American Journal of Analytical Chemistry, 2018, 9, 124-133
http://www.scirp.org/journal/ajac
ISSN Online: 2156-8278
ISSN Print: 2156-8251

Discriminant Analysis of Liquor Brands Based on Moving-Window Waveband Screening Using Near-Infrared Spectroscopy

Jie Zhong[1], Jiemei Chen[2], Lijun Yao[1,3], Tao Pan[1*]

[1]Department of Optoelectronic Engineering, Jinan University, Guangzhou, China
[2]Department of Biological Engineering, Jinan University, Guangzhou, China
[3]Guangzhou SonDon Network & Technology Co., Ltd, Guangzhou, China
Email: *466945939@qq.com

How to cite this paper: Zhong, J., Chen, J.M., Yao, L.J. and Pan, T. (2018) Discriminant Analysis of Liquor Brands Based on Moving-Window Waveband Screening Using Near-Infrared Spectroscopy. *American Journal of Analytical Chemistry*, **9**, 124-133.
https://doi.org/10.4236/ajac.2018.93011

Received: February 2, 2018
Accepted: March 9, 2018
Published: March 12, 2018

Abstract

Partial least squares discriminant analysis (PLS-DA) with integrated moving-window (MW) waveband screening was applied to the discriminant analysis of liquor brands with near-infrared (NIR) spectroscopy. Luzhou Laojiao, a popular liquor with strong fragrant flavor, was used as the identified liquor brand (160 samples, negative, 52 vol alcoholicity). Liquors of 10 other brands with strong fragrant flavor were used as the interferential brands (200 samples, positive, 52 vol alcoholicity). The Kennard-Stone algorithm was used for the division of modeling samples to achieve uniformity and representativeness. Based on the MW-PLS-DA, a simplified optimal model set with 157 wavebands was further proposed. This set contained five types of wavebands corresponding to the NIR absorption bands of water, ethanol, and other micronutrients (*i.e.*, acids, aldehydes, phenols, and aromatic compounds) in liquor for practical choice. Using five selected simple models with 4775 - 4239, 7804 - 6569, 6264 - 5844, 9435 - 7896, and 12066 - 10373 cm^{-1}, the validation recognition rates were obtained as 99.3% or higher. Results show good prediction performance and low model complexity, and also provided a valuable reference for designing small dedicated instruments. The proposed method is a promising tool for large-scale inspection of liquor food safety.

Keywords

Liquor Brands, Near-Infrared Spectroscopy, Partial Least Squares Discriminant Analysis, Moving-Window Waveband Screening, Simplified Optimal Model Set

1. Introduction

Chinese liquor is a distilled spirit mainly made from grain and obtained using distiller's yeast, which is a complex mixture and composed mainly of water and ethanol as well as micronutrients and active ingredients, including acids, aldehydes, phenols, and aromatic compounds. Unfortunately, because of the huge market share and high profitability, many fake liquors are being sold in the market, which not only causes economic losses to the producers of liquor brands but also poses a threat to consumers' health. At present, identification of liquor brands usually requires the determination of various components and their content recipes using traditional analysis methods (e.g., high-performance liquid chromatography), which are complex and costly. Another method is the sensory judgment of tasters, which has great subjectivity and relies on the experience. The above methods are difficult to conduct in large-scale promotions.

With the developments of chemometric and sensor technology, near-infrared (NIR) spectroscopy has been proven to be a significant potential tool in the rapid and reagent-less measurement of various fields. It is reported that NIR quantitative analysis has been applied to determine the main components of liquor, such as ethanol [1], ethyl acetate [2], and aldehydes [3]. However, the components and contents are different in liquors of various brands. Therefore, it is still difficult to identify the liquor brands by quantitative analysis of the above conventional components.

Spectral discriminant analysis uses the spectral overall features to identify and to classify samples; its bases are that the spectral similarities of the samples of the same types and the spectral differences among samples of different types. Principal component analysis-linear discriminant analysis (PCA-LDA) is the commonly well-performed method for spectral discriminant analysis, which has been applied in the identification of liquor brands [4] [5]. Partial least squares-discriminant analysis (PLS-DA) is more effective than the PCA-LDA method in theory and practice [6] [7] [8], which has been applied in the identification of liquor brands [9]. However, neither of the literatures [4] and [5] strictly used the liquor brands of the same flavor and ethanol content for discriminant analysis. Literature [9] shown the experimental result of identifying liquor brands with the same flavor and ethanol content, but model only used the entire spectral region without any waveband selection and the prediction recognition rates required further improved.

Appropriate wavelength selection is essential for mitigating disturbance, improving prediction accuracy and simplifying the model, especially for the complex samples with multiple components. However, the above works [1] [2] [3] [4] [5] [9] on liquor brands identification are still based on the whole spectral region because of algorithm complexity. In the quantitative analysis of the NIR spectrum, moving-window waveband screening [10]-[15] combined with the PLS method can extract information effectively, eliminate noise disturbances, and improve predictive capability.

In the current study, moving-window (MW) waveband screening was integrated to PLS-DA (MW-PLS-DA) and employed for the NIR spectral discriminant analysis of liquor brands. Furthermore, the optimal model set and its simplified method were further proposed, and the simple models with high accuracy were obtained.

The spectra of liquor samples of different flavors (or different ethanol contents) are remarkably diverse [1] [2] [3] [4]. This work is focused on the identification for liquor brands with the same flavor and ethanol content. Although difficult, such a method is important and essential.

2. Materials and Methods

2.1. Experiment

Luzhou Laojiao, a popular liquor with strong fragrant flavor in China, was used as the identified brand (negative). A total of 160 bottles of Luzhou Laojiao Danya Erqu liquor (52 vol alcoholicity) were collected. Liquors of 10 other brands with strong fragrant flavor and the same ethanol content (52 vol) were used as the interferential brands (200 bottles, positive), which were composed of 20 bottles from each liquor brand, including 1) Bainian Hutu, 2) Dukang Taibai, 3) Shixiantaibai, 4) Luzhou Laojiufang, 5) Tangchao Laojiao, 6) Wudang Xiaojiufan, 7) Jingjuyuan Pingjian, 8) Tianxiafu, 9) Guifeizuijiu, and 10) Kongfujia Shengshidatao. A total of 360 liquor samples were extracted and used for spectral measurement.

The instrument was a VERTEX 70 FT-NIR Spectrometer (Bruker, Germany) equipped with a transmission accessory and a 1 mm cuvette. An InGaAs detector was used. Twelve scans were added to each spectrum. The entire scanning region was 14,994 - 3996 cm^{-1} at a wavenumber interval of 3.857 cm^{-1}, with 2852 wavenumbers. Each sample was measured twice, and the mean value was used for modeling and validation. The spectra were obtained at 25˚C ± 1˚C and 45% ± 1% RH.

2.2. Sample Division

Initially, 60 negative and 80 positive samples were randomly selected into the independent validation set (140 samples), while the rest of 100 negative and 120 positive samples were used for modeling set (220 samples). Then, using the Kennard-Stone algorithm [16], the negative and positive modeling samples were further equally divided into calibration and prediction samples, to fulfill uniformity and representativeness.

2.3. Integrated MW-PLS-DA Method

All sub-waveband were traversed for modeling, using the following two parameters [11] [12] [13] [14] [15]: (1) initial wavenumber (I) and (2) number of wavenumbers (N). The search range was 14,994 - 3996 cm^{-1} with 2852 wavenumbers, and I was set to $I \in \{14994, 14990, \cdots, 3996\}$. To cut down the volume of work and guarantee representativeness, N was set to

$N \in \{1,2,\cdots,50\} \cup \{50,60,\cdots,500\} \cup \{500,550,\cdots,2800\} \cup \{2852\}$.

The obtained wavebands were used to establish the calibration and prediction models of PLS-DA. The process can refer to [6] [7] [8] [9]. Here, the positive and negative samples were assigned to the category value 1 and 0 respectively, then the quantitative calculation was carried out, the samples are classified by the 0.5 as the threshold. Where, the number of PLS factors (*F*) was set to $F \in \{1,2,\cdots,20\}$.

On the basis of the predicted category of the samples and their genuine brand type, it is easy to calculate the prediction recognition rate denoted as *P*_REC. According to the maximum *P*_REC, the optimal parameters (*i.e. I*, *N* and *F*) were selected and then the optimal MW-PLS-DA models were obtained.

2.4. Optimal Model Set and Its Simplification

Given that the optimal MW-PLS-DA models corresponded to the maximum *P*_REC (denoted as *P*_REC*) were not usually unique, the optimal model set and its simplification method were further proposed for the appropriate selection of wavebands. The optimal waveband set can be expressed as follows:

$$\Lambda^* = \left\{ \left(I^{(q)}, E^{(q)} \right) \middle| q = 1,2,3,\cdots,Q \right\}, \tag{1}$$

where I and E are the initial and ending wavenumbers, respectively, and Q is the number of optimal wavebands. If a containing relationship exists between two optimal wavebands, then

$$\left(I^{(i)}, E^{(i)} \right) \subset \left(I^{(j)}, E^{(j)} \right),\ 1 \le i \ne j \le Q. \tag{2}$$

The latter contained redundant wavenumbers, which must be removed from the optimal model set. The same processing was performed for each optimal waveband. Accordingly, the simplified optimal model set (denoted as Ω^*) can be obtained. In the set of Ω^*, no containing relationship existed between any two wavebands.

2.5. Model Validation

The validation group that containing 60 negative and 80 positive samples (total 140 samples) as well as out of the modeling optimization procedure were used for verifying the selected models screened using MW-PLS-DA method. According to the predicted category of validation samples and their genuine brand type, it is easy to calculate the validation recognition rate denoted as *V*_REC. Furthermore, the validation recognition rates of negative and positive samples can be calculated and were denoted as V_REC^- and V_REC^+, respectively.

The computer platform was developed using Matlab 2012a.

3. Results and Discussion

3.1. Full PLS-DA Model

The NIR spectra ranging from 14,994 to 3996 cm^{-1} of liquor samples for 160

Luzhou Laojiao (negative, upper) and 200 Non-Luzhou Laojiao (positive, lower) are plotted, as shown in Figure 1. There were no apparent differences of spectra for direct discriminant analysis, since the given spectra of negative and positive samples were overlapping.

The PLS-DA model based on the entire scanning region (14,994 - 3996 cm^{-1}), called full PLS-DA, was first established. The optimal F was 7, and the corresponding P_REC was 99.1%. However, the adopted waveband contained a large number of wavenumbers ($N = 2852$), which may include redundant wavenumbers. Therefore, the model complexity must be further reduced.

3.2. Simplified Optimal Model Set with MW-PLS-DA

The waveband selection was performed using the MW-PLS-DA method. The maximum P_REC achieved 100% (P_REC*), and the optimal waveband set Λ^* contained 37,870 wavebands. The corresponding 2D diagram for initial and ending wavenumbers is shown in Figure 2(a). In the set of Λ^*, a large amount of containing relationship was easily observed. Therefore, Λ^* must be further simplified. Using the simplification method mentioned above, the simplified optimal model set Ω^* contained only 157 models. The 2D diagram is shown in Figure 2(b). No containing relationship was observed in Ω^*. In the average spectra, 157 wavebands were marked to clearly observe their position, as in Figure 3.

The wavebands of the simplified optimal model set Ω^* could be divided into two parts, as follows.

The first part was associated mainly with the NIR characteristic absorption bands of water and ethanol. At 4347 cm^{-1}, the absorption band related to the characteristic absorption of ethanol [3] could be applied to a quantitative analysis of ethanol in liquor. Two wavebands existed in Ω^* containing the absorption band. These wavebands were 4775 - 4239 and 4772 - 4235 cm^{-1}, the number

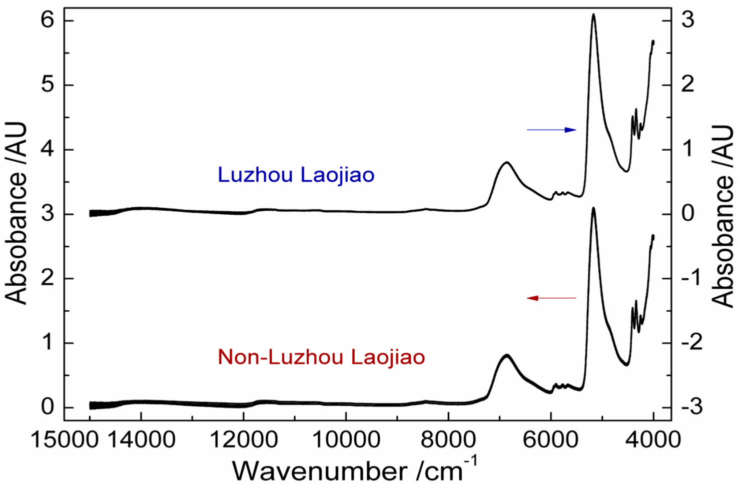
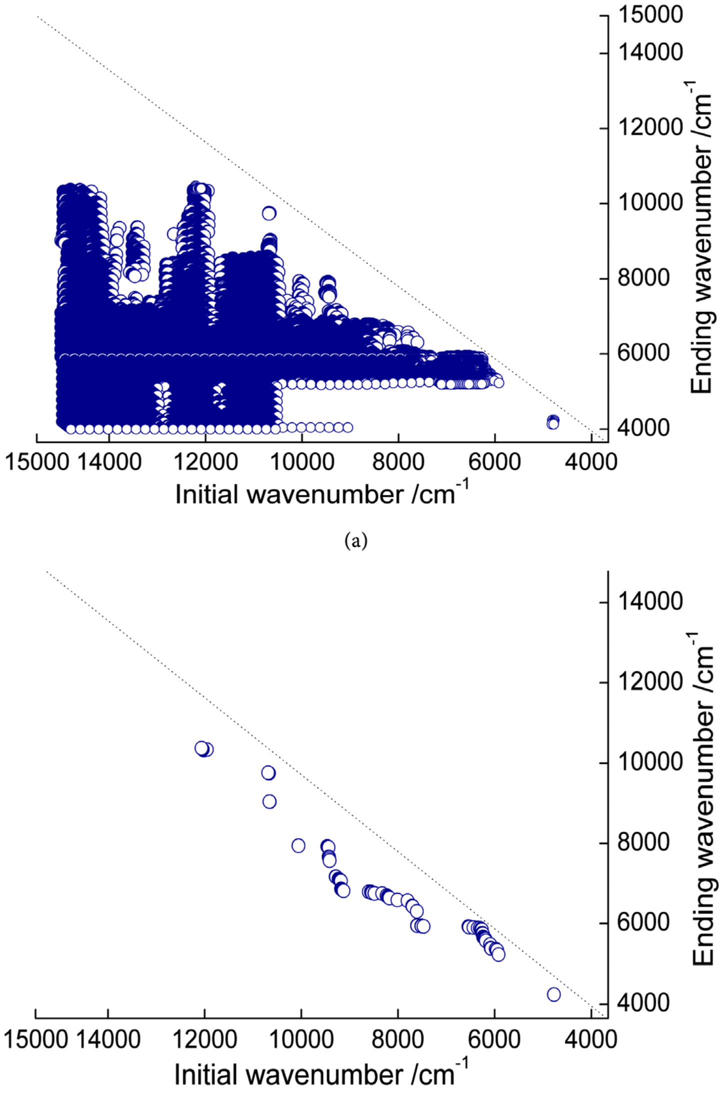


Figure 1. NIR spectra of liquor samples for Luzhou Laojiao (160 negative, upper) and Non-Luzhou Laojiao (200 positive, lower).



(a)

(b)

Figure 2. Two-dimensional diagrams for initial and ending wavenumbers of (a) Entire optimal waveband set and (b) Simplified optimal waveband set.

of wavenumbers *N* were both 140, and the corresponding optimal *F* were 7 and 8.

At 5128 and 6896 cm^{-1}, the absorption bands were related to the O-H stretch first overtone and second overtone of water [17]. A total of 55 wavebands in Ω^* was around the band at 6896 cm^{-1}. Among them, the waveband (7804 - 6569 cm^{-1}) was of low model complexity ($N = 320$) with the corresponding *F* of 8.

The second part was associated mainly with the NIR characteristic absorption bands of other micronutrients (*i.e.*, acids, aldehydes, phenols, and aromatic compounds) in liquor.

Both varieties showed absorption bands at 5586 cm^{-1} related with the C-H stretch first overtone of acids, aldehydes, phenols; bands at 5917 cm^{-1} were

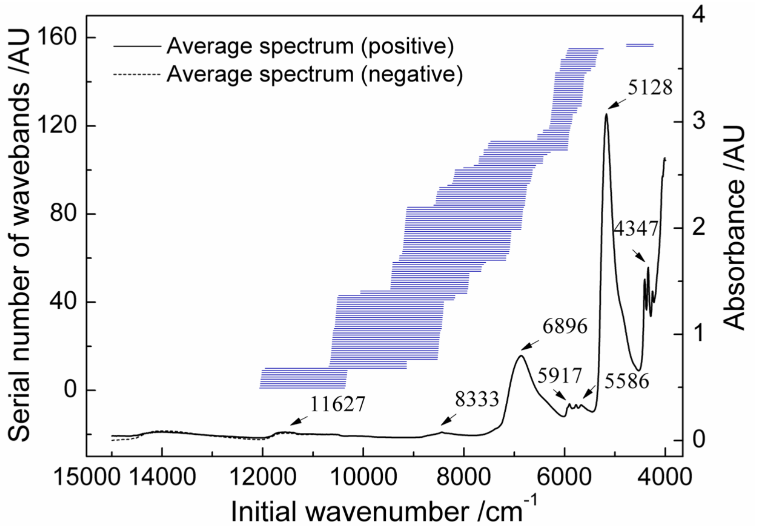


Figure 3. The position of 157 simplified optimal wavebands in average spectra of the positive and negative samples.

related with either the C-H^3 stretch first overtone or the C-H first overtone of aromatic groups [17]. These bands were contained in 42 wavebands of Ω^*. Among them, the waveband (6264 - 5844 cm^{-1}) was of low model complexity (N = 110) with the corresponding F of 7.

The small peak around 8333 cm^{-1} arose from the second overtones of C-H with stretching vibrations of acids, aldehydes, and phenols [4]. A total of 49 wavebands in Ω^* were around the band at 8333 cm^{-1}. Among them, the waveband (9435 - 7896 cm^{-1}) was of low model complexity (N = 400) with the corresponding F of 9.

The band at 11,235 cm^{-1} related with the C-H third overtone of aromatic compounds was contained by the remaining 9 wavebands in Ω^*. Among them, the waveband (12,066 - 10,373 cm^{-1}) was of low model complexity (N = 440) with the corresponding F of 7.

Liquor samples with different ethanol contents have significantly different water contents. Hence, the wavebands in the first part of Ω^* were suitable to identify those samples. In this study, ethanol contents of samples were the same. Therefore, the wavebands in the second part of Ω^* were appropriate for the discriminant analysis.

3.3. Independent Validation

The randomly selected validation samples excluded from the modeling optimization process were used to validate the five selected simple wavebands (4775 - 4239, 7804 - 6569, 6264 - 5844, 9435 - 7896, and 12,066 - 10,373 cm^{-1}). The corresponding parameters and validation effects are summarized in **Table 1**. The validation recognition rates (V_RECs) were 99.3% or higher, and the number of wavenumbers (N) were 440 or less. Furthermore, the full PLS-DA model (14,994 - 3996 cm^{-1}) was used to validate for comparison. The V_REC and N were 98.6%

Table 1. Parameters and validation effects of the five selected models screened using the MW-PLS-DA method.

Waveband (cm^{-1})	N	F	V_REC	V_REC^{-}	V_REC^{+}
4775 - 4239	140	7	99.3%	100%	98.8%
7804 - 6569	320	8	99.3%	100%	98.8%
6264 - 5844	110	7	100%	100%	100%
9435 - 7896	400	9	100%	100%	100%
12,066 - 10,373	440	7	99.3%	100%	98.8%

and 2852, respectively (see also in [9]). The results indicate that, the validation effects of the five selected models were superior to the full PLS-DA model in two aspects of prediction performance and model complexity.

4. Conclusions

Most of the fake liquors are usually made into the products with the same flavor and ethanol content as regular brand, so the identification for such liquor samples is essential. However, it is also difficult because their components are very similar.

In the present study, the MW-PLS-DA was integrated and successfully applied to the NIR spectral discriminant analysis of liquor brands with the same flavor and ethanol content. A simplified optimal model set with 157 wavebands was further proposed based on the MW-PLS-DA. The five types of wavebands in the simplified optimal model set corresponded to the NIR absorption bands of water, ethanol, and other micronutrients (*i.e.*, acids, aldehydes, phenols, and aromatic compounds) in liquor. According to the differences in components and NIR absorption features of objects, an appropriate model can be selected from them.

The experimental results indicate that the selected models achieved high prediction recognition rates with low model complexity, and provide a valuable reference for designing small dedicated instrument. The proposed method is a promising tool for large-scale inspection of liquor food safety.

Acknowledgements

This work was supported by the Science and Technology Project of Guangdong Province of China (No. 2014A020213016, No. 2014A020212445).

References

[1] Qu, F.F., Ren, D., Wang, J.H., Zhang, Z., Lu, N. and Meng, L. (2016) An Ensemble Successive Project Algorithm for Liquor Detection Using Near Infrared Sensor. *Sensors*, **16**, 89-102. https://doi.org/10.3390/s16010089

[2] Liu, J.X., Zhang, W.W., Han, S.H., Li, X., Li, P.Y., Yang, G.D., Yang, Y., Xu, B.C. and Luo, D.L. (2016) Rapid Detection of Caproic Acid and Acetic Acid in Liquor Base Based on Fourier Transform Near-Infrared Spectroscopy. *Food Science*, **37**,

181-185.

[3] Zhang, W.W., Liu, J.X., Han, S.H., Pan, Y.O., Li, X., Li, P.Y. and Xu, B.C. (2016) Determination of Aldehydes in Liquor Base Based on Fourier Transform Near-Infrared Spectroscopy. *Food Science*, **37**, 111-115.

[4] Li, Z., Wang, P.P., Huang, C.C., Shang, H., Pan, S.Y. and Li, X.J. (2014) Application of Vis/NIR Spectroscopy for Chinese Liquor Discrimination. *Food Analytical Methods*, **7**, 1337-1344. https://doi.org/10.1007/s12161-013-9755-9

[5] Tan, C., Chen, H., Lin, Z., Wu, T., Wang, L. and Zhang, K.S. (2015) Classification of Liquor Using Near-Infrared Spectroscopy and Chemometrics. *Analytical Letters*, **48**, 291-300. https://doi.org/10.1080/00032719.2014.938343

[6] Chiang, L.H., Russell, E.L. and Braatz, R.D. (2000) Fault Diagnosis in Chemical Processes Using Fisher Discriminant Analysis, Discriminant Partial Least Squares, and Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **50**, 243-252. https://doi.org/10.1016/S0169-7439(99)00061-1

[7] Barker, M. and William, R. (2003) Partial Least Squares for Discrimination. *Journal of Chemometrics*, **17**, 166-173. https://doi.org/10.1002/cem.785

[8] Miguel, P.E. and Michel, T. (2003) Prediction of Clinical Outcome with Microarray Data: A Partial Least Squares Discriminant Analysis (PLS-DA) Approach. *Human Genetics*, **112**, 581-592.

[9] Yang, B., Yao, L.J. and Pan, T. (2017) Near-Infrared Spectroscopy Combined with Partial Least Squares Discriminant Analysis Applied to Identification of Liquor Brands. *Engineering*, **9**, 181-191. https://doi.org/10.4236/eng.2017.92009

[10] Jiang, J.H., Berry, R.J., Siesler, H.W. and Ozaki, Y. (2002) Wavelength Interval Selection in Multicomponent Spectral Analysis by Moving Window Partial Least-Squares Regression with Applications to Mid-Infrared and Near-Infrared Spectroscopic Data. *Analytical Chemistry*, **74**, 3555-3565. https://doi.org/10.1021/ac011177u

[11] Chen, H.Z., Pan, T., Chen, J.M. and Lu, Q.P. (2011) Waveband Selection for NIR Spectroscopy Analysis of Soil Organic Matter Based on SG Smoothing and MWPLS Methods. *Chemometrics and Intelligent Laboratory Systems*, **107**, 139-146. https://doi.org/10.1016/j.chemolab.2011.02.008

[12] Pan, T., Liu, J.M., Chen, J.M., Zhang, G.P. and Zhao, Y. (2013) Rapid Determination of Preliminary Thalassaemia Screening Indicators Based on Near-Infrared Spectroscopy with Wavelength Selection Stability. *Analytical Methods*, **5**, 4355-4362. https://doi.org/10.1039/c3ay40732b

[13] Pan, T., Li, M.M. and Chen, J.M. (2014) Selection Method of Quasi-Continuous Wavelength Combination with Applications to the Near-Infrared Spectroscopic Analysis of Soil Organic Matter. *Applied Spectroscopy*, **68**, 263-271. https://doi.org/10.1366/13-07088

[14] Long, X.L., Liu, G.S., Pan, T. and Chen, J.M. (2014) Waveband Selection of Reagent-Free Determination for Thalassemia Screening Indicators Using Fourier Transform Infrared Spectroscopy with Attenuated Total Reflection. *Journal of Biomedical Optics*, **19**, 087004. https://doi.org/10.1117/1.JBO.19.8.087004

[15] Chen, J.M., Yin, Z.W., Tang, Y. and Pan, T. (2017) Vis-NIR Spectroscopy with Moving-Window PLS Method Applied to Rapid Analysis of Whole Blood Viscosity. *Analytical and Bioanalytical Chemistry*, **409**, 2737-2745. https://doi.org/10.1007/s00216-017-0218-9

[16] Kennard, R.W. and Stone, L.A. (1969) Computer Aided Design of Experiments. *Technometrics*, **11**, 137-148. https://doi.org/10.1080/00401706.1969.10490666

[17] Chen, H., Tan, C., Wu, T., Wang, L. and Zhu, W. (2014) Discrimination between Authentic and Adulterated Liquors by Near-Infrared Spectroscopy and Ensemble Classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **130**, 245-249. https://doi.org/10.1016/j.saa.2014.03.091