ELSEVIER

Journal of Chromatography A, 958 (2002) 35–49

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Peak deconvolution in one-dimensional chromatography using a two-way data approach

G. Vivó-Truyols, J.R. Torres-Lapasió*, R.D. Caballero, M.C. García-Alvarez-Coque

Departamento de Química Analítica, Universitat de València, c/Dr. Moliner 50, 46100 Burjassot, Spain

Received 1 February 2002; received in revised form 4 April 2002; accepted 5 April 2002

Abstract

A deconvolution methodology for overlapped chromatographic signals is proposed. Several single-wavelength chromatograms of binary mixtures, obtained in different runs at diverse concentration ratios of the individual components, were simultaneously processed (multi-batch approach), after being arranged as two-way data. The chromatograms were modelled as linear combinations of forced peak profiles according to a polynomially modified Gaussian equation. The fitting was performed with a previously reported hybrid genetic algorithm with local search, leaving all model parameters free. The approach yielded more accurate solutions than those found when each experimental chromatogram was fitted independently to the peak model (single-batch approach). The improvement was especially significant for those chromatograms where the peaks were severely affected by the tails of the preceding compounds. Peak shifts among chromatograms, which are a usual source of non-bilinearity, were modelled in a continuous domain instead of in a discrete way, which avoided some drawbacks associated with latent variable methods. An experimental design involving simulated chromatograms was applied to check the method performance. Five main factors affecting the deconvolution were examined: concentration pattern, chromatographic resolution, number of batches and replicates, and noise level, which were evaluated using first- and second-order figures of merit. The method was also tested on three real samples containing compounds showing different overlap. Four multi-batch deconvolution methods were considered differing in the nature of the processed information and kind of peak matching among chromatograms. In all cases, the multi-batch deconvolution yielded better performance than the single-batch approach.

Keywords: Peak deconvolution; Second-order data; One-dimensional chromatography

1. Introduction

Chromatographic techniques often give rise to situations where reaching complete resolution is not possible. This is more frequent with compounds belonging to the same family, which undergo similar interactions with the separation environment. There is thus a greater probability that they coelute under a given experimental condition. Fortunately, even in such cases, coelution is not always complete, and finding a situation where at least a partial separation exists is feasible.

The selectivity is conventionally improved by modifying the chromatographic conditions to isolate the individual peaks. When the peaks remain unresolved, the quantification can be faced by enriching the information obtained within a single chromato-

*Corresponding author.

E-mail address: jrtorres@uv.es (J.R. Torres-Lapasió).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00409-0

gram. This is the case of hyphenated techniques, such as high-performance liquid chromatography–diode array detection (HPLC–DAD), which yield the so-called second-order data, e.g. a collection of spectra measured at different times [1]. Through the application of multivariate analysis, such as self-modelling curve resolution methods, the chromatographer is able to retrieve the individual contributions of each compound. Selective regions should, however, exist for each compound in the sample. When this condition is not satisfied sufficiently, the solutions found are not unique and can be more or less biased.

When second-order instruments are not available, true second-order data cannot be obtained [1]. However, a collection of measurements of several related samples prepared in different conditions and obtained with a first-order instrument, share the same properties of second-order data, and can be considered as a particular case of two-way data. Each data order is related to a source of variation of the signal. Application of multivariate analysis to the whole data set instead of treating independently each experiment is often beneficial. Some examples that exploit this kind of ''pseudo second-order data'' can be found in different fields, such as kinetics studies [2,3], spectrometric titrations [4–6], or industrial process monitoring [7].

In the chromatographic literature, some reports where different one-dimensional chromatograms are treated altogether have been published. Synovec and co-workers [8] deconvolved overlapped peaks treating data from different runs, obtained using single-channel detection. In that study, the relative change in analyte concentration was measured by computing the ratio of two sequential chromatograms. Later, Hämäläinen and co-workers [9] applied an eigenanalysis technique (HELP) to a set of associated one-dimensional chromatograms, processing all of them as a single second-order experiment. In subsequent studies, this strategy was applied to the assessment of peak identity [10].

For a successful application of any eigenanalysis technique, the matrix to be processed must be bilinear [11]. This condition is not always well fulfilled, especially when the two-way data come from a set of one-dimensional chromatograms obtained in different conditions, where usually the retention times do not agree. In such a situation, a previous correction should be performed to compensate the retention time shifts that are produced among injections, in order to make the chromatographic profiles comparable [9]. Since the data are discrete, this peak matching is limited by the instrument data acquisition rate. Also, the relative peak positions should be constant among injections. An imperfect correction of these peak shifts produces a lack of bilinearity, and consequently, a wrong solution. The accommodation of the whole signal set to a combination of forced elution profiles allows a better minimisation of these sources of error.

On the other hand, the treatment of data measured at low signal-to-noise ratios, or the use of low instrumental data acquisition frequencies, are problematic. Processing of replicates is considered in this work to overcome this problem. Diverse alternatives have been published, generally applied to fast chromatographic modes. Some examples are the application of signal averaging in gas chromatography [12,13] and high-speed open tubular liquid chromatography [14], and the use of adaptive digital filters in multi-input chromatography [15].

In this work, artificial and real binary samples exhibiting different overlaps were deconvolved to test the performance of the proposed methodology. The samples were mixtures of oxytetracycline–tetracycline, sulfathiazole–sulfachloropyridazine and sulfisoxazole–sulfapyridine. In each case, two-way data corresponding to chromatograms at different concentration ratios were fitted simultaneously (multi-batch deconvolution) to a linear combination of modelled signals. This approach requires a powerful tool, able to explore and fit the peak model parameters accurately. A local optimised genetic algorithm (LOGA) previously described was used [16].

2. Theory

2.1. Modelling a set of chromatograms

Many published peak models [17–21] can be applied to deconvolution problems. Most of them, however, involve a high number of parameters or are difficult to fit. In this work, a simple semi-empirical model is used [19], which is a Gaussian function

where the standard deviation varies quadratically (PMG_2) with the distance to the retention time, t_R:

$$\hat{y} = h \exp\left[-\frac{1}{2}\left(\frac{t - t_R}{\sigma_0 + \sigma_1(t - t_R) + \sigma_2(t - t_R)^2} \right)^2 \right] \quad (1)$$

$\hat{y}$ being the predicted signal at time t, h the peak height and σ_i coefficients related to the width and asymmetry of the chromatographic peak. One drawback of this model is the strong growth of the baseline outside the peak region [19,20]. This problem was solved by finding the minima of the function at both sides of the peak, and setting these values for times outside the peak window. Also, the chromatograms of the mixtures were cut to avoid unnecessary baseline points.

Deconvolution procedures based on peak models are usually performed in two consecutive steps [20]. In the first step, chromatograms of standards are fitted to determine the σ_i coefficients for each compound. In the second step, these parameters are used (keeping them constant) to fit t_R and h in the deconvolution of the sample chromatogram(s). Leaving all parameters free in the fitting has the advantage of avoiding the requirement of a full previous knowledge on the individual compounds. However, this approach (which will be called single-batch deconvolution) may lead to unstable and biased solutions for strongly overlapped peaks.

Adapting Eq. (1) to a multi-batch treatment can attenuate these problems. A two-way data can be obtained from a set of chromatograms involving *ns* solutes, obtained in *nb* batches (i.e. runs at different solute concentration ratios), each one replicated *nr* times. The predicted signal for a chromatogram of a sample, defined by its replicate and batch identification number (r and b), can be calculated as follows:

$$\hat{y}_{r,b} = \sum_{s=1}^{ns} h_{s,b} \times \exp\left[-\frac{1}{2}\left(\frac{t - t_{R,s} + t_{M,r,b}}{\sigma_{0,s} + \sigma_{1,s}(t - t_{R,s} + t_{M,r,b}) + \sigma_{2,s}(t - t_{R,s} + t_{M,r,b})^2} \right)^2 \right] \quad (2)$$

In this equation, some parameters describe the peak position and shape, and are specific for each solute (s) (i.e. common for all replicates and batches): $t_{R,s}$, $\sigma_{0,s}$, $\sigma_{1,s}$ and $\sigma_{2,s}$. In contrast, the matching time, $t_{M,r,b}$, which models the shift in retention time among runs, takes a characteristic value for each chromatogram. This parameter is introduced to compensate the random fluctuations in pump pressure and sample injection. A reference value of $t_{M,r,b}$ is required to get a single solution from the deconvolution. Accordingly, one of the injections was taken as the origin of time shifts, setting its $t_{M,r,b}$ value to zero. Finally, the peak height ($h_{s,b}$) is specific for each batch.

Although a linear standard deviation yields acceptable results, the quadratic model (Eq. (1)) is able to fit more accurately strongly asymmetrical peaks, using a reasonable number of parameters. The use of a small set of parameters is interesting in situations where a large number of individual signals have to be fitted simultaneously. However, higher order polynomials (i.e. including additional terms: $\sigma_{3,s}$, $\sigma_{4,s}$, etc.) can be used to enhance the descriptions of experimental peaks.

The multi-batch deconvolution of chromatograms using peak models involves an important advantage with regard to latent variable methods: some sources of non-bilinearity can be modelled in a continuous domain. As commented, when some flow-rate irregularities are present during the elution of the compounds of interest, the relative retention time of a given peak (i.e. referred to the other peaks in the chromatogram) is not constant among injections. In such a situation, latent variable methods will produce biased results. However, fitting the experimental chromatograms to Eq. (2) will not assure a correct answer to the problem either, since the matching time compensates the retention time shifts of the whole chromatogram as a block, displacing all the peaks simultaneously. A solution that overcomes this problem is the following equation:

$$\hat{y}_{r,b} = \sum_{s=1}^{ns} h_{s,b} \times \exp\left[-\frac{1}{2}\left(\frac{t - t_{R,s,r,b}}{\sigma_{0,s} + \sigma_{1,s}(t - t_{R,s,r,b}) + \sigma_{2,s}(t - t_{R,s,r,b})^2} \right)^2 \right] \quad (3)$$

Note that the parameter $t_{M,r,b}$ is not present in Eq. (3), and the retention time of each solute, ($t_{R,s,r,b}$) varies independently for each replicate and batch.

In both treatments (Eqs. (2) and (3)), the number of peaks must be known and a hypothesis is implicitly accepted: the peak shape of each compound does not vary significantly among injections (i.e. for a solute *s*, only $h_{s,b}$ may change). This assumption holds well in most cases, and is the key that explains the greater benefits achieved from a simultaneous deconvolution of chromatograms of compounds present at different concentration ratios.

2.2. Optimisation method

The multi-batch deconvolution of two-way data is based on Eqs. (2) or (3), which involve a large number of parameters. In addition, both equations are non-linear, and therefore, an iterative fitting procedure is required. In this work, a summation of squared residuals among experimental and predicted chromatograms (SSR), extended to all experimental points in all replicates and batches, was selected as objective function to be minimised. This multi-batch fitting is especially difficult in situations where a high correlation in both data orders exists. In this case, non well-defined error surfaces (i.e. SSR plots as a function of the model parameters) are obtained, which is translated in more than one solution of similar quality. In order to overcome this problem, the selected algorithm must show a large exploration capability, having nevertheless a high precision to allow an accurate fitting of the model parameters.

Natural computation has experienced a significant expansion in recent years [22]. This idea, which is inspired in the imitation of natural world mechanisms, has given rise to powerful optimisation methods, that have been adapted to a wide diversity of problems, such as retention modelling in HPLC [23] or peak deconvolution [24]. In a recent work, a hybrid genetic algorithm with local search, called LOGA, was proposed for peak deconvolution [16]. LOGA fully hybridises local (Gauss–Newton) and global (genetic algorithm) optimisation techniques. This algorithm was also applied to combinatorial problems with excellent results [25]. It is especially useful in optimisations having high dimensional search spaces (i.e. involving a large number of parameters), with non well-defined error surfaces, as those found in this work. LOGA is here applied to deconvolve sets of artificial and real binary mixtures.

2.3. First- and second-order figures of merit

The multi-batch deconvolution of two-way information entails benefits only if the involved data orders are at least partially uncorrelated. This means, on the one hand, that no full coelution should happen between solutes (time—or chromatographic—order), and on the other, the concentration ratio should vary among the processed batches for each solute (batch—or concentration—order). Besides this, the interaction between both orders must be considered. A low correlation between the concentration patterns is especially recommended for those solutes presenting a high correlation in the chromatographic order (i.e. small peak distances). On the other hand, solutes showing a low correlation in the chromatographic order can be deconvolved with good results even at high correlation in the concentration order. In general, the higher the correlation, the greater the risk of existence of non-unique solutions, and hence, of convergence in biased deconvolved profiles.

Analytical figures of merit, such as selectivity and sensitivity, are useful to evaluate the difficulty of any deconvolution problem. One fundamental concept in the calculation of these figures is the net analyte signal (NAS), which can be defined for a given analyte as the part of its signal that is orthogonal to the contribution of its interferents [26].

It is important to consider that the NAS dimension must agree with the order of the tensor obtained by the instrument when a single sample is measured [27]. In this work, a chromatograph provided with a single-channel detector was used, and therefore, any signal (i.e. chromatogram) is a row vector of t time measurements (mathematically, a first-order tensor). Consequently, the NAS of any compound will be also a $1\times t$ vector. However, the whole set of chromatograms is formally treated as second-order data. Accordingly, second-order figures of merit (such as the second-order NAS) can be calculated. Note that, in this context, the concept of signal is not straightforward: a signal $\mathbf{S}_i$ is defined as a $b\times t$ matrix, since b chromatograms are folded, each one containing t measurements. Although this definition is artificial, the second-order NAS is a useful tool to give information about the correlation in both chromatographic and concentration orders.

Unfortunately, the calculation of the second-order

NAS is controversial and several definitions can be found in the literature [28–30]. In this work, the definition proposed by Messick and co-workers [30] is used, because the interaction between both data orders is implicitly included. According to this definition, the NAS of the *i*th compound (from the *c* available), considering *b* batches, each one with *t* measurements, is calculated as follows:

$$\mathbf{NAS_i} = \mathbf{S_i} - \mathbf{P_i}(\mathbf{S_i}) \quad (4)$$

where $\mathbf{S_i}$ is the $b \times t$ matrix obtained from the injection of the *i*th pure compound at *b* different concentration levels, and $\mathbf{P_i}(\mathbf{S_i})$ is the projection of the $\mathbf{S_i}$ matrix on the vector space spanned by the $(c-1)$ *b*-fold chromatograms of the remaining compounds: $\{\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_{i-1}}, \mathbf{S_{i+1}}, \ldots, \mathbf{S_c}\}$. Mathematical details on the computation of $\mathbf{P_i}(\mathbf{S_i})$ are given elsewhere [27].

A useful derived measurement is the Frobenius norm of the **NAS** matrix, which is the sensitivity at unity concentration in a second-order calibration. In this work, rigorously speaking, the second-order sensitivity cannot be calculated since one of the data orders is the concentration itself. However, the Frobenius norm reduces the **NAS** matrix to a single value, which quantifies the fraction of orthogonal signal considering the two data orders.

The transition from second- to first-order figures of merit is easily done by applying Eq. (4) to vectors instead of matrices:

$$\mathbf{nas_i} = \mathbf{s_i} - \mathbf{p_i}(\mathbf{s_i}) \quad (5)$$

Note that here $\mathbf{s_i}$ can be both the ''chromatographic'' signal (a vector $t \times 1$ sized containing the peak profile of compound *i*), or the ''peak height'' signal (a vector $1 \times b$ sized containing the peak heights of the *i*th solute in all batches). Accordingly, first-order ''chromatographic selectivity'' and ''peak height selectivity'' for a given *i* compound are defined as the ratios of the corresponding $\mathbf{nas_i}$ and $\mathbf{s_i}$ Euclidean norms. In order to avoid confusions between chromatographic peak profile and peak height profile, we have adopted the more intuitive term ''concentration selectivity'' instead of ''peak height selectivity'', although figures of merit were actually calculated using peak heights. The selectivity quantifies the relative amount of information that is orthogonal to the interferents in a given data order. The chromatographic selectivity of any solute can be viewed as a measurement of the chromatographic resolution [27,30], because it evaluates which part of the *i*th peak profile cannot be modelled as a linear combination of the peak profiles of its interferents. The higher the selectivity, the lower the correlation among peak profiles, and therefore, the greater the chromatographic resolution.

Two drawbacks arise from the use of first- and second-order NAS and their derived figures of merit. The first one is the fact that these measurements are sensitive to the tensor size, which means that they depend on the number of points in the chromatogram and the number of batches in the set. The second drawback is the requirement of knowing accurately the true background corrected signals (i.e. both peak and concentration profiles) in advance, for each compound. In the experimental practice, figures of merit are calculated from the deconvolved peak profiles, which can be more or less biased. It is evident that only studies based on simulated signals will lead to unbiased figures of merit. This is the reason of not including replicates in this study since peak shapes of replicates are strictly identical.

3. Experimental

3.1. Apparatus

An Agilent (Model 1100, Palo Alto, CA, USA) chromatograph, equipped with an isocratic pump, a UV-visible detector and an autosampler, was used. A PC computer was connected to the chromatograph through an Agilent integrator (Model 3396A). Signal acquisition was made with the PEAK-96 software (Hewlett-Packard, Avondale, PA, USA). ODS-2 Spherisorb column (125×4.6 mm I.D., 5 μm particle size) and precolumn (35×4.6 mm I.D., 5 μm particle size) (Scharlab, Barcelona, Spain) were used. The detection wavelength was 364 nm for tetracyclines and 275 nm for sulfonamides. The flow-rate was 1.0 ml/min, and the injection volume, 20 μl. The whole study was carried out at room temperature.

3.2. Reagents

Mobile phases were prepared with sodium dodecyl sulphate (SDS, 99% purity, Merck, Darmstadt, Germany), and acetonitrile or 1-butanol (Scharlab), buffered at pH 3 with citric acid and sodium hydroxide (Panreac, Barcelona, Spain). The mobile phases and solutions to be injected were vacuum filtered through 0.45 μm Nylon membranes (Micron Separations, Westboro, MA, USA), except sulfachloropyridazine, for which celulose acetate filters were used, owing to adsorption problems in the Nylon filters. The standards were prepared from 100 μg/ml stock solutions, dissolving the pure compounds in a few milliliters of 95% (v/v) ethanol (Prolabo, Fontenay, France). The working solutions were obtained by dilution with aqueous 1% acetic acid (Panreac). The probe compounds were oxytetracycline chlorhydrate, tetracycline chlorhydrate, sulfachloropyridazine, sulfapyridine, sulfathiazole and sulfisoxazole (Sigma, St. Louis, MO, USA). Nanopure water (Barnstead, Boston, MA, USA) was used to prepare all solutions.

3.3. Software

Home-built in routines, written in MATLAB 4.2c (The Mathworks), were developed for data treatment.

4. Results and discussion

4.1. Factors affecting the deconvolution

The performance of the deconvolution approach was first studied using artificial chromatograms of binary mixtures, which were built from individual experimental peaks previously fitted to the PMG_2 model (Eq. (1)). For this purpose, sulfisoxazole and sulfapyridine were chromatographed with a 0.10 *M* SDS–6% (v/v) acetonitrile mobile phase at pH 3. In these conditions, the asymmetry factors measured at 10% peak height were 1.6 and 2.1, respectively. Mixtures of different complexity were obtained by keeping the retention time of the sulfisoxazole peak, and shifting gradually the sulfapyridine peak to cross the former (from +29 to −27 s from its original position, stepped by 8 s). The overlapping degree depended on the identity of the preceding peak, since the asymmetry factors of both compounds were different. Artificial samples were built by varying the peak heights of both components between 3 and 5 units. The final signal for the binary mixture was obtained by adding normally distributed noise to the sum of individual chromatograms.

An experimental design (Table 1) was applied to establish the effects of five experimental factors on the results: concentration selectivity, chromatographic selectivity, number of batches and replicates, and noise. The total number of experiments was $3\times8\times3\times2\times3=432$, which allowed a comprehensive study of the proposed approach. The studied chromatographic selectivity values corresponded to the following difference between the retention times of the two peaks: $f_2=0.185$(−2 s), 0.475(6 s), 0.616(−10 s), 0.834(14 s), 0.867(−18 s), 0.960(−26 s), 0.965(22 s), 0.994(30 s). It should be noted that the chromatographic selectivity depends not only on peak separation but also on the elution order, due to the different peak asymmetries. The 3-batch case is useful to understand the meaning of the concentration selectivity values: $f_1=0$ corresponds to the height values (5, 5), (4, 4) and (3, 3) for batches 1, 2 and 3; $f_1=0.21$ to (4, 5), (4, 4), and (4, 3); and $f_1=0.41$ to (3, 5), (4, 4) and (5, 3). The 2- and 4-batch cases were similarly established. Those configurations presenting the same correlation in concentration but a different number of batches (and consequently, different concentration selectivity), were computed as the mean value. The experiments were processed treating each batch independently and considering all batches simultaneously, in order to compare the performance of the single- and multi-batch deconvolution.

Table 1
Factors and levels in the experimental design

Factor	Levels
Concentration selectivity (f_1)[a]	0, 0.21, 0.41
Chromatographic selectivity (f_2)[a]	0.185, 0.475, 0.616, 0.834, 0.867, 0.960, 0.965, 0.994
Number of batches (f_3)	2, 3, 4
Number of replicates (f_4)	1, 2
Noise (f_5)[b]	0.01, 0.03, 0.05

[a] See Section 2.2 for mathematical definition.
[b] Measured as standard deviation.

The use of artificial signals allows an accurate control of the studied factors, which is not possible with real samples. This also permits removing or controlling undesirable sources of error, such as the lack of fit or the lack of bilinearity. The lack of fit was completely suppressed by using the same peak model to build and deconvolve the signals. The relative shifts in retention time of the peaks belonging to the same batch were kept constant, which eliminated partially the lack of bilinearity. A practical methodology should be, however, able to face some non-idealities always present in real samples. For this reason, the retention times in each chromatogram were shifted as a block with regard to a chromatogram taken as reference. Accordingly, Eq. (2) was selected for multi-batch deconvolution.

The sum of squared residuals for each individual peak profile (SSR_i) was used as global measurement of the deconvolution error [16]. These residuals were computed by subtracting, for each compound, the deconvolved peak profile from the true one. Unfortunately, the calculation of SSR_i values requires real individual peak profiles to be available, which is only possible with simulated chromatograms. When working with real signals, the true profiles are unknown, and the analyst is limited to less exhaustive measurements, such as the comparison of peak areas or heights from standards.

The results obtained from the deconvolution of the chromatograms in the 432 configurations of the experimental design are plotted in Fig. 1a (single-batch deconvolution) and Fig. 1b (multi-batch deconvolution). In the *X*-axis, the average SSR_i($\overline{SSR_i}$), is represented as error measurement. This is calculated as the sum of SSR_i for each solute in all chromatograms included in a given configuration, divided by the number of experimental points extended to all chromatograms (considering batches and replicates). Defined in this way, $\overline{SSR_i}$ is independent of the number of points, batches and replicates in the experimental design. The *Y*-axis depicts a signal-to-noise measurement: the Frobenius norm of the second-order **NAS** (Eq. (4)), multiplied by the square root of the number of replicates and divided by the standard deviation of the added noise. This measurement summarises the five studied factors and will be called relative NAS (r-NAS). It includes the interaction between both concentration and chromatographic correlations, and indicates the signal fraction that is free of overlap in both data orders.

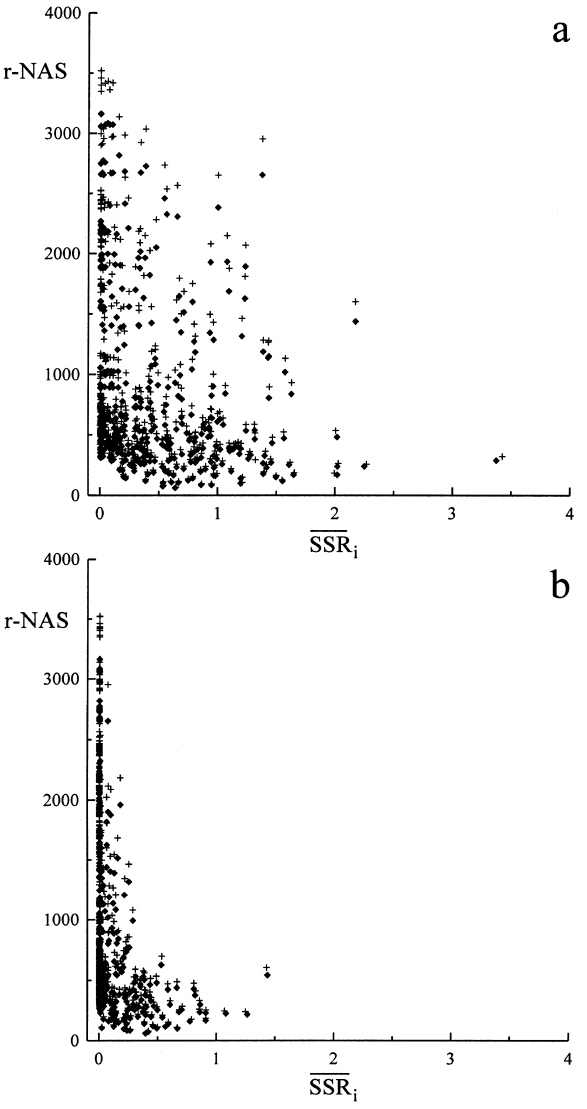

Fig. 1. Performance of: (a) the single- and (b) multi-batch deconvolution of the chromatograms included in the experimental design. Compounds: sulfapyridine (+) and sulfisoxazole (♦). See Section 4.1 for r-NAS and $\overline{SSR_i}$ definitions.

As expected, the errors obtained with the single-batch deconvolution were larger than in the simultaneous treatment of a set of chromatograms. The multi-batch deconvolution reports therefore benefits with regard to the single-batch treatment. Since $\overline{SSR_i}$

for both compounds in the deconvolved mixture were similar, in further studies the sum of these two values was used for simplicity. Fig. 1 shows that r-NAS is negatively correlated with the deconvolution error. The lower the selectivity in the concentration or chromatographic order (i.e. the higher the correlation in any data order), the lower the r-NAS value, and the greater the risk of obtaining a biased solution. This negative correlation is more apparent in the multi-batch approach. Also, in this case, for a given r-NAS, the probability of obtaining the right solution was appreciably greater. In the multi-batch approach, the risk of reaching a biased solution increased significantly for about r-NAS$<$ 1000, whereas the single-batch deconvolution did not guarantee a right convergence even at r-NAS$=$3000.

Although r-NAS summarises in a single measurement the effects of the five factors considered in this study, finding the significance of each one is interesting. Fig. 2 shows multiple box-and-whisker plots of the deconvolution errors, for each factor, in both single- (Fig. 2a–e) and multi-batch (Fig. 2f–j) approaches. As indicated, the error was calculated as the sum of $\overline{\mathrm{SSR}_i}$ extended to the two compounds.

The significance of each factor was also evaluated by building a linear model of $\Sigma\,\overline{\mathrm{SSR}_i}$ as a function of the five factors and their first-order interactions:

$$\sum \overline{\mathrm{SSR}_i} = \beta_0 + \sum_{i=1}^{5} \beta_i f_i + \sum_{i=1}^{5} \sum_{j=i+1}^{5} \beta_{ij} f_i f_j \qquad (6)$$

In this equation, f_i are factor levels and β_i model parameters. Previously to the computation, the $\Sigma\,\overline{\mathrm{SSR}_i}$ and f_i values were normalised. When all configurations in the experimental design were included in the fitting, no clear conclusions could be obtained, owing to the strong influence of the extreme outliers. For this reason, these experiments were removed (24 from the 432 experiments). The results are presented in Table 2 together with the 95% confidence intervals.

Fig. 2f shows that, in the multi-batch deconvolution, the higher the concentration selectivity (f_1), the smaller the errors (see also the significant negative value of β_1 in Table 2). The observed outliers for $f_1=0.21$ and 0.41 correspond to highly overlapped peaks. In these cases, the influence of the concentration selectivity on the error was smaller. Since single-batch treatments resolve each chromatogram independently, no relevant influence of f_1 on the deconvolution errors should be expected. However, the error increased significantly with f_1 (Fig. 2a and Table 2). This result can be understood if one considers that a low concentration selectivity is obtained with chromatograms showing peaks of similar height, which have more selective regions in the chromatogram. The opposite situation will yield a highly biased result: if the difference in height is too large, the probability of modelling the chromatogram of the mixture by only considering the peak of the major component is greater, since less selective regions exist for the smaller peak. This effect will be also observed in the deconvolution of real chromatograms in the next section.

The factor having the greater impact on the error in both single- and multi-batch deconvolution is the peak resolution (Fig. 2b,g and Table 2), which is here measured as the chromatographic selectivity (f_2). This factor is observed to be inversely correlated with the deconvolution error. Note that at intermediate resolutions (i.e. $f_2=0.62$–0.87), the multi-batch approach yielded not only smaller errors than the single-batch treatment, but also more successful convergences (i.e. the number of cases giving errors close to zero was greater). The results obtained with the multi-batch approach were reliable for $f_2\geq$ 0.62, whereas the single-batch approach only succeeded for $f_2\geq 0.96$.

The number of batches was also inversely correlated with the error in the multi-batch deconvolution, whereas as expected, no significant correlation was found in the single-batch treatment (see Fig. 2c,h and confidence intervals in Table 2). The inclusion of more replicates did not produce any improvement in the results (Fig. 2d,i and Table 2), due to the greater influence of the other studied factors (especially in the single-batch deconvolution). Finally, noisy peaks gave wrong solutions, which is clearly observed in the multi-batch deconvolution. In this approach, the error increment was higher when going from $f_5=$ 0.01 to 0.03 than from 0.03 to 0.05 (standard deviation of the normally distributed noise). It seems that for $f_5>0.03$, the noise was too high and the deconvolution yielded similar errors. In the single-batch approach, all the considered noise levels were too high and the method failed in most cases. The

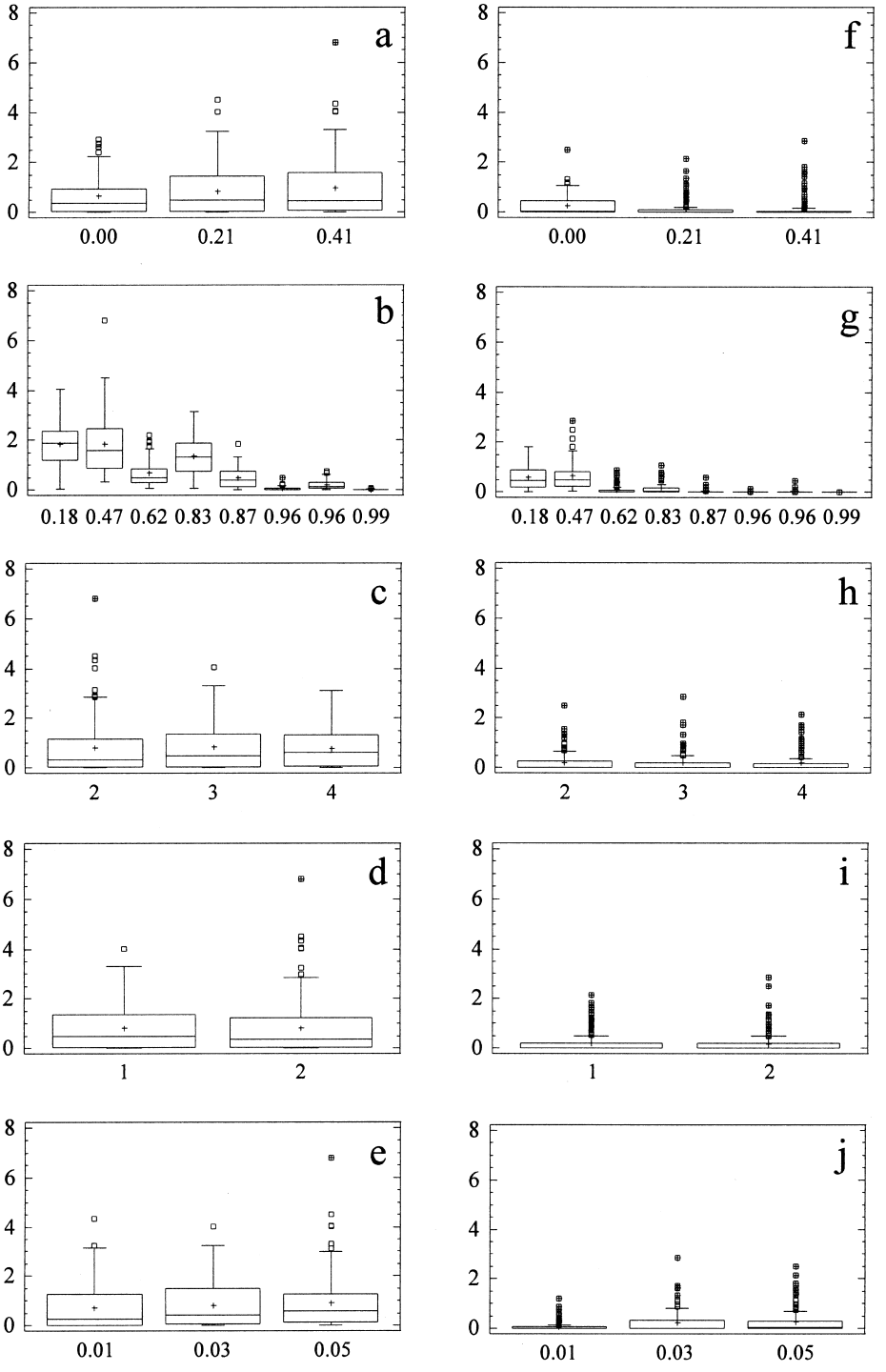

Fig. 2. Multiple box-and-whisker plots of the deconvolution errors $\left(\Sigma\overline{SSR_i}\right)$ obtained with: (a–e) the single- and (f–j) multi-batch approaches, for the 432 configurations in the experimental design (Table 1). The factors (abscise scale) are: (a, f) concentration selectivity, (b, g) chromatographic selectivity, (c, h) number of batches, (d, i) number of replicates, and (e, j) noise.

Table 2
Values and confidence intervals of the parameters in Eq. (6) for the single- and multi-batch deconvolution of the simulated signals in the experimental design[a]

Factors	Coefficients	Single-batch deconvolution	Multi-batch deconvolution
Independent term	β_0	**0.31**±0.10	**0.57**±0.11
Concentration selectivity	β_1	**0.28**±0.11	−**0.33**±0.12
Chromatographic selectivity	β_2	−**0.31**±0.11	−**0.59**±0.12
Number of batches	β_3	−0.01±0.11	−**0.14**±0.14
Number of replicates	β_4	−0.01±0.09	−0.07±0.09
Noise added	β_5	0.04±0.11	**0.23**±0.12
Interactions	β_{12}	−**0.20**±0.11	**0.32**±0.12
	β_{13}	−**0.09**±0.08	−0.01±0.09
	β_{14}	0.01±0.07	0.04±0.07
	β_{15}	−0.07±0.08	−**0.11**±0.09
	β_{23}	0.06±0.11	**0.14**±0.12
	β_{24}	−0.02±0.09	0.04±0.09
	β_{25}	0.01±0.11	−**0.14**±0.12
	β_{34}	0.01±0.07	0.02±0.07
	β_{35}	0.02±0.08	−0.01±0.09
	β_{45}	0.01±0.07	0.02±0.07

[a] Bold numbers point out the significant factors or interactions. 95% confidence intervals are given.

multi-batch deconvolution was more robust in fitting noisy signals.

According to the magnitude of the main effects, β_{12} (interaction between both first-order selectivities) is the strongest cross-product term, followed by β_{23}, β_{25} and β_{15} (multi-batch) and β_{13} (single-batch) (Table 2).

4.2. Deconvolution of real chromatograms

Three binary mixtures yielding chromatograms of different complexity were considered to check the performance of the proposed methodology on real data: case (i), oxytetracycline–tetracycline; case (ii), sulfathiazole–sulfachloropyridazine; and case (iii), sulfisoxazole–sulfapyridine. Mobile phase compositions were 0.05 *M* SDS–5% 1-butanol at pH 3 for case (i), and 0.10 *M* SDS–6% acetonitrile at pH 3 for cases (ii) and (iii). Calibrates were performed using standards of the pure compounds. The concentration ranges (μg/ml) of the standards were: 0.5–5.0 (oxytetracycline and tetracycline), 0.6–2.8 (sulfathiazole), and 1.0–5.0 (sulfachloropyridazine, sulfisoxazole and sulfapyridine).

Table 3 summarises the peak parameters of the six compounds (t_R, N and B/A), which were calculated

Table 3
Peak parameters for the probe compounds and resolution of peak pairs

Compound	t_R, min	N[a]	B/A[b]	R_S[c]	f_2[d]
Oxytetracycline	8.97	284	2.4	0.64	0.971
Tetracycline	10.15	194	2.8		
Sulfathiazole	3.60	554	1.8	0.43	0.926
Sulfachloropyridazine	3.85	591	1.7		
Sulfisoxazole	5.05	813	1.6	0.14	0.579
Sulfapyridine	5.15	521	1.9		

[a] Plate counts were obtained according to Foley and Dorsey [17].

[b] Peak asymmetry was measured at 10% of peak height; B and A are the distances from the centre to the tailing and leading edge of the peak, respectively.

[c] Measured as $R_S = \dfrac{t_{R,2} - t_{R,1}}{B_2 - A_1}$.

[d] f_2 is the chromatographic selectivity.

from standard solutions. The R_S values and chromatographic selectivities (f_2) are also given for the binary mixtures. The number of batches was 6, 2 and 3 for cases (i), (ii) and (iii), respectively. Three replicated injections by batch were made in case (ii), and two in cases (i) and (iii), for both samples and calibration standards. The replicates were measured on 2 different days in case (i), in order to test the methodology under less favourable conditions. In this example, the data obtained each day were treated independently, and the mean of the two concentration values was given as final result. Fig. 3 depicts the experimental chromatograms for each mixture.

Five methods were tested, which can be classified attending to the amount of information processed in the deconvolution. In method (i), the multi-batch approach was applied considering the samples and all the available standards altogether. This means that the standards of each compound were processed as additional batches included in the data matrix. Method (ii) considered only one standard by compound together with the samples. The concentration of these standards was (μg/ml): 2.0 (oxytetracycline and tetracycline), 2.2 (sulfathiazole), 4.0 (sulfachloropyridazine), and 3.0 (sulfisoxazole and sulfapyridine). Methods (iii) and (iv) only processed the samples. Finally, method (v) corresponds to the single-batch deconvolution. Eq. (2) was used in methods (i), (ii) and (iii), whereas Eqs. (1) and (3) were applied in methods (v) and (iv), respectively.

Calibrates were built in all methods using the deconvolved signals of the pure standards. In method (i), the calibration values were obtained after a simultaneous processing of all available standards and samples. The calibration models were built in methods (ii)–(iv) by applying the multi-batch treatment to the whole set of standards. In method (v), each standard was fitted separately. Peak heights and areas were used in all cases for sample quantification, but only the results obtained using heights are given (Tables 4–6).

The single-batch deconvolution—method (v)—gave the poorest results (Table 4). In the mixture of oxytetracycline–tetracycline, this is especially noteworthy for the second compound, which eluted at longer times and was affected by the tail of the preceding peak. As can be seen, enriching the information by arranging the data vectors in a matrix (multi-batch deconvolution) solved this drawback, at least partially. For the five strategies using the PMG_2 model, the smallest errors were associated with samples 3 and 4, where the areas of both peaks were similar, and the largest errors were observed for samples 1 and 6, where the concentration ratios were 0.1 and 10, respectively. In these cases, the risk of modelling the whole signal by considering only the major component was higher. This behaviour, which is especially troublesome in the single-batch deconvolution, was also observed with the simulated signals in Section 4.1.

The PMG_3 model (i.e. a cubic polynomial is used as standard deviation) yielded better accuracy in the fitting of peaks showing long tails, as is the case of oxytetracycline and tetracycline ($B/A=2.4$ and 2.8, respectively). However, when the peaks are strongly overlapped, the higher flexibility of this model can be translated into wrong solutions (overfitting). As can be observed in Fig. 3a, the peaks involved in the mixture of oxytetracycline–tetracycline were moderately resolved ($R_S=0.64$). It seems that the use of the PMG_3 model did not constitute any problem, since the results in Table 4 were improved. Observe that the inclusion of standards in the multi-batch deconvolution—compare methods (i), (ii) and (iii)—usually yielded similar results, since the resolution of the mixture of oxytetracycline and tetracycline was large enough to allow a satisfactory retrieval of the individual elution profiles. The error obtained with the PMG_2 model for tetracycline deserves some attention. In this case, the inclusion of the whole set of calibration standards led to slightly better results. As indicated, this compound was affected by the tail of oxytetracycline, and the retrieval of the true peak profile was likely more difficult. Note that PMG_3 solved this problem, meanwhile the use of Eq. (3) instead of Eq. (2) did not modify significantly the results.

Although the coelution in the mixture of sulfathiazole–sulfachloropyridazine was stronger, with $R_S=0.43$ (Fig. 3b), the results were better than in the previous example (Table 5). It should be noted that the differences between peak heights of both compounds in the analysed samples were not so large. On the other hand, peak tails were smaller (see Table 3), and PMG_2 yielded sufficiently good results, making the use of PMG_3 unnecessary. The single-

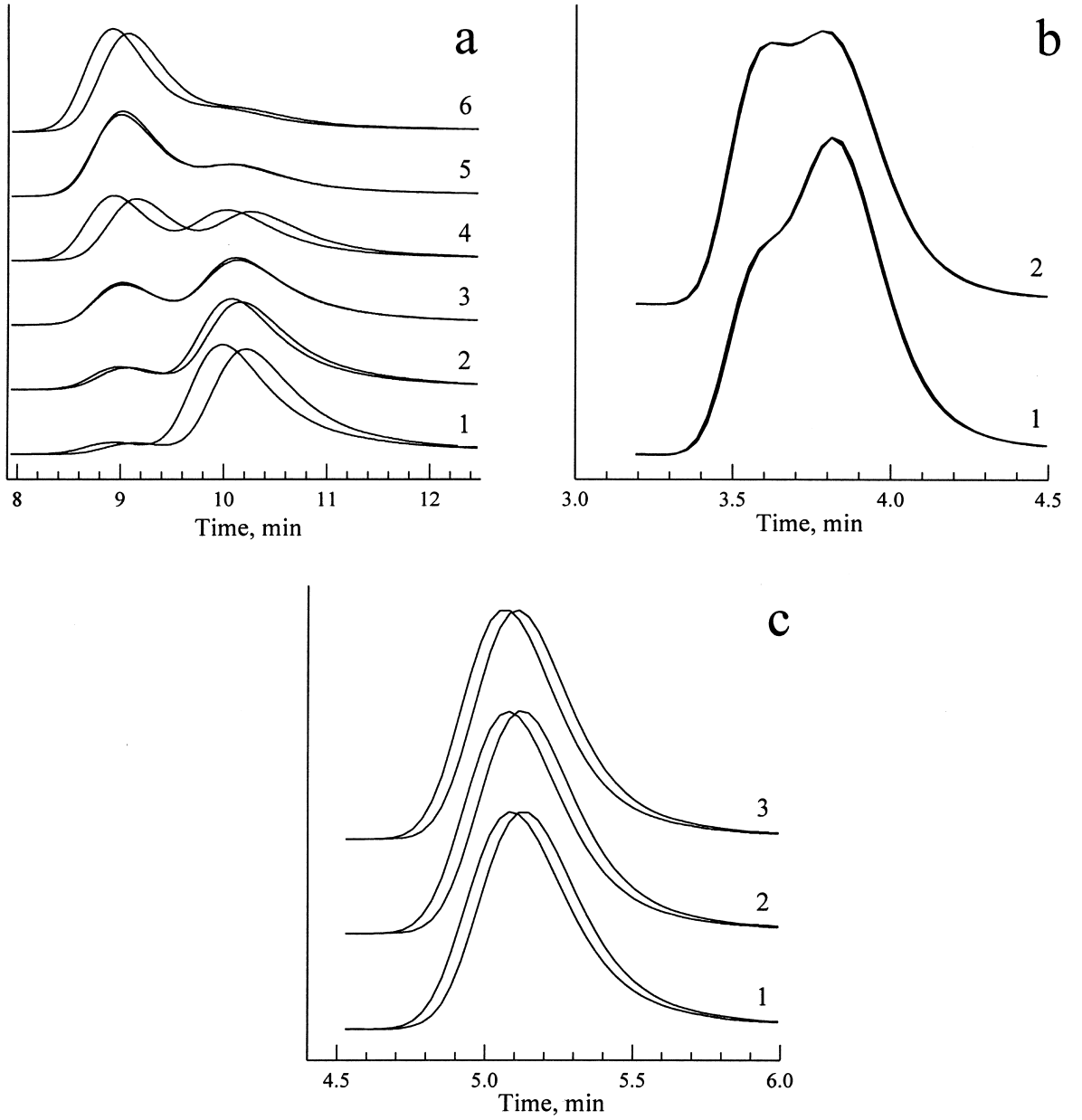

Fig. 3. Experimental chromatograms of binary mixtures of: (a) oxytetracycline–tetracycline (six batches with two replicates); (b) sulfathiazole–sulfachloropyridazine (two batches with three replicates); and (c) sulfisoxazole–sulfapyridine (three batches with two replicates). Tables 4–6 give the composition of each sample.

batch deconvolution gave again poorer results for the compound eluting at longer time, which was strongly affected by the tail of the preceding peak. Again, the multi-batch deconvolution solved this problem. The use of standards enhanced the method performance (compare methods (i), (ii) and (iii)).

As can be deduced from Fig. 3c, the mixture of sulfisoxazole–sulfapyridine yielded practically full

Table 4
Concentrations (μg/ml) obtained in the deconvolution of oxytetracycline and tetracycline real mixtures using the PMG_2 and PMG_3 models[a]

Compound	Model	Method[b]	Samples						Error[c]
			1	2	3	4	5	6	
Oxytetracycline		Theoretical value	0.50	1.00	2.00	3.00	4.00	5.00	
	PMG_2	i	0.62	1.14	2.00	2.98	3.86	4.68	0.74
		ii	0.62	1.13	2.00	2.98	3.86	4.67	0.75
		iii	0.62	1.13	2.00	2.98	3.87	4.68	0.73
		iv	0.63	1.14	1.99	2.97	3.85	4.65	0.81
		v	0.60	1.11	1.98	2.97	3.87	4.59	0.81
	PMG_3	i	0.61	1.12	1.99	2.98	3.87	4.69	0.70
		ii	0.60	1.11	1.98	2.96	3.85	4.66	0.77
		iii	0.60	1.11	1.98	2.96	3.85	4.66	0.77
		iv	0.61	1.12	1.97	2.95	3.83	4.64	0.83
		v	0.62	1.12	1.97	2.96	3.84	4.63	0.85
Tetracycline		Theoretical value	5.00	4.00	3.00	2.00	1.00	0.50	
	PMG_2	i	4.91	4.02	2.89	2.03	1.11	0.65	0.52
		ii	4.90	4.03	2.91	2.08	1.17	0.72	0.68
		iii	4.90	4.03	2.92	2.09	1.19	0.74	0.73
		iv	4.90	4.02	2.91	2.07	1.16	0.70	0.64
		v	4.72	3.97	2.88	2.08	1.23	0.87	1.12
	PMG_3	i	4.91	4.02	2.87	2.01	1.07	0.60	0.41
		ii	4.91	4.02	2.88	2.01	1.08	0.60	0.42
		iii	4.91	4.02	2.88	2.01	1.08	0.60	0.43
		iv	4.90	4.00	2.86	1.99	1.05	0.57	0.38
		v	4.88	3.94	2.91	2.01	1.12	0.84	0.74

[a] Heights were used for calibration.
[b] See Section 4.2 for method definition.
[c] Sum of all deviations (absolute values) with regard to the theoretical concentration.

coelution ($R_S = 0.14$). The single-batch deconvolution was unable to retrieve the peak profiles and failed dramatically (Table 6). The multi-batch deconvolution also failed, except when the calibration standards were treated together with the samples (methods (i) and (ii)). Surprisingly, method (ii) gave better results than method (i), which contains more information on the pure peak profiles. This can be explained by considering that highly overlapped peaks lead to badly defined error surfaces, and minor changes in the deconvolution problem (such as the inclusion of a different number of standards) can be translated in different solutions. However, the most remarkable is that, in spite of the complexity of the problem, the results were satisfactory.

Computation of peak areas instead of heights is more problematic, due to the difficulty in establishing properly the baseline. The results obtained using areas were usually similar to those with heights, except when method (v) was applied to the samples of oxytetracycline–tetracycline exhibiting very different concentrations (samples 1 and 6). In this case, the errors for the most diluted compound were significantly larger. For the samples of sulfisoxazole–sulfapyridine using methods (iii), (iv) and (v), the results obtained using areas were poorer than those using heights.

In general, the differences observed between both calibration strategies (i.e. heights and areas) denote a disagreement in the retrieved peak profiles from the chromatograms of mixtures and calibration set. For example, with method (i) no differences were obtained by using heights or areas, except those associated with peak area computation. This result reflects a perfect agreement in peak profile among mixtures and calibration standards, forced by the inclusion of the pure standards as additional batches in the data matrix. As the standards are removed

Table 5
Concentrations (μg/ml) obtained in the deconvolution of sulfathiazole and sulfachloropyridazine real mixtures using the PMG_2 model[a]

Compound	Method[b]	Samples		Error[c]
		1	2	
Sulfathiazole	Theoretical value	2.20	2.80	
	i	2.19	2.78	0.04
	ii	2.18	2.78	0.04
	iii	2.17	2.76	0.07
	iv	2.14	2.74	0.12
	v	2.18	2.81	0.02
Sulfachloropyridazine	Theoretical value	4.00	3.00	
	i	3.94	2.98	0.09
	ii	3.93	2.97	0.10
	iii	3.85	2.86	0.29
	iv	3.91	2.93	0.15
	v	3.77	2.80	0.44

[a] Heights were used for calibration.
[b] See Section 4.2 for details.
[c] Sum of all deviations (absolute values) with regard to the theoretical concentration.

from the processed matrix in the multi-batch approach, the individual peak profiles in the samples become less comparable to those in the standard solutions, especially for strongly overlapped mixtures. As the resolution increases, this contrast between processing heights and areas is only observed with the single-batch treatment, applied to samples showing large differences in peak height.

Table 6
Concentrations (μg/ml) obtained in the deconvolution of sulfisoxazole and sulfapyridine real mixtures using the PMG_2 model[a]

Compound	Method[b]	Samples			Error[c]
		1	2	3	
Sulfisoxazole	Theoretical value	3.00	4.00	5.00	
	i	2.17	3.37	4.59	1.87
	ii	2.82	3.75	4.82	0.61
	iii	5.06	5.34	5.72	4.13
	iv	2.75	1.59	2.42	5.24
	v	2.52	7.45	7.94	6.87
Sulfapyridine	Theoretical value	5.00	4.00	3.00	
	i	5.86	4.70	3.46	2.02
	ii	5.19	4.31	3.20	0.70
	iii	3.82	3.65	3.32	1.85
	iv	5.24	6.32	5.57	5.13
	v	6.27	0.74	0.43	7.09

[a] Heights were used for calibration.
[b] See Section 4.2 for details.
[c] Sum of all deviations (absolute values) with regard to the theoretical concentration.

5. Conclusions

Building two-way data from one-dimensional chromatograms obtained in different runs yielded enhanced results, with regard to those achieved from an independent deconvolution of each experiment. This is especially noteworthy in the analysis of compounds strongly affected by the peak tails of preceding compounds. The methodology proposed in this work allows to model in a continuous way the lack of bilinearity associated to peak shifts, which is undesirable in the application of latent variable methods.

Second-order figures of merit are useful for testing the difficulty of multi-batch deconvolution problems. These measurements should be, however, taken with care because they vary with the data dimensionality and are defined for bilinear data. Consequently, non-idealities (such as the lack of fit or lack of bilinearity) cannot be appropriately measured. These limitations prevent the second-order figures of merit to be applied to the comparison of deconvolution problems. Also, these values can only be computed when the true peak profiles and concentration patterns are known.

The factors that showed a stronger influence on the results were the chromatographic and concentration selectivities. The probability of success in the multi-batch deconvolution was greater than in the single-batch approach, which was especially noteworthy for peaks exhibiting a moderate or strong coelution.

The studies carried out in this work show that the simultaneous deconvolution of the chromatograms of several samples enhances the information and leads to satisfactory results, even for highly overlapped peaks. The smaller the chromatographic resolution, the greater the requirement to introduce information from pure compounds. In the experimental practice, when several samples of two overlapped compounds at different concentration ratios should be analysed, the simultaneous treatment of all samples is rec-

ommended. When only one sample is analysed, the application of the multi-batch approach requires the inclusion of the chromatograms of standards to get the enrichment effect.

When large tails are observed and the overlapping is moderate, the use of PMG_3 instead of PMG_2 is recommended. However, the former model can lead to overfitting in the deconvolution of strongly overlapped peaks. On the other hand, the similar results obtained with Eqs. (2) and (3) indicate that, at least for the studied problems, the relative shifts in peak position are non-significant, and matching of different chromatograms can be performed by displacing the two overlapped peaks as a block.

The proposed deconvolution approach may give rise—depending on the available information—to several particular cases of smaller difficulty, where the analyst may set one or more peak parameters (corresponding to one or more solutes), to get an easier deconvolution [20,31]. The most simplified procedure would include fixed values of the peak shape parameters obtained from external standards. Such an approach can yield good results, provided pure standards of all individual compounds are available. This is not always the case in the experimental practice (e.g. in some cases only some of the pure compounds are available; in others, certified materials with mixed composition are used as standards to consider matrix effects).

The multi-batch approach presents an additional advantage apart from quantitative applications: it can be used in peak purity assays of chromatographic signals. Thus, when the analyst suspects that an impurity is present under a peak, spiked samples can be prepared with the known analyte and a tentative deconvolution can be carried out searching two compounds, to check whether a secondary compound actually remains under the main peak, without DAD detection.

Acknowledgements

This work was supported by Project BQU2001-3047 (Ministerio de Ciencia y Tecnología of Spain). JRTL thanks the MCYT for a Ramón y Cajal position.

References

[1] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[2] Y.L. Xie, J.J. Baeza-Baeza, G. Ramis-Ramos, Chemom. Intell. Lab. Syst. 32 (1996) 215.
[3] J.M. García, A.I. Jiménez, J.J. Arias, K.D. Khalaf, Analyst 120 (1995) 313.
[4] R.M. Dyson, S. Kaderli, G.A. Lawrance, M. Maeder, A.D. Zuberbühler, Anal. Chim. Acta 353 (1997) 381.
[5] A. Izquierdo-Ridorsa, J. Saurina, J. Hernández-Cassou, Chemom. Intell. Lab. Syst. 38 (1997) 183.
[6] R. Gargallo, R. Tauler, A. Izquierdo-Ridorsa, Anal. Chem. 69 (1997) 1785.
[7] R. Tauler, B.R. Kowalski, S. Fleming, Anal. Chem. 65 (1993) 2040.
[8] R.E. Synovec, E.L. Johnson, T.J. Bahowick, Anal. Chem. 62 (1990) 1597.
[9] M.D. Hämäläinen, Y. Liang, O.M. Kvalheim, R. Anderson, Anal. Chem. 271 (1993) 101.
[10] Y. Liang, M.D. Hämäläinen, O.M. Kvalheim, R. Andersson, J. Chromatogr. A 662 (1994) 113.
[11] O.M. Kvalheim, Y. Liang, Anal. Chem. 64 (1992) 936.
[12] L.H. Ghaoui, J. High. Resolut. Chromatogr. 15 (1992) 449.
[13] L. Ghaoui, L.D. Rothman, J. High. Resolut. Chromatogr. 15 (1992) 36.
[14] C.A. Monnig, D.M. Dohmeier, J.W. Jorgenson, Anal. Chem. 63 (1991) 807.
[15] S. Mitra, T. Bose, J. Chromatogr. Sci. 30 (1992) 256.
[16] G. Vivó-Truyols, J.R. Torres-Lapasió, A. Garrido-Frenich, M.C. García-Alvarez-Coque, Chemom. Intell. Lab. Syst. 59 (2001) 107.
[17] J.P. Foley, J.G. Dorsey, Anal. Chem. 55 (1983) 730.
[18] J. Grimalt, H. Iturriaga, J. Olivé, Anal. Chim. Acta 201 (1987) 193.
[19] J.R. Torres-Lapasió, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, Anal. Chem. 69 (1997) 3822.
[20] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, J. Chromatogr. A (2001) 13.
[21] V.B. Di Marco, G.G. Bombi, J. Chromatogr. A 931 (2001) 1.
[22] D. Corne, M. Dorigo, F. Glover (Eds.), New Ideas in Optimisation, McGraw-Hill, London, 1999.
[23] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, A. Zitrou, J. Chromatogr. A 942 (2002) 93.
[24] W. Cai, F. Yu, X. Shao, Z. Pan, Anal. Lett. 33 (2000) 373.
[25] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Chemom. Intell. Lab. Syst. 59 (2001) 89.
[26] A. Lorber, Anal. Chem. 58 (1986) 1167.
[27] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 419.
[28] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 52 (1980) 1071.
[29] Y.D. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, J. Chemom. 7 (1993) 117.
[30] N.J. Messick, J.H. Kalivas, P.M. Lang, Anal. Chem. 68 (1996) 1572.
[31] R.D. Caballero, M.C. García-Alvarez-Coque, J.J. Baeza Baeza, J. Chromatogr. A 954 (2002) 59.