Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1057 (2004) 21–30
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

New background correction method for liquid chromatography with diode array detection, infrared spectroscopic detection and Raman spectroscopic detection

Hans F.M. Boelens[a,*], Reyer J. Dijkstra[b], Paul H.C. Eilers[c], Fiona Fitzpatrick[d], Johan A. Westerhuis[a]

[a] *Process Analysis and Chemometrics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[b] *Department of Applied Physics, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands*
[c] *Department of Medical Statistics, Leiden University Medical Centre, P.O. Box 9604, 2300 RC Leiden, The Netherlands*
[d] *Substance Analysis, AstraZeneca R&D Mölndal, SE-431-83 Mölndal, Sweden*

Received 19 March 2004; received in revised form 8 September 2004; accepted 10 September 2004

Abstract

A new method to eliminate the background spectrum (EBS) during analyte elution in column liquid chromatography (LC) coupled to spectroscopic techniques is proposed. This method takes into account the *shape* and also *intensity* differences of the background eluent spectrum. This allows the EBS method to make a better estimation of the background eluent spectrum *during* analyte elution. This is an advantage for quantification as well as for identification of analytes. The EBS method uses a two-step procedure. First, the baseline spectra are modeled using a limited number of principal components (PCs). Subsequently, an asymmetric least squares (asLS) regression method is applied using these principal components to correct the measured spectra during elution for the background contribution. The asymmetric least squares regression needs one parameter, the asymmetry factor *p*. This asymmetry factor determines relative weight of positive and negative residuals. Simulations are performed to test the EBS method in well-defined situations. The effect of spectral noise on the performance and the sensitivity of the EBS method for the value of the asymmetry factor *p* is tested. Two applications of the EBS method are discussed. In the first application, the goal is to extract the analyte spectrum from an LC–Raman analysis. In this case, the EBS method facilitates easy identification of unknown analytes using spectral libraries. In a second application, the EBS method is used for baseline correction in LC–diode array detection (DAD) analysis of polymeric standards during a gradient elution separation. It is shown that the EBS method yields a good baseline correction, without the need to perform a blank chromatographic run.

Keywords: Background correction; LC–DAD; LC–Raman; Library search; Spectral identification; Gradient elution

1. Introduction

When column liquid chromatography (LC) is coupled to spectroscopic techniques, such as UV–vis diode array detection (DAD), Fourier transform infrared (FT-IR) or Raman, it is often difficult to completely remove the interfering spectrum of the eluent. The spectral response of the eluent is usually much larger than the contribution from the analyte and the composition of the eluent may not be constant. To overcome this problem, it is common practice to select an eluent that has spectral bands outside the spectral range of the analytes. However, this is not always possible. For instance, in the case of on-line LC–Raman and on-line LC–IR the spectrum of the solvent always contains distinct bands that overlap or partially overlap with the bands of the analyte. In

* Corresponding author. Present address: Biosystems Data Analysis Group, Sammerdam Institute of Life Sciences (SILS), Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands. Tel.: +31 20 5256547; fax: +31 20 5256971.
E-mail address: hans.boelens@science.uva.nl (H.F.M. Boelens).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.09.035

addition, in LC–DAD, an eluent that is suitable from a separation point of view can sometimes be rejected because it has an unacceptable spectroscopic response. Having a background correction method that can remove the spectral contribution of the eluent from the detector response would broaden the eluent choice in LC–DAD. One example where a good background correction procedure would increase efficiency is in solvent gradient elution in LC–DAD. Normally, the background is removed by subtraction of a blank chromatographic run. If the EBS method could be used to calculate the background spectrum during analyte elution, a blank run would no longer be needed. This would save significant time and effort.

Why does simple subtraction of the eluent spectrum just before analyte elution (auto-zeroing) not always give good results? These are two main reasons for this. The first one is the change in spectral intensity of the eluent spectrum during a chromatographic run. For instance, the intensity of the eluent spectrum during analyte elution is smaller than the intensity of the spectrum before elution. Straightforward subtraction will, therefore, lead to overcorrecting. The second reason is that small spectral shape changes of the eluent spectrum might occur. These changes may have several causes, such as wavelength shift, spectral drift and/or offset. In addition, in solvent gradient elution the solvent composition is deliberately changed, which causes both shape and intensity changes in the eluent spectrum.

Here a new correction method, called elimination of background spectrum (EBS), is described, that can take into account these shape and intensity differences of the eluent spectra. Some simulations are discussed that show the advantage of this method with respect to straightforward subtraction of the eluent spectrum. Subsequently, the EBS method is evaluated for two test cases. In the first application, an LC–Raman separation is evaluated. In this case, the analyte spectrum is completely overwhelmed by the strong spectrum of the eluent. A companion publication discusses the application of the EBS method to these LC–Raman data in more detail [1]. In the second application, a gradient LC–DAD is evaluated. Compared to the first application, the analyte spectrum is more intense. It is shown that the EBS method removes the baseline well, without requiring a blank chromatographic run.

Several multivariate approaches to background correction in hyphenated chromatography are known [2–8]. The EBS method has some differences and advantages compared to these methods. Second order calibration methods can be used to estimate concentrations and uncover the spectra of unknown compounds when measurement of reference mixtures is feasible [2–5]. In case the chromatographic run is used to find what analytes are present in the sample such an approach would be begging the question. This is because the analytes for which the calibration should be done are not yet known. Even if such a second order approach is feasible, the disadvantage is that some second order methods require the retention times of the additional chromatographic run to be well reproducible. The EBS method only needs data of a single chromatographic run and does not presuppose knowledge of the analyte spectra. Additional measurements on standards are not needed and therefore retention time stability is no longer crucial.

Another method based on (adaptive) Kalman filters and derivative spectroscopy is used by Gerow and Rutan [6]. This method requires knowledge about the spectral response of the individual analytes. Although this knowledge does not need to be fully accurate, it is not always available. The method proposed by Liang and Kvalheim [7] does only use the measurements from one chromatographic run. Major principal components (PCs) are extracted from spectra measured in zero component regions before and after analyte elution. Essentially, their correction procedure assumes that the eluent contribution during analyte elution is constant. Gemperline et al. [8] relieve this assumption and automate the whole procedure. They allow the scores of major principal components (and thereby in fact the concentration of the eluent compounds) to change according to a cubic polynomial model. In contrast with this type of approach, the EBS method corrects each spectrum measured during elution separately and does *not* assume a predefined model for the concentrations of the eluent compounds in time. Another drawback of an *interpolation* approach is that it cannot account for the decrease of the spectral contribution of the eluent during analyte elution caused by the presence of the analyte. The EBS method is able to do this.

Additionally, the EBS method only needs some representative background spectra to work. All of these spectra could very well be collected during the same chromatographic run before (or after) analyte elution. These features make the EBS method a more versatile tool for the user.

2. Theory

2.1. Notation

Bold-face capital letters represent matrices, bold-face lowercase characters represent vectors, and italic, lower case letters represent scalars values. The subscript ‘b’ indicates quantities at retention times when only a chromatographic baseline is present (eluent). The subscript ‘ab’ indicates quantities at retention times at which also some analyte is present. And finally the subscript ‘a’ indicates that quantities only refer to the analyte.

2.2. The EBS method

The measured data is collected in a data matrix $\mathbf{X}$ ($n_{\text{chan}} \times n$) having n columns. Each column contains a spectrum measured at a number of channels (n_{chan}). These channels may be wavelengths or wavenumbers. Also assume that the matrix $\mathbf{X}$ can be split into two parts. One part is the matrix $\mathbf{X}_{\text{b}}$ ($n_{\text{chan}} \times n_{\text{b}}$), where only the eluent is present, the

other part is $\mathbf{X}_{ab}$ ($n_{chan} \times n_{ab}$), where eluent *and* analyte are present. In practical use the matrix, $\mathbf{X}_b$, will often contain spectra before and after elution of the analyte.

The EBS method accounts for intensity and shape differences between the eluent spectra and the spectroscopic contribution of the eluent to the spectra measured during elution. The EBS method is a two-step procedure: first, all variation of the eluent spectra at baseline level is modeled in a background spectral subspace (B-space) constructed by principal component analysis [9,10]. Secondly, the spectra measured during analyte elution are corrected by performing an asymmetric least squares regression (asLS) with respect to the B-space found.

2.3. Step (i)

The matrix $\mathbf{X}_b$ is modeled with a limited number (n_{pc}) of its principal components:

$$\mathbf{X}_b = \mathbf{P}\mathbf{K}_b + \mathbf{E}_b \tag{1}$$

The columns of the ($n_{chan} \times n_{pc}$) matrix $\mathbf{P}$ form an orthonormal basis that allows a faithful description of the changes in shape and size of the eluent spectrum. The spectral space spanned by this orthonormal basis is called the B-space. The columns of the ($n_{pc} \times n_b$) matrix $\mathbf{K}_b$ contain the coordinates of all eluent spectra in this B-space. The ($n_{chan} \times n_b$) matrix $\mathbf{E}_b$ describes all spectral variation of the eluent spectra that is not modeled in the B-space. Ideally, this matrix $\mathbf{E}_b$ contains only spectral noise. There are many methods and algorithms to determine the number of principal components that span the B-space. An overview and comparison of these methods can be found in literature [11–13]. Here the *IND* [14] algorithm is used.

2.4. Step (ii)

Assume that this orthonormal basis (i.e. $\mathbf{P}$) is also valid for the eluent spectrum that is present during elution of the analyte. Each spectrum (vector $\mathbf{x}_{ab}$) measured during elution of the analyte, can now be written as:

$$\mathbf{x}_{ab} = \mathbf{P}\mathbf{q}_{ab} + \mathbf{s}_a \tag{2}$$

The first part, namely $\mathbf{P}\mathbf{q}_{ab}$, gives the contribution of the eluent to the measured spectrum and $\mathbf{s}_a$ represents the spectrum of the analyte. The ($n_{pc} \times 1$) coefficient vector $\mathbf{q}_{ab}$ is still unknown. Exactly, this vector allows us to describe shape and intensity changes of the eluent spectrum during analyte elution. A (too) simple approach to estimate the vector $\mathbf{q}_{ab}$ would be by using ordinary linear regression. Such a regression yields positive and negative deviations around the fitted model (viz. $\mathbf{P}\mathbf{q}_{ab}$). It is known, however, that all elements of the analyte spectrum are positive. A solution to this problem is to use *asymmetric* least squares [15–18] for the estimation of the coefficients $\mathbf{q}_{ab}$.

2.5. Asymmetric least squares

In asymmetric least squares positive residuals and negative residuals do not receive the same weights. The ($n_{chan} \times 1$) residual vector $\mathbf{r}$ is introduced as:

$$\mathbf{r} = \mathbf{x}_{ab} - \mathbf{P}\mathbf{q}_{ab} \tag{3}$$

AsLS now minimizes the quantity f by changing $\mathbf{q}_{ab}$:

$$f = \sum_{i=1}^{n_{chan}} w_i r_i^2 \tag{4}$$

The weight w_i in this equation depends on the sign of the corresponding residuals r_i. For a residual larger than zero the weight is set to p, while for a residual smaller than or equal to zero the weight is set to $1 - p$. This p is called the *asymmetry* factor ($0 < p < 1$). It is clear that if p is near zero, the positive residuals get much less weight than the negative ones; hence the coefficients $\mathbf{q}_{ab}$ will be such that the vast majority of residuals are positive.

Once the n_{chan} weights in $\mathbf{w}$ are given it is easy to estimate $\mathbf{q}_{ab}$ by a weighted linear regression [19]. And once the coefficients $\mathbf{q}_{ab}$ are known, it is trivial to set the weights $\mathbf{w}$. This suggests an iterative algorithm, starting with all weights equal and set to 1. One can show that this iterative algorithm is gradient-following [15]. The goal function is convex and convergence must follow [18]. Practice shows that convergence is nearly always obtained in 10 or less iterations.

2.6. Requirements of the method

Some requirements of the EBS method can be stated in advance. The spectra that are used to calculate the B-space should be representative of the spectral variation that is caused by the eluent and the instrument. If not, the EBS method will fail. As an example, consider a case where the eluent consists of two spectroscopic active species (B_1 and B_2). If only one of these species, say B_1, is captured in the B-space, the EBS method will yield an estimated "analyte" spectrum that is a combination of the real analyte spectrum and the spectrum of B_2. This is clearly undesirable. In most cases, however, enough baseline spectra are available to model the spectral variation of the eluent. Moreover, the selection of the spectra to be used is not critical.

Another limitation is posed by the amount of spectral overlap of analyte and eluent spectrum. If the shape of the eluent spectrum and the analyte spectrum are very similar, then the amount of specific analyte information in the measured spectrum is only limited. This will hamper the method and results in errors. In practice, this will not be a real limitation.

Finally, in deriving the method the spectral noise was disregarded. If the p value is set closer to 0, the negative residuals are more and more punished. In the end all residuals will thus

forced to be zero or positive. In the presence of spectral noise on the measured spectra this may lead to an (small) offset in the reconstructed analyte spectrum ($\mathbf{s}_a$). However, it is easy to correct for this type of offset. On the other hand for values of p close to 0.5 the positive and negative residuals will receive nearly equal weights. The asymmetric least square solution then will approach the ordinary least squares solution. The reconstructed analyte spectra ($\mathbf{s}_a$) will contain negative phases, which is obviously wrong.

2.7. Comparing the estimated and the true analyte spectra

To evaluate the EBS method, it is necessary to compare the background corrected spectrum with the true analyte spectrum. Two indicators are used for this purpose: the correlation coefficient (ρ) which is a measure of the similarity in shape between the estimated ($\hat{\mathbf{s}}_a$) and true analyte spectrum ($\mathbf{s}_a$). Also the amount of remaining spectral variation is calculated. This is expressed by the mean square error (MSE) of the spectral residuals.

The spectral residuals ($\mathbf{e}$) are calculated by modeling the estimated analyte spectrum as:

$$\hat{\mathbf{s}}_a = c_1\mathbf{s}_a + c_2 + \mathbf{e} \tag{5}$$

in which c_1 and c_2 are constants determined by a linear regression. The residuals are:

$$\mathbf{e} = \hat{\mathbf{s}}_a - c_1\mathbf{s}_a - c_2 \tag{6}$$

The MSE is defined as:

$$\text{MSE} = \frac{1}{n_{\text{chan}} - 2}||\mathbf{e}||^2 \tag{7}$$

A correlation coefficient that is close to unity and a low value for MSE are indicators for a good similarity between the extracted and the real analyte spectrum. The lowest value that the MSE can take is the variance of the instrumental spectral noise. A good estimate of the analyte spectrum will have MSE's close to this variance.

2.8. Reference method

A straightforward subtraction of the background spectrum using a measurement just before ($\mathbf{x}_{\text{before}}$) and just after elution ($\mathbf{x}_{\text{after}}$) of the analyte is used as a reference method (REF). The analyte spectrum ($\hat{\mathbf{s}}_a$) during elution at time t is calculated using the following equation:

$$\hat{\mathbf{s}}_a = \mathbf{x}_{ab} - \mathbf{x}_{\text{before}} - \frac{t - t_{\text{before}}}{t_{\text{after}} - t_{\text{before}}}(\mathbf{x}_{\text{after}} - \mathbf{x}_{\text{before}}) \tag{8}$$

Note that this reference method is an advanced version of an "auto-zero" approach in which only the eluent spectrum measured just before elution of the analyte is subtracted.

3. Experimental

3.1. LC–Raman

The Raman spectra were recorded using a LC system coupled to a Raman spectrometer via a liquid-core waveguide (LCW). The eluent composition was an aqueous 10 mM HCl solution with 5% (v/v) methanol. The flow was set to 0.4 ml/min. After passing through an on-line UV absorbance detector, the effluent was led into the LCW. The UV absorbance signal was used to determine the start and end times of the analyte elution.

The spectroscopic resolution of the Raman spectrometer was 20 cm^{-1} (FWHM, Full Width at Half Maximum). Adenosine 5′-monophosphate disodiumsalt (AMP; Fluka, Buchs, Switzerland) was used as model compound. An aqueous solution (40 μL) of 20 mg/mL AMP was injected into the LC system. Other chemicals were of analytical-grade quality. The standard deviation of the instrumental noise is estimated to be 350 counts. The spectrum of an aqueous solution of AMP was measured separately. This spectrum was corrected for water contribution. A spectral baseline correction was also performed. The resulting spectrum is the AMP reference spectrum. More details can be found in Dijkstra et al. [20,21].

3.2. LC–DAD

Experiments were carried out on a Waters 2690 LC system. Gradient control, data acquisition and analysis were controlled by Waters Millennium 3.2 software. The stationary phase was Supelco Discovery C_{18}, particle size 5 μm, pore diameter 180 Å, column dimensions were 150 mm × 2.1 mm i.d. and column temperature was maintained at 25 °C. The solvents were THF (Biosolve, Valkenswaard, The Netherlands) and acetonitrile (Rathburn, Walkerburn, UK), both were HPLC grade. The flow rate was 0.2 mL/min. Samples consisted of low-dispersity polystyrene standards (Polymer Labs., Church Stretton, UK, Pressure Chemical, Pittsburgh, PA, USA and Polymer Standards Service, Mainz, Germany). The sample-injection volume was 10 μL and sample concentrations were 1.5 mg/mL each. UV–vis spectra (resolution 1.2 nm) were collected each second. Details can be found in Fitzpatrick et al. [22]. UV–vis spectra of polystyrene standards in THF were recorded separately in a cuvette on a Hewlett-Packard 8453 spectrophotometer (path length: 2 mm, spectral resolution: 1 nm). Two full repeats were measured of each polystyrene standard. The mean spectrum was used for comparison.

3.3. Simulations

The simulations and all processing of the measured spectra were performed in Matlab 6.1 (MathWorks, Natick, USA, version 6.1, 2001). The Matlab code to perform correction with the EBS method is available from the author.

Table 1
Overview of simulations

Simulation	Number of eluent species	Baseline drift	Spectral noise	ρ_{REF}	ρ_{EBS}	MSE_{REF} ($\times10^{-6}$)	MSE_{EBS} ($\times10^{-6}$)
1	1	No	No[a]	0.899	>0.999	15.8	0.00007
2	2	No	No	0.877	0.995	67.7	0.75
3[b]	2	No	No	0.605	>0.999	128	0.085
				0.535	>0.999	198	0.032
				0.517	>0.999	230	0.002
4	2	Yes	Yes	0.743	0.976	63.7	3.04

ρ_{EBS} is the correlation coefficient between estimated and true analyte spectrum for the EBS method at peak apex. ρ_{REF} is same for the REF method. The MSE values are calculated according to Eq. [7].

[a] Results of simulation 1 with noise are given in Table 2.

[b] The entries of simulation 3 are supplied for three estimated spectra at the maximum elution of the three analytes (see Fig. 4 for analyte spectra).

Table 2
Results simulation 1; spectral noise added

$(S/N)_{tot}$	$(S/N)_a$	ρ_{REF}	ρ_{EBS}	MSE_{REF} ($\times10^{-6}$)	MSE_{EBS} ($\times10^{-6}$)
10^6	6.10^4	0.899	>0.999	15.8	0.00007
10^5	6.10^3	0.899	>0.999	15.8	0.00013
10^4	6.10^2	0.899	>0.999	15.8	0.0280
10^3	60	0.898	0.998	16.4	3.21
10^2	6	0.720	0.746	68.6	33.9

$(S/N)_{tot}$ is the S/N ratio calculated as the maximum total spectral intensity measured during the run (~0.5) divided by the standard deviation of the white noise added. $(S/N)_a$ is the S/N ratio for the analyte signal, calculated as the maximum spectral intensity of the analyte measured during a run divided by the same standard deviation of the noise. Explanation of other symbols is supplied in Table 1.

4. Results and discussion

4.1. Simulations

Simulations were performed to illustrate the EBS method. An overview of the simulations is supplied in Table 1. In the first set of simulations it is assumed that apart from the analyte, only one eluent species is present. The noiseless spectra are shown in Fig. 1A. It is assumed that the concentration of the analyte at elution maximum is about 3% (v/v) of the flow through the detection cell. The REF method clearly overcorrects for the presence of the eluent spectrum (Fig. 1C), because the intensity of the eluent spectrum at peak-start (and peak-stop) is higher than during peak elution. Too much background signal is subtracted by the REF method. The EBS method on the other hand, yields a near perfect estimate of the analyte spectrum (Fig. 1D). The correlation between the estimated and the true analyte spectrum is larger than 0.999 (Table 1). The difference between the REF and EBS method is also reflected in the MSE values. Because this simulation is noiseless, it is expected that the MSE value is equal to zero. This means that the estimated analyte spectrum in such a simple case should be identical to the true analyte spectrum. The MSE for the EBS method is indeed close to zero, but for the REF method, this is clearly not the case. To investigate the sensitivity of the method the noise level is varied. At each noise level 500 runs are performed. In each run a different noise realization is used. Normally distributed white noise was used in these runs. Table 2 shows that the MSE value for the EBS and the REF method increases with a decreasing signal-to-noise (S/N) level as may be expected. For both methods the correlation between the true analyte spectrum and the estimated analyte spectrum at peak apex decreases. At a S/N level of 6 with respect to the maximum level of the analyte spectrum (0.03) both methods break down. For this S/N level the analyte spectrum is getting blurred totally by spectral noise and both the methods are failing. For higher S/N levels the EBS method outperforms the REF method. The difference between the methods becomes smaller for decreasing S/N ratio's.

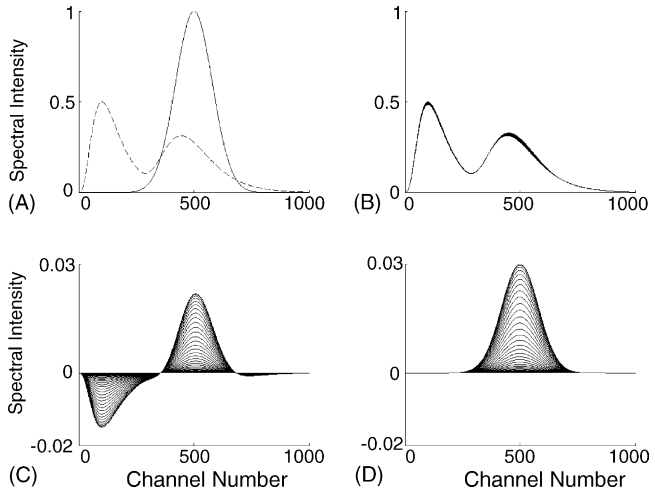

Fig. 1. Results of simulation 1. (A) Spectrum of analyte (solid) and background species (dashed). (B) Spectra during elution of analyte. (C) Reconstructed analyte spectra using the REF method. (D) Same for the EBS method.

In the second simulation, another eluent species is added. Fig. 2 shows that the eluent composition changes slightly during the chromatographic run, as it is the case for gradient

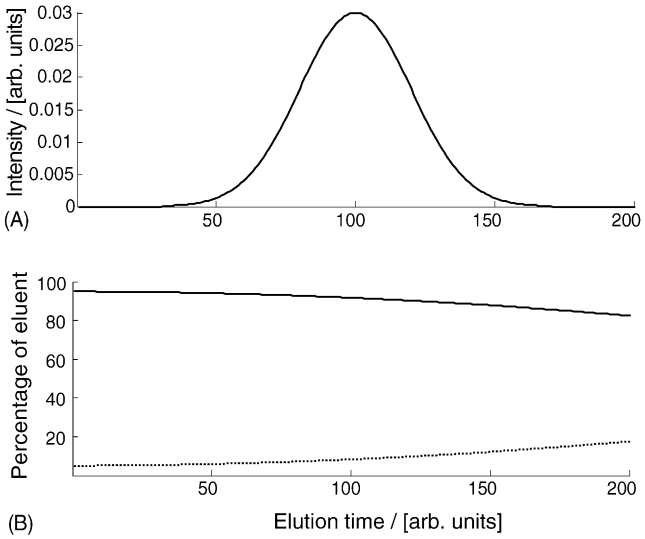

Fig. 2. Simulation 2. (A) Elution profile of analyte. (B) Change in eluent composition: first eluent species (solid), second eluent species (dotted).

elution. The concentration of the second eluent species increases from 5% to about 18% during elution. Fig. 3C shows that the REF method has severe difficulties in returning a correct estimate of the analyte spectrum. The estimated spectra contain large spurious bands that do not originate from the analyte. The intensity of the analyte spectrum (at the correct band position) is also incorrect. In this case, the EBS method (Fig. 3D) is not perfect, – a small positive band shows up at about 350 nm, but the overall picture remains rather satisfactory. The correlation between the estimated and the true analyte spectrum is still high (0.995, Table 1).

A more difficult situation is considered in simulation 3. Instead of only one analyte as in simulation 2, three co-eluting

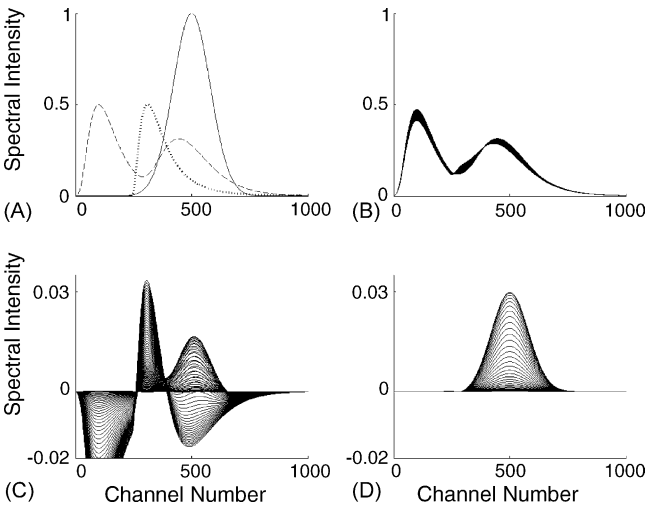

Fig. 3. Results of simulation 2: analyte and two eluent species, eluent composition is changing (see: Fig. 2). (A) Spectrum of analyte (solid), first background species (dashed) and second background species (dotted). (B) Spectra during elution of analyte. (C) Reconstructed analyte spectra using the REF method. (D) Same for the EBS method.

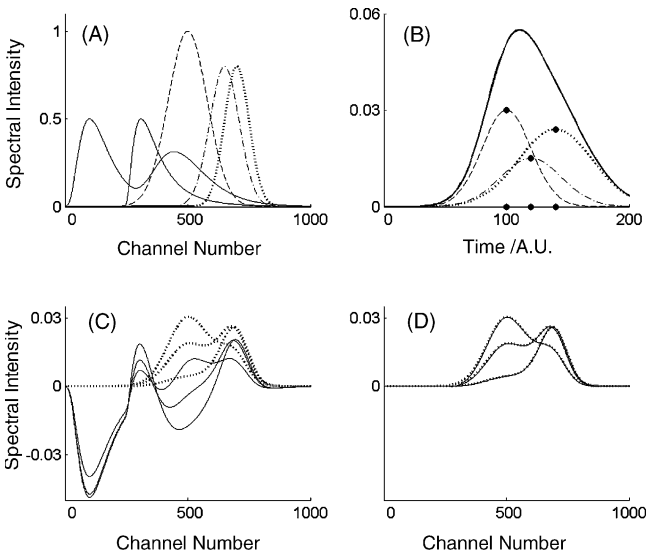

Fig. 4. Results of simulation 3: three co-eluting analytes and two eluent species. (A) Spectra of the three analytes (1: dashed, 2: dash-dotted, 3: dotted) and of both background species (solid). (B) Elution profiles of the three analytes separately (1: dashed, 2: dash-dotted, 3: dotted) and of the overall elution profile (solid bold line). (C) The spectra (solid) at the three times of maximum elution (indicated by dots in (B)) reconstructed with the REF method. True spectra at the same points in time (dotted). Note that each true spectrum shown is the sum of the three analyte spectra because of co-elution of the analytes. (D) Same as (C), but for the EBS method.

analytes with overlapping spectra are used. The spectra of the compounds (three analytes and two background species) are shown in Fig. 4A, the separate chromatographic profiles of the analytes are shown in Fig. 4B. Fig. 4C and D shows some of the estimated analyte spectra recovered with the REF and the EBS method. It can be seen that the true spectral profiles at elution maximum of analytes can be recovered very well with the EBS method, but not with the REF method. For the EBS method only slight deviations of the true spectra occur at channel numbers between 300 and 500. This is also reflected in Table 1 (3rd–5th entry) in which the figures of merit are collected. Correlation of estimated spectra with true spectra is high for the EBS method (>0.999) and low (∼0.5–0.6) for the REF method. The MSE values are very low for the EBS method (target value is zero) indicating no residual spectral variation. The MSE values are higher for the REF method.

Finally, in the fourth simulation, a varying spectral baseline and spectral noise is added to the data generated in the simulation 2. The varying spectral baseline and noise are shown in Fig. 5A. The generated spectral noise is white normal distributed noise ($\sigma = 10^{-4}$). Fig. 5B shows the reconstructed analyte spectra using the REF method. The disturbing spectral baselines have the shape of a second order polynomial. Fig. 5C and D shows that the overall results for the EBS method are still very similar to simulation 2. Adding noise and a varying baseline does not hamper the EBS method. In fact, the EBS method seems to be able to suppress spectral baseline fluctuations. The correlation between

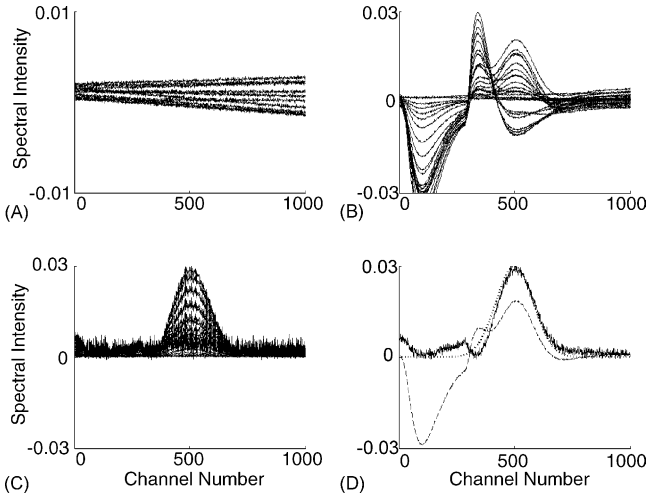

Fig. 5. Results of simulation 4. (A) Spectral baseline and spectral noise added. (B) Reconstructed analyte spectra using the REF method. (C) Same for the EBS method. (D) True analyte spectrum (dotted), estimated analyte spectrum with the EBS method (solid) and the REF method (dashed) at maximum elution.

the estimated and the true analyte spectrum is only slightly lower (0.976) than in simulation 2 (Table 1). When the reconstructed analyte spectrum would perfectly match the true analyte spectrum the presence of spectral noise determines the lowest MSE value. In that case the limiting MSE would be 10^{-8} (σ^2). It can be seen that both methods have an MSE larger than this lowest possible value. The REF method, however, has a much higher MSE value than the EBS method.

To establish the effect of a different choice of the *p*-value, the correlation between true and reconstructed analyte(s) spectral contribution is calculated as a function of the *p*-value. The *p*-value is varied between 10^{-1} and 10^{-5}. Fig. 6 shows the results for simulations 1–3 that have a different amount

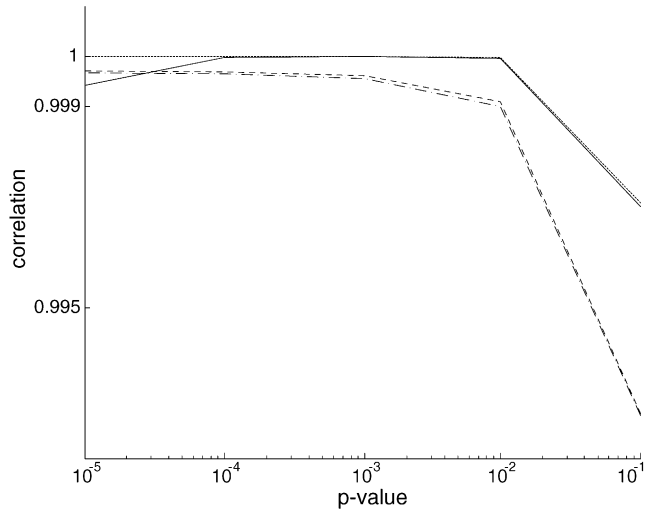

Fig. 6. Correlation between reconstructed and true spectral contribution of analyte(s) as a function of the *p*-value of the EBS method. Data of simulation 1 (dotted), simulation 2 (dash-dot), simulation 3 (dash–dash) and data of simulation 1 disturbed by noise, S/N = 100 (solid).

of spectral overlap between analyte and eluent spectra. It can be seen that for *p*-values smaller than 10^{-2} the correlation is high (>0.999) and slightly varying. Taking the *p*-value larger than 10^{-2} decreases the correlation. Overall the correlation is not very sensitive to the *p*-value.

4.2. Application 1: extracting analyte spectrum in LC–Raman

In LC–Raman both the analyte and the solvent (eluent) contribute to the Raman spectrum. Fig. 7A shows the Raman spectrum of the eluent. The bands at 1025, 1120, 1171 and 1471 cm^{-1} are from methanol (dotted vertical lines) and the band at 1647 cm^{-1} (vertical dashed) is from water. Fig. 7B shows the reference spectrum of AMP that was measured separately. The figure also shows the spectra that are recorded during analyte elution (Fig. 7C). The inset shows the wavenumber range where strong analyte bands are present. No distinct analyte information can be found. Obviously, the eluent spectrum overwhelms the smaller analyte bands.

The EBS method uses all spectra measured at baseline level to determine the B-space. The points in time that are used to determine this space are shown in Fig. 8 (solid fat line showing the UV detector trace of the separation). The *IND* method finds four principal components. The asymmetry factor (*p*) of the EBS method was set to 0.01.

Fig. 9 shows the estimated analyte spectra at maximum elution for both methods. Fig. 9C shows the analyte spectrum and Fig. 9A the residual spectral variation for the REF method. Fig. 9B and D shows the results for the EBS method.

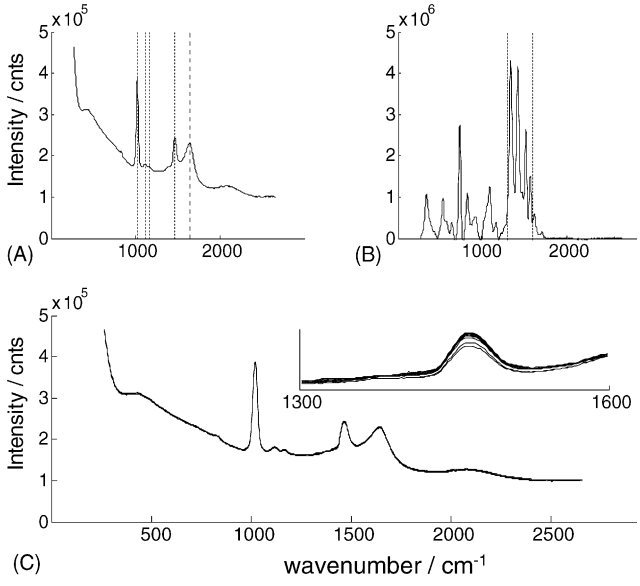

Fig. 7. (A) Raman spectrum of solvent (eluent). Major bands of methanol (dotted vertical lines) and of water (dashed vertical line) are shown. (B) Raman spectrum of analyte (AMP). (C) Raman spectra measured during elution of AMP. Inset magnifies wavenumber range that should contain large analyte bands.

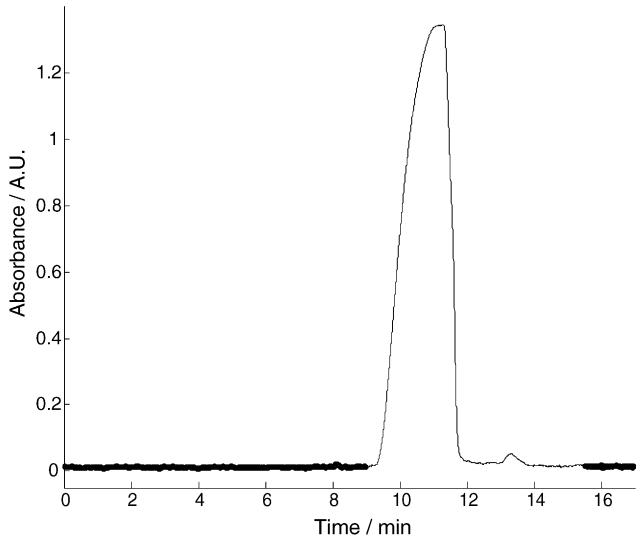

Fig. 8. UV-detector signal of AMP: (solid fat) points in time considered to be part of baseline. Spectra measured at these time points are used to determine the B-space.

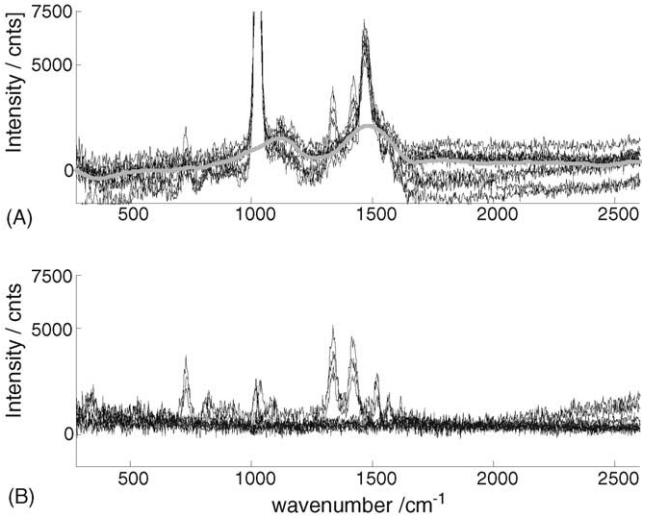

Fig. 10. (A) Estimated analyte spectra during elution (the REF method). The gray solid line indicates a slow drift-like estimation artifact. (B) Same for the EBS method.

The residual spectral variation is calculated using equation [6]. It can be seen that the REF method fails in two ways. In the first place, two spurious bands can be seen in the estimated spectrum. These bands are located at approximately 1124 and 1468 cm^{-1} (dotted vertical lines). At these positions strong bands are present in the spectrum of the eluent (Fig. 7). These bands can be assigned to methanol. It appears that the concentration of methanol during AMP elution was not completely constant. Secondly, the analyte spectrum (the REF method) is more disturbed by baseline drift. This can be seen more clearly in Fig. 10, where six estimated analyte spectra during AMP elution are overlaid. The spectral baseline drift is much smaller for the EBS method, although it has not been fully removed.

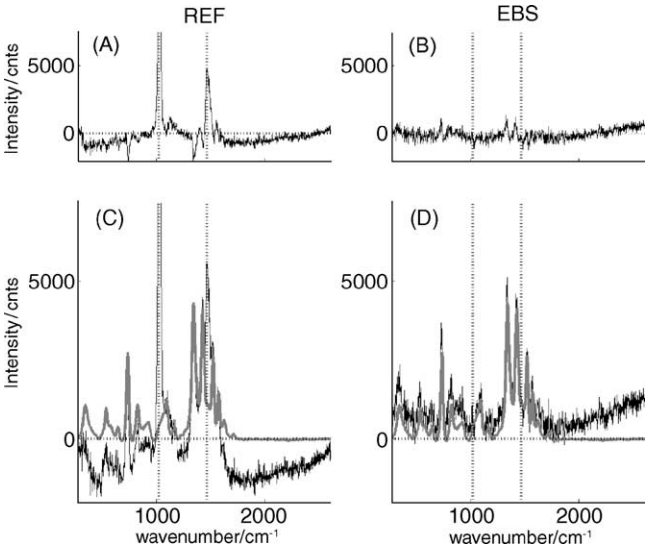

Fig. 9. Results of the REF method (A and C). (A) Residuals calculated using Eq. (6). (C) Estimated AMP spectrum (solid) at elution maximum and 'true' analyte spectrum (grayed solid). Results of the EBS method (B and D). The vertical dotted lines are drawn at strong bands of the eluent (1124 and 1468 cm^{-1}).

In Fig. 11A, the correlation coefficient of the estimated AMP spectrum and the known spectrum of AMP is plotted as a function of elution time. The EBS method has a fairly high correlation (>0.9) at the elution maximum and has near zero correlations at the peak-start and the peak-stop times. On the other hand, for the REF method, the maximum correlation is smaller (∼0.6) and at the peak-start and peak-stop a small correlation (∼0.2) still exists. The small correlations at peak-start and peak-stop can be traced back to the combination of the broad 'band-like' features (see the gray line in Fig. 10A) and the spurious bands at 1124 and 1468 cm^{-1} in the estimated analyte spectra. When both effects are removed from the estimated analyte spectra, the correlation curve for the REF method starts and ends at approximately zero correlation (Fig. 11A, dash-dotted curve), but also the maximum

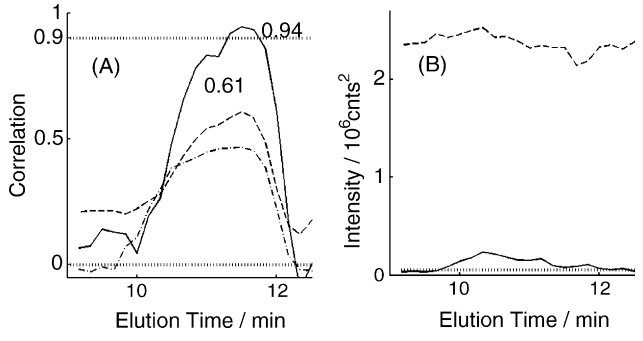

Fig. 11. (A) Correlation coefficient between 'true' AMP spectrum and estimated AMP spectrum as a function of elution time: the EBS method (solid line), the REF method (dashed line) and curve (dash-dotted) of corrected (see text) analyte spectra (also the REF method). (B) MSE (Eq. (7)) between of residuals of estimated AMP spectrum as a function of elution time: the EBS method (solid) and the REF method (dashed). The dotted horizontal line ($\sim 5 \times 10^4$) is at instrumental noise level.

correlation drops to a lower value (<0.5). In Fig. 11B shows the variance of spectral residuals for the EBS method is close to the variance of the spectral noise. It is two orders of magnitude smaller than for the REF method. It was also checked whether the number of spectra used for estimating the B-space is critical. To a certain degree, changing the number of spectra does not affect the shape of the estimated analyte spectrum. Of course always more spectra should be used than the number of independent phenomena hidden in the baseline spectra (matrix $\mathbf{X}_b$). Using more spectra improves the estimate of the B-space and this stabilizes the estimated analyte spectrum.

4.3. Application 2: background correction in LC–DAD

To further illustrate the feasibility of the EBS method, two (different) LC–DAD separations were performed. A mixture of four polystyrene standards was injected in each sample run. Immediately, after that, a blank run was recorded. Fig. 12 shows the total spectral intensity in the UV–vis spectrum in the wavelength range from 200 to 300 nm for the sample and the blank runs.

The settings for the EBS method were as follows. The spectra used to determine the B-space are indicated by dots. The asymmetry factor (p) was 0.01. Similar results were obtained for p-values between 10^{-5} and 10^{-1}. No dependency on the value of p could be detected. For the REF method, the spectra just before and just after elution of the polystyrenes were selected (circles in Fig. 12). For both LC–DAD runs the adequate number of PCs was 3. The B-space thus has dimension 3. Two eluent species change in relative concentration during the chromatographic run. Therefore, it is expected to find at least a two dimensional B-space. The third dimension is apparently required to describe small disturbances in the eluent spectra during the run.

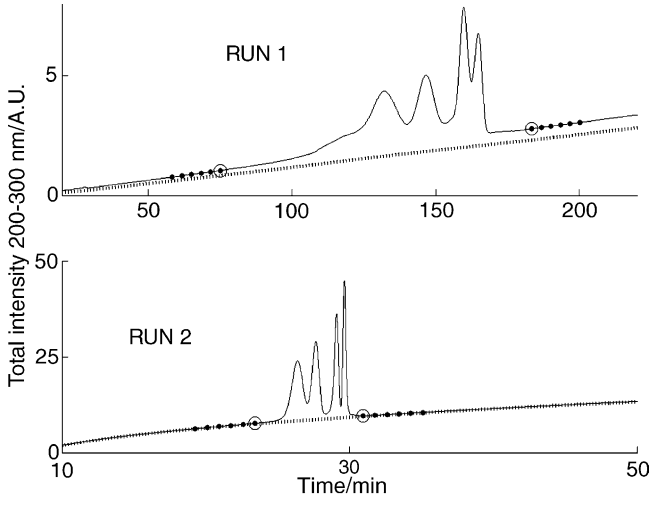

Fig. 12. Total spectral intensity between 200 and 300 nm as a function of elution time. Dotted line is blank chromatographic run. Run 1: sample is mixture of four polystyrene standards (M_p = 10 900, 17 600, 30 000 and 39 200) in tetrahydrofuran–acetonitrile. Run 2: sample is mixture four polystyrene standards (M_p = 10 900, 17 600, 39 200 and 76 600) also in tetrahydrofuran–acetonitrile; gradient is different.

Table 3
Results of HPLC–DAD measurements

Run	M_p of polystyrene	ρ_{REF}	ρ_{EBS}	MSE_{REF} ($\times 10^{-4}$)	MSE_{EBS} ($\times 10^{-4}$)
1	10.900	0.90	0.95	4.65	1.29
1	17.600	0.89	0.94	7.95	2.01
1	30.000	0.90	0.94	26.9	7.32
1	39.200	0.90	0.95	14.2	4.22
2	10.900	0.85	0.93	322	94.1
2	17.600	0.87	0.92	498	195
2	30.000	0.91	0.94	632	321
2	39.200	0.95	0.96	547	322

Comparison of off-line measured spectra of the standards with the reconstructed spectra using the EBS and the REF method for runs 1 and 2.

Table 3 shows the results when the reconstructed analyte at peak maxima are compared with off-line measured UV–vis spectra of the standards. The correlation for the EBS method is somewhat higher than for the REF method, MSE of the EBS method are lower. In Fig. 13 for run 2 the reconstructed (the REF and the EBS methods) spectra at the apex of the last peak are shown together with the off-line measured spectrum of the 39.200 standard. It can be seen that the REF method yields a too high intensity of the estimated spectrum at low wavelengths (near 200 nm). This artifact is located at the position of strong THF band (200–210 nm). Thus, the change in the relative concentration of THF during the run will, for the REF method, still affect the overall spectral intensity of the estimated polystyrene spectrum. This explains the somewhat lower correlation coefficients and higher MSE values for the REF method (Table 3).

In order to show that the EBS method is more flexible in use than the REF method the next processing was performed on the collected data. Only the spectra measured before elu-

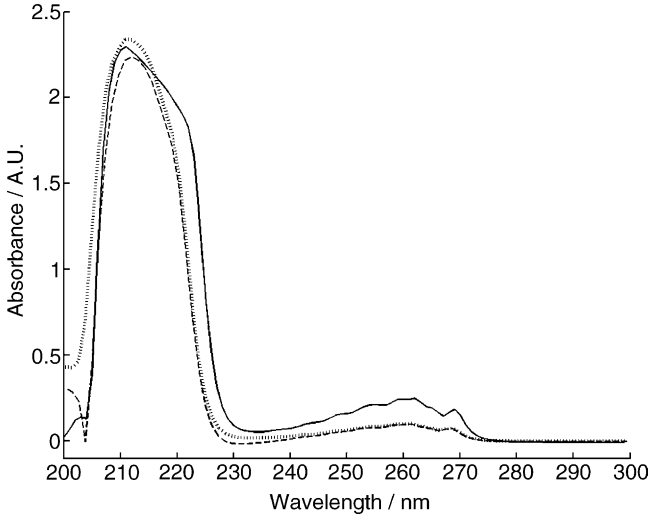

Fig. 13. *Mean* spectrum of polystyrene standard 39.200 (solid line) measured off-line on HP8453 diode array spectrometer. For run 2 the reconstructed polystyrene spectra for the apex of the corresponding peak is shown: the REF method (dotted line) and the EBS method (dashed line).

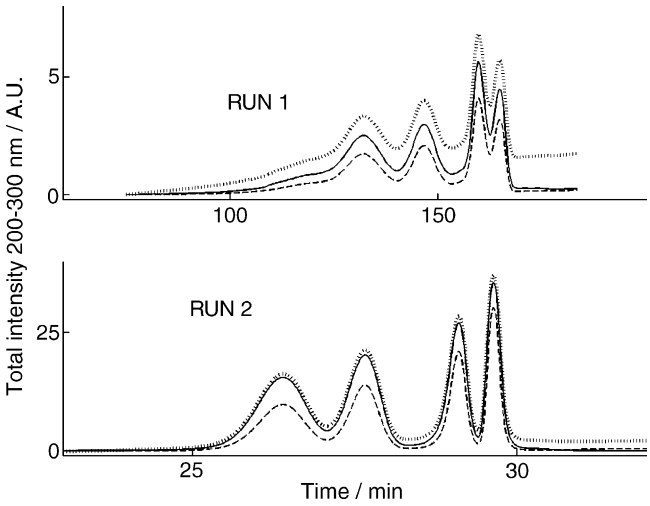

Fig. 14. Total spectral intensity between 200 and 300 nm is plotted as a function of elution time. Sample run signal corrected for blank run signal (solid line); signal corrected using auto-zero method method (dotted line), signal corrected using the EBS method (dashed line). Only the background spectra (shown in Fig. 12) before elution of standards are used for correction.

tion of the polystyrene peak cluster (Fig. 12), are used for correcting the contribution of the background signal (establishing the B-space). In effect, this means that the REF method now truly is an auto-zero method. Results in Fig. 14 shows that the EBS method succeeds in nearly fully correcting the baseline of the runs, while the auto-zero method cannot fully correct for the baseline and some additional baseline correction procedure is needed. For both the REF and the EBS method the peak area ratios are the same as for the blank corrected chromatographic signal.

5. Conclusions

A new method (EBS) based on asymmetric least squares is proposed to eliminate the spectral contribution from the eluent in the hyphenated chromatography. Simulations and first tests of this background correction method on LC–Raman and LC–DAD data show its feasibility. Advantage of the method are that it only needs the data of one single chromatographic run and that each spectrum during analyte elution can be analyzed separately, without using relation to other spectra during elution. From simulations it can be concluded that the new method performs better than a straightforward spectral subtraction method (REF). This is also true in the presence of spectral noise. Furthermore, setting the value of the only parameter of the method (asymmetry factor) is easy and turned out not to be critical. The practical significance of the method is shown for an LC–Raman and an LC–DAD application. In LC–Raman the EBS method can extract the analyte spectrum much better the REF method. A more extensive comparison between the methods is made in a companion publication [1]. In the LC–DAD examples the performance advantage of the EBS method is smaller than for LC–Raman. However, the flexibility of the EBS method allows one the correct for spectral background using only some spectra measured on one side of the eluting peak cluster.

References

[1] R.J. Dijkstra, H.F.M. Boelens, J.A. Westerhuis, F. Ariese, U.A.Th. Brinkman, C. Gooijer, Anal. Chim. Acta 519 (2004) 129.
[2] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[3] Y.Z. Liang, R. Manne, O.M. Kvalheim, Chem. Intell. Lab. Syst. 14 (1992) 175.
[4] Y.D. Wang, O.S. Borgen, B.R. Kowalski, J. Chemom. 7 (1993) 439.
[5] A.K. Smilde, Y.D. Wang, B.R. Kowalski, J. Chemom. 8 (1994) 21.
[6] D.D. Gerow, S.C. Rutan, Anal. Chim. Acta 184 (1986) 53.
[7] Y.Z. Liang, O.M. Kvalheim, A. Rahmani, R.G. Brereton, Chem. Intell. Lab. Syst. 18 (1993) 265.
[8] P.J. Gemperline, J.H. Cho, B. Archer, J. Chemom. 13 (1999) 153.
[9] H. Martens, T. Naes, Multivariate Calibration, Wiley, New York, 1989.
[10] I. Jolliffe, Principal Component Analysis, Springer, New York, 2002.
[11] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 2002 (Chapter 4).
[12] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.
[13] M. Meloun, J. Apek, P. Miksik, R.G. Brereton, Anal. Chim. Acta 423 (2000) 51.
[14] E.R. Malinowski, Anal. Chem. 49 (1977) 612.
[15] W.K. Newey, J.L. Powell, Econometrica 55 (1987) 819.
[16] P.H.C. Eilers, Kwantitatieve Methoden 23 (1988) 45.
[17] P.H.C. Eilers, Anal. Chem. 76 (2004) 404.
[18] P. Groenen, personal communication.
[19] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1998.
[20] R.J. Dijkstra, A.N. Bader, G.P. Hoornweg, U.A.Th. Brinkman, C. Gooijer, Anal. Chem. 71 (1999) 4575.
[21] R.J. Dijkstra, G.J. Slooten, A. Stortelder, J.B. Buijs, F. Ariese, U.A.Th. Brinkman, C. Gooijer, J. Chromatogr. A 918 (2001) 25.
[22] F. Fitzpatrick, B. Staal, P. Schoenmakers, J. Chromatogr. A, submitted for publication.