Study on the Near Infrared Model Development of Mixed Liquid Samples by the Algorithm of OSC-PLS*

Dong Wang[1,2,**], Zhihong Ma[2], Shengfeng Ye[1], and Shungeng Min[1,***]

[1] Department of Applied Chemistry, College of Science, China Agricultural University, No. 2, Yuanmingyuan West Road, Haidian District, Beijing 100193, P.R. China
[2] Beijing Research Center for Agri-food Testing and Farmland Monitoring, Beijing Research Center for Information Technology in Agriculture, Beijing, 100097, P.R. China
wangd@nercita.org.cn, minsg@cau.edu.cn

Abstract. The orthogonal signal correction of Wold algorithm was introduced in this thesis, which was combined with the partial least square algorithm. Calibration model of five components of benzene, methylbenzene, chlorobenzene, benzaldehyde and acetophenone in the mixed liquid samples were developed and validated by the method of independent . The result indicated that the calibration models of the components of benzene, methylbenzene, chlorobenzene and benzaldehyde will be developed well after the 1-dimension OSC filter; while a good calibration model of the component of acetophenone will be acquired after the 2-dimension OSC filter.

Keywords: Orthogonal Signal Correction, Partial Least Square, Calibration Model, Near Infrared Spectroscopy, Liquid Samples.

1 Introduction

Near-infrared (NIR) spectroscopy is a kind of non-destructive [1], rapid [2] and environmental friendly analysis method, with the characteristics of easy-acquirement optical source, abundant spectral information and strong penetrability. NIR spectroscopy analysis has been applied in the fields of quality analysis of agricultural products [3-5], petrochemical industry [6], pharmaceutical analysis [7-9] and so forth.

However, the high-strength and complicative background exists in NIR spectroscopy analysis. With the characteristics of serious overlapping band, there are non-correlative information with the target component and the information of the instruments in NIR spectra, which will influence the precision of thecalibration model.

The common multi-signal calibration methods have the risk of eliminating the information correlated with the specified values from spectra, which may decrease the precision of the calibration model.

* Supported by National Science Foundation of China, No. 20575076.
** Major: near-infrared spectroscopy.
*** Corresponding author.

D. Li, Y. Liu, and Y. Chen (Eds.): CCTA 2010, Part III, IFIP AICT 346, pp. 592–597, 2011.

The Algorithm of Orthogonal Signal Correction (OSC) was raised by Wold [10] in 1998 firstly, which is a method of signal processing. The basic idea of OSC is that the non-correlative part of the spectral matrix (X) to the specified values matrix (Y) was eliminated by the mathematical method of orthogonality, after which, a new matrix Xnew was acquired. The calibration model will developed by the matrix Xnew, in order to acquire robust calibration model. Since OSC algorithm eliminated the orthogonal part of the specified values, the valuable information will be kept in spectra matrix, in addition, it may simplify the calibration model for the complicative system.

After Wold raising OSC algorithm, many scholars raised the improvement to OSC. Tom Fearn [11] improved the OSC algorithm in 2000. Johan A. Westerhuis [12] et. al. raised the Direct-OSC algorithm. Robert N. Feudale [13] et. al. raised the Piecewise OSC algorithm. OSC algorithm has been applied in the calibration of NIR data successfully at present [14, 15].

In this thesis, OSC-PLS algorithm will be applied in the calibration model of the NIR spectra of 30 mixed liquid samples. The filtering effect of different dimension of OSC was compared.

2 Experiment

2.1 Instrument and Reagent

Instrument: FT-NIR spectrometer (Spectrum One NTS, Perkin Elmer, U.S.);

Reagents: Benzene (A.R.), methylbenzene (A.R.), chlorobenzene (A.R.), benzaldehyde (A.R.), acetophenone (A.R.).

2.2 Procedure

30 mixed liquid samples, including benzene, methylbenzene, chorobenzene, benzaldehyde and acetophenone, were made up. The information of the samples was shown in Table 1.

The range of NIR spectra collected was 10000 cm^{-1} - 4000 cm^{-1}. The air was regarded as reference in order to collect the background spectrum.

Table 1. The information of the components in the liquid mixed samples

	Benzene	Methylbenzene	Chlorobenzene	Benzaldehyde	Acetophenone
Max (%)	50.0	30.0	10.0	10.0	5.0
Min (%)	1.0	1.0	1.0	0.5	0.1
Mean (%)	9.9	19.4	5.3	4.3	2.0
Std	12.3	8.6	3.0	3.0	1.7

2.3 Data Processing

Software: Matlab v6.5, preprocessing and OSC-PLS program written by self.

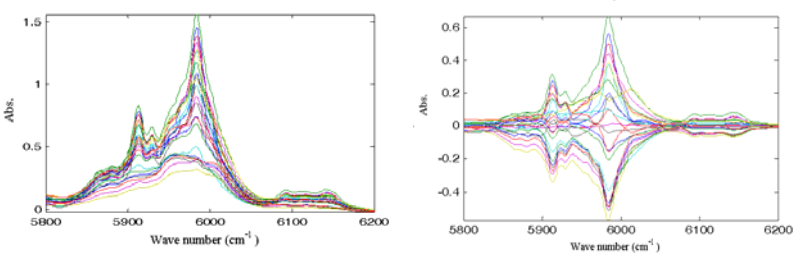
Calibration method: Independent validation, of which, 6 spectra selected according to the concentration gradient were regarded as the validation set for each component;

smooth by Savitzki-Golay (with the width of 5); data normalization method: SNV; OSC-PLS regression.

3 Result and Discussion

3.1 OSC Filter

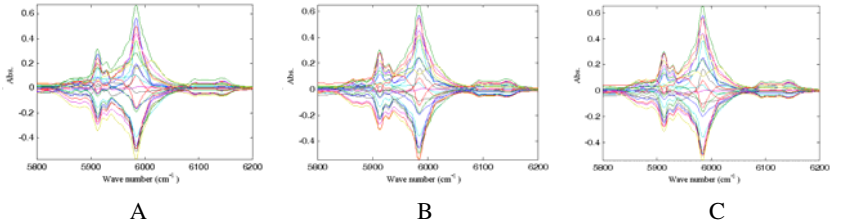
Take the specified values of benzene for example, Fig. 1A is the raw spectra of the 30 mixed liquid samples in the range of 6200 cm^{-1} -5800 cm^{-1}, which is the 1st overtone of C-H group. Centralization of the spectra was done, as is shown in Fig. 1B.



A B

Fig. 1. The raw spectra (A) and the spectra after centralization processing (B)

After the spectra having been processed of centralization, OSC filter with 1-, 2- and 3-dimension was applied to the spectra data. The spectra after OSC filter were shown as Fig. 2.



A B C

Fig. 2. The spectra processed by OSC filter with the specified values of benzene, OSC dimension: 1(A), 2(B), 3(C)

3.2 The Calibration Models of the 5 Components in the Mixed Liquid Samples

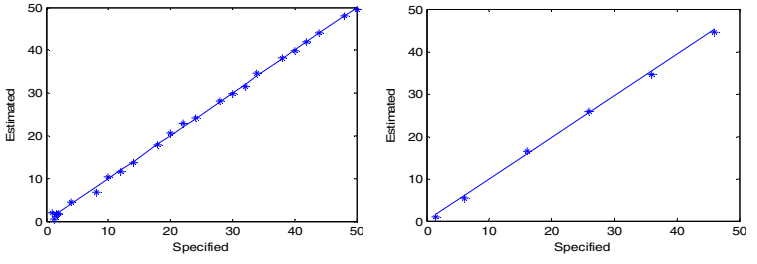
Partial Least Square (PLS) regression was applied to the specified values of benzene and the spectra filtered by 1-, 2- and 3-dimension OSC with the specified values of benzene, the result of which was shown as Tab. 3, in which, R^2_c represents the determination coefficient of the calibration set, while R^2_v represent the determination coefficient of the validation set.

Table 2. The calibration result of benzene after 1-, 2-, 3-dimension OSC filter

OSC dimension	1	2	3
R^2_c (%)	99.90	99.90	99.90
R^2_v (%)	99.69	99.69	99.69
RMSEC	0.54	0.54	0.54
RMSEV	0.97	0.97	0.97
Model dimension	2	2	2

It can be seen from Table 2 that there is no remarkable difference among the results of 1-, 2- and 3-dimension OSC filter. The calibration after 1-dimension OSC filter had the enough precision.

Fig. 3 is the correlation of the estimated values and specified values of the calibration set (A) and the validation set (B) of benzene after 1-dimension OSC filter.



A B

Fig. 3. The correlation of the estimated and specified values of benzene, calibration set: A, validation set: B

It can be seen from Fig. 3 that a high precision can be acquired for benzene after 1-dimension OSC filter and PLS regression.

The calibration result of the other 4 components after 1-, 2- and 3-dimension OSC filter and PLS regression were shown as Table 3.

Table 3. The calibration result after each OSC filter

	Methylbenzene			Chlorobenzene			Benzaldehyde			Acetophenone		
N_{OSC}	1	2	3	1	2	3	1	2	3	1	2	3
R^2_c (%)	99.75	92.89	92.89	99.45	94.23	94.23	99.94	95.55	95.55	99.78	99.78	99.78
R^2_v (%)	99.42	97.10	97.10	96.34	91.45	91.45	84.94	78.15	78.15	97.47	97.97	97.97
RMSEC	0.52	2.65	2.65	0.23	0.74	0.74	0.09	0.66	0.66	0.09	0.09	0.09
RMSEV	0.77	1.71	1.71	0.55	0.84	0.84	1.17	1.41	1.41	0.3	0.27	0.27
N_C	4	2	2	5	4	4	8	3	3	5	5	5

In Tab. 3, NOSC represents the dimension of OSC filter, NC represents the dimension of calibration model.

It can be seen from Table 2 and Table 3 that 1-dimension OSC filter is good for the components of benzene, methylbenzene, chlorobenzene and benzaldehyde since RMSEV is minimum among 1-, 2- and 3-dimension of OSC filter; while 2-dimension OSC filter is good for the component of acetophenone since RMSEV is minimum among the three dimensions of OSC filter but will not decrease when 3-dimension OSC filter being applied.

4 Conclusion

OSC-PLS algorithm was applied to benzene, methylbenzene, chlorobenzene, benzaldehyde and acetophenone in the mixed liquid samples in this thesis. The result indicated that OSC is a good filter for PLS, which can extract the useful information efficiently and acquire precise calibration model.

Acknowledgment

This thesis is supported by National Science Foundation of China, No. 20575076.

References

1. Wu, G., Xiang, B.: Nondestructive Determination of MDMA and MA in Ecstasy by Near Infrared Spectroscopy. Journal of Instrumental Analysis 26(5), 689–701 (2007)
2. Ye, S., Min, S., Qin, F., Li, N., Zhang, M.: Simultaneous Determination of Glucose, Fructose and Sucrose in Aqueous Solution by Near- infrared Spectroscopy. Journal of Instrumental Analysis 22(3), 89–91 (2003)
3. Yang, C.l., Zhao, X., Chen, W., Tian, J., Li, D.: Analyses of Oil Content and Erucic Acid in Small Sample of Rapeseed by Near Infrared Reflectance Spectroscopy. Journal of Instrumental Analysis 26(1), 29–33 (2007)
4. Xiao, X., Xie, X., Mao, X., Luo, W., Chen, J., Liu, Y., Li, X.: Determination of Protein in Milled Rice by Near Infrared Transmittance Spectroscopy. Journal of Instrumental Analysis 23(4), 43–45 (2004)
5. Li, J., Li, H., Liang, J., Wang, Z., Liu, W., Hou, J., Zhao, J., Gao, S.: Analysis of Ash, Total Volatile Acids and Total Volatile Bases in Tobacco with Near Infrared Spectroscopy. Journal of Instrumental Analysis 26(5), 655–657 (2007)
6. Qian, P., Sun, G., Zhang, C.: A Novel Fuzzy Neural Network Method for Diesel Quantitative Analysis with Near Infrared Spectroscopy and Spectral Analysis. Spectroscopy and Spectral Analysis 28(12), 2851–2854 (2008)
7. Liu, M., Meng, Y., Ren, Y., Zhang, H.: Nondestructive Quantitative Analysis of Cofrel Medicines by Improved Partial Least Squares - NIR Spectroscopy. Spectroscopy and Spectral Analysis 27(6), 1098–1101 (2007)
8. Fang, L., Lin, M.: Prediction of Active Substance Contents in Pharmaceutical Tablet Using ICA and NIR. Acta Chimica Sinica 66(15), 1791–1795 (2008)
9. Zhao, W., Dou, Y., Tang, Z., Liu, F., Liu, M., Wang, Y., Ren, Y.: Determination of cimetidine in tablets by near-infrared spectroscopy and PLS. Chinese Journal of Analysis Laboratory 23(10), 85–88 (2004)

10. Wold, S., Antti, H., Lindgren, F., Öhman, J.: Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent Laboratory Systems 44, 175–185 (1998)
11. Fearn, T.: On orthogonal signal correction. Chemometrics and Intelligent Laboratory Systems 50, 47–52 (2000)
12. Westerhuis, J.A., De Jong, S., Smilde, A.K.: Direct orthogonal signal correction. Chemometrics and Intelligent Laboratory Systems 56, 13–25 (2001)
13. Feudale, R.N., Tan, H., Brown, S.D.: Piecewise Orthogonal Signal Correction. Chemometrics and Intelligent Laboratory Systems 63, 129–138 (2002)
14. Blanco, M., Coello, J., Montoliu, I., Romero, M.A.: Orthogonal signal correction in near infrared calibration. Analytica Chimica Acta 434, 125–132 (2001)
15. Bertran, E., Iturriaga, H., Maspoch, S., Montoliu, I.: Effect of orthogonal signal correction on the determination of compounds with very similar near infrared spectra. Analytica Chimica Acta 431, 303–311 (2001)