Analyst

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2013, **138**, 4483

Morphological weighted penalized least squares for background correction

Zhong Li,†[a] De-Jian Zhan,†[b] Jia-Jun Wang,[c] Jing Huang,[a] Qing-Song Xu,*[b] Zhi-Min Zhang,[d] Yi-Bao Zheng,[d] Yi-Zeng Liang*[d] and Hong Wang[b]

Received 13th April 2013
Accepted 20th May 2013

DOI: 10.1039/c3an00743j

www.rsc.org/analyst

Backgrounds existing in the analytical signal always impair the effectiveness of signals and compromise selectivity and sensitivity of analytical methods. In order to perform further qualitative or quantitative analysis, the background should be corrected with a reasonable method. For this purpose, a new automatic method for background correction, which is based on morphological operations and weighted penalized least squares (MPLS), has been developed in this paper. It requires neither prior knowledge about the background nor an iteration procedure or manual selection of a suitable local minimum value. The method has been successfully applied to simulated datasets as well as experimental datasets from different instruments. The results show that the method is quite flexible and could handle different kinds of backgrounds. The proposed MPLS method is implemented and available as an open source package at http://code.google.com/p/mpls.

Introduction

Chromatography and Raman spectroscopy have been increasingly used in a wide range of fields. However, background drift often occurs because of instrument fluctuation in chromatography or fluorescent background from fluorescent substances in Raman spectra. The background always blurs or even swamps spectral signals and then impairs the accuracy of calibration and classification. In order to improve the prediction ability of the model in multivariate calibration, classification or clustering, it is necessary to eliminate, or at least reduce, the background in the analytical signals.[1–4]

Recently, more and more attention has been paid to background correction by researchers using chromatography[2,5,11] and Raman spectroscopy[3,4] for quantitative analysis. Backgrounds can be rectified through either experimental methods or chemometric methods. For the methods utilizing experimental techniques to reject backgrounds, excitation with specific wavelength(s),[6–8] standard addition based derivative spectra[9] and preprocessing the measured samples[10] are the common choices. These methods either involve experimental techniques that work for certain sample types or are expensive to implement, which have limited their extensive applications. So, researchers have proposed large numbers of chemometric methods for background correction based on its sample-independent smooth curve characteristics. The studies with regard to background correction methods are mainly scattered in the literature of chromatography and Raman spectroscopy.

For the chromatographic dataset, background is a signal caused by the chromatographic instrument fluctuations and/or experimental conditions, which is not related to the analytes. Background correction of chromatograms is an important issue for chromatographic analysis. David modeled the background by the abstract spectra obtained from factor analysis and adaptive Kalman filter.[11] Liang *et al.* proposed a two-way procedure for background correction based on least-squares fit of the zero-component regions.[12] The wavelet had also been applied to correct background drift in HPLC by Pan *et al.*[13] and Shao *et al.*[14,15] The orthogonal projections were used to correct the background and resolve the overlapping peaks simultaneously.[16] Boelens *et al.* proposed a new background correction method for HPLC-DAD using PCA modeling and asymmetric least squares regression.[17] Quintás *et al.* have introduced a series of background correction methods for on-line liquid chromatography infrared spectrometry (LC-IR).[18–22] An iteratively reweighted procedure and penalized least squares were combined together by Zhang *et al.* for background correction of HPLC.[23,24] Komsta employed the quantile regression for background elimination.[25] Singular value decomposition-based background correction (SVD-BC) was also proposed for online comprehensive two-dimensional liquid chromatography (LC × LC) data.[26] Filgueira *et al.* applied 1D background correction methods to 2D backgrounds in LC × LC orthogonally, which could remove extremely

[a]*Yunnan Academy of Tobacco Science, Kunming 650106, P. R. China*
[b]*School of Mathematics and Statistics, Central South University, Changsha 410083, P. R. China. E-mail: qsxu@csu.edu.cn; Tel: +86-13787192436*
[c]*Hongyunhonghe Tobacco (Group) Co., Ltd., Kunming 650106, P. R. China*
[d]*College of Chemistry and Chemical Engineering, Research Center of Modernization of Chinese Medicines, Central South University, Changsha 410083, P. R. China. E-mail: yizeng_liang@263.net; Tel: +86-731-8830824*
† These authors contributed equally to this paper.

serious background disturbances from the change in the refractive index of the mobile phase in very fast gradients.[27]

Background correction is also a crucial step which must be performed prior to peak detection or any quantification procedure in Raman spectroscopy. Linear correlation was used for both laser-induced breakdown and Raman spectra, which has successfully eliminated the continuum background without compromising spectral integrity.[28] An automatic method based on modified least-squares polynomial curve fitting was also proposed for fluorescence subtraction in Raman spectroscopy for biological applications.[29] Jirasek *et al.* investigated the accuracy and precision of manual background determination and removal with 16 volunteers, which found that the users' experience had a great influence on the correction results.[30] By designing and minimizing a non-quadratic cost function, a fast and simple method was devised.[31] Gan *et al.* fitted the background with automatic thresholds and polynomial fitting in the iterative processes.[32] By splitting Raman spectra into different frequency components by wavelet transformation, and then the background can be corrected *via* removing the varying low-frequency background.[33,34] Based on peak detection, smoothing, interpolation and linear programming, two simple background elimination methods were proposed by Baek *et al.*[35,36] BaselineWavelet,[37] which consists of peak detection by wavelet and penalized least squares, was used for correcting the fluorescence background and improving the clustering analysis results of medicine tablets.[38] With a customizing wrapper to rescale the spectrum abscissa, customized background correction could behave right in all parts of the spectra.[39] Recently, SVD-based,[40] automated iterative moving averaging,[41] large-window moving average, small-window moving average,[42] iterative background modeling,[43] geometric approach,[44] range-independent background subtraction,[45] hybrid least squares,[46] and adaptive least-squares thresholding[47] have been proposed for background correction.

From previous literature, (penalized) least squares, iteration with threshold, moving averaging, and peak detection are techniques which are commonly used for background estimation. In this paper, a novel, fast and flexible background fitting algorithm is proposed based on the morphological opening operation and weighted penalized least squares (MPLS). The morphological opening operation can search the local minimum value intelligently. Then the weight vector is constructed from the local minimum value by using the penalized least squares. The vector can fit the background accurately without iteration.

The paper is organized as follows. Mathematical concepts relevant to the MPLS algorithm are presented and investigated in the Theory section. Then, the MPLS algorithm is applied to the simulated data, chromatograms and Raman spectra to demonstrate its performance. Results of the above applications are discussed together with discussions about the proposed algorithm. Finally, some conclusions and perspectives are given in the Conclusion section.

Theory

The most important step in MPLS is background fitting *via* morphological opening operation. But it will introduce flaws in the peak region that will change the shape. In order to compensate the shortcomings of the opening operation, the rough background fitted by the morphological opening operation and the local minimum value are used as weight vectors of penalized least squares, respectively, for background refinement.

Mathematical morphology

Mathematical morphology is a theory and technique for the analysis and processing of geometrical structures, based on classical set operations. Mathematical morphology is the foundation of morphological image processing, which consists of a set of operators that transform images. Recently, mathematical morphology has been used in different research fields where the shape is the most important characteristic. Also, it can be applied to the processing of one-dimensional signals such as spectra. Rosanna *et al.* introduced morphology-based automated background removal for Raman spectra of artistic pigments.[48]

Most morphological operations are based on simple expanding and shrinking operations. These transformations involve the interaction between a signal A (the object of interest) and a structuring set B, called the structuring element. Typically the structuring element B is a circular disc in the plane, and it can be of any shape. The two basic morphological set transformations are erosion and dilation, respectively. These operators are the basis of any other morphological transformations.

In order to adapt the above operator definitions to one-dimensional signal processing, the structuring element B is defined as a plane structuring element, a window centered at x and half the window width at w.

Erosion of the object A by the structuring element B is defined with the familiar function as:

$$\text{erosion}(x_i) = \min(x_{i+j}),\ j = -w,\ \ldots,\ w \tag{1}$$

Dilation of the object A by a structuring element B is given by

$$\text{dilation}(x_i) = \max(x_{i+j}),\ j = -w,\ \ldots,\ w \tag{2}$$

The opening of A by B, denoted by $opening(x_i)$, is given by the erosion by B, followed by the dilation by B, that is

$$\text{opening}(x_i) = \text{dilation}[\text{erosion}(x_i)] \tag{3}$$

Fig. 1(a) shows the erosion and dilation of a signal by a half window of 15. The red line is the result obtained by the erosion operation (see also eqn (1)), while the blue line shows the result of the dilation operation (see also eqn (2)). At first glance, both erosion and dilation operations are not the need as the contours of the spectrum change too much after transformation by dilation or erosion. Fortunately, the opening operation, consisting of the dilation and erosion, can generate a rough background, which can be used for fulfilling the task of background removal. It can be seen that if the result obtained by erosion operation (Fig. 1(b)) continues to be operated by dilation operation, say the dilation operation acting upon the red line obtained by erosion operation, one can get the blue line shown in Fig. 1 (c), which is

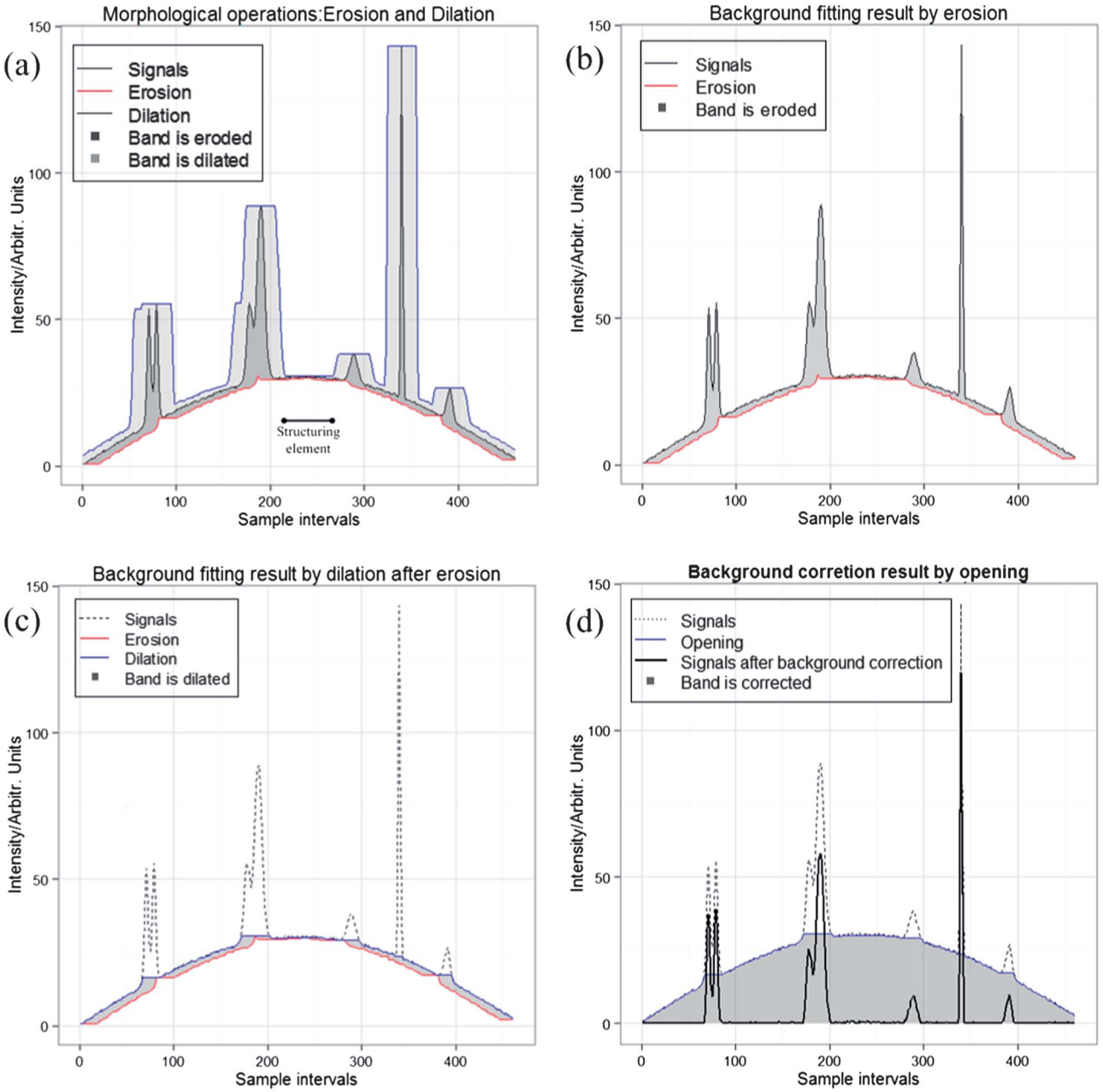

Fig. 1 Illustration of morphological operations: (a) erosion and dilation; (b) the red line is the background fitting result by erosion operation and the gray band shows erosion; (c) the blue line was background fitting result by dilation after erosion operation and the gray band shows dilation; (d) opening operation for background correction fitting and the gray band shows correction.

the result for opening operation. With the help of mathematic morphology, the background can be removed by morphological opening operation (see also Fig. 1(d)).

Penalized least squares algorithm

The opening operation can generate a rough background of the spectrum. However, it will introduce flaws in the peak region that will change the shape (see also Fig. 2(a)). So, in order to get the balance between the fidelity and the roughness, one needs to combine the opening operation and the penalized least squares.

Assuming $\boldsymbol{x}$ is the vector of the analytical signals, and $\boldsymbol{z}$ is the fitted vector, both with m elements. The roughness of the fitted data $\boldsymbol{z}$ can be written as its squared and summed differences and denoted as R:

$$R = \sum_{i=2}^{m} (z_i - z_{i-1})^2 = \sum_{i=1}^{m-1} (\Delta z_i)^2 \quad (4)$$

In most cases, the square of second difference penalties can be a natural way to quantify the roughness. One also could perform first, second or even higher difference penalties by adjusting the parameter for the special purpose.

The fidelity of $\boldsymbol{z}$ to $\boldsymbol{x}$ can be expressed as the sum of square errors between them and denoted as F:

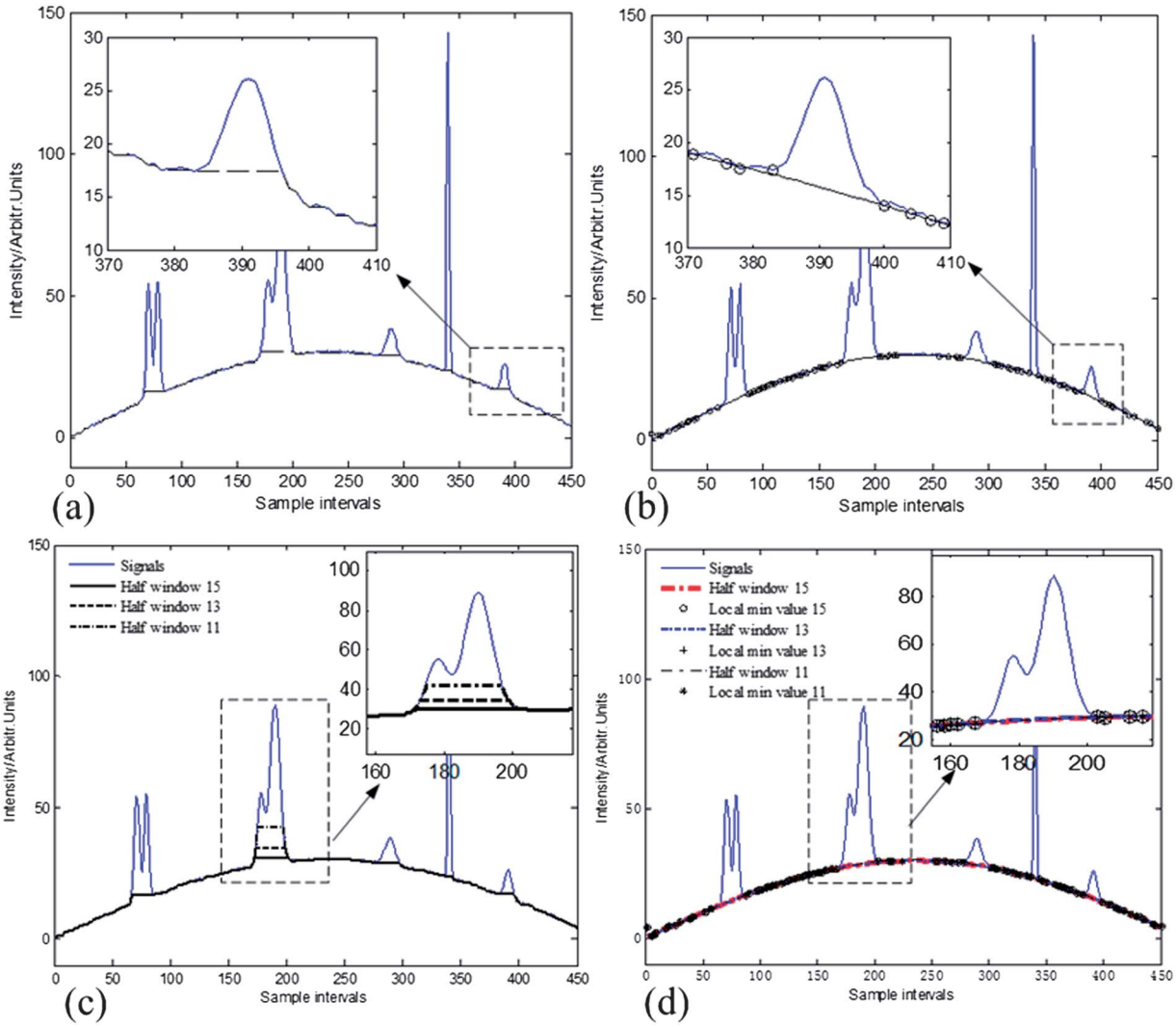

Fig. 2 (a) Background correction results using only morphological opening operation; (b) background correction results with morphology weighted penalized least squares; (c) background correction results using only the morphological opening operation at different window sizes; (d) background correction results with morphology weighted penalized least squares at different window sizes.

$$F = \sqrt{\sum_{i=1}^{m} (x_i - z_i)^2} \tag{5}$$

The balanced combination of these two goals can be measured as the fidelity and the roughness, which can be described by the following equation:

$$Q = F + \lambda R = \|\boldsymbol{x} - \boldsymbol{z}\|^2 + \lambda \|\boldsymbol{D}\boldsymbol{z}\| \tag{6}$$

where λ is an adjustable parameter. A larger λ brings a smoother fitted vector. Balance of fidelity and smoothness can be achieved by tuning this parameter. $\boldsymbol{D}$ is the derivative of the identity matrix such that $\boldsymbol{D}\boldsymbol{z} = \Delta\boldsymbol{z}$.

Let the partial derivatives be equal to 0 $\left(\frac{\partial Q}{\partial \boldsymbol{z}} = 0\right)$. We can get the linear system of equations:

$$(\boldsymbol{I} + \lambda \boldsymbol{D}'\boldsymbol{D})\boldsymbol{z} = \boldsymbol{x} \tag{7}$$

where $\boldsymbol{I}$ is the identity matrix.

In order to use penalized least squares for background correction, the weighted penalized least squares can be adopted with the rough background and the local minimum value by opening.

Actually, after the peak removal procedure is completed with opening operation, one can find a series of flat regions in the *opening*(x). For each flat region, the start and end points are denoted as s_k and e_k, respectively. The local minimum value is then defined as the minimum point between e_k and s_{k+1}, which can be mathematically described as follows:

$$M_k = \arg\min(x_j), j \in (e_k, \ldots, s_{k+1}) \tag{8}$$

where j is the region between flat regions k and $k + 1$. M_k is the k^{th} local minimum value. The elements of the weight vector $\boldsymbol{w}$ at the local minimum value is set to 1, and the non-local minimum value of $\boldsymbol{w}$ is set to an arbitrary small value, say 0. The fidelity of z to the spectrum x is changed to:

$$F = \sum_{i=1}^{m} w_i (x_i - z_i)^2 = (\boldsymbol{x} - \boldsymbol{z})' W (\boldsymbol{x} - \boldsymbol{z}) \quad (9)$$

where $\boldsymbol{W}$ is a diagonal matrix with w_i on its diagonal.

Solving the above linear equations, the fitted vector can be obtained easily:

$$\boldsymbol{z} = (\boldsymbol{W} + \lambda \boldsymbol{D}'\boldsymbol{D})^{-1} \boldsymbol{W}\boldsymbol{x} \quad (10)$$

Based on the discussions above, the calculation procedure of MPLS is follows:

(1) To obtain the rough background profile by mathematical morphological opening operation.

(2) The rough background profile obtained as well as the local minimum values are used as the input vector for penalized least squares to refine the background profile.

(3) By subtracting the refined background profile from the original signal, the background corrected signal is achieved.

Fig. 2 shows the rough background profile by using morphological opening operation and the proposed MPLS method. One can see obviously (Fig. 2(a)) that the rough background profile obtained by morphological opening operation is not smooth. And the background profile shape is changed in the peak region. Moreover the window size of the structuring element has a great influence on the obtained rough background profile. If the window is too small, the peak shape will change. In contrast, if the window is too big, some background information will be ignored. It was worse when the peak widths are different one by one. Thus it is difficult to obtain a suitable window size for all peaks. Fig. 2(c) shows the obtained background profile when different window sizes of the structuring element are used.

For MPLS, after the rough background profile is refined by weighted penalized least squares, the background profile obtained is much more accurate and reasonable. Fig. 2(b) shows these better results. Furthermore, MPLS could avoid the influence of the size of the structuring element. It can be seen from Fig. 2(d) that the local minimum and background profile are stable when different window sizes are used.

Experimental

A simulated chromatographic dataset and two experimental data on chromatography and Raman, respectively, were used to validate the performance of the proposed method for elimination of varying background and noise.

Simulated data

Simulated data consist of linear or curved backgrounds, analytical signals and random noise, which can be mathematically described as follows:

$$s(x) = p(x) + b(x) + n(x)$$

here, $s(x)$ denotes the simulated data, $p(x)$ the pure analytical signal, $b(x)$ the linear or curved background and $n(x)$ the random noise.

Pure signals are three Gaussian peaks with different heights (listed in Table 1), means and variances. The curved background is a sine curve. Random noise $n(x)$ is generated using the random number generator, whose intensity is about 2 percent of the simulated signals. Signals with the linear and curved backgrounds are illustrated in Fig. 3(a) and (b), respectively.

Chromatographic data

The raw tobacco leaves are from Honghe district, Yunnan province. The weight of each cigarette is 0.700 ± 0.015 g, which is filled by a CMB-120 cigarette tube filling machine (Burghart, Germany). The plant perfumes from herb extractions are injected into the cigarette *via* a CIJECTOR cigarette injection machine (Burghart, Germany). 20 cigarettes are smoked simultaneously by the smoking machine (Borgwaldt, Germany), and the cigarette smoke is gathered by a Cambridge filter. The extraction solvent (80 mL, dichloromethane–methanol = 2 : 1 (v/v)) is used to elute the compounds enriched by the Cambridge filter. After extraction, evaporation and concentration to 1 mL, the sample is injected into GCT Premier™ GC/oa-TOF-MS. The chromatographic column is DB-35MS (30 m × 0.25 mm, 0.25 μm) with the split ratio of the injector being 1 : 30 at 250 °C.

Helium carrier gas was used at a constant flow rate of 1.5 mL min^{-1}. The column temperature was programmed from 50 °C to 280 °C. Mass spectra from 40 to 400 amu were collected. The ionization voltage was 70 eV and the ion source temperature was 220 °C. One can see the chromatograms with various

Table 1 Analysis of the influence of different sizes of morphological structuring elements on the fitted background

		Peak height				
Background type	Peak no.	Uncorrected	Expected	Size = 10	Size = 20	Size = 30
Linear	Peak 1	95.40	80.10	79.91	79.91	79.91
	Peak 1	94.16	48.56	47.57	48.02	48.02
	Peak 1	93.77	17.87	16.95	16.55	16.32
Curved	Peak 1	94.75	80.10	79.88	79.88	79.88
	Peak 1	78.52	48.56	46.68	48.31	48.32
	Peak 1	35.09	17.87	15.68	16.98	16.98
	Peak 1	95.40	80.10	79.91	79.91	79.91

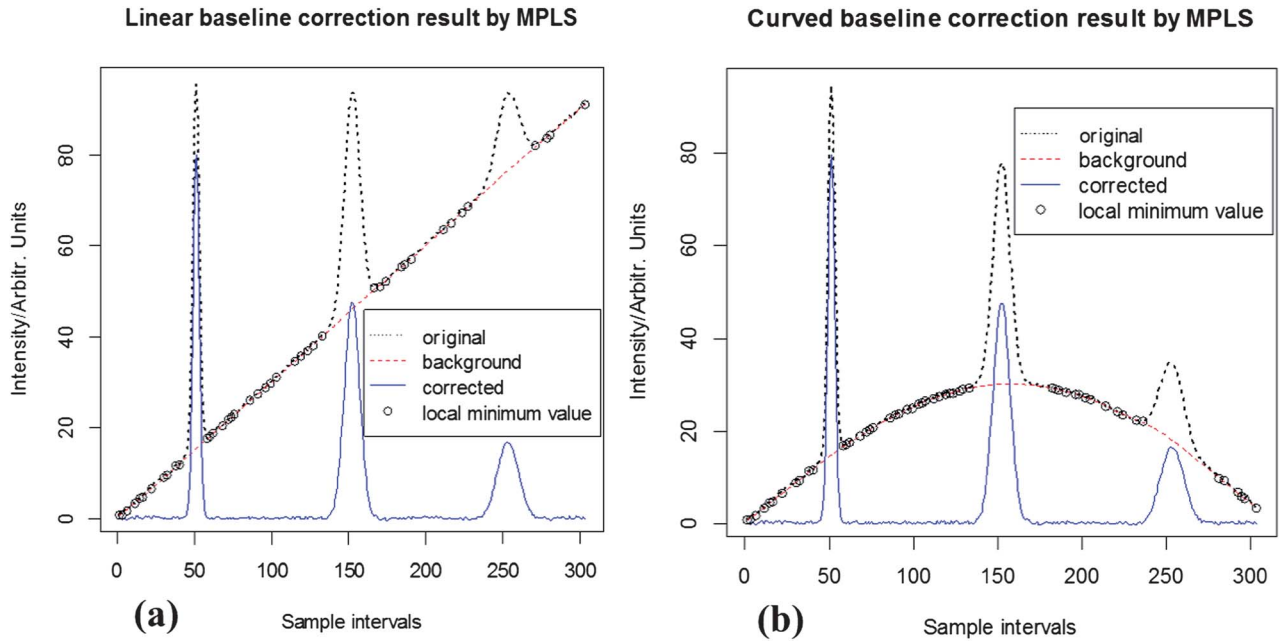

Fig. 3 Background correction results for both linear and curved background signals by MPLS. (a) Result for linear background and (b) result for curved background.

backgrounds of GC/oa-TOF-MS analysis of tobacco smoke from Fig. 3(a).

Raman spectra data

Prednisone Acetate Tablets (PATs) and Glibenclamide Tablets (GTs) were measured using a laser of 785 nm wavelength for excitation by a BWTEK i-Raman-785 spectrometer with 2048 element thermoelectric cooled linear charge-coupled device (TEC-CCD) arrays. PATs, from 10 different pharmaceutical factories, were recorded using 5000 ms integration times. GTs, from 6 different pharmaceutical factories, were also recorded using 5000 ms integration times to obtain comparable spectra. Since we measured 3 tablets for each pharmaceutical factory, there are 48 Raman spectra in total. These Raman spectra are used to demonstrate that the MPLS background correction method can improve the classification accuracy.

Results and discussion

Results of simulated data

The background correction results of simulated, chromatographic and Raman datasets are firstly presented in this section. The influence of parameters of morphological structuring elements has also been discussed. Then the MPLS method is compared with ALS, airPLS and Morphology. Finally, the classification model is built on the background corrected signals to show that classification accuracy can be achieved by the MPLS method.

Background correction results

The MPLS method was applied to fit the backgrounds of simulated datasets. The correction results are displayed in Fig. 3(a) and (b). In both linear and curved background situations, one can see from Fig. 3 that the local minimum value (black circles) was detected accurately and intelligently with the mathematical morphological operations. With the series of local minimum values found by morphological operations, the rough background profile is fitted by weighted penalized least squares (the red dashed line in Fig. 3). After subtracting the background from the original signal, one can obtain the background corrected signal (the blue line in Fig. 3) successfully. The MPLS method is flexible enough to fit both linear and curved baselines without peak detection and iteration.

For the chromatograms of tobacco smoke, the background was also corrected with MPLS. This time the results were analyzed by principal component analysis (PCA). By comparing Fig. 4(a) with 4(b), we can observe that the MPLS method is flexible enough to remove the backgrounds varied from chromatogram to chromatogram. Further, PCA was applied to evaluate the validity of the MPLS method. Since all the eight samples are parallel samples, they should locate close to each other in principal component spaces if the influence of the background is ignored. In Fig. 4(c), the circle represents a chromatogram without background correction by MPLS in principal component spaces. The samples difference is mainly reflected in the first principal component direction, as the first principal component takes 93% of the total variance. Let $\boldsymbol{D}$ denote the range of these corrected samples (chromatograms) in the first principal component direction. The range of original chromatograms in the first principal component direction is $2.3 \times \boldsymbol{D}$. This indicates that MPLS can reduce the variance greatly originating from the background.

As is known to all, numerical differentiation can remove the slowly changing background,[49,50] so PCA has been also applied to the original and corrected chromatograms with first-order

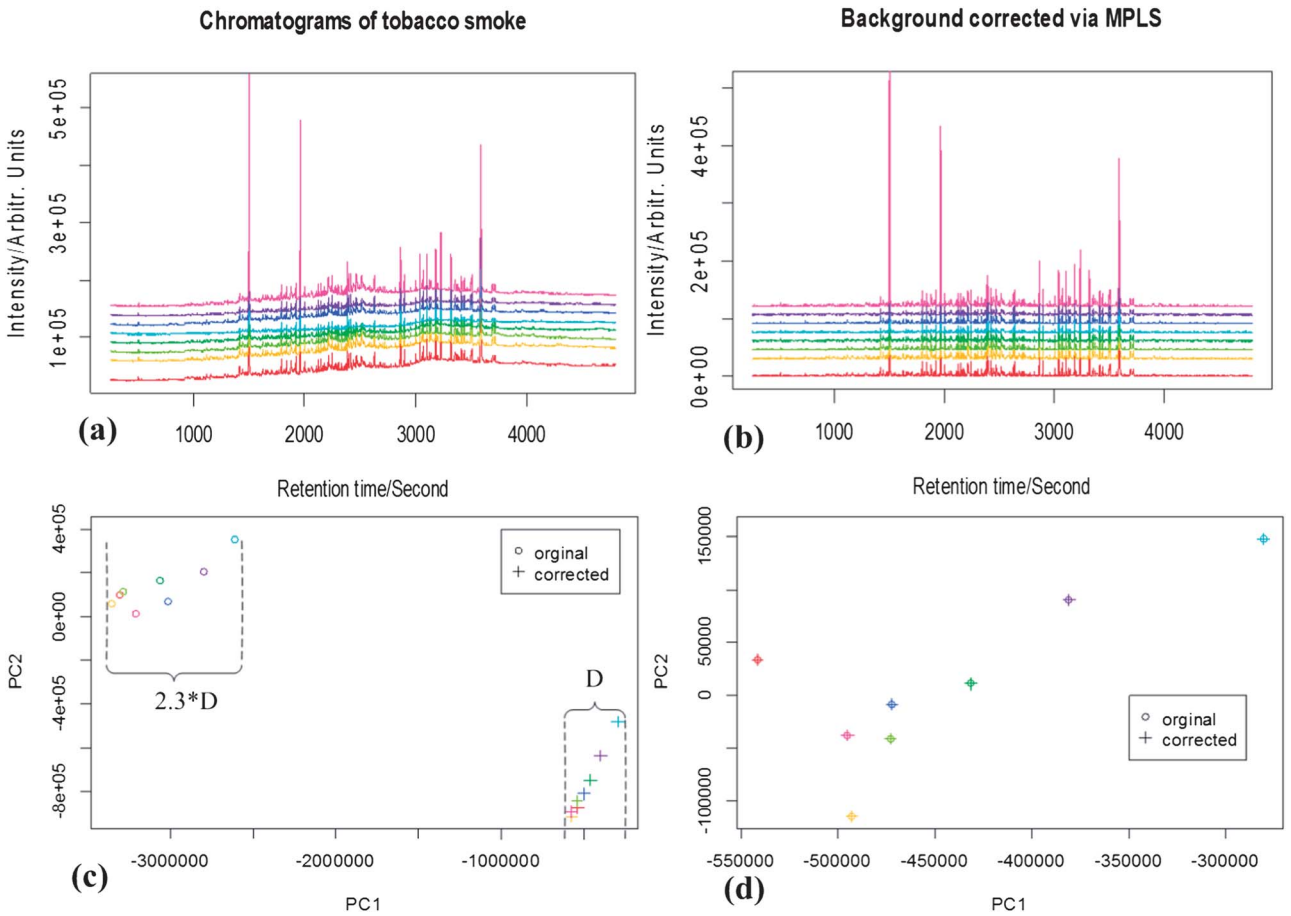

Fig. 4 Background correction results for the GC/oa-TOF-MS dataset of tobacco smoke by MPLS. (a) Original dataset with various backgrounds; (b) backgrounds have been corrected by the MPLS method; (c) compare the distribution of samples before and after background correction in principal component spaces; (d) first two principal components of the original and corrected chromatograms with first-order numerical differentiation preprocessing.

numerical differentiation preprocessing. The circle and cross in Fig. 4(d) have good matching in principal component spaces, which means that the MPLS method will not remove useful information from chromatograms. From the above analysis, one can infer that the larger variation of original chromatograms in the first principal component direction is due to the varied backgrounds from chromatogram to chromatogram. The proposed MPLS method can remove this background variation among chromatograms without losing useful information.

Influence of morphological structuring elements

This MPLS background correction method consists of morphological operation and penalized least squares. The parameters for penalized least squares are the order of the penalty and the smooth parameter λ. In most cases, the square of the second difference is a natural penalty to quantify the roughness.[24,37] Thus the second order of the penalty is used. The smooth parameter λ is also easy to adjust, if the fitted background is too flexible, the λ parameter is increased, and *vice versa*.

For the morphological operation part, the opening operation is sensitive to the size of the morphological structuring element. In order to overcome the influence of the size of structuring elements, the local minimum values of the rough background around the peak region are used as weighted vectors for the weighted penalized least squares. This technique reduces the dependence in the peak region on the size of the morphological structuring element.

In order to investigate the influence of the size of morphological structuring elements, we have completed the correction algorithm with size equal to 10, 20 and 30 respectively using the simulated dataset with known peak heights. Since the heights of three Gaussian peaks are already known, the background correction can be processed with different sizes of morphological structuring elements. Then the difference between peak heights before and after background correction by MPLS with different sizes can be utilized to evaluate the methods used. As shown in Table 1, the corrected peak heights are very similar when sizes of morphological structuring elements are different. The MPLS seems to be not sensitive to the size of morphological structuring elements.

Comparison with other methods

For the simulated dataset, the height of each peak is already known. Thus it can be used as a quantitative measurement for

Table 2 Compare ALS, airPLS, Morphology with the MPLS method by peaks of known peak heights[a]

		Peak height					
Background type	Peak no.	Uncorrected	Expected	ALS	airPLS	Morphology	MPLS
Linear	Peak 1	95.64	80.34	79.28	80.23	79.77	80.20
	Peak 1	93.51	47.91	45.70	47.71	47.36	47.70
	Peak 1	93.22	17.32	19.24	17.12	18.53	17.02
Curved	Peak 1	94.99	80.34	79.55	80.29	80.46	80.22
	Peak 1	77.87	47.91	47.27	47.86	49.04	48.02
	Peak 1	34.54	17.32	17.10	17.28	17.75	17.18

[a] For the ALS method, the parameters are as follows: $\lambda = 25, p = 0.001, d = 2$. For the airPLS method, the parameters are as follows: $\lambda = 96\ 100, d = 2$. For the Morphology method, the parameters are as follows: half window = 30. For the MPLS method, the parameters are as follows: $\lambda = 62\ 100$, $d = 2$, half window = 30.

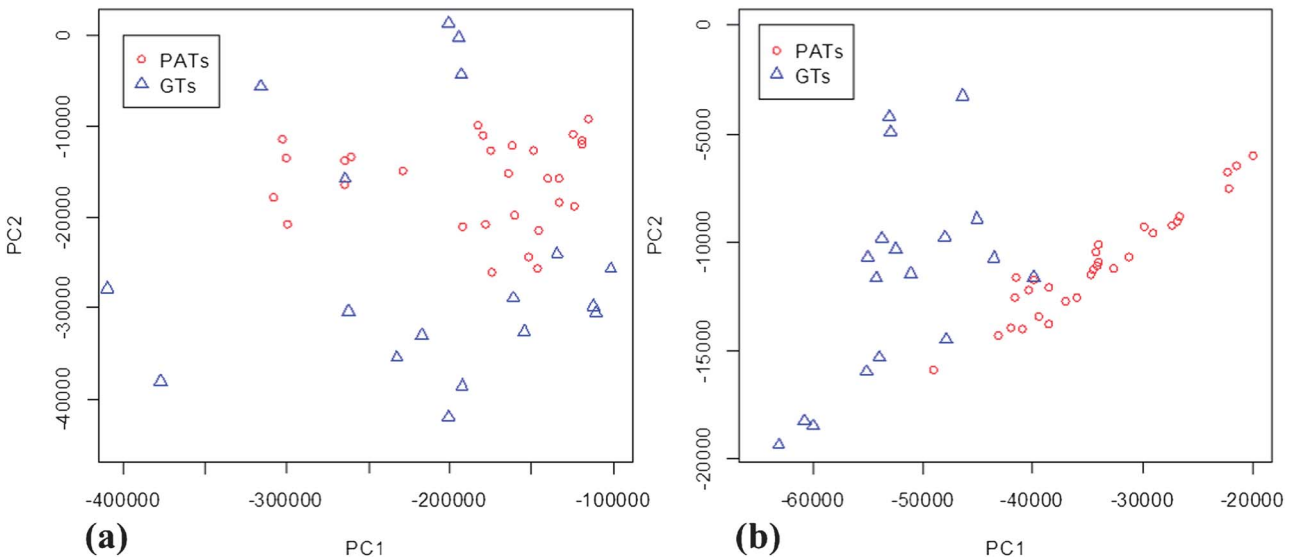

Fig. 5 Analysis of the validity of MPLS with PCA clustering. (a) Clustering results of Raman spectra of PAT and GT samples without any pre-processing; (b) clustering results of Raman spectra of PAT and GT samples after background correction by MPLS.

the background correction results. In order to compare the proposed MPLS with Asymmetric Least Squares (ALS) background correction of Eilers *et al.*,[17] adaptive iteratively reweighted Penalized Least Squares (airPLS) of Zhang *et al.*[23] and morphology-based (Morphology) of Rosanna *et al.*,[48] the heights of three Gaussian peaks before and after correction by each background correction method are listed in Table 2. After pre-processing by ALS, airPLS, Morphology and MPLS, all the peak heights are closer to the expected height, which means that all the three methods can remove the background from signal well. The MPLS method achieves almost the best results in both large and small peaks. That is the MPLS method can correct the background more accurate than ALS, airPLS and Morphology for signal swamp by either a linear or curved background in this example.

Improvement of clustering results

Since the background corrected spectra are commonly used for building a clustering, classification or calibration model, the improvement of clustering results are also discussed in this section to demonstrate the validity of the proposed method. The Raman spectra of PATs and GTs have 1715 wavelengths. The dimension reduction method PCA was applied to the spectral matrix which consists of Raman spectra of PATs and GTs without any pre-processing, and the first two principal components of each sample were used for analysis (Fig. 5(a)). Due to the influence of backgrounds, PAT (blue triangles) and GT (red circles) samples are mixing together without any simple decision boundary. For the corrected spectra, the results are shown in Fig. 5(b). After the backgrounds were corrected by MPLS, the samples of PATs and GTs could be separated into two categories.

Conclusion

In this research, an intelligent background correction method based on morphological opening operations and weighted penalized least squares without peak detection or an iteration procedure has been proposed. The morphological opening operation is applied to obtain a rough background. All the local

minimum values found are used as weight vectors for background fitting *via* penalized least squares. This method is more accurate and intelligent when compared with Morphology, airPLS and ALS based on simulated datasets with peaks of known heights. The compactness in principal component spaces means that the baseline of chromatographic datasets of tobacco smoke is successfully corrected by MPLS. The clustering results of Raman spectra from two kinds of medicines are improved when the fluorescent background is removed through MPLS. These examples have demonstrated the accuracy, practicability and validity of MPLS. MPLS is a valuable pre-processing method for chromatograms, Raman spectra and other analytical instrument signals.

Acknowledgements

This work is financially supported by the Major science and technology project of China tobacco Yunnan industrial Co., Ltd (grant no. 2012JC03), National Nature Foundation of China (grant no. 21275164, 11271374 and 21075138). The studies meet with the approval of the university's review board. We are grateful to all employees of this institute for their encouragement and support of this research.

Notes and references

1 J. M. Amigo, T. Skov and R. Bro, *Chem. Rev.*, 2010, **110**, 4582–4605.
2 M. Daszykowski and B. Walczak, *TrAC, Trends Anal. Chem.*, 2006, **25**, 1081–1096.
3 N. K. Afseth, V. H. Segtnan and J. P. Wold, *Appl. Spectrosc.*, 2006, **60**, 1358–1367.
4 K. H. Liland, T. Almoy and B. H. Mevik, *Appl. Spectrosc.*, 2010, **64**, 1007–1016.
5 P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom and T. Moritz, *Anal. Chem.*, 2004, **76**, 1738–1745.
6 D. L. A. de Faria and M. A. de Souza, *J. Raman Spectrosc.*, 1999, **30**, 169–171.
7 A. O'Grady, A. C. Dennis, D. Denvir, J. J. McGarvey and S. E. J. Bell, *Anal. Chem.*, 2001, **73**, 2058–2065.
8 B. Auguié, A. Reigue, E. C. Le Ru and P. G. Etchegoin, *Anal. Chem.*, 2012, **84**, 7938–7945.
9 N. Li, X.-Y. Li, Z.-X. Zou, L.-R. Lin and Y.-Q. Li, *Analyst*, 2011, **136**, 2802–2810.
10 A. M. Macdonald and P. Wyeth, *J. Raman Spectrosc.*, 2006, **37**, 830–835.
11 D. D. Gerow and S. C. Rutan, *Anal. Chem.*, 1988, **60**, 847–852.
12 Y.-Z. Liang, O. M. Kvalheim, A. Rahmani and R. G. Brereton, *Chemom. Intell. Lab. Syst.*, 1993, **18**, 265–279.
13 Z. Pan, X. Shao, H. Zhong, W. Liu, H. Wang and M. Zhang, *Chin. J. Anal. Chem.*, 1996, **24**, 149–153.
14 X. G. Shao, W. S. Cai and Z. X. Pan, *Chemom. Intell. Lab. Syst.*, 1999, **45**, 249–256.
15 X. G. Shao, A. K. M. Leung and F. T. Chau, *Acc. Chem. Res.*, 2003, **36**, 276–283.
16 X. T. Zhang and H. Y. Zhu, *Chin. J. Anal. Chem.*, 1999, **27**, 1324–1328.
17 H. F. M. Boelens, R. J. Dijkstra, P. H. C. Eilers, F. Fitzpatrick and J. A. Westerhuis, *J. Chromatogr., A*, 2004, **1057**, 21–30.
18 G. Quintás, B. Lendl, S. Garrigues and M. de la Guardia, *J. Chromatogr., A*, 2008, **1190**, 102–109.
19 J. Kuligowski, G. Quintás, S. Garrigues and M. de la Guardia, *J. Chromatogr., A*, 2009, **1216**, 3122–3130.
20 J. Kuligowski, D. Carrión, G. Quintás, S. Garrigues and M. de la Guardia, *J. Chromatogr., A*, 2010, **1217**, 6733–6741.
21 J. Kuligowski, G. Quintás, S. Garrigues and M. de la Guardia, *Talanta*, 2010, **80**, 1771–1776.
22 J. Kuligowski, G. Quintás, S. Garrigues, B. Lendl and M. de la Guardia, *TrAC, Trends Anal. Chem.*, 2010, **29**, 544–552.
23 Z. M. Zhang, S. Chen and Y. Z. Liang, *Analyst*, 2010, **135**, 1138–1146.
24 Z. M. Zhang and Y. Z. Liang, *Chromatographia*, 2012, **75**, 313–314.
25 Ł. Komsta, *Chromatographia*, 2011, **73**, 721–731.
26 R. C. Allen, M. G. John, S. C. Rutan, M. R. Filgueira and P. W. Carr, *J. Chromatogr., A*, 2012, **1254**, 51–61.
27 M. R. Filgueira, C. B. Castells and P. W. Carr, *Anal. Chem.*, 2012, **84**, 6747–6752.
28 I. B. Gornushkin, P. E. Eagan, A. B. Novikov, B. W. Smith and J. D. Winefordner, *Appl. Spectrosc.*, 2003, **57**, 197–207.
29 C. A. Lieber and A. Mahadevan-Jansen, *Appl. Spectrosc.*, 2003, **57**, 1363–1367.
30 A. Jirasek, G. Schulze, M. M. L. Yu, M. W. Blades and R. F. B. Turner, *Appl. Spectrosc.*, 2004, **58**, 1488–1499.
31 V. Mazet, C. Carteret, D. Brie, J. Idier and B. Humbert, *Chemom. Intell. Lab. Syst.*, 2005, **76**, 121–133.
32 F. Gan, G. H. Ruan and J. Y. Mo, *Chemom. Intell. Lab. Syst.*, 2006, **82**, 59–65.
33 Y. G. Hu, T. Jiang, A. G. Shen, W. Li, X. P. Wang and J. M. Hu, *Chemom. Intell. Lab. Syst.*, 2007, **85**, 94–101.
34 R. M. Rao, M. A. Slamani, T. H. Chyba and D. K. Emge, in *Signal and Data Processing of Small Targets 2010*, ed. O. E. Drummond, Spie-Int Soc Optical Engineering, Bellingham, 2010, vol. 7698.
35 S.-J. Baek, A. Park, J. Kim, A. Shen and J. Hu, *Chemom. Intell. Lab. Syst.*, 2009, **98**, 24–30.
36 S.-J. Baek, A. Park, A. Shen and J. Hu, *J. Raman Spectrosc.*, 2011, **42**, 1987–1993.
37 Z. M. Zhang, S. Chen, Y. Z. Liang, Z. X. Liu, Q. M. Zhang, L. X. Ding, F. Ye and H. Zhou, *J. Raman Spectrosc.*, 2010, **41**, 659–669.
38 S. Chen, X. N. Li, Y. Z. Liang, Z. M. Zhang, Z. X. Liu, Q. M. Zhang, L. X. Ding and P. Ye, *Spectrosc. Spectral Anal.*, 2010, **30**, 2157–2160.
39 K. H. Liland, E. O. Rukke, E. F. Olsen and T. Isaksson, *Chemom. Intell. Lab. Syst.*, 2011, **109**, 51–56.
40 J. Palacky, P. Mojzes and J. Bok, *J. Raman Spectrosc.*, 2011, **42**, 1528–1539.
41 B. D. Prakash and Y. C. Wei, *Analyst*, 2011, **136**, 3130–3135.
42 H. G. Schulze, R. B. Foist, K. Okuda, A. Ivanov and R. F. B. Turner, *Appl. Spectrosc.*, 2012, **66**, 757–764.

43 Q. J. Bao, J. W. Feng, F. Chen, W. P. Mao, Z. Liu, K. W. Liu and C. Y. Liu, *J. Magn. Reson.*, 2012, **218**, 35–43.
44 N. Kourkoumelis, A. Polymeros and M. Tzaphlidou, *Spectrosc. Int. J.*, 2012, **27**, 441–447.
45 H. Krishna, S. K. Majumder and P. K. Gupta, *J. Raman Spectrosc.*, 2012, **43**, 1884–1894.
46 D. Van de Sompel, E. Garai, C. Zavaleta and S. S. Gambhir, *PLoS One*, 2012, **7**, 6971–6974.
47 A. T. Weakley, P. R. Griffiths and D. E. Aston, *Appl. Spectrosc.*, 2012, **66**, 519–529.
48 R. Perez-Pueyo, M. J. Soneira and S. Ruiz-Moreno, *Appl. Spectrosc.*, 2010, **64**, 595–600.
49 M. N. Leger and A. G. Ryder, *Appl. Spectrosc.*, 2006, **60**, 182–193.
50 D. M. Zhang and D. Ben-Amotz, *Appl. Spectrosc.*, 2000, **54**, 1379–1383.

