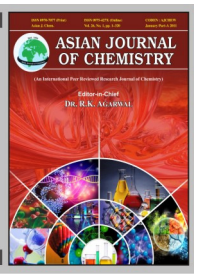
Asian Journal of Chemistry; Vol. 26, No. 2 (2014), 467-471

ASIAN JOURNAL OF CHEMISTRY

http://dx.doi.org/10.14233/ajchem.2014.15447



Raman Spectrum Online Monitoring in Aspirin Synthesis Process

HUI SU, KUO SUN, ZHIXIANG YAO*, PEIXIAN HUANG and LIU LIU

Department of Biology and Chemical Engineering, Guangxi University of Science and Technology, Liuzhou 545006, Guangxi P.R. China

*Corresponding author: E-mail: hailfellow512@163.com

Received: 7 March 2013; *Accepted*: 31 July 2013; *Published online*: 15 January 2014; AJC-14568

In this work, the Raman spectra were obtained directly in the synthesis process of aspirin as prediction set. The concentration sample according to the amount of material changes before and after reaction and it's extracted Raman information made up training set and a quantitative analysis model of the salicylic acid-acetic anhydride-aspirin mixed system was established by applying partial least squares (PLS). The values of correlation coefficient squared between measured and regression (R^2), root mean square error of prediction set (RMSEP), mean deviation (MD) indicate that the influence of the prediction precision by the baseline drift can be effectively reduced in the method of spectrum pretreatment *via* mormalize + savitzky-go1ay smoothing first derivative. After the prediction set was used in this analysis mode, the mass per cent between salicylic acid and aspirin of online monitoring was 4.59 %, while the off-line Raman analysis and high performance liquid chromatography analysis results were 5.66 and 2.75 %. The relative deviations were -1.07 and 1.84 %. This method provides effective data support for synthesis of aspirin Raman spectroscopy online monitoring.

Keywords: Pretreatment, Partial least squares, Online monitoring, Raman spectroscopy, Aspirin synthesis.

INTRODUCTION

Increasing demands on process safety, product quality and cost reduction are the main reasons for the growing interest in the online monitoring whose features are real-time, non-destructive, rapid analysis. The traditional analysis methods (titration, high performance liquid chromatography (HPLC)) cannot achieve requirements for online monitoring. Raman spectroscopy analysis has advantages whose properties are non-destructive monitoring, high sensitivity, short detection time *etc.* and Raman spectrum information on reaction medium can be rapidly obtained with a fiber optic probe. Simultaneous determination of multicomponent can be performed in this method[1-4]. Due to the presence of multiple characteristic peaks in the Raman spectrum data of measured components and the characteristic information of measured components were overlapping[5], it is difficult to separate characteristic peaks to determine the component concentration of the synthesis process. In order to overcome the influence of multiple correlation caused by overlapping information, multivariate calibration techniques need to be used in Raman spectrum analysis[6-8]. To determine the content of known object component that is white analytical system[9], both principal component regression (PCR) and partial least squares (PLS) can extract component variable characteristic information[10,11], when the latter decompose spectral matrix, it makes the region of the spectrum with the measured remarkable attribute greater weight and the measured component attribute matrix is considered[12]. Thomas and Haaland[13] compared classical least squares (CLS), PCR with partial least squares, they draw the conclusion that partial least squares was the optimal method for multicomponent prediction.

However, during the application of Raman spectroscopy, fluorescence of organic compounds in the samples, sometimes several orders of magnitude more intense than the weak Raman scatter and interfere with the Raman signals[14-16]. So in order to eliminate fluorescence influence, to remove a portion of the noise and increase the signal to noise ratio[17], pretreatment should be adopted in normal conditions before regression. The normal pretreatment methods include normalization, smoothing, derivative, *etc.*[18-26].

This work chose sulphamic acid catalytic synthesis of aspirin system as a study object and adopted the Raman spectrum analysis method based on the partial least squares algorithm *via* spectrum pretreatment to test and verify online monitoring of three components in aspirin synthesis process[27].

Theory

Partial least squares: The spectra are collected as rows in a matrix X, with n rows and k columns. Each row represents a spectrum and each column a single wavelength, the corresponding, concentration of major components are placed as rows in a matrix y.

Partial least squares is a regression method with a matrix X and a vector y (or matrix). Partial least squares calculates

components that both capture the variation in X and correlates with the variation in y. The loading weight vector w_a, where a denotes partial least squares component number, shows the spectral features in X that correlate with y, calculated as in eqn. 1. The score vector, which can be interpreted, is calculated as in eqn. 2. A scalar coefficient is then calculated (eqn. 4) which in turn together with constitutes an equation for y, where f is a residual with non modeled variation in y (eqn. 4).

$$w_a = \frac{X'_{a-1}y_{a-1}}{X_{a-1}y_{a-1}} \quad (1)$$

$$t_a = X_{a-1}w_a \quad (2)$$

$$c_a = \frac{t'_a y_{a-1}}{t_a t_a} \quad (3)$$

$$y = c_1t_1 + c_2t_2 + \ldots + c_At_A + f \quad (4)$$

The number of partial least squares components to be used is often selected with cross validation and/or independent test sample sets. For partial least squares, the sign of the loading vector is determined by the direction of the change in y.

The predictive performance was assessed on the basis of the root mean squared error of calibration and prediction, calculated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n}} \quad (5)$$

Normalization: Normalization methods normally include area normalization, maximum normalization and vector normalization. Vector normalization is commonly used as a Raman spectral data processing method to correct the spectral changes caused by the small optical path distinction, which is following equation:

$$x_{i,norm} = \frac{x_i}{\sqrt{\sum_{i=1}^{k} x_i^2}} \quad (6)$$

where $x_{i,norm}$ is the transformed Raman intensity for wavenumber x_i, k wavenumbers is the spectrum. Eqn. 6 shall be repeated for all k wavenumbers in the spectrum.

Smoothing: Smoothing can effectively eliminate the high frequency component of the spectrum while retaining low frequency components, can effectively improve signal to noise ratio. The weakness in this method was that useful high frequency signal data may therefore suffer losses. In this connection Gorry proposed the convolution least squares smoothing method based on Gram polynomial recursive nature[21-24]. We define the order n polynomial model function f(x) as eqn. 7.

$$f(x) = a_0 + a_1x + a_xx^2 + \ldots a_{k-1}x^{k-1} = \sum_{j=0}^{n} a_jx^j \quad (7)$$

$$f(x_i) = g_i + \varepsilon_i \quad (8)$$

$$\begin{pmatrix} f(-m) \\ f(-m+1) \\ \vdots \\ f(m) \end{pmatrix} = \begin{pmatrix} 1 & -m & \ldots & (-m)^n \\ 1 & -m+1 & \ldots & (-m+1)^n \\ \vdots & \vdots & \vdots & \vdots \\ 1 & m & \ldots & m^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad (9)$$

$$BA = g \quad (10)$$

$$A = (B^TB)^{-1}B^Tg \quad (11)$$

$$\hat{g} = BA = B(B^TB)^{-1}B^Tg = Bg \quad (12)$$

$$K = B(B^TB)^{-1}B^T \quad (13)$$

$$i = 2m+1 \quad (14)$$

here, B is a matrix whose rows provide the terms of the polynomial f(x) which take the coefficients in a and g is same as f(x). $\hat{g}$ is the filtering value. K is the filtering coefficient matrix, which is only relevant to both filtering window and polynomial order n.

Derivative: Spectrum of the first order and second order derivative are commonly used in pretreatment methods of spectral analysis to correct baseline drift and extract spectral characteristic value. Compared with the original spectrum signal, derivative spectrum can effectively eliminate the interference of baseline drift and background[26]. But the derivation may reduce the signal to noise ratio. Thus original spectra need to be pretreated in the method of smoothing before derivation. Savitzky-Golay convolution first order derivative is relatively commonly used as eqn. 15

$$\frac{\partial f}{\partial x} = a_1 \quad (15)$$

EXPERIMENTAL

Salicylic acid (AR), acetylsalicylic acid (AR), acetic anhydride (AR), acetic acid (AR), amino sulfuric acid (AR) were used as received. Laser Aperture (Laser-785, Danger, America). Raman Spectrometer (Scientific-grade QE65000, Ocean optics, America). Fiber Optic Probe (BAC100-785-OEM, Ocean optics, America). LC (1200 SL, Agilent, America). Electronic balance (ALC-210.4, Sartorius, German). Quartz cuvettes(1cm).Three-necked round-bottomed flask (100 mL). Water bath, *etc*.

Raman platform setting: A laser with the wavelength of 785 mm was used as the excitation light source and Raman information was obtained through the optical fiber probe, the setting of ocean optics spectra suite was that the x axis on workstation menu was selected Raman shift; selected integral time was 1/s to obtain the Raman spectrum of 0-2000 cm^{-1} spectral range.

Preparation of training sample set: Aspirin synthetic route are shown in Fig. 1.

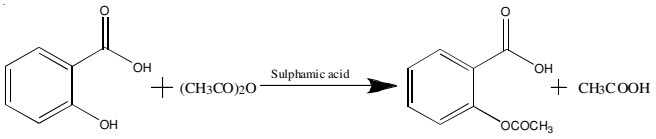


Fig. 1. Aspirin synthetic route

As Fig. 1 shows, suppose that salicylic acid completely reacted at the end of the reaction, before and after the reaction, the mole ratio of salicylic acid, acetic anhydride, aspirin are n: 2 n: 0 and 0: n: n. 41 standard samples were prepared by mixing appropriate amounts of salicylic acid, acetic anhydride, aspirin and acetic acid. According to mole number range of salicylic acid, acetic anhydride, aspirin, respectively was 5-0, 10-5 and 0-5 mmol, the three materials were weighed to 10 mL volumetric flasks by decreasing and increasing every 0.125 mmol and acetic acid was used as solvent to 10 mL constant volume.

The samples were, respectively transferred to 1 cm optical path quartz cuvettes to collect and save its Raman spectrum information as the training spectroscopy set.

Preparation of prediction sample set: According to the literature[27], we precisely weighed acetic anhydride 41 g, salicylic acid 27.7 g, sulphamic acid 0.5 g. When the temperature rise to about 81 ºC using a water bath, acetic anhydride, salicylic acid and sulphamic acid was transferred to 100 mL three-necked round-bottomed flask one after another, the reaction was lasting for 18 min by magnetic stirring at 81 ºC. Raman spectrum was saved every minute as prediction sample through scanning reaction system. The online monitoring device is shown in Fig. 2.

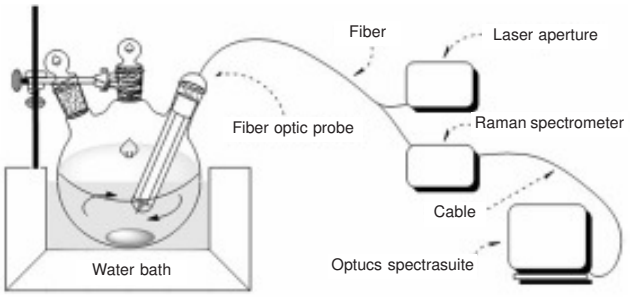


Fig. 2. Online monitoring device

Preparation of off-line sample: Because of short reaction period of this reaction process, off-line sample can not be prepared per minute. We sampled at equilibrium of reaction approximately 10 min after sulphamic acid was added. 1 g reaction sample was taken to a 10 mL volumetric flask using acetic acid to volume and mixed. Its Raman information as the off-line spectroscopy set was extracted and saved three times, the relative content of salicylic acid and aspirin was determined by Agilent 1200 HPLC system (Agilent Technologies, USA) equipped with a vacuum degasser, a quaternary pump, an autosampler, a thermostatic column compartment, a diode array detector (DAD). Separation was performed on an Agilent Zorbax SB-C_{18} column (250 mm × 4.6 mm, 5 μm), using methanol 0.5 % acetic acid (30:70, v/v) as mobile phase. The detection wavelength was set at 278 nm. The flow rate of the mobile phase was 1 mL/min.

RESULTS AND DISCUSSION

Spectrum pretreatment: It was investigated that derivatives can reduce peak overlap and eliminate constant and linear baseline drift[28]. The principle is that inflection points in close peaks become turning points in the derivatives. To eliminate the spectral differences from the baseline shifts, first derivative spectra were adopted as pretreatment model. Raman spectrum data of both Training set and Prediction set were preprocessed *via* normalize + Savitzky-Go1ay smoothing first derivative (Fig. 3).

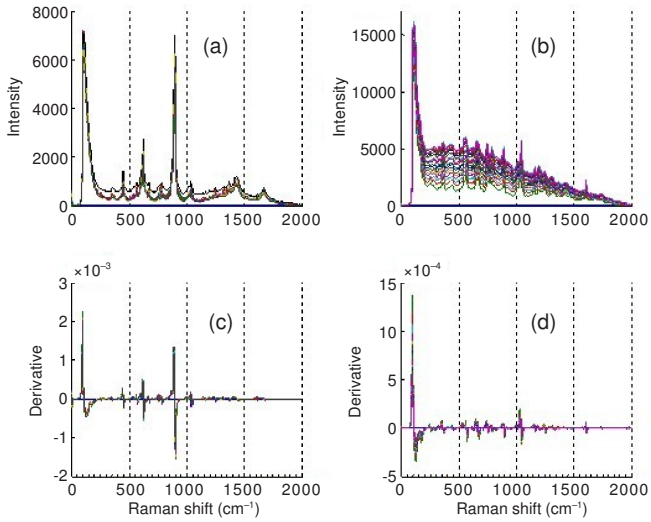


Fig. 3. Raman spectrum of before and after pretreatment. (a) Raman spectrum of training set; (b) Raman spectrum of prediction set; (c) Raman spectrum of training set *via* pretreatment; (d) Raman spectrum of prediction set *via* pretreatment

Fig. 3(a-b) show that there are baseline drift in the Raman spectrum of both training set and prediction set, especially in prediction set, degree of baseline drift is larger as Fig. 3(b) shown. Fig. 3(a)(c) display the Raman feature of acetic acid background in training set exist sharp and well-defined peaks at 600 and 900-800 cm^{-1} and observed Raman peaks of both acetic acid and measured components (salicylic acid, acetic anhydride and aspirin) exit in 200-50 cm^{-1}. There are Raman features of measured components clearly shown in Fig. 3(b-d).

Form Fig. 3 it can be deduced that the influence of the baseline drift can be reduced and the Raman spectrum differences caused by chemical component enhanced *via* normalize + Savitzky-Go1ay smoothing first derivative pretreatment.

Establishment of the training set: Ten samples used as test set were selected every 4 counts from 3rd to the 39th in 41 samples of the training set, others were used as calibration set. Principal factors that main spectrum variables caused by changes of sample in training set, they are acetylsalicylic acid, salicylic acid, acetic acid and acetic anhydride. Selected number of partial least squares factors was 4.

Evaluation of the training set: Component concentration and Raman spectrum of all wavelength ranges were regressed by both direct partial least squares regression and partial least squares regression *via* normalize + Savitzky-Go1ay smoothing (polynomial = 2, points = 5) first derivative. In order to assess regression effect, some parameters of analysis were made for the partial least squares model, they were values of R^2, RMSECV, RMSEP and MD[29,30]. Regression effects were shown in the Fig. 4.

TABLE-1
EVALUATION PARAMETERS COMPARISON OF TRAINING MODEL

Component	R^2		RMSEP (mg mL^{-1})		MD (mg mL^{-1})	
	Un-pretreatment	Pretreatment	Un-pretreatment	Pretreatment	Un-pretreatment	Pretreatment
Salicylic acid	0.9910	0.9970	2.1032	1.1228	-0.6960	-0.1904
Acetic anhydride	0.9850	0.9960	2.1377	1.0966	-1.0750	0.6271
Aspirin	0.9900	0.9970	2.8227	1.5266	0.8574	0.1762

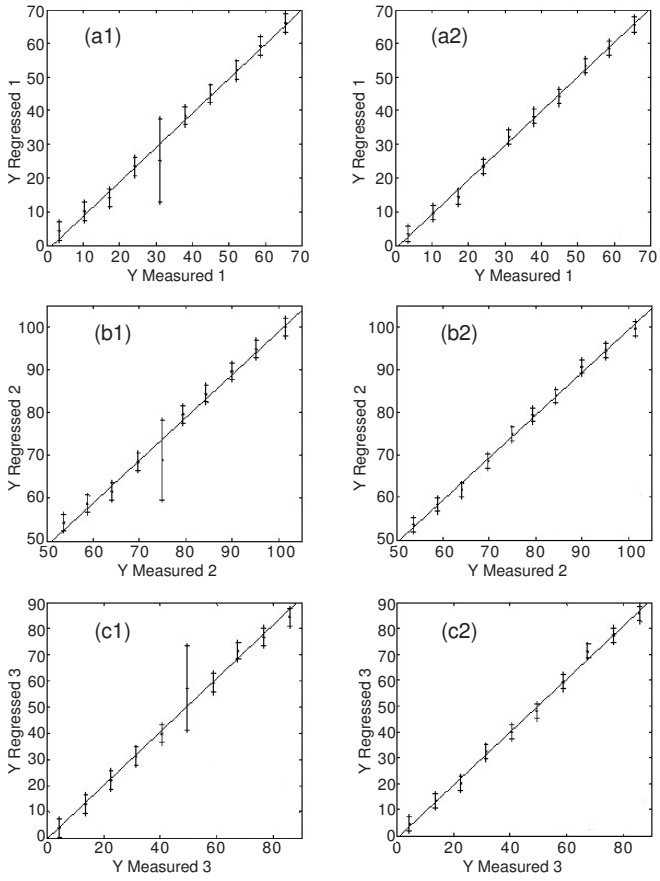


Fig. 4. Evaluation of training model. (a1) Salicylic acid correlogram by direct partial least squares regression; (b1) Acetic anhydride correlogram by direct partial least squares regression; (c1) Aspirin correlogram by direct partial least squares regression; (a2) Salicylic acid correlogram by partial least squares regression *via* spectrum pretreatment; (b2) Acetic anhydride correlogram by partial least squares regression *via* Spectrum pretreatment; (c2) Aspirin correlogram by partial least squares regression *via* Spectrum pretreatment

In Fig. 4, (a1) (b1) (c1) show correlation between measured and regressed by direct partial least squares regression and there are most of residual levers of the sample point evenly distributing near the fit line but the 6th sample point. From Fig. 3a the conclusion can be drawn that the large residual lever of the 6th sample point was caused by its Raman spectrum obviously drifting in the training set. While (a2) (b2) (c2) show correlation between measured and regressed *via* normalize + Savitzky-Go1ay smoothing (polynomial = 2, points = 5) first derivative. All residual levers of the sample point evenly distribute near the fit line, no obvious high lever exist.

Fig. 4 indicates above that the influence of the prediction precision by the baseline drift can be effectively reduced *via* normalize + Savitzky-Go1ay smoothing (polynomial = 2, points = 5) first derivative pretreatment.

Related evaluation parameters of training model are shown in Table-1.

Table-1 resumes the effect of partial least squares regression *via* normalize + Savitzky-Go1ay smoothing (polynomial = 2, points = 5) first derivative pretreatment in the training set are optimistic compared with unpretreatment. Fig. 4 (Table-1) deduced that the influence of the prediction precision by the baseline drift can be effectively reduced in this pretreatment method.

Application of analysis model

Results of off-line determination: The off-line spectroscopy set was applied in the calibration set, after the form of each component concentration in calibration set was translated into the mass percent by the principle of mass conservation. The results are shown in the Table-2.

TABLE-2
RESULTS OF OFF-LINE RAMAN ANALYSIS

No.	Salicylic acid (%, m/m)	Acetic anhydride (%, m/m)	Aspirin (%, m/m)
1	2.7624	31.9216	48.9898
2	2.7841	31.9376	48.9614
3	2.7710	31.9279	48.9785
Mean value	2.7725	31.9291	48.9766

The offline Raman analysis result of percentage between salicylic acid and aspirin was 5.66 % and it was 2.75 % in the method of HPLC.

Application in online monitoring: Due to concentration range of components between calibration set and reaction solution is inconsistent, the form of each component concentration need to translated into the mass percent by principle of mass conservation. Mass percent of each component in synthetic process were obtained after spectral data of prediction set was applied in the calibration set. The results are shown in Table-3.

The data in Table-3 multiplied by the total mass of reactant, the mass of each component every minute in synthetic process were obtained and they were shown in the Fig. 5.

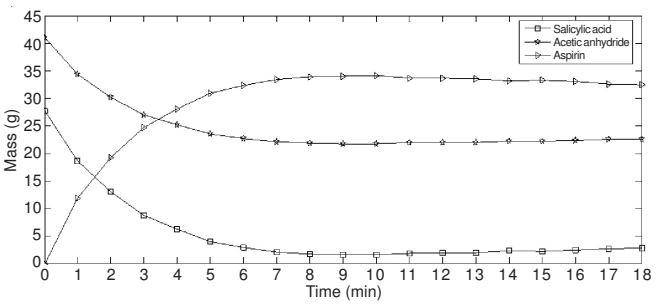


Fig. 5. On-line results of each component

TABLE-3
MASS PER CENT OF PREDICTION SET

Time (min)	Salicylic acid (%, m/m)	Acetic anhydride (%, m/m)	Aspirin (%, m/m)	Time (min)	Salicylic acid (%, m/m)	Acetic anhydride (%, m/m)	Aspirin (%, m/m)
1	27.1054	49.9122	17.2378	10	2.2768	31.5598	49.6253
2	18.9049	43.8504	27.9351	11	2.7394	31.9019	49.0217
3	12.7266	39.2836	35.9943	12	2.7902	31.9395	48.9554
4	9.0669	36.5784	40.7682	13	2.8844	32.0091	48.8326
5	5.8254	34.1823	44.9967	14	3.3036	32.3191	48.2857
6	4.1926	32.9755	47.1265	15	3.2374	32.2702	48.3720
7	3.1053	32.1719	48.5448	16	3.4971	32.4622	48.0332
8	2.4921	31.7187	49.3447	17	3.9492	32.7966	47.4433
9	2.3484	31.6126	49.5320	18	4.0881	32.8992	47.2622

Fig. 5 indicates that, in this synthetic process, 10 min after initial reaction 0.5g sulphamic acid was added, the reaction reached equilibrium. Then the online result of percentage between salicylic acid and aspirin was 4.59 %. Compared with results of offline Raman analysis and HPLC, the relative deviations are -1.07 and 1.84 %.

Conclusion

This work used Raman spectroscopy characteristics of salicylic acid, acetic anhydride and aspirin in the synthesis system of aspirin under sulphamic acid catalytic, based on a study in the changes of the material in the reaction process for model optimization, to establish a rapid simultaneous quantitative analysis model for salicylic acid, acetic anhydride and aspirin. In this analysis model, by using partial least squares regression *via* Normalize + Savitzky-Go1ay smoothing (polynomial = 2, points = 5) first derivative pretreatment, selected number of partial least squares factors was 4. Online determination of the three components in aspirin synthesis process can be achieved. This analytical method can be used in process monitoring of other synthetic system and offer base data for a dynamic optimization control.

ACKNOWLEDGEMENTS

The authors acknowledged the foundation support from Science and Technology Planning Project of Guangxi (129825-20 and 1355010-15).

REFERENCES

1. C. Wang, T.J. Vickers, J.B. Schlenoff and C.K. Mann, *Appl. Spectrosc*., **46**, 1729 (1992).
2. C. Wang, T.J. Vickers and C.K. Mann, *Appl. Spectrosc*., **47**, 928 (1993).
3. J.B. Cooper, K.L. Wise and B.J. Jensen, *Anal. Chem*., **69**, 1973 (1997).
4. N. Everall and B. King, *Macromol. Symp*., **141**, 103 (1999).
5. G.V.L.N. Murty and T.R. Seshadri, *Proc. Math. Sci*., **19**, 17 (1944).
6. J.B. Cooper, *Chemom. Intell. Lab. Syst*., **46**, 231 (1999).
7. F. Estienne, D.L. Massart, N. Zanier-Szydlowski and P. Marteau, *Anal. Chim. Acta*, **424**, 185 (2000).
8. H. Witjes, M. van den Brink, W.J. Melssen and L.M.C. Buydens, *Chemom. Intell. Lab. Syst*., **52**, 105 (2000).
9. Y. Liang, Multivariate Calibration Methods for White Analytical Systems, In: Q. Chen, editor. White, Grey and Black Multicomponent Systems and Their Chemometric Algorithms. Changsha, China: Hunan Publishing House of Science and Technology, pp. 29-36 (1996).
10. F. Estienne and D.L. Massart, *Anal. Chim. Acta*, **450**, 123 (2001).
11. J.B. Cooper, K.L. Wise, W.T. Welch, M.B. Sumner, B.K. Wilt and R.R. Bledsoe, *Appl. Spectrosc*., **51**, 1613 (1997).
12. P. Geladi and B.R. Kowalski, *Anal. Chim. Acta*, **185**, 1 (1986).
13. E.V. Thomas and D.M. Haaland, *Anal. Chem*., **62**, 109l (1990).
14. C.L. Stanford, B.A. Mantooth and B.T. Jones, *J. Chem. Educ*., **78**, 1221 (2001).
15. Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang and J. Hu, *Chemom. Intell. Lab. Syst*., **85**, 94 (2007).
16. R.L. McCreery, In ed.: J.D. Winefordne, Magnitude of Raman Scattering, Raman Spectroscopy for Chemical Analysis, Wiley-Interscience, New York, USA, pp. 25-26 (2000).
17. O.E. De Noord, *Chemom. Intell. Lab. Syst.*, **23**, 65 (1994).
18. Y.G. Shi, X.L. Feng and Z.C. Li, In ed.: G.Y. Zhang, Chemical Pattern Recognition, Chemometrics.Beijing, China: China Petrochemical Press, pp. 141-143 (2002).
19. K. Kachrimanis, D.E. Braun and U.J. Griesser, *J. Pharm. Biomed. Anal*., **43**, 407 (2007).
20. W. Wu, Q. Guo, D. Jouan-Rimbaud and D.L. Massart, *Chemom. Intell. Lab. Syst.*, **45**, 39 (1999).
21. A. Savitzky and M.J.E. Golay, *Anal. Chem*., **36**, 1627 (1964).
22. H.H. Madden, *Anal. Chem*., **50**, 1383 (1978).
23. J. Steinier, Y. Termonia and J. Deltour, *Anal. Chem*., **44**, 1906 (1972).
24. P.A. Gorry, *Anal. Chem*., **62**, 570 (1990).
25. C.D. Brown and P.D. Wentzell, *Chemometrics*, **13**, 133 (1999).
26. L.S. Mendes, F.C.C. Oliveira, P.A.Z. Suarez and J.C. Rubim, *Anal. Chim. Acta*, **493**, 219 (2003).
27. S. Yang, *J. Kunming Teachers College*, **29**, 108 (2007).
28. W.J. McCarthy, TQ Analyst User's Guide, Thermo Nicolet Corp., Madison, WI, p. 155 (2003).
29. S. Macho and M.S. Larrechi, *TrAC-Trends Anal. Chem*., **21**, 799 (2002).
30. J.M. Andrade, S. Garrigues, M. de la Guardia, M. Gómez-Carracedo and D. Prada, *Anal. Chim. Acta*, **482**, 115 (2003).