the spectra as variables to establish a quantitative analysis model of NIRS. PLS regression is the most commonly used multivariate calibration method in quantitative analysis, which overcomes the collinearity, band overlap, data interference, and other common problems. All samples were randomly divided into a calibration set and validation set at a ratio of 3:1. The calibration model was established using samples in the calibration set, where internal cross-validation was performed, and then external verification was performed with samples in the validation set. The optimal model was determined based on indicators such as the coefficient of determination ($R^2$), root mean square errors of cross-validation (RMSECV), and root mean square errors of prediction (RMSEP). A good model should have lower RMSECV and RMSEP, and higher $R^2$, with a small disparity between RMSECV and RMSEP (Aleixandre-Tudo et al., 2018). $R^2$, RMSECV, and RMSEP were calculated using Eq.1, Eq.2 and Eq.3, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$  \hspace{1cm} (Eq. 1)

where, $n =$ number of samples; $y_i =$ reference measurement value of sample $i$; and $\hat{y}_i =$ predicted value of sample $i$.

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$  \hspace{1cm} (Eq. 2)

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$  \hspace{1cm} (Eq. 3)

where, $n =$ number of samples in the validation set; $y_i =$ reference measurement value of sample $i$; and $\hat{y}_i =$ predicted value of sample $i$.

**Results and discussion**

**Chemical analysis**

The 162 liquor samples were analysed by GC-FID, and the analysis results of the distribution of the ethyl octanoate content in liquor are shown in Table 1. Since the liquor samples were from different fermentation pools, the ethyl octanoate content in the samples covered a relatively wide range. The range of the validation set was included in the larger range of the calibration set. In addition, different concentrations of ethyl octanoate were evenly distributed in the two data set, which was suitable for building a good NIR model.

**NIR spectra of Chinese liquor**

The original spectra of the liquor samples are shown in Figure 1(A). There were a total of 2025 variables (data points) with the wavenumbers ranging from 12,000 to 4000 cm$^{-1}$. The spectral curve could be explained by the overtones of different functional groups in the samples. The spectral changes of all the samples were basically the same, and no abnormal values could be found with naked eyes. Although PLS could tolerate full-band modelling, the models’ short-wavelength regions were relatively flat with no sufficient spectral information. If an excessively wide band range is selected, it contains much redundant information. This will damage the performance of the model, as it will increase the calculation time, and even reduce the prediction effect of the model. Therefore, avoiding short-wavelength regions is recommended for spectral region selection. The spectra had two large absorption bands at 6896 and 5128 cm$^{-1}$, which were ascribed to the O-H first overtone and O-H

<table>
<thead>
<tr>
<th>Subset</th>
<th>Number of sample</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration set</td>
<td>121</td>
<td>15.50 - 89.00</td>
<td>56.13</td>
<td>17.68</td>
</tr>
<tr>
<td>Validation set</td>
<td>41</td>
<td>24.16 - 88.83</td>
<td>54.29</td>
<td>17.05</td>
</tr>
</tbody>
</table>

SD: standard deviation.
combination band in water molecules, respectively. There was also a small absorption band at 10,416 cm⁻¹, which was related to the O-H second overtone (Chu et al., 2014; Wu et al., 2015c). In order to obtain a model with high stability and accuracy, this segment needs to be avoided in the process of selecting the optimal spectral regions. The absorption bands corresponding to ethyl octanoate appeared at 5900 cm⁻¹, correlated with the first overtone of the methyl (-CH₃) group; while at 5770 and 5670 cm⁻¹, it is associated with the first overtone of the methylene (-CH₂) group (Han et al., 2016). Therefore, this spectral region should be taken into consideration for establishing the NIR model of ethyl octanoate. Figure 1(B) illustrates the original spectra pre-processed with SNV, with a selected spectral region from 6101.7 to 5449.8 cm⁻¹, which removed irrelevant information, and improved the operation efficiency and model stability.

**Calibration and validation using NIR spectroscopy**

Based on spectral variables and chemical values, the NIR mathematical model was established using PLS regression. The optimal pre-processing method and optimal spectral regions were selected based on the "optimisation function" of the OPUS software. Table 2 lists the effects of eight models built with different pre-processing methods and varied spectral regions. It can be seen from the table that most models were set up in the spectral region of 6101.7 - 5449.8 cm⁻¹. This is consistent with the above NIR spectrum analysis results, which showed that 6101.7 - 5449.8 cm⁻¹ might be the characteristic spectral region of the ethyl octanoate model, and could achieve good results by directly using the original spectrum variables (none pre-processing) to participate in modelling. The model built after pre-processing by SNV, MSC, and MMN could achieve better evaluation results than the original spectrum model, because these pre-processing methods could eliminate spectral differences between samples caused by scattering. However, the modelling effect after derivative pre-processing (D1 or D2 pre-processing) was not so good as the effect without pre-processing, as it might be attributed to signals that were not amplified, but spectral noises increased after derivative pre-processing. By comparison, the pre-processing method selected in the final modelling was SNV, and the optimal spectral region was 6101.7 - 5449.8 cm⁻¹.

The number of factors that participated in regression should be properly selected for constructing a PLS model. If there are too many or too few PLS factors, the calibration model may not accurately depict the behaviours of the components. The optimal number of PLS factors for each model was determined by internal cross-validation using the calibration model. A total of 20 PLS factors were introduced into the PLS model for factor optimisation. Figure 2 presents how $R^2$ and RMSECV of the calibration model built with the SNV pre-processing method interact with increasing PLS factors. The most ideal curve is that RMSECV first decreases rapidly with the increase in the number of PLS factors, the minimum value is reached, and then RMSECV increases slightly. Usually, the number of PLS factors with the smallest RMSECV is chosen. The relationship between $R^2$ and the number of PLS factors was just opposite to the relationship between RMSECV and the number of PLS factors (reverse curve). The ideal situation was to increase it with the increase in the number of PLS factors at first, and then gradually decrease after the maximum value appeared (Moser et al., 2015).
Therefore, it can be seen from Figure 2 that the optimal number of PLS factors should be set as 11 in the calibration model with the SNV pre-processing method.

To illustrate the robustness of the model, the predicted values were plotted against the measured values. Figure 3(A) reveals the correlation between the predicted values and measured values obtained after the calibration set samples were internally cross-validated. It can be observed that the calibration set samples were evenly distributed on both sides of the regression line, indicating that the NIR spectra of ethyl octanoate had a strong linear correlation with its chemical values. $R^2$ of the model was 0.9507, and RMSECV was 3.91 mg L$^{-1}$. These results showed that the mathematical statistics of the model reached the most expected value, with $R^2$ being 0.9507 and RMSECV being 3.91 mg L$^{-1}$.

<table>
<thead>
<tr>
<th>Pre-processing method</th>
<th>Spectral region</th>
<th>Number of PLS factor</th>
<th>Calibration set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>6101.7 - 5449.8</td>
<td>17</td>
<td>0.9397</td>
<td>0.9475</td>
</tr>
<tr>
<td>SNV</td>
<td>6101.7 - 5449.8</td>
<td>11</td>
<td>0.9507</td>
<td>0.9537</td>
</tr>
<tr>
<td>MSC</td>
<td>6101.7 - 5449.8</td>
<td>11</td>
<td>0.9496</td>
<td>0.9525</td>
</tr>
<tr>
<td>MMN</td>
<td>6101.7 - 5449.8</td>
<td>11</td>
<td>0.9492</td>
<td>0.9520</td>
</tr>
<tr>
<td>SLS</td>
<td>7501.7 - 5449.8</td>
<td>16</td>
<td>0.9124</td>
<td>0.9421</td>
</tr>
<tr>
<td>D1</td>
<td>6101.7 - 5449.8</td>
<td>16</td>
<td>0.9388</td>
<td>0.9197</td>
</tr>
<tr>
<td>D2</td>
<td>7501.7 - 5449.8</td>
<td>15</td>
<td>0.8592</td>
<td>0.9001</td>
</tr>
<tr>
<td>D1+SNV</td>
<td>9403.2 - 7497.9</td>
<td>14</td>
<td>0.8956</td>
<td>0.9177</td>
</tr>
</tbody>
</table>

Table 2. Results of NIR models with different pre-processing methods and spectral regions.

SNV: standard normal variate; MSC: multiplicative scatter correction; MMN: min-max normalisation; SLS: straight line subtraction; D1: first derivative; D2: second derivative.

![Figure 2. Results of calibration models based on different numbers of PLS factors.](image)

Therefore, it can be seen from Figure 2 that the optimal number of PLS factors should be set as 11 in the calibration model with the SNV pre-processing method.

![Figure 3. Correlation between predicted values and measured values of ethyl octanoate in the calibration set (A) and validation set (B).](image)
close to 1, and RMSECV being small.

The regression coefficients of the optimal PLS model for ethyl octanoate were plotted in Figure 4. There were strong peaks and troughs for the selected effective wavenumbers (6101.7 - 5449.8 cm\(^{-1}\)). As can be seen, the correlation of the chemical values measured by GC-FID with those estimated for ethyl octanoate by the calibration model was excellent. It was found that a high regression coefficient was observed at the wavenumber of 5900 cm\(^{-1}\), which might be linked to the first overtone of \(-\text{CH}_3\), while the wavenumbers of 5770 and 5670 cm\(^{-1}\) could be linked to the first overtone of \(-\text{CH}_2\). These conclusions are consistent with the analysis results of the original spectra of the liquor samples. The results demonstrated that NIR spectroscopy provided a good prediction model for the ethyl octanoate content in Chinese liquor samples.

After the calibration model was established, external validation by the validation set was required to ensure the applicability of the model. Figure 3(B) shows the correlation of the predicted value with the measured value by external validation. \(R^2\) of the validation set was 0.9537, and RMSEP was 3.62 mg L\(^{-1}\). The correlation and deviation between the NIRS results in of all the samples, and the results of the reference methods met the reproducibility requirements. The predicted value was basically consistent with the chemical value measured by GC-FID, and the prediction effect of the model was satisfactory, which could meet the test requirements in the production of the liquor industry.

Conclusion

In the present work, 162 Chinese liquor samples were scanned by FT-NIR spectroscopy, and the contents of ethyl octanoate were determined by GC-FID to establish a NIR model for rapid detection of ethyl octanoate in Chinese liquor. The experimental results confirmed the FT-NIR capability to predict the content of ethyl octanoate in Chinese liquor. When compared with traditional GC methods, the NIR technique is a non-destructive and rapid method with no need for previous treatment of samples, and having no contamination issues. It is an ideal rapid detection technology suitable for online quality control of industrial liquor production.

Acknowledgement

The present work was financially supported by the National Natural Science Foundation of China (Grant No. 31471658), Natural Science Foundation of Henan Province, China (No. 162300410074), Key Scientific and Technological Research Project of Henan Province, China (No. 182102110300 and 172102310694), and Key Science and Technology Program of Educational Commission of Henan Province, China (No. 19A550001).

References


Wang, X., Fan, W. and Xu, Y. 2014. Comparison on aroma compounds in Chinese soy sauce and strong aroma type liquors by gas chromatography–olfactometry, chemical quantitative and
odor activity values analysis. European Food Research and Technology 239(5): 813-825.


Ye, M., Yue, T., Yuan, Y. and Zhao, L. 2014. Application of FT-NIR spectroscopy to apple wine for rapid simultaneous determination of soluble solids content, pH, total acidity, and total ester content. Food and Bioprocess Technology 7(10): 3055-3062.
