RESEARCH ARTICLE

Rapid and simultaneous analysis of multiple wine quality indicators through near-infrared spectroscopy with twice optimization for wavelength model

Jiemei CHEN[1], Sixia LIAO[1], Lijun YAO[2], Tao PAN (✉)[2]

1 Department of Biological Engineering, Jinan University, Guangzhou 510632, China
2 Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China

© Higher Education Press 2020

Abstract Alcohol, total sugar, total acid, and total phenol contents are the main indicators of wine quality detection. This study aims to establish simultaneous analysis models for the four indicators through near-infrared (NIR) spectroscopy with wavelength optimization. A Norris derivative filter (NDF) platform with multiparameter optimization was established for spectral pretreatment. The optimal parameters (i.e., derivative order, number of smoothing points, and number of differential gaps) were (2, 9, 3) for alcohol, (1, 19, 5) for total sugar, (1, 17, 11) for total acid, and (1, 1, 1) for total phenol. The equidistant combination-partial least squares (EC-PLS) was used for large-scale wavelength screening. The wavelength step-by-step phase-out PLS (WSP-PLS) and exhaustive methods were used for secondary optimization. The final optimization models for the four indicators included 7, 10, 15, and 13 wavelengths located in the overtone or combination regions, respectively. In an independent validation, the root mean square errors, correlation coefficient for prediction (i.e., SEP and R_P), and ratio of performance-to-deviation (RPD) were 0.41 v/v, 0.947, and 3.2 for alcohol; 1.48 g/L, 0.992, and 6.8 for total sugar; 0.68 g/L, 0.981, and 5.1 for total acid; and 0.181 g/L, 0.948, and 2.9 for total phenol. The results indicate high correlation, low error, and good overall prediction performance. Consequently, the established reagent-free NIR analytical models are important in the rapid and real-time quality detection of the wine fermentation process and finished products. The proposed wavelength models provide a valuable reference for designing small dedicated instruments.

Keywords wine, quality indicators, near-infrared (NIR) spectroscopy, Norris derivative filter (NDF) platform, wavelength model optimization

1 Introduction

Wine is an alcoholic beverage with mild alcohol content, diversified taste, and high popularity among consumers. The appropriate amount of wine is good for one's health and can reduce the risk of coronary heart disease and atherosclerosis [1]. Vinification includes the selection of raw grapes, juice extraction, and alcohol fermentation [2].

Alcohol and sugar contents are the most basic indicators of wine used to characterize its unique taste and odor. As a substrate, sugar is gradually consumed under the action of *Saccharomyces cerevisiae* during the fermentation process. As a product, alcohol is gradually increased [3]. Finished wine usually requires the control of the alcohol and sugar contents to keep the concentrations stable within a certain range. In addition, acid is derived from tartaric and malic acid in grapes and succinic, lactic, and acetic acid produced during fermentation. Moderate acid can promote appetite, help digestion, and benefit human health. The polyphenols in red wine can eliminate free radicals in the human body; hence, health effects (e.g., antioxidation and lowering of blood lipids) are observed. The astringency and the ruby color of red wine are closely related to polyphenols. The total phenol is the total amount of polyphenols.

Therefore, the alcohol, total sugar, total acid, and total phenol contents are the main indicators for monitoring the wine fermentation process and detecting the product quality. Distillation, neutralization, and colorimetric reactions are traditional chemical analysis methods for the abovementioned indicators [4–6]. These methods require sample preparation and are highly specialized and time consuming. In addition, different indicators require different measurement methods and reagents. Thus, they are unsuitable for the rapid and real-time quality detection of the wine fermentation process and finished wine.

Near-infrared (NIR) spectroscopy is based on the nonresonant molecular vibration associated with transitions

from the ground state to the high-energy level. It primarily reflects the overtone absorption and vibration combinations of hydrogen-containing (X-H) functional groups. NIR spectroscopy can directly measure multiple indicators in a sample and features a fast, real-time, and online measurement [7]. NIR spectroscopy has also been effectively used in numerous fields, such as the agricultural [8–11], food [12–15], environmental [16,17], and biomedical fields [18–24].

NIR spectroscopy combined with partial least squares (PLS) regression has been applied to the rapid quantitative analysis of wine quality indicators. Several preliminary works involve the indicators of alcohol, total sugar, total acid, and total phenol in wine [5,25–27]. Reference [25] shows a relevant comparison and jointly used NIR and mid infrared (MIR) spectroscopy combined with the PLS method to perform a quantitative analysis of the seven indicators of alcoholic degree, volumic mass, total acidity, glycerol, total polyphenol index, lactic acid, and total sulfur dioxide in the finished wine. Reference [5] also used NIR and MIR spectroscopy combined with the PLS method to quantitatively analyze the sugar content, ethanol, glycerol, and phenolic compounds of the fermentation sample of wine. Meanwhile, Ref. [26] used NIR spectroscopy combined with principal component analysis (PCA) and PLS methods to perform a cluster analysis of red wine from different grape varieties. Reference [27] used UV–Vis–NIR spectroscopy combined with principal component regression (PCR) and PLS methods to determine seven different phenolic compounds in red wine. These works investigated the feasibility of NIR spectroscopy to rapidly measure the quality indicators of finished wine or wine fermentation samples. However, studies on the methods of spectral modeling and prediction accuracy, especially proper spectral preprocessing and wavelength model optimization methods, must be further performed. The integrated application research of related chemometric methods will help improve the detection effect of NIR spectroscopy in the field of wine processing. The wavelength selection of the NIR analysis model can avoid noise interference, extract useful information, improve analysis accuracy, and provide a required reference for designing dedicated instruments.

In this study, the calibration–prediction models for the rapid and simultaneous quantitative analysis of alcohol, total sugar, total acid, and total phenol in wine are established with the NIR spectra. In addition, several novel chemometric methods are integrated to apply further modeling optimization. Samples of various commercial and homemade wines with wide ranges of total sugar and acid were adopted to improve the results' representativeness.

First, the famous Norris derivative filter (NDF) [28,29], which is an algorithm group with various parameters, is used to pretreat the wine spectra and improve the spectral signal-to-noise ratio. The appropriate Norris parameters cannot be preset on based on experience, but are selected in accordance with the modeling effect [24,30]. A large-scale parameter optimization platform for the NDF algorithm combined with PLS is established herein based on the modeling effect. The optimal NDF parameters for the NIR analysis of the four wine quality indicators are then determined.

Wavelength model optimization is another core chemometric method that plays a crucial role in the NIR spectral analysis application. The recently proposed equidistant combination PLS (EC-PLS) method [11,20] considers the advantages of continuous and discrete wavelength models. The initial wavelength, wavelength number, and wavelength gap number were adopted as the cyclic parameters for the quasicontinuous wavelength combination selection. EC-PLS optimized not only the wavelength position and number, but also the wavelength gaps. In addition, it has been applied in various objects [11,20,21,31–33]. The EC-PLS model set properly includes that of the famous moving-window PLS (MW-PLS) [19,20,22,34]. Therefore, it is a strict extension from an algorithm point of view. Here, EC-PLS was also used to establish the wavelength model for the NIR analysis of the four indicators.

NIR spectroscopy involves data for hundreds or thousands of wavelengths. At present, the scientific computing power cannot exhaustively optimize any wavelength combination. Thus, the wavelength combination obtained by any strategy must be subjected to a secondary optimization. Redundant wavelengths are difficult to avoid in the equidistant wavelength combination; thus, the wavelength step-by-step phase-out PLS (WSP-PLS) was proposed to correct the EC-PLS models [24,35]. The WSP-PLS and exhaustive methods were used herein as the secondary optimization for correcting the EC-PLS models of the four indicators.

2   Materials and methods

2.1   Samples

A total of 52 bottles of finished red wines covering 21 commercial brands were purchased, and 49 bottles of homemade red wines were collected. After placing the bottles at room temperature for a sufficient amount of time, a small amount of wine (i.e., approximately 2 mL) was taken from each bottle as a sample for the spectral measurement. Accordingly, 101 wine samples were obtained. The spectrum of the double distilled water was also measured for a spectral comparison. The appropriate amount of wine sample was taken from each bottle as a sample for the measurements of alcohol, total sugar, total acid, and total phenol through standard chemical analysis methods.

2.2 Analysis of reference values for four wine quality indicators

The alcohol content of each wine sample was analyzed in accordance with the standard method (alcohol meter method, GB/T 15038-2006) [4]. The method removed nonvolatile substances in each wine sample by distillation, measured the alcohol volume percentage value of the distillate with the alcohol meter, and corrected the temperature to obtain the sample's alcohol content.

The total sugar content was determined with 3,5-dinitrosalicylic acid (DNS) colorimetry [6]. Each wine sample was hydrolyzed with hydrochloric acid at a constant temperature and neutralized with a NaOH solution. The color reaction was performed using a DNS reagent. Absorbance was measured by an ultraviolet spectrophotometer. Subsequently, the result was used to calculate the total sugar content by comparing it with the standard curve of the glucose solution.

The total acid concentration was analyzed in accordance with the standard method (indicator method, GB/T 15038-2006) [4], where phenolphthalein was used as an indicator. Acid–base titration was performed with the NaOH standard solution. The total acid content can be calculated according to the consumption amount of the NaOH solution.

The total phenol concentration was determined through the Folin–Ciocalteau method [5,25,26]. The color reaction was performed by adding the Folin–Ciocalteau reagent to each wine sample under alkaline conditions. The absorbance was measured with an ultraviolet spectrophotometer. The results were used to calculate the total phenolic content by comparing it with the standard curve of pure gallic acid.

All of the abovementioned measurements were performed in triplicate and averaged. The average measured values were used as the reference for the modeling and validation of the NIR spectroscopic analysis. Table 1 presents the statistical analysis of the actual values of alcohol, total sugar, total acid, and total phenol of the 101 samples.

2.3 Near-infrared spectra acquisition

The instrument used herein was an XDS Rapid Content™

Table 1 Statistical analysis of the actual alcohol, total sugar, total acid, and total phenol values of all wine samples

indicator	min	max	mean	SD
alcohol/(v·v^{-1})	10.4	15.5	12.4	1.2
total sugar/(g·L^{-1})	1.0	55.9	7.0	10.5
total acid/(g·L^{-1})	4.1	16.4	7.6	3.9
total phenol/(g·L^{-1})	0.47	3.15	1.72	0.59

Note: SD, standard deviation.

Liquid Grating Spectrometer (FOSS, Denmark) with 1 mm cuvette. The spectra were acquired over 780–2498 nm with 2 nm wavelength gap and covered the whole NIR region. Si and PbS detectors were used to detect the 780–1100 and 1100–2498 nm wavebands, respectively. Every sample was measured thrice. The average spectra were then used. The spectral measurement was conducted at $(25\pm1)°C$ and $(46\pm1)\%$ relative humidity.

2.4 Calibration–prediction–validation based on multiple sample partitioning

Given the randomness of a sample, the sample partitioning differences may result in a parameter fluctuation. The modeling samples were randomly divided for multiple times into calibration–prediction sets. In addition, the PLS model for each division was established. The parameters were then optimized based on the comprehensive modeling effect of all divisions to ensure model stability and objectivity.

First, 101 samples were randomly divided into the modeling (71 samples) and validation (30 samples) sets. The modeling set was further divided for 20 times into the calibration (41 samples) and prediction (30 samples) sets.

The root mean square error and the correlation coefficient for prediction for each division i were calculated and denoted as SEP_i and $R_{P,i}$, respectively, where $i = 1, 2,…, 20$. The mean values (SEP_{Ave} and $R_{P,Ave}$) and the standard deviations (SEP_{SD} and $R_{P,SD}$) of all the divisions were further calculated. The comprehensive indicator $SEP^+ = SEP_{Ave} + SEP_{SD}$ for the prediction accuracy and stability was used to determine the modeling parameters.

Finally, the selected models were validated using independent validation samples not used in the modeling. The corresponding SEP, R_P, and ratio of the performance-to-deviation (RPD, where $RPD = \dfrac{SD}{SEP}$) were further determined. A high RPD value represented a good overall predicted performance. The SDs of the actual alcohol, total sugar, total acid, and total phenol values for the 30 validation samples were 1.3 v/v, 10.1, 3.5, and 0.53 g/L, respectively. Figure 1 shows the schematic of the calibration–prediction–validation process with sample multi-partitioning and evaluation indicators.
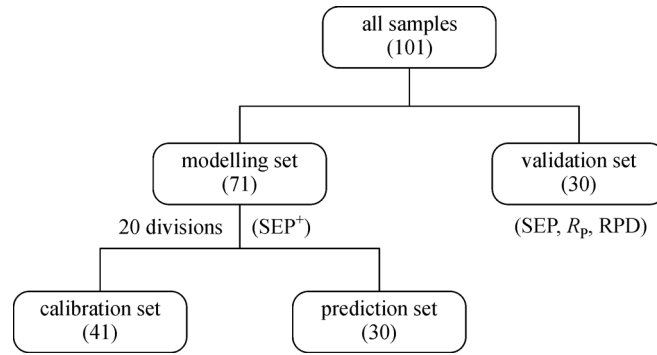
2.5 Norris derivative filter

The NDF algorithm included two steps, namely moving average smoothing and differential derivation, which used the parameter derivative order (d), number of smoothing points (s), and number of differential gaps (g). The NDF is an algorithm group with various parameters and modes for spectral preprocessing [24,30].

Each set of NDF spectra was used to build a PLS model,

Fig. 1 Calibration–prediction–validation process with sample multi-partitioning

called the Norris-PLS model. The global optimization selection for the NDF modes was performed according to the predicted effect (SEP$^+$) of the PLS model. The number of PLS latent variables (LV) was set as LV $\in \{1, 2, ..., 15\}$. Parameters d, s, and g were set as $d = 0, 1, 2$, $s \in \{1, 3, ..., 31\}$, and $g \in \{1, 2, ..., 30\}$, respectively. A total of 976 NDF modes were obtained on the basis of all parameter combinations of (d, s, and g). The optimal Norris parameters are then selected as follows in accordance to the predicted effect:

$$\mathrm{SEP}_*^+ = \min_{\substack{d \in \{0, 1, 2\} \\ s \in \{1, 3, ..., 31\} \\ g \in \{1, 2, ..., 30\}}} \mathrm{SEP}^+(d, s, g). \tag{1}$$

s and g are important Norris parameters. s was the only variable parameter available when $d = 0$. The prediction effect is denoted as SEP$^+$ (0, s). Two variable parameters (s and g) were used when $d = 1, 2$. The prediction effects of the local optimal models for all single parameters are as follows:

$$\mathrm{SEP}^+(s) = \min_{\substack{g \in \{1, 2, ..., 30\} \\ d \in \{0, 1, 2\}}} \mathrm{SEP}^+(d, s, g), s = 1, 3, ..., 31, \tag{2}$$

$$\mathrm{SEP}^+(g) = \min_{\substack{s \in \{1, 3, ..., 31\} \\ d \in \{1, 2\}}} \mathrm{SEP}^+(d, s, g), g = 1, 2, ..., 30. \tag{3}$$

2.6 Equidistant combination-partial least squares method

The EC-PLS method used all equidistant wavelength models in a specific wavelength range to establish PLS models, which adopted the initial wavelength (I), number of wavelengths (N), number of wavelength gaps (G), and LV as the cyclic parameters. Please see Refs. [31] and [32] for a detailed description. The global optimal parameters were selected as follows in accordance with the predicted effect (SEP$^+$):

$$\mathrm{SEP}_*^+ = \min_{I, N, G, \mathrm{LV}} \mathrm{SEP}^+(I, N, G, \mathrm{LV}). \tag{4}$$

The choice of the wavelength model may be limited by practical conditions. Therefore, in addition to the global optimal wavelength model, the local optimal model corresponding to each single parameter also has a practical value. The corresponding local optimal model for every fixed single parameter I (N or G) was selected according to the following equations:

$$\mathrm{SEP}^+(I) = \min_{N, G, \mathrm{LV}} \mathrm{SEP}^+(I, N, G, \mathrm{LV}), \tag{5}$$

$$\mathrm{SEP}^+(N) = \min_{I, G, \mathrm{LV}} \mathrm{SEP}^+(I, N, G, \mathrm{LV}), \tag{6}$$

$$\mathrm{SEP}^+(G) = \min_{I, N, \mathrm{LV}} \mathrm{SEP}^+(I, N, G, \mathrm{LV}). \tag{7}$$

The spectral region of 400–2498 nm was used for EC-PLS screening. Parameters I, N, G, and LV were set as $I \in \{780, 782, ..., 2498\}$, $N \in \{1, 2, ..., 200\}$, $G \in \{1, 2, ..., 20\}$, and LV $\in \{1, 2, ..., 15\}$, respectively.

2.7 Secondary wavelength optimization

The WSP-PLS method can be used to correct any continuous or discrete wavelength model with N wavelengths through the following steps: first, wavelengths were subjected to a backward elimination, that is, the lowest prediction error was obtained each time a wavelength was eliminated and until only one wavelength remained; and second, the optimal model was selected through the WSP model. Please see Refs. [24,33,35] for a detailed description.

Notably, if the exhaustive method is used to optimize any one wavelength combination with N wavelengths, the $2^N - 1$ PLS models must be calculated. On the contrary, the WSP-PLS algorithm only required $N(N + 1)/2$ operations. At a large N, the exhaustive method cannot be implemented given its heavy calculation load. However, the amount of WSP-PLS calculation is still moderate and can be implemented quickly. Here, the WSP-PLS and exhaustive methods were used for the secondary optimiza-

tion when $N \leqslant 17$ for the EC-PLS model ($2^N - 1 \leqslant 131071$). When $N > 17$, the WSP-PLS method was first used for the secondary optimization, and the exhaustive method was then used for further optimization.

The computer algorithms for all the above-mentioned methods were designed using MATLAB version 7.6 software.

3   Results and discussion

3.1   Optimization of Norris parameters

Figure 2(a) depicts the NIR spectra of the 101 wine samples over the whole scanning region (780–2498 nm). Figure 2(b) also illustrates the average spectrum of all the wine samples and the double distilled water spectrum for a comparison. The baseline drift and tilt were observed near the two water absorption peaks (i.e., 1400–1500 and 1850–2050 nm) of the NIR overtone region. In addition, the wine and water spectra in waveband 2200–2350 nm of the NIR combination region showed differences.

The analyses of the alcohol, total sugar, total acid, and total phenol contents of the wine samples were independently modeled. First, the full PLS models were established based on the whole NIR region (780–2498 nm, $N = 860$). Table 2 summarizes the optimal parameters (e.g., LV) and the prediction effects (i.e., SEP_{Ave}, $R_{P,Ave}$, SEP_{SD},

$R_{P,SD}$, and SEP^+) for the four indicators. The SEP^+ values of the four indicators were 0.57 v/v, 2.46, 2.41, and 0.253 g/L, respectively.

The spectra were then preprocessed using the NDF method, and the corresponding 976 Norris-PLS models were established. Figures 3–6 show the prediction effects ($SEP^+(s)$ and $SEP^+(g)$) of the local optimal models for each s and g for the four wine indicators. A significant difference in the modeling effects, which corresponded to different Norris-PLS models, was observed. The Norris parameters cannot be selected on the basis of experience and must be optimized.

The results also imply that the global optimal parameters were $d = 2$, $s = 9$, and $g = 3$ for alcohol; $d = 1$, $s = 19$, and $g = 5$ for total sugar; $d = 1$, $s = 17$, and $g = 11$ for total acid; and $d = 1$, $s = 1$, and $g = 1$ for total phenol. The corresponding SEP^+ were 0.47 v/v, 2.22, 1.43, and 0.238 g/L, respectively. Table 3 summarizes the optimal parameters (d, s, g, and LV) and the prediction effects. In comparison with the results in Table 2, SEP^+ of alcohol, total sugar, total acid, and total phenol decreased by 17.5%, 9.8%, 40.7%, and 5.9%, respectively. The optimal Norris-PLS models were significantly better than the full PLS models without pretreatment. Therefore, the wine spectrum pretreatment was necessary and can improve the prediction effect of the spectra.

Notably, the Norris parameters selected according to the modeling effects were not the same for different indicators.
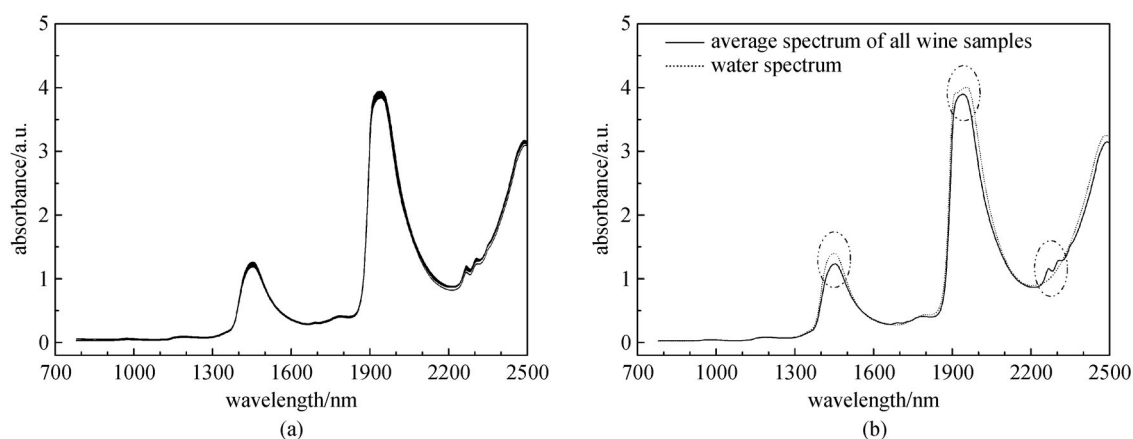

Fig. 2   NIR spectra of wine and water samples. (a) Spectra of all wine samples. (b) Average spectrum of all wine samples and water spectrum

Table 2   Parameters and prediction effects of the full PLS models for the four wine indicators

| indicator	LV	SEP_{Ave}	SEP_{SD}	$R_{P,Ave}$	$R_{P,SD}$	SEP^+
alcohol/(v·v^{-1})	7	0.50	0.07	0.904	0.036	0.57
total sugar/(g·L^{-1})	8	2.17	0.29	0.981	0.006	2.46
total acid/(g·L^{-1})	5	2.09	0.32	0.876	0.038	2.41
total phenol/(g·L^{-1})	3	0.231	0.022	0.933	0.015	0.253
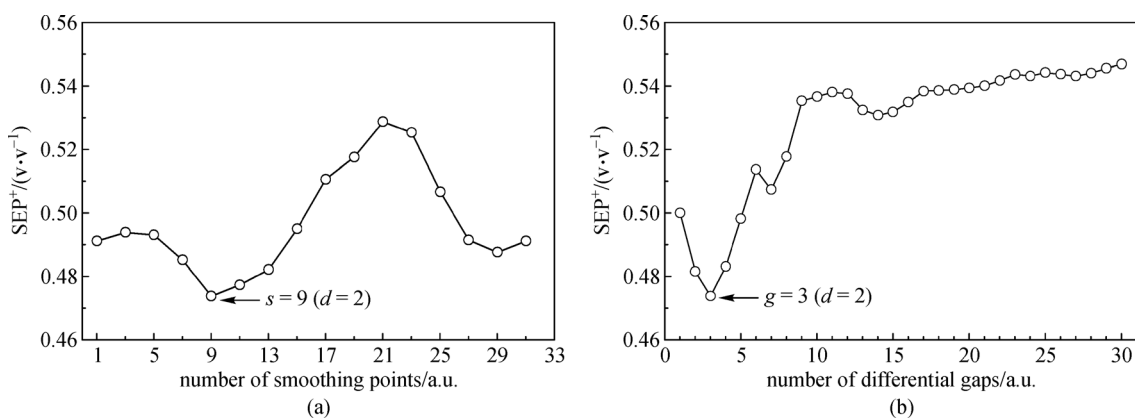
Fig. 3 SEP$^+$ of the local optimal Norris-PLS model (alcohol) for each parameter. (a) Number of smoothing points. (b) Number of differential gaps
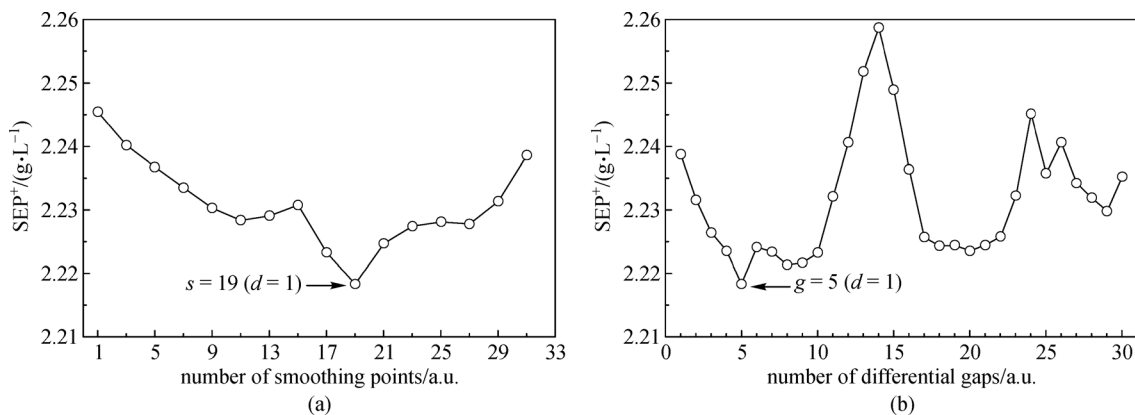
Fig. 4 SEP$^+$ of the local optimal Norris-PLS model (total sugar) for each parameter. (a) Number of smoothing points. (b) Number of differential gaps
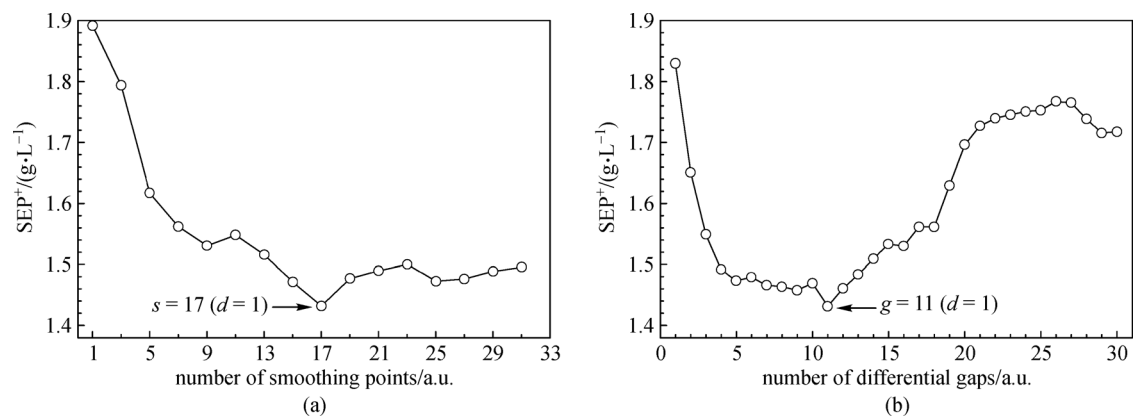
Fig. 5 SEP$^+$ of the local optimal Norris-PLS model (total acid) for each parameter. (a) Number of smoothing points. (b) Number of differential gaps
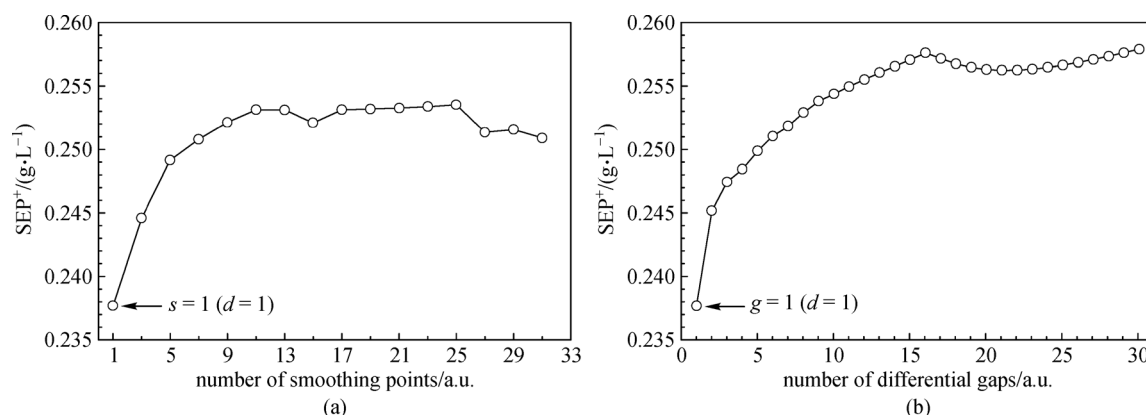
Fig. 6 SEP$^+$ of the local optimal Norris-PLS model (total phenol) for each parameter. (a) Number of smoothing points. (b) Number of differential gaps

Table 3 Parameters and prediction effects of the optimal Norris-PLS models for the four wine indicators

indicator	d	s	g	LV	SEP$_{Ave}$	SEP$_{SD}$	$R_{P,Ave}$	$R_{P,SD}$	SEP$^+$
alcohol/(v·v^{-1})	2	9	3	3	0.43	0.04	0.931	0.017	0.47
total sugar/(g·L^{-1})	1	19	5	6	1.99	0.23	0.984	0.004	2.22
total acid/(g·L^{-1})	1	17	11	15	1.23	0.20	0.960	0.012	1.43
total phenol/(g·L^{-1})	1	1	1	2	0.216	0.022	0.941	0.016	0.238

Figure 7 illustrates the Norris derivative spectra of all wine samples based on the optimal Norris parameters of the four indicators.

3.2 Optimal equidistant combination-partial least squares models

Further wavelength model optimization was performed using the abovementioned EC-PLS method based on the NDF spectra. The optimal EC-PLS models for the four indicators were selected. The obtained optimal parameters of I, N, G, and LV were 2158 nm, 11, 1, and 5 for alcohol; 1642 nm, 15, 5, and 12 for total sugar; 1618 nm, 19, 6, and 12 for total acid; and 1530 nm, 24, 4, and 7 for total phenol. The corresponding SEP$^+$ was further reduced to 0.40 v/v, 1.68, 0.53, and 0.175 g/L, respectively. Table 4 summarizes the optimal parameters and prediction effects. In comparison with the Norris-PLS model of the whole NIR region (780–2498 nm, $N = 860$), the SEP$^+$ values of the four indicators significantly decreased by 14.9%, 24.3%, 62.9%, and 26.5%, respectively. Moreover, the number of adopted wavelengths (N) greatly decreased to 11, 15, 19, and 24, respectively. The results show that the wavelength models were considerably simplified.

Figure 8 depicts the prediction effects (SEP$^+$(I)) of the local optimal models for all I for the four indicators. Different modeling effects were observed for the different wavelength positions.

3.3 Secondary optimization models

The WSP-PLS method was used to remove unavoidably redundant wavelengths in the optimal EC-PLS model and further improve the spectral prediction effect. Table 5 summarizes the optimal parameters (N, LV) and the prediction effects for the obtained optimal EC-WSP-PLS models of the four indicators.

In comparison with those of the optimal EC-PLS models, SEP$^+$ of the optimal EC-WSP-PLS models improved for the four indicators. Moreover, the adopted N for alcohol, total sugar, total acid, and total phenol greatly decreased to 7, 10, 15, and 17, respectively. Therefore, the redundant wavelengths must be eliminated through the WSP-PLS method.

The corresponding wavelength combinations for alcohol, total sugar, total acid, and total phenol were 2158, 2160, 2162, 2168, 2170, 2174, and 2178 nm of the combination region; 1642, 1652, 1672, 1682, 1712, 1732, 1742, 1762, 1772, and 1782 nm of the overtone region; 1618, 1642, 1654, 1690, 1702, 1714, 1738, 1750, 1762, 1774, 1786, 1798, 1810, 1822, and 1834 nm of the overtone region; and 1538, 1546, 1562, 1578, 1594, 1602, 1626, 1642, 1650, 1658, 1666, 1674, 1682, 1690, 1698, 1706, and 1714 nm of the overtone region, respectively.

The N values of the optimal EC-PLS models for alcohol and total sugar were 11 and 15 ($N < 17$), respectively. Section 2.7 showed that the WSP-PLS and exhaustive
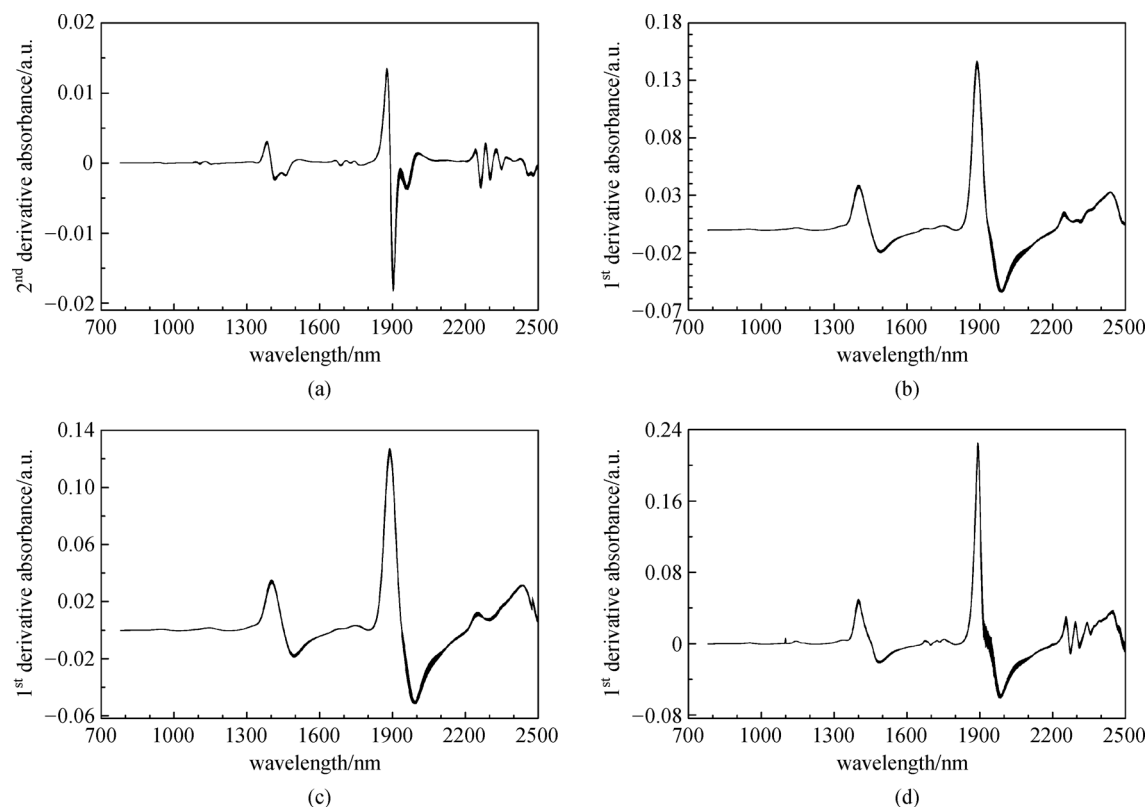
Fig. 7　NDF spectra of all wine samples based on the optimal Norris parameters. (a) $d = 2$, $s = 9$, and $g = 3$ for alcohol. (b) $d = 1$, $s = 19$, and $g = 5$ for total sugar. (c) $d = 1$, $s = 17$, and $g = 11$ for total acid. (d) $d = 1$, $s = 1$, and $g = 1$ for total phenol

Table 4　Parameters and prediction effects of the optimal EC-PLS models for the four wine indicators

indicator	I/nm	N	G	LV	SEP_{Ave}	SEP_{SD}	$R_{P,Ave}$	$R_{P,SD}$	SEP^{+}
alcohol/(v·v^{-1})	2158	11	1	5	0.38	0.02	0.950	0.008	0.40
total sugar/(g·L^{-1})	1642	15	5	12	1.46	0.22	0.992	0.003	1.68
total acid/(g·L^{-1})	1618	19	6	12	0.48	0.05	0.994	0.001	0.53
total phenol/(g·L^{-1})	1530	24	4	7	0.157	0.018	0.969	0.008	0.175

methods were used to further optimize the selected optimal EC-PLS models. For the two indicators, the results showed that the optimal models of the WSP-PLS method were exactly similar to those of the exhaustive method.

The N values of the optimal EC-PLS model for total acid and total phenol were 19 and 24 ($N > 17$), respectively. According to Section 2.7, the WSP-PLS method was first used to optimize the selected optimal EC-PLS models. The exhaustive method was then used to further optimize the selected optimal WSP-PLS models. The results for the total acid indicated that the optimal WSP-PLS model was also exactly similar to the optimal model of the exhaustive method. For the total phenol, the effect of the optimal model of the exhaustive method improved slightly compared to the optimal WSP-PLS model. Parameters N and LV for the optimal model of the exhaustive method were 13 and 7, respectively. Moreover, SEP_{Ave}, SEP_{SD}, $R_{P,Ave}$, $R_{P,SD}$, and SEP^{+} were 0.153, 0.013, 0.971, 0.006,

and 0.166, respectively. The corresponding wavelength combination ($N = 13$) was 1538, 1546, 1562, 1594, 1602, 1626, 1650, 1674, 1682, 1690, 1698, 1706, and 1714 nm.

In summary, the final optimization models for alcohol, total sugar, and total acid were the optimal WSP-PLS models, while that for total phenol was the optimal model of the exhaustive method. The results of the four data sets showed that the WSP-PLS method as a secondary optimization method almost approached the global optimization effect. Moreover, its calculation load was much lower than that of the exhaustive method and can be quickly implemented at the existing calculation level.

3.4　Independent validation

The 30 validation samples excluded from the modeling were used to evaluate the final optimization model for alcohol, total sugar, total acid, and total phenol. The PLS
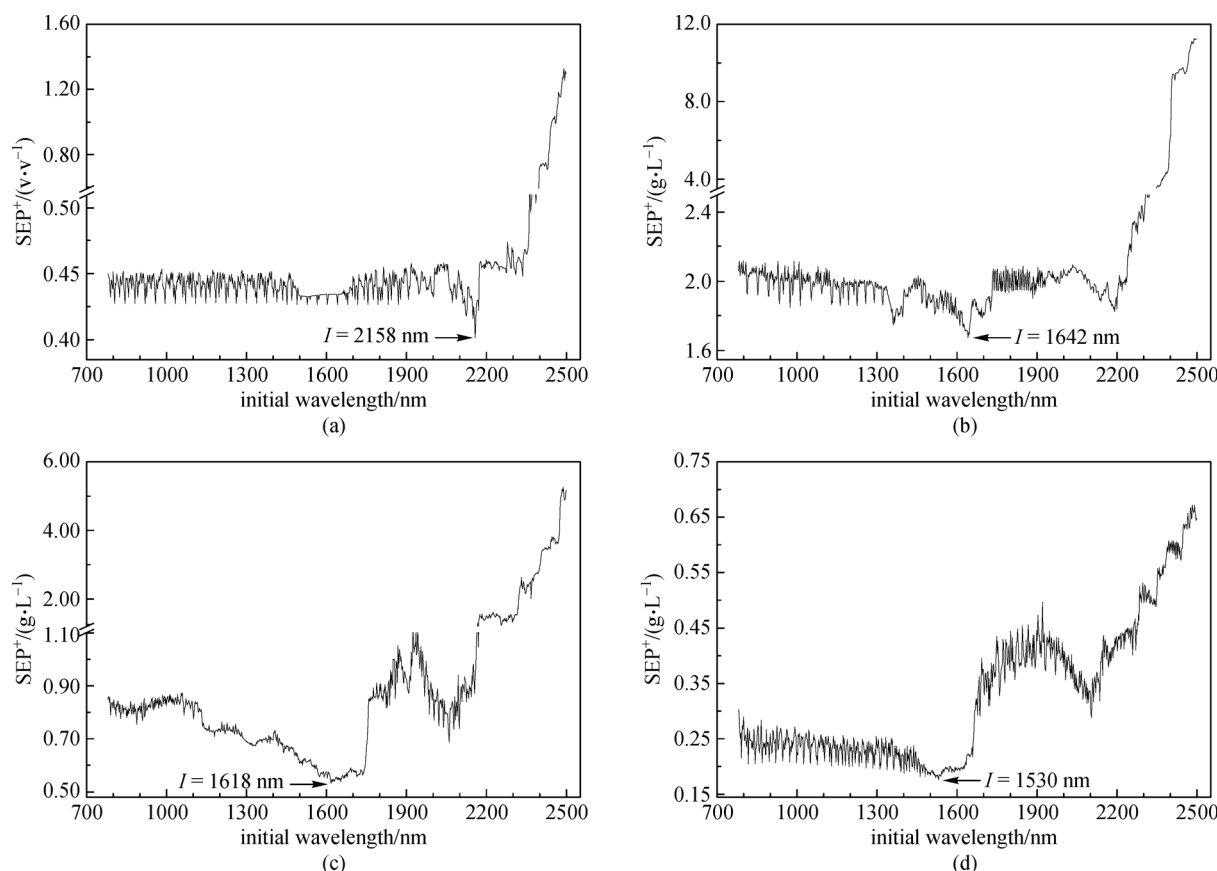
Fig. 8 SEP$^+$ values of the local optimal models corresponding to each initial wavelength for (a) alcohol, (b) total sugar, (c) total acid, and (d) total phenol

Table 5 Parameters and prediction effects of the optimal EC-WSP-PLS models for the four wine indicators

| indicator	N	LV	SEP$_{Ave}$	SEP$_{SD}$	$R_{P,Ave}$	$R_{P,SD}$	SEP$^+$
alcohol/(v·v^{-1})	7	5	0.37	0.02	0.951	0.007	0.39
total sugar/(g·L^{-1})	10	10	1.39	0.21	0.992	0.002	1.60
total acid/(g·L^{-1})	15	12	0.47	0.04	0.994	0.001	0.51
total phenol/(g·L^{-1})	17	7	0.153	0.014	0.972	0.006	0.167

regression coefficients were determined using the spectra and actual values of all modeling samples based on the model parameters (i.e., wavelengths and LV). The predicted values of the four indicators were then calculated using the obtained regression coefficients and the spectra of the validation samples. Table 6 summarizes the evaluation values for the validation (i.e., SEP, R_P, and RPD). The RPD values representing the overall predicted performance were 3.2, 6.8, 5.1, and 2.9, respectively.

Figure 9 presents the numerical relationships between the predicted and actual values of the validation samples for the four indicators. The results showed that the correlations between the spectral predicted and actual values were high, and the errors were low. The overall predicted performance of the RPD values was also high.

NIR spectroscopy usually contains hundreds to thousands of wavelength variables, and its spectral bands overlap flatly and have a poor absorption interpretation. Optimizing the wavelength model according to the prediction effect is more realistic because of the interference of other unknown components and noise. The obtained wavelength model still had a certain correlation with the NIR absorption band of the functional group. In fact, the obtained wavelength models ($N = 10, 15, 13$) for total sugar, total acid, and total phenol were all located in the NIR overtone region and were basically consistent with the NIR overtone region (1660–1800 nm) of the C–H functional group mentioned in the literature ([36], pp. 263). Moreover, the selected optimal wavelengths ($N = 7$) for alcohol were located in the NIR combination region. The

Table 6 Validation effects of the final optimization models for the four wine indicators

indicator	N	LV	SEP	R_P	RPD
alcohol/(v·v^{-1})	7	5	0.41	0.947	3.2
total sugar/(g·L^{-1})	10	10	1.48	0.992	6.8
total acid/(g·L^{-1})	15	12	0.68	0.981	5.1
total phenol/(g·L^{-1})	13	7	0.181	0.948	2.9
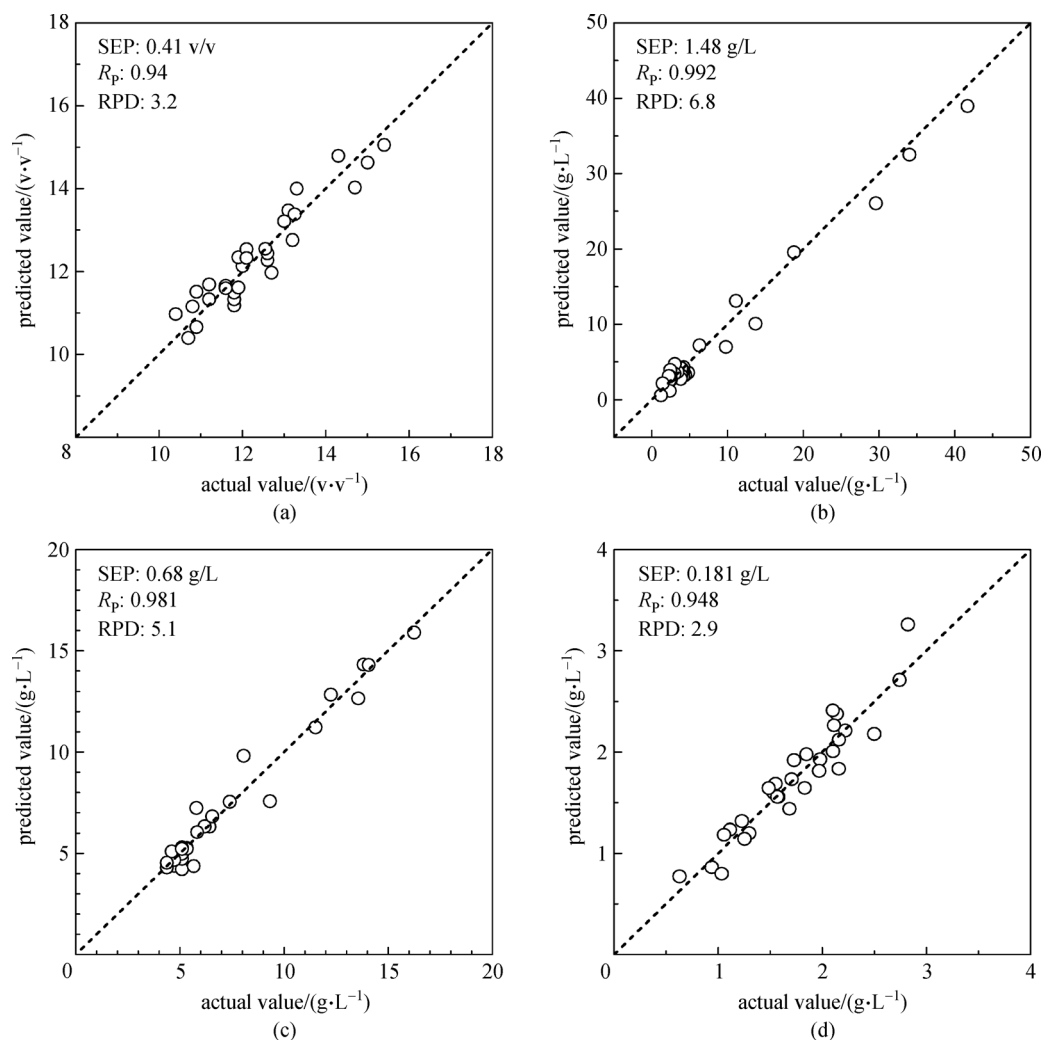
Fig. 9 Numerical relationship between the predicted and actual values of the validation samples for (a) alcohol, (b) total sugar, (c) total acid, and (d) total phenol

wavelengths were contained in the combined absorption band (1800–2200 nm) of the O–H, C–H, and C–O functional groups mentioned in the literature ([36], pp. 268). The literature also indicated that the combined absorption band (1800–2200 nm) can be used to quantify the alcohol in water ([36], pp. 269). Hence, the selected optimal wavelengths and the absorption band in Ref. [36] were also consistent.

The experiment results confirmed the feasibility of the simultaneous detection of the four indicators (i.e., alcohol,

total sugar, total acid, and total phenol) in wine through reagent-free NIR spectroscopy combined with wavelength optimization.

4   Conclusions

In this study, the simultaneous analysis of the four quality indicators in wine performed through reagent-free NIR spectroscopy is important in the rapid and real-time quality

detection of the wine fermentation process and the finished wine.

The novel integrated chemometrics approaches were the core of the technologies used herein. The multiparameter-NDF optimization platform was established and used to select the most suitable spectral preprocessing modes for the four wine indicators. The EC-PLS adopted three cyclic parameters for the large-scale screening of the wavelength models of equidistant combination. The WSP-PLS and exhaustive methods, which are secondary optimization methods, were used to correct the obtained equidistant wavelength models to further enhance the modeling effects and simplify the wavelength models. Four simplified wavelength models ($N = 7$, 10, 15, and 13) located in the NIR overtone or combination regions were obtained to analyze the four indicators. The independent validation results indicate that each model achieved a high correlation and a low error between the spectral predicted and actual values. The overall predicted performance RPD values were also high. Moreover, this modeling process was based on the multiple divisions of the calibration–prediction samples. Thus, the results obtained were stable and reliable.

Notably, prediction effects close to global optimization were achieved through the integrated optimization process based on the EC-PLS, WSP-PLS, and exhaustive methods. The wavelength optimization strategy can be applied to other analysis objects. The proposed wavelength models provide a valuable reference for designing small dedicated instruments.

References

1. Chiva-Blanch G, Arranz S, Lamuela-Raventos R M, Estruch R. Effects of wine, alcohol and polyphenols on cardiovascular disease risk factors: evidences from human studies. Alcohol and Alcoholism (Oxford, Oxfordshire), 2013, 48(3): 270–277

2. Bartowsky E. Microbiology of winemaking. Microbiology Australia, 2017, 38(2): 76–79

3. Park H, Choi W I, Park J, Jeong C, Kim S, Yoon H S. Brewing and quality characteristics of new grape cultivar 'okrang' wine in fermentation process. Journal of the Korean Society of Food Science and Nutrition, 2017, 46(5): 622–629

4. National Standards of People's Republic of China. GB/T 15038–2006. Analytical methods of wine and fruit wine. Beijing: Standards Press of China, 2006 (in Chinese)

5. Di Egidio V, Sinelli N, Giovanelli G, Moles A, Casiraghi E. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. European Food Research and Technology, 2010, 230(6): 947–955

6. Li H, Jiao A, Xu X, Wu C, Wei B, Hu X, Jin Z, Tian Y. Simultaneous saccharification and fermentation of broken rice: an enzymatic extrusion liquefaction pretreatment for Chinese rice wine production. Bioprocess and Biosystems Engineering, 2013, 36(8): 1141–1148

7. Holroyd S E, Prescott B, McLean A. The use of in-and on-line near infrared spectroscopy for milk powder measurement. Journal of Near Infrared Spectroscopy, 2013, 21(5): 441–443

8. Cozzolino D, Morón A. A potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil & Tillage Research, 2006, 85(1–2): 78–85

9. Viscarra Rossel R A, Walvoort D J J, McBratney A B, Janik L J, Skjemstad J O. Near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 2006, 131(1–2): 59–75

10. Chen H Z, Pan T, Chen J, Lu Q P. Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods. Chemometrics and Intelligent Laboratory Systems, 2011, 107(1): 139–146

11. Pan T, Li M, Chen J. Selection method of quasi-continuous wavelength combination with applications to the near-infrared spectroscopic analysis of soil organic matter. Applied Spectroscopy, 2014, 68(3): 263–271

12. Chen J Y, Zhang H, Matsunaga R. Rapid determination of the main organic acid composition of raw Japanese apricot fruit juices using near-infrared spectroscopy. Journal of Agricultural and Food Chemistry, 2006, 54(26): 9652–9657

13. Liu Z, Liu B, Pan T, Yang J. Determination of amino acid nitrogen in tuber mustard using near-infrared spectroscopy with waveband selection stability. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2013, 102(2): 269–274

14. Guo H S, Chen J M, Pan T, Wang J H, Cao G. Vis-NIR wavelength selection for non-destructive discriminant analysis of breed screening of transgenic sugarcane. Analytical Methods-UK, 2014, 6(21): 8810–8816

15. Lyu N, Chen J, Pan T, Yao L, Han Y, Yu J. Near-infrared spectroscopy combined with equidistant combination partial least squares applied to multi-index analysis of corn. Infrared Physics & Technology, 2016, 76: 648–654

16. Sousa A C, Lucio M M L M, Neto O F B, Marcone G P S, Pereira A F C, Dantas E O, Fragoso W D, Araujo M C U, Galvão R K H. A method for determination of COD in a domestic wastewater treatment plant by using near-infrared reflectance spectrometry of seston. Analytica Chimica Acta, 2007, 588(2): 231–236

17. Pan T, Chen Z H, Chen J M, Liu Z Y. Near-infrared spectroscopy with waveband selection stability for the determination of COD in sugar refinery wastewater. Analytical Methods-UK, 2012, 4(4): 1046–1052

18. Xie J, Pan T, Chen J M, Chen H Z, Ren X H. Joint optimization of Savitzky-Golay smoothing models and partial least squares factors for near-infrared spectroscopic analysis of serum glucose. Chinese Journal of Analytical Chemistry, 2010, 38: 342–346

19. Pan T, Liu J M, Chen J M, Zhang G, Zhao Y. Rapid determination of preliminary thalassaemia screening indicators based on near-infrared spectroscopy with wavelength selection stability. Analytical Methods-UK, 2013, 5(17): 4355–4362

20. Han Y, Chen J M, Pan T, Liu G S. Determination of glycated

hemoglobin using near-infrared spectroscopy combined with equidistant combination partial least squares. Chemometrics and Intelligent Laboratory Systems, 2015, 145: 84–92

21. Yao L, Lyu N, Chen J, Pan T, Yu J. Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2016, 159: 53–59

22. Chen J, Yin Z, Tang Y, Pan T. Vis-NIR spectroscopy with moving-window PLS method applied to rapid analysis of whole blood viscosity. Analytical and Bioanalytical Chemistry, 2017, 409(10): 2737–2745

23. Pan T, Yan B R, Chen J M, Yao L J. Discrete combination method based on equidistant wavelength screening and its application to near-infrared analysis of hemoglobin. Frontiers of Optoelectronics, 2018, 11(3): 296–305

24. Yang Y H, Lei F F, Zhang J, Yao L J, Chen J M, Pan T. Equidistant combination wavelength screening and step-by-step phase-out method for the near-infrared spectroscopic analysis of serum urea nitrogen. Journal of Innovative Optical Health Sciences, 2019, 12 (06): 1950018

25. Urbano Cuadrado M, Luque de Castro M D, Pérez Juan P M, Gómez-Nieto M A. Comparison and joint use of near infrared spectroscopy and Fourier transform mid infrared spectroscopy for the determination of wine parameters. Talanta, 2005, 66(1): 218–224

26. Guggenbichler W, Huck C W, Kobler A. Near infrared spectroscopy, cluster and multivariate analysis-contributions to wine analysis. Journal of Food Agriculture and Environment, 2006, 4: 98–106

27. Martelo-Vidal M J, Vázquez M. Determination of polyphenolic compounds of red wines by UV–VIS–NIR spectroscopy and chemometrics tools. Food Chemistry, 2014, 158: 28–34

28. Norris K H, Williams P C. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. Cereal Chemistry, 1984, 8: 99–110

29. Norris K H. Understanding and correcting the factors which affect diffuse transmittance spectra. NIR News, 2001, 12(3): 6–9

30. Pan T, Zhang J, Shi X W. Flexible vitality of near-infrared spectroscopy-talking about Norris derivative filter. NIR News, 2019, 9: 1–4

31. Yao L J, Tang Y, Yin Z W, Pan T, Chen J M. Repetition rate priority combination method based on equidistant wavelengths screening with application to NIR analysis of serum albumin. Chemometrics and Intelligent Laboratory Systems, 2017, 162: 191–196

32. Chen J, Peng L, Han Y, Yao L, Zhang J, Pan T. A rapid quantification method for the screening indicator for β-thalassemia with near-infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 193: 499–506

33. Zhang J, Li M L, Pan T, Yao L J, Chen J M. Purity analysis of multi-grain rice seeds with non-destructive visible and near-infrared spectroscopy. Computers and Electronics in Agriculture, 2019, 164: 104882

34. Ozaki Y. Near-infrared spectroscopy–its versatility in analytical chemistry. Analytical Sciences, 2012, 28(6): 545–563

35. Chen J, Li M, Pan T, Pang L, Yao L, Zhang J. Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 219: 179–185

36. Chu X L. Molecular Spectroscopy Analytical Technology Combined With Chemometrics and Its Applications. Beijing: Chemical Industry Press, 2011 (in Chinese)

Jiemei Chen is an associate professor and graduate tutor in Department of Biological Engineering, Jinan University, China. She received her B.S. and M.S. degrees in Microbiology from Sichuan University and Guangxi University, China, respectively. And she received her Ph.D. degree in Biology from Mie University, Japan. She is engaged in studies of microbiology, spectroscopy, biomedical information, etc. She has published more than 60 peer reviewed papers.

Sixia Liao is a master student majored in Microbiology in Department of Biological Engineering, Jinan University, Guangzhou, China.

Lijun Yao is an associate professor in Department of Optoelectronic Engineering, Jinan University, China. He received his B.S. degree in electronic Technique from The First Aviation Academy of Chinese Air Force, China, his M.S. degree in Condensed Matter Physics and Ph.D. degree in Biological Information Engineering from Jinan University, China. He is engaged in studies of spectroscopy, biomedical information, chemometrics, and pattern recognition. He has published more than 40 peer reviewed papers.

Tao Pan is a professor and doctorial tutor in Department of Optoelectronic Engineering, Jinan University, China. He received his B.S. degree in Mathematics from Sichuan University, China, his Ph.D. degree in Biological Information Engineering from Mie University, Japan. He is director of Applied Spectroscopy Laboratory at College of Science and Engineering, Jinan University, China. He is engaged in studies of spectroscopy, biomedical information, chemometrics, pattern recognition and partial differential equations. He has published more than 100 peer reviewed papers.