2012 International Workshop on Information and Electronics Engineering (IWIEE)

Study on the Best Analysis Spectral Section of NIR to Detect Alcohol Concentration Based on SiPLS

Xiaofei Wang[a], Yanfei Bao[a],*, Guili Liu[a], Gang Li [b], Ling Lin [b]

[a] School of Optoelectronic Information & Telecommunication Engineering of Beijing Information Science& Technology University, BeiJing 100192, China

[b] Tianjin Key Laboratory of Biomedical Detection Techniques & Instruments, Tianjin University, Tianjin 300072, China

Abstract

The alcohol concentration is usually measured through analyzing the multiple frequency absorption peak or co-frequency absorption peak of alcohol's C-H and O-H by NIRS. Based on the normalized transmittance spectrogram and Synergy interval partial least squares algorithm (SiPLS), the best analysis spectrum band detecting alcohol concentration is decided at the range of 1100nm-1330nm and 2110nm-2400nm or so, which is the second-harmonic peak of C-H and bending and expansion of vibration spectral section of C-H&O-H. The SiPLS model is established to analyzing 53 samples with different alcohol concentration, and obtained the root mean square error of prediction (RMSEP) is 0.0060, the correlation coefficient (R) is 0.9997, the better measuring result is obtained and this is hopeful be used to detect alcohol concentration during the Production process.

Keywords: Alcohol concentration; the best spectrum band; SiPLS; NIRS

1. Introduction

The near infrared reflectance spectroscopy(NIRS) analysis technology is widely used in detecting alcohol concentration of kinds of wines, medical alcohol and industrial alcohol because of the superiority of rapid, undamaged, simple and convenient. But which NIR spectral section should be used to analyse alcohol concentration on earth? In the present research, some researchers use 1700nm or so bands, which

---

* Corresponding author. Tel.: +86-15210663675.
E-mail address: baoyanfei27361@163.com.

is the absorption band of first-harmonic peak of C-H[1]; some researchers use the absorption band of second-harmonic peak of C-H or the absorption band of first-harmonic peak of O-H (1450nm or so) besides the absorption band of first-harmonic peak of C-H[2]; some researchers use 960nm wave band, which is the absorption band of third -harmonic peak of C-H[4]; some researchers use 2300nm wave band, which is the absorption band of Co-frequency of C-H and CH2[5,6]; and others also use the full spectrum of 1000nm-2500nm[7,8] or use 1250nm-2500nm spectrum combine with some algorithm of spectrum selection to get rid of the spectrum overlap seriously or contain interference information[9].

In a word, the research thinking of present mainly based on analyzing the multiple frequency absorption peak or Co-frequency absorption peak of alcohol's C-H and O-H to determine the alcohol concentration. The paper has some different thinking from others, using the normalized transmittance spectrogram and Synergy interval partial least squares algorithm (SiPLS), the best analysis spectrum band detecting alcohol concentration is decided at the range of 1100nm-1330nm and 2110nm-2400nm or so, which is the second-harmonic peak of C-H and bending and expansion of vibration spectral section of C-H&O-H.

2. Materials and methods

2.1. Experimental instruments

Experiment adopts Avantes Company's NIR spectrograph with 1000-2500nm wavelength range; Air-filled type of tungsten halogen light source of Avantes Company, which is stable in 360nm-2000nm wavelength; 10×10mm silica-cuvette and support of cuvette to ensure keeping measurement seat always in same place every time. The graph of NIR test facility for detecting alcohol concentration is shown as Fig. 1:

Adopt MATLAB R2009b software platform and IPLS toolbox write by Nørgaard to data analysis and establish analysis model program.
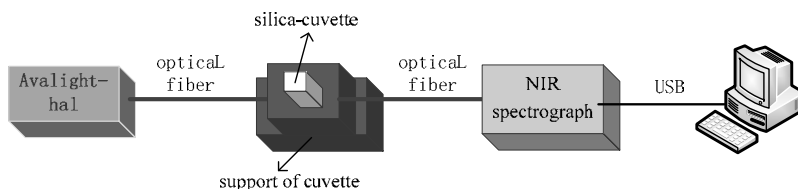

Fig.1. The system of NIR test facility

2.2. Experimental samples

There were two sources of alcohol solution samples: one is respectively allocate, 43 alcoholic solution with different concentration got from proportioning 99.7% alcohol with distilled water according to volume; the other is bought from market, 10 kinds of alcoholic solution (white wine, red wine, beer, medical alcohol and industrial alcohol) with different alcohol degree, and demarcated alcohol degree by alcoholmeter. 53 test samples were picked totally.

2.3. Spectral collection

Avasoft spectroscopy software of Avantes Company for collection platform, collected in transmission mode, empty silica-cuvette for dark background, integral time was 5.55ms, average scan times was 90

and smoothness was 0. The normalized transmittance spectrogram of 53 wine samples gave above were collected, then deleted the noise date of front and back, got 1100-2400nm wavelength range contain 211 wavelengths, shown as Fig.2.
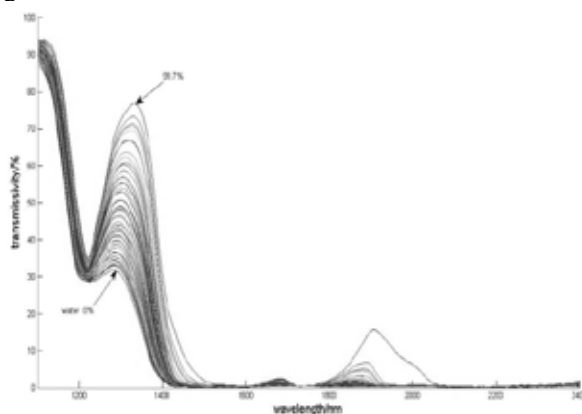

Fig.2. Normalized transmittance spectrogram

3. Spectrogram analysis

NIRS is mainly due to the anharmonicity of the molecular vibrations from the ground state to a higher energy level transition, it records the information of multiple frequency and co-frequency of fundamental frequency vibration of molecular single chemical bond, it always is dominated by overlap of the multiple frequency and co-frequency of groups containing hydrogen(such as C-H, O-H, N-H, S-H, P-H), so NIRS mainly reflects the absorption of multiple frequency and co-frequency which is caused by stretching, vibration, curve of groups containing hydrogen.

Bases of the test samples are ethanol and water, molecular formula of ethanol is C2H6O, structural formula is CH3-CH2-OH. Three kinds of hydrogen atoms are contained in the molecular structure: methyl (-CH3), methylene (-CH2-) and hydroxyl (-OH); five kinds of chemical bonds: a C-C, three C-H, a C-O and a O-H. Water molecule only has O-H, but have no group containing C.

Shown as Fig.2, nearby 1220nm is the second-harmonic peak of C-H, Peng bangzhu[2] used this spectrum section and the third-harmonic absorption peak of C-H 960nm, combined with PLS to analyze the alcohol degree of white wine; the first combined frequency spectrum section of water molecular symmetric and antisymmetric O-H stretching vibration, and the first co-frequency spectrum section of ethanol molecular unimer and multimer O-H stretching vibration presents to 1600nm nearby, Feng shangkun[3] used these spectrum sections to analyze alcohol degree of beer by BP neural network; nearby 1720nm is the first-harmonic absorption band of methyl hydrogen and methylene hydrogen's symmetric and antisymmetric O-H stretching vibration and absorption band of hydroxyl hydrogen O-H flexural vibration, Li daixi[[1] used this spectrum section to analyze alcohol degree of beer and vinous by PLS; nearby 1950nm is co-frequency spectrum of O-H, nearby 2080nm is co-frequency spectrum of C-O-H, 2200-2400nm is co-frequency spectrum section of C-H, C-H2 and O-H bending and expansion vibration, Geng chaoxi[6] used these to analyze alcohol degree of vinous fermentation liquor; Liu hongxin[7] used the whole spectrum 1000-2500nm to analyze alcohol degree of beer by PLS, which has heavy computation; Chen bin[9] used 1250-2500nm long spectrum to analyze alcohol degree of beer by genetic

algorithm(GA), effectively filter out overlap seriously and interferential regions, enhanced the reliability of data, but its computation is maximum, operation time is longer.

Only investigated the normalized transmittance spectrogram shown at Fig.2, it is obviously that sensitive spectrum of alcohol solution NIRS which reacted different alcohol degree present to 1200-1400nm, this spectrum section was between the C-H second-harmonic absorption peak and the O-H first-harmonic absorption peak, this spectrum section shown the alcohol degree higher then spectral transmittance higher. At the follows we established model using SiPLS algorithm to test and verify which spectrum section containing the best characteristic NIRS information for detecting alcohol degree.

4. SiPLS model analysis and discussion

4.1. SiPLS modeling

Synergy interval partial least squares algorithm (SiPLS)is the expansion of interval partial least squares algorithm(iPLS), first of all, it equal-divides the whole spectrum into several subintervals, then do partial least squares regression through permutation and combination with different number and different subintervals, at the last, using the synergy intervals spectrum which has the smallest root mean square error of cross validation(RMSECV) to establish the SiPLS model and prediction.

39 train set samples were selected randomly from 53 experimental samples, and the rest 14 samples as test set. The whole spectrum with 211 wavelength of 39 train set were equal-divided into 5-20 subintervals, and 2,3,4 or more subintervals were separately used to establish the PLS models. Contrasted those PLS models then worked out when the whole spectrum was divided into 6 equal subintervals and modeled with [1 6] synergy interval could obtain the smallest RMSECV, it is 0.01845. Table 1 shown several correction models with better RMSECV who were established by different equal-divides and different synergy interval selected by SiPLS.

Table 1. The results of SiPLS correction model established by different parts and different synergy interval

sequence number	equal-divide number	synergy interval number	synergy interval	spectrum section(nm)	components	RMSECV
1	-	-	whole	1100~2400	5	0.03572
2	6	2	[1 6]	1100~1340, 2200~2400	7	0.01845
3	8	3	[1 7 8]	1100~1280, 2110~2400	6	0.01913
4	8	4	[1 4 7 8]	1100~1280, 1630~1780, 2110~2400	7	0.01915
5	16	4	[1 2 13 14]	1100~1300, 2170~2320	7	0.01849

From Table 1 we can find that the suitable spectrum sections for modeling mainly focus on 1100-1330nm or so and 2110-2400nm,which is the second-harmonic peak of C-H and bending , co-frequency of C-O-H and expansion of vibration spectral section of C-H&O-H. Compares with group 3,group 4 increase the subinterval of 1630-1780nm,but the precision of model's RMSECV descend，and others which are not listed are worse. The courses of above maybe the first-harmonic peak of water molecular O-H pulled down the 1280-1450nm spectrum which are can't be absorbed rapidly, makes alcohol concentration and its NIR transmittance of each sample exists relationship of non-linear, thus causes the unsatisfactory prediction effect of PLS model who is based on linear analysis. Therefore, from the model

results of spectrogram combines with SiPLS algorithm, we can see the sensitive spectrum sections shown at spectrogram not always contain the best characteristic information for model, but the best analysis spectrum section may different because of kinds of experimental facilities(such as type of spectrograph, thickness of cuvette and so on) and methods of quantitative analysis, yet its basic law still can be traced accords to the NIRS absorption characteristic.

4.2. Model validation and evaluation

Trying out using several synergy intervals above to establish PLS model and predict 14 test set separately, then found out the model predicted precision got the best when modeled with 8 equal-division,3 synergy subintervals and 6 components. The dash area of Fig.3 shown the selected spectrum sections for modeling, Fig.4 shown the correlation of model's measured value and predicted value. From Fig.4, we can see the measured value and predicted value basically identical, the correlation coefficient (R) was 0.9996, the root mean square error of prediction (RMSEP) was 0.0067, visible the model's predicted effect was very good. Also shown that the model was reliable, had good generalization and predicted precision.
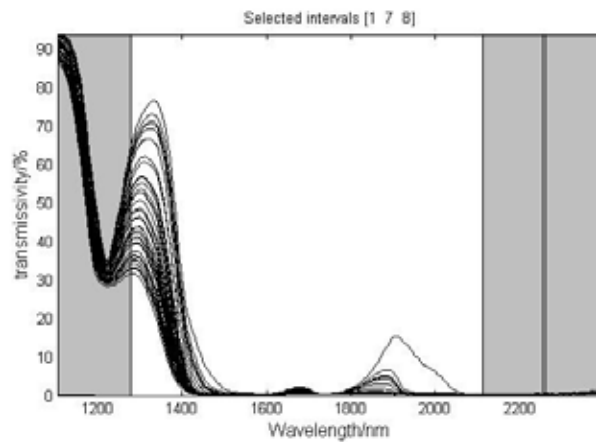

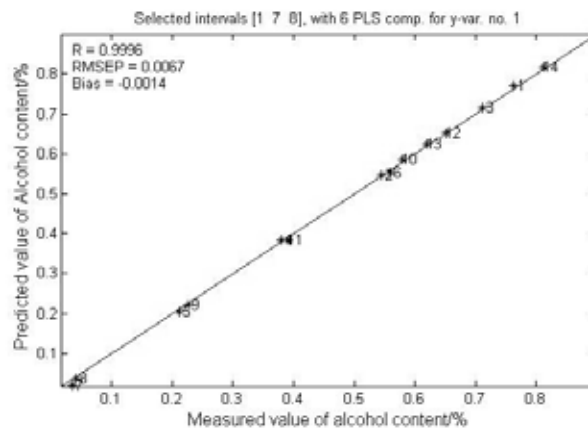Fig.3. the spectrum sections for modeling selected by SiPLS

Fig.4. the correlation of model's measured value and predicted value

5. Conclusion

Through analyzed the evaluate result of NIRS SiPLS model, we can see the SiPLS model who was established by 1100-1330nm and 2110-2400nm spectrum sections had good predictive ability and stability. In liquor-making industry, routine analysis detection in national standard code is very time-consuming, hard sledding, complicated and can't realize on-line analysis during production and provide data timely. The paper experiment model was shown the feasibility of detecting alcohol degree by NIR analysis technology real-timely during brewing process.

References

[1] Daixi Li, Zhiyong Wu, Ruijun Xu, dai-xi Li, Yuanzhi Xu. Measurement of the Principal Compoents in Beer by Means of Near Infrared Spectroscopy. Chinese Journal of Analytical Chemistry, 2004, 32: 1070-1073.

[2] Bang-zhu Peng, Ming-hua Long, Tian-li Yue, Yahong Yuan, Zhihua Zhao. Determination of alcohol concenteation in distilled spirit based on PLS-FT-NIRS.Transactions of the Chinese Society of Agricultural Engineering, 2007, 23: 233-237.

[3] Shang-kun Feng, Hai-ju Xu. Study on Content of Beer Alcohol with Near Infrares Spectroscopy Based on Back-Propagation Neural Network. Infrared Technology, 2008, 30: 58-60.

[4] Cozzolino.D, Kwiatkowski.M.J, Waters.E.J. A feasibility study on the use of visible and short wavelengths in the near-infrared region for the non-destructive measurement of wine composition. Analytical and Bioanalytical Chemistry, 2007, 387: 2289-2295.

[5] Iñón.Fernando A, Llario. Rafae, Garrigues.Salvadorl. Development of a PLS based method for determination of the quality of beers by use of NIR: Spectral ranges and sample-introduction considerations. Analytical & Bioanalytical Chemistry, 2005, 382: 1549-1561.

[6] Chao-xi Geng, Qian Sun, Lei Tian, Donghai Han. Feasibility of Using Ethanol Liquor NIRS Model to Predict Ethanol in Vinous Ferment Liquid. Spectroscopy and Spectral Analysis, 2009, 29: 1805-1808.

[7] Hong-xin Liu, Jun Zhang, Fu-rong Huang, Yong Bing, Lijuan He, Yuping Zeng, et al. In vestigation on the Quality Indicators of Beers Using NIR. Spectroscopy and Spectral Analysis, 2008, 28: 313-316.

[8] Xiao-bang Hu, Jin Lv, Hui-jun Liu, Yang Shi, Tiebing Liu, Bobin Li. Quantitative Model of Alcohol and Acid in Chinese Rice Wine Based on Near-infrared Transmission Spectroscopy. Opto-Electronic Engineering, 2010, 37: 122-127.

[9] Bin Chen, Hao Wang, Song Lin, Jiewen Zhao. Analysis of near infrared spectroscopy of beer alcohol content by correlation coefficients and genetic algorithms. Transactions of the Chinese Society of Agricultural Engineering, 2005, 21: 99-102.